

Automatic foliar spot detection from low-cost RGB digital images using a hybrid approach of convolutional neural network and random forest classifier

Jaime Carlos Macuácu^{1,2} - ORCID: 0000-0002-4822-6098

Jorge António Silva Centeno¹ - ORCID: 0000-0002-2669-7147

Caisse Amisse³ - ORCID: 0000-0001-9458-5510

Mário Ernesto Jijón-Palma¹ - ORCID: 0000-0003-4890-2997

Kauê de Moraes Vestena¹ - ORCID: 0000-0003-1225-2371

¹ Departamento de Geomática, Universidade Federal do Paraná, Curitiba, Brasil.

E-mail: macua.gis@gmail.com, centeno@ufpr.br, majijpa@hotmail.com, kauemv2@gmail.com

² Universidade Eduardo Mondlane. Avenida Julius Nyerere, 3453, Maputo - Moçambique.

E-mail: macua.gis@gmail.com

³ Universidade Rovuma, Faculdade de Ciências, Avenida Josina Machel nº 256, Nampula - Moçambique.

E-mail: caamisse@gmail.com

Received in 08th December 2022.

Accepted in 16th November 2023.

Abstract:

Tomatoes are widely cultivated, both by family farmers and corporate producers. During the tomato growth cycle, several diseases can affect the plant. The identification of these diseases through short-range images is significant, and computer vision techniques are commonly used to identify diseases in plant leaves. In this paper, a hybrid model that combines a convolutional neural network (CNN) and a Random Forest (RF) decision tree is used for foliar spot detection in tomato leaves. High-level features learned and extracted from CNN are used as input for the RF classifier. To evaluate the proposed model's performance for plant disease identification, a case study of 2480 low-cost digital RGB images collected in actual field conditions, under different intensities of light exposure, were used, including healthy tomato leaves and leaves with visible symptoms of powdery mildew fungus, which attacks the tomato leaf. The results were compared with six conventional machine learning classifiers: Logistic Regression (LR), Linear Discriminant Analysis (LDA), K- Nearest Neighbors (KNN), Naive Bayes (NB), Support Vector Machine (SVM) and Random Forest (RF). The results show that the proposed model outperformed conventional classifiers, reaching an accuracy of 98%. The results highlight the importance of fusing models to improve the detection plant's diseases.

Keywords: Short range imaging; Deep learning; Random Forest Classifiers; Disease Identification.

How to cite this article: MACUÁCUA JC, CENTENO JAS, AMISSE C, JIJÓN-PALMA ME, VESTENA KM. Automatic foliar spot detection from low-cost RGB digital images using a hybrid approach of convolutional neural network and random forest classifier. *Bulletin of Geodetic Sciences*. 30: e2024001, 2024.



This content is licensed under a Creative Commons Attribution 4.0 International License.

1. Introduction

Tomato is one of the most important oleraceae due to its high consumption demand and its contribution to the generation of jobs and income, in addition to its significant participation in agribusiness. Tomato is one of the most consumed vegetables, both fresh and processed. However, tomato production has faced problems caused by microorganisms such as bacteria, fungi, nematodes and viruses that attack tomato plants. Imbalances in essential factors that affect plant growth, such as nutrients, water and light (Moore and Bradley, 2018), also have a significant impact on production yield, agricultural product quality and plant mortality (Liakos, 2018). These physiological processes negatively affect the country's economy, especially in the case of family farmers. Although producers are striving to reduce the impacts of these diseases by using pesticides and insecticides, the excessive use of these products can negatively affect the environment and human health and incorrect disease diagnosis can lead to the inappropriate use of these substances.

The automatic diagnosis of diseases in tomato plants plays a key role in the implementation of precautionary measures. By detecting diseased tomato plants, the collected information can be used to monitor large areas of cultivation. Accurate diagnosis of plant diseases allows for taking preventive measures to reduce production loss, improve product quality and increase farmers' income (Golhani et al., 2018; Hu et al., 2019).

Common methods for diagnosing plant diseases include visual estimation by an expert who identifies a disease based on characteristic disease symptoms or visible signs of a pathogen. Visual methods are considered accurate and reliable but are subject to one's experience (Hu et al., 2019) and, as they demand well-developed skills in disease diagnosis, they are considered a laborious, time-consuming and subjective task (Ma et al., 2018).

There is a need for methods that enable a rapid and cheaper diagnosis, which consists of using information technology to monitor the various factors present in the agricultural process to help farmers, especially in less-developed regions. An alternative is to use computer vision-assisted digital short-range imaging to help evaluate, identify, detect and, if possible, automatically diagnose plant disease based on image analysis. Automated methods are often questioned due to the disturbances that can occur during "in situ" imaging, such as variations in lighting (Liu and Wang, 2021). Such an uncontrolled condition can introduce colour variations associated with the background that can make the use of image analysis difficult (Barbedo, 2019; Liu and Wang, 2021), especially for automatic detection in large areas, although such methods constitute a rich field of research due to their advantages in terms of costs and the increasing processing power of small computers and cloud computing. Among the automated detection methods, some are based on segmentation with contrast elongation (Barbedo, 2019), segmentation using K-Mean clustering (Abdu et al., 2020), comprehensive colour combination with a growing region (Ma et al., 2018), and it is also possible to highlight the use of the correlation coefficient to separate infected regions (Khan et al., 2018). Their results can be combined with a variety of classification schemas, such as colour coherence and HSV (Barbedo, 2019; Ma et al., 2018; Abdu et al., 2020), the use of K-nearest neighbour (Abdu et al., 2020), SVM (Hu et al., 2019; Abdu et al., 2020) and multiclass SVM (Khan et al., 2018; Sharif et al., 2018), Bayesian classification network (Abdu et al., 2020), and artificial neural network (Sladojevic et al., 2016; Ferentinos, 2018; Zhang et al., 2019). The accuracy of these methods depends on the accurate extraction and selection of the visible features in the plant leaf and decreases when dealing with complex images.

Advanced approaches, based on so-called sensor-based methods, can detect and identify plant diseases. The sensors assess the optical properties of plants within different regions of the electromagnetic spectrum and then use the information beyond the visible range to observe different plant parameters, as exemplified in Mahlein (2016). Other promising approaches for plant disease assessment rely on observing how leaves reflect light. In this approach, hyperspectral imaging is used to detect subtle changes in plants' spectral reflectance. Kuswidiyanto et al. (2022) provide an overview of the literature regarding the use of hyperspectral aerial images for disease detection.

Image analysis for the accurate diagnosis of plant diseases is a challenging task in the agricultural sector that can be solved using image processing and deep learning techniques (Liu and Wang, 2021). In recent decades, the use of deep learning methods in the context of digital processing and computer vision has facilitated the identification of plant diseases in agriculture (Sladojevic et al., 2016; Too et al., 2019). The introduction of these technologies in rural areas can help small producers in the decision-making process, which is historically based on experience and intuition. Thus, the fusion of artificial intelligence with agriculture defines what is known as digital agriculture, which emerges as a new scientific field that uses large amounts of data, assisting in strategic planning and in decisions to increase agricultural production and productivity (Liakos, 2018; Golhani et al., 2018). Deep learning convolution neural networks (CNNs) have achieved great performance in automatic plant disease identification and classification. A detailed survey of existing studies that rely on CNNs to automatically identify crop diseases can be found in Boulent et al. (2019). The CNN models are capable of automatically extracting rich high-dimensional spatial features from raw data through the training process (Yamashita et al., 2018; Zhang et al., 2021). Despite this advantage, there is a considerable barrier in the use of CNN. It is often difficult to estimate the optimal model parameters for the extraction of high-level spatial features when similar and smaller training datasets are used (Hu et al., 2015; Kamilaris, and Prenafeta-Boldú, 2018); they sometimes take much longer to train; and there is a need for large, annotated datasets (i.e., hundreds/thousands of images). In the agricultural sector, there are not many publicly available datasets and, in many cases, there is the need to collect one's own datasets (Kamilaris, and Prenafeta-Boldú, 2018). The performance of plant disease identification drops significantly when the amount of training data is smaller. Thus, CNN models require a larger amount of data that must necessarily contain images captured in as many different conditions as possible. For plant disease identification, available solutions involving CNN models include the use of hybrid models (Tuncer, 2021; Bedi and Gole, 2021; Singh et al., 2022; Kaur et al., 2022; Rezk et al., 2022), however, there are still challenges because similar spatial and spectral characteristics between plant diseases make it difficult to extract sufficient features from CNN models.

In this paper, we propose a hybrid approach that fuses CNNs and Random Forest classifiers for the automatic identification of plant diseases using low-cost digital RGB images. The approach takes advantage of CNN's recognised ability to extract complex and hierarchical features directly from raw data, and RF's capability of dealing with classification and regression problems, as well as providing data interpretability. In the context of this study, CNN was trained to specialise in extracting high-level features from images, while RF plays a complementary role, handling a wide variety of features, including those that might not be easily learned by CNN. By combining these two models, we capture a broader range of information and improve the model's overall performance. The proposed approach was tested in our dataset, which contains a vast number of labelled images of tomato leaves collected in real field conditions, under different intensities of light exposure, including both healthy leaves and leaves with disease symptoms. In this case, powdery mildew fungus, which attacks the tomato leaf, is used as an example.

The results reported herein are part of a major study with the goal of detecting plant diseases and mapping the frequency of their occurrence based on the positional information provided by the cameras. So, at the end, a map would be produced to help farmers with disease management. The mapping step will be the theme of a future paper. The present study concentrates on the detection stage.

The paper is organised as follows: It begins with a discussion of related work in Section 2. In Section 3, the proposed method is described in detail. Section 4 evaluates the performance of the presented framework on a real dataset. Finally, section 5 summarises the most important findings and provides an outlook on future work.

2. Related works

There are several studies with deep learning approaches applying RGB digital image processing techniques for the identification and classification of diseases in plants, using the characteristics of symptoms visible in the

plant leaf or the fruit, as summarized in Table 1. Deep learning-based methods such as Mohanty et al. (2016), Brahimi et al. (2017), Amara et al. (2017); Too et al. (2019), Ferentinos (2018), Abdu et al. (2020), Khan et al. (2018), Kamal et al. (2019), Karthik et al. (2020), Afifi et al. (2020), Singh et al. (2022), Chen et al. (2022), and Saberi Anari (2022), produced robust results for diagnosing plant diseases, with accuracy levels above 98%.

Such studies focus on the analysis of RGB images, with certain ones leveraging publicly available datasets. These datasets encompass a diverse range of species, resulting in varying spatial resolutions to accommodate leaf fitting within the images. Additionally, the images are resized to a standardized dimension, thus modifying their resolutions.

Table 1: Main characteristics of studies on plant disease identification.

Author	Crop	Dataset	Sensor, Resolution and Input image	Model	Accuracy (%)
Kawasaki et al., 2015	Cucumber	Own	RGB- 800 images, Varied sizes; Resized (224 × 224px)	Custom CNN	94.90
Fujita et al., 2016	Cucumber	Own	RGB- 7,520 images; Resized (224 × 224px)	CNN1 and CNN2	83.20
DeChant et al., 2017	Maize	Own	RGB - 1,834 images; Varied sizes; Resized (224 × 224 px)	Pipeline of CNN	96.70
Ramcharan et al., 2017	Cassava	Own	RGB - 11,670 images; Varied sizes	Inception V3	93.00
Picon et al., 2019	Wheat	Own	RGB- 8,178 images; Resized (224 × 224px)	ResNet50	87.00
Selvaraj et al., 2019	Banana	Own	RGB - 700 images	ResNet, Inception V2, MobileNetV1	90.00
Hu et al., 2019	Tea leaf	Own	RGB- 120 images	C-DCGAN	90.00
Ma et al., 2018	Cucumber	Own	RGB-, 14,208 images, Size: 2592 × 1944px; Resized (800 × 600px)	DCNN	93.40
Zhang et al., 2019	Cucumber	Own	RGB- 800 images, Size: 2456 × 2058px; Resized (240 × 240px)	GPDCNN, combined with the AlexNet model	95.18
Saberi Anari, 2022	Apple, corn, cotton, grape, pepper, and rice	PVD ¹⁺ UCI ²	RGB- 90,000 images; Varied sizes; Resized (416 × 416px)	Transfer learning and ensemble learning	99.10
Chen et al., 2022	Peanuts, potatoes, apples, and 15 other crops	PVD + Own	RGB - 6033 images; Varied sizes; Resized (224×224px) RGB -54634 images; Varied sizes; Resized (224×224px)	CACPNET (channel attention and channel pruning)	99.70
Afifi et al., 2020	Coffee leaf	Own + PVD	- RGB- 54,305 images; Varied sizes	ResNet18, ResNet34, and ResNet50	99.00
Sharif et al., 2018	Citrus	Own + PVD	RGB, 580 images; Varied sizes RGB - 1000 images -Size: 100x150px	Hybrid method, scoring PCA, entropy and covariance vector given M-SVM	97.0 -PVD, 89-PVD + Own, 90.4- Own
Lu et al., 2017	Rice	PVD	RGB- 9,230 images; Varied sizes; Resized (224 × 224px)	Custom CNN	95.48

continue...

Table 1: Continuation.

Author	Crop	Dataset	Sensor, Resolution and Input image	Model	Accuracy (%)
Singh et al., 2022	Apple, corn, potato, tomato, and rice plants	PVD	RGB- 37,315 images; Varied sizes	LeNet, ShuffleNet, AlexNet, EffNet, and MobileNet,	96.10
Sladojevic et al., 2016	13 crop species	Websites	RGB- 30880- training and 2589-validation; Resized (256 × 256px)	CaffeNet	96.30
Mohanty et al., 2016	14 crop species	PVD	RGB - 54,306 images; Varied sizes; Resized (256 × 256px)	AlexNet and GoogleNet	99.34
Brahimi et al., 2017	Tomato	PVD	RGB- 14,828 images; Varied sizes; Resized (256 × 256px)	AlexNet and GoogleNet	99.18
Amara et al., 2017	Banana	PVD	RGB- 3700 images; Resized (60 x 60px)	LeNet	99.72
Too et al., 2019	14 crop species	PVD	RGB- 54,306 images; Varied sizes; Resized (224×224px)	ResNet with 50, 101, 152 Layers, VGG16, DenseNet and InceptionV4	99.75
Ferentinos, 2018	25 crop species	PVD	RGB- 87,848 images; Varied sizes; Resized (256 × 256px)	AlexNet, AlexNetOWTBn, GoogLeNet, Overfeat, VGG	99.53
Abdu et al., 2020	Potato	PVD	RGB- 1,400 images; Varied sizes; Resized (256 × 256px)	Proposed Algorithm	99.0
Karthik et al., 2020	Tomato	PVD	RGB- 95999 training images and 24001; validation images; Varied sizes	CNN	98.0
Khan et al., 2018	Citrus	CASC-IFW ³	RGB - (Exp 1: 2688 images; Exp 2: 2679 images; Exp 3, 4, 5, and 6: 1200 images); Varied sizes	VGG16, Caffe AlexNet	98.6
Kamal et al., 2019	55 crop species	PVD	RGB- 82,161 images; Varied sizes; Resized (224 x224) and (150x150)	Modified MobileNet and Reduced MobileNet	98.34

PVD¹ - Plant Village Dataset.

UCI² - University of California Irvine Machine Learning Repository.

CASC-IFW³ – Comprehensive Automation for Specialty Crops - Internal Feeding Worm Database.

The challenge in using deep learning models that work under the learn-by-example principle lies in having a sufficiently large set of sample images to compute enough features to describe the desired object. To facilitate this task, some research centres such as UCI, PVD, CASC and IFW have public repositories and allow access to databases containing samples of different objects, including leaves. This enables the conduction of several studies about the detection and classification of plant diseases based on publicly available datasets. On the other hand, some researchers use their own datasets, aiming to obtain a better description of the leaf and the specific cases they are studying. There are also other authors, such as Sharif et al. (2018), Afifi et al. (2020), Saberi Anari (2022), and Chen et al. (2022), who carried out studies using public and specific datasets for diagnosing plant diseases.

Studies that use public datasets report better results in terms of accuracy compared to those using specific datasets. For example, Sharif et al. (2018) used both types of dataset, public and specific. The same trained model

produced different results, showing greater accuracy (97%) in experiments with public datasets, less accuracy (90.4%) with specific datasets and even less accuracy (89%) with the combined use of both types of datasets. Similar accuracy values with a specific dataset were found in Kawasaki et al. (2015), Fujita et al. (2016), DeChant et al. (2017), Ramcharan et al. (2017), Picon et al. (2019), Selvaraj et al. (2019), Hu et al. (2019), Ma et al. (2018), Zhang et al. (2019), and Singh et al. (2022). The probable reason is that the images from public datasets are obtained under better conditions or otherwise that they are taken under controlled conditions, unlike the user-specific datasets.

The advantages of using a user-specific dataset are related to the possibility of developing specific solutions adapted to local climatic conditions, as well as including the knowledge of the conditions under which the images were collected. According to Dhaka et al. (2021), plants are influenced by several factors, including those associated with climatic conditions, such as temperature, humidity and precipitation, factors that contribute to the growth of bacteria, viruses, fungi, nematodes and other microorganisms. Collecting their own datasets allows researchers to increase the description of the image collection and environmental conditions and avoid unwanted environmental data that may be included in the image and may interfere with the image segmentation process (Zhang et al., 2019; Magsi et al., 2020), an important stage for classification. In our research, a specific dataset was used, and the precision values obtained in similar experiments were considered for comparison purposes.

3. Material and Methods

3.1 Data source

For the present study, images were collected using an iPhone Xr smartphone digital camera with 12 Megapixels, in uncontrolled field conditions, without flash or optical zoom, at a distance of 1.5 meters and taken under lighting conditions without a higher incidence of sunlight, allowing the identification of different pathologies in the plant leaf in a field environment. The data was collected from small gardens within the experimental field of the Federal University of Paraná's farm, located in Canguiri, where tomato cultivation is carried out in an open field, following soil irrigation and weed control practices. The images were collected during the spring, in the middle phase of plant development, approximately 7 to 8 weeks after transplanting the seedlings in the field and before the first maturation phase. This stage is notable for high leaf vigour, giving it significant importance for the analysis. For the experiments, a specific dataset composed of 2480 RGB images was created, in which 1240 images show plants with healthy leaves and 1240 images contain plants with leaf disease symptoms (Figure 1). Note that the study does not aim to perform a detailed genesis of the disease, concentrating on symptom recognition. The images were labelled based on their visual content, taking into account different characteristics representing distinctions between the classes. This manual labelling process was crucial for providing the correct labels in the training data and enabling the development of an accurate classification model. The dataset was randomly partitioned into two groups, one containing 70% of the images (1736) for training and the other containing 20% of the images (496), which were randomly selected to compose the test set. To impartially evaluate the model's performance on new examples, a separate dataset consisting of 248 images (10%), 124 from each class, was used. This test data was selected independently from the training and validation sets. This approach ensured that the model was evaluated in an unbiased manner, without being familiarized with these specific examples during training and validation.



Figure 1: Tomato crop leaf samples: (a) healthy, (b) spots with visible symptoms of a disease.

Before applying the classification models, the RGB digital images were submitted to a pre-processing step, namely: resizing to 150x150 pixels, binarization, and image noise removal. The image resizing process maintains the uniformity of images in terms of size, reduces computational costs, and improves image processing efficiency. Binarization allows for background removal, separating healthy green patches from the colour image, and maintaining the region of interest (Abdu et al., 2019). After image binarization, small spurious regions were removed by applying mathematical morphology operators, such as opening and closing.

3.2 Methods

The experiments were carried out in the cloud, within the Google Colab environment, using Python language and exploring Tensorflow and Keras API.

Deep models are composed of several layers, stacked hierarchically, to be explored for feature learning, pattern analysis and classification. Examples of deep models include Recurrent Neural Networks (RNNs), Deep Neural Networks (DNNs), and Convolutional Neural Networks (CNN). Among them, CNN is one of the most used models for image processing, particularly for the detection and identification of plant diseases (Golhani et al., 2018; Zhang et al., 2019; Abdu et al., 2020; Jaiganesh et al., 2020; Qi et al., 2021).

This paper proposes fusing two deep learning techniques for the identification and classification of plant diseases based on visible spots in leaf images. The proposed model consists of two modules, a CNN as a feature extractor and a machine learning classifier (eg. RF) for disease diagnosis. The constructed CNN was composed of 32, 64 and 128 filters defined in the 3x3 size layers using ReLu activation functions. The proposed structure uses MaxPool2D to decrease image dimensions by the pool2 size parameter, to a 128*128*3 output size. The total number of convolutional layers is 3*stacks. Each stack has a convolutional layer, a batch normalization layer, and a ReLU layer. The image size is 150*150*3, in which 150*150 is the size in pixels, and the number 3 represents the depth of the image. The hyperparameter configuration was: Learning rate: 0.00001; Number of epochs: 100; Steps per epoch: 50; Loss function: Binary cross-entropy; Activation functions: ReLU and Sigmoid. Furthermore, the CNN feature extraction was carried out and the defined model was validated.

Thus, the most prominent features produced in the previous operation are converted from a multidimensional matrix to a 1D matrix through the Flatten process. Finally, the fully connected layer is responsible for grouping all the collected information in a single descriptor that serves as input for the image classification process. The proposed

flowchart of the suggested method is summarised in Figure 2. When a set of images is presented to the algorithm, CNN extracts deep features, which are refined for disease identification and classification by an RF machine learning classifier. The output is the confirmation of the disease on the leaf and the probability of correct identification. Formally, the filtering performed on each convolutional neural layer can be written according to equations 1 and 2.

$$s = \sum_{i=1}^{nv} x_i p_i + b \quad (1)$$

$$y = f(s) \quad (2)$$

where x_i represents the value of a pixel in the neighborhood, p_i its respective weight, and b an added constant (bias).

The successful detection of significant image features is highly dependent on the proper determination of the most appropriate weights for the desired task. In a convolutional network, these weights are estimated based on training samples in a supervised phase. Therefore, weights depend on training samples.

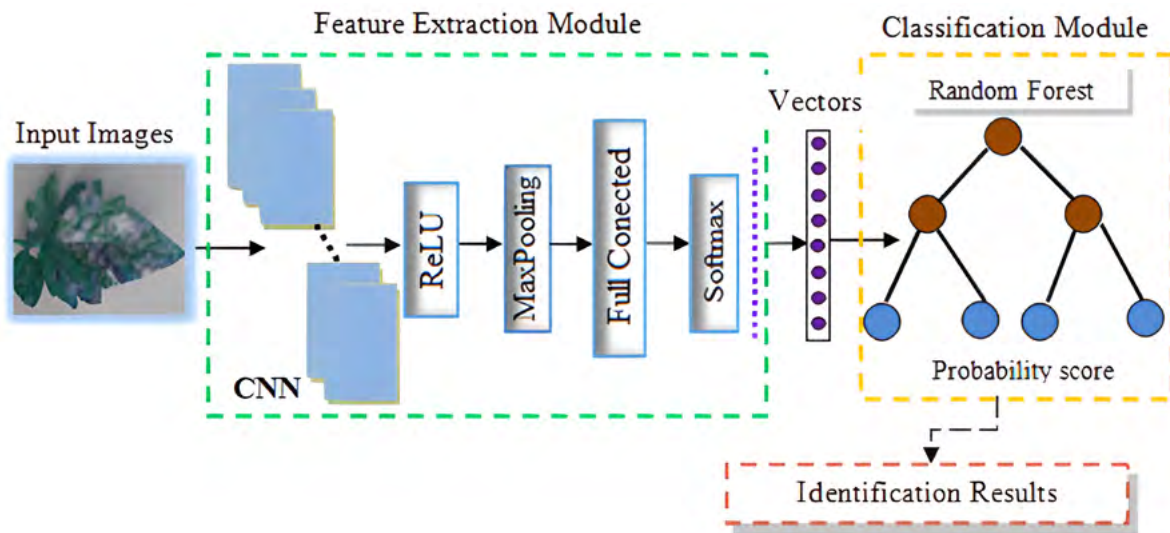


Figure 2: Flowchart of the proposed method that integrates CNNs and machine learning classifiers.

To evaluate the model's performance, the following metrics were computed: precision, recall, F-score measurement and accuracy (equations 3-6). These metrics were computed from the number of true positive (TP), true negative (TN), false positive (FP), and false-negative (FN) results.

The precision and accuracy measurements verify the classifier's capacity to replicate the same precise results in continuous iterations (Ma et al., 2018; Turkoglu and Hanbay, 2019; Abdu et al., 2020). Recall assesses the method's ability to successfully detect results classified as positive for all observations (the producer's accuracy). F-score is the harmonic mean of recall and precision (Tharwat, 2020), commonly used to evaluate binary rating systems, which rank examples as "positive" or "negative". The accuracy and F-score metrics are comprehensive indicators, so a larger value means higher accuracy, while recall and precision are related to omission and commission errors, respectively.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

$$\text{F - score} = \frac{2TP}{2TP + FP + FN} \quad (5)$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \quad (6)$$

3.3 Comparison with other classifiers

For comparison purposes, the same dataset was used as input for the six other conventional classifiers, described as follows:

Logistic Regression: a linear classifier that learns the characteristics of the sample to obtain hypothetical functions between labels, training the positive and negative examples in the data. For classification, a cost function is established, the optimisation method is iterated to obtain the optimal model parameters, and then the model is validated (Qi et al., 2021). In this case study, disease prediction usually has a finite number of outcomes, such as yes or no.

Linear Discriminant Analysis: the supervised method, related to principal component analysis (PCA) and factor analysis. The method looks for linear combinations of variables that best explain the data. It explores sample class labels to identify attribute projections that potentially maximise class discrimination (Lasisi and Attoh-Okine, 2018). The method learns a linear transformation that minimises the within-class distance and maximises the between-class discrepancy.

Naive Bayes: a family of simple probabilistic classifiers based on a common assumption that all attributes are independent of each other, given the category variable, often used as the baseline in classification (Xu, 2018). In the case of plant disease identification and classification, it is based on the conditional probability of each attribute corresponding to a given label during training, following a probabilistic independence rule to predict the class label with the highest probability.

K-nearest Neighbour: characterised as a conventional non-parametric classifier, which has been used as the baseline classifier in many pattern classification problems. This method is commonly used in machine learning, image processing and statistical estimation, classifying data based on distance metrics (e.g., Euclidean, Minkowski and Manhattan distance) from existing learning data, assigning newly input data to the cluster closest to the established sample (Hu et al., 2016; Sengur et al., 2018; Rehman, 2019; Abdu et al., 2020). The pixel-based method is used to divide the leaf image into groups of pixels, easily used to detect spots with disease symptoms.

Support Vector Machine: this method was developed by Vapnik, based on statistical estimation and the concept of decision planes (kernels), defining decision limits in a high-dimensional space for classification (Rehman, 2019; Xian and Ngadiran, 2021). This method allows for the existence of a linear discriminant function with the largest marginal, separating the classes from each other, and the model can classify linearly distinguishable and indistinguishable datasets (Turkoglu and Hanbay, 2019; Xian and Ngadiran, 2021). For the present article, the training and testing samples in the model were vectors of healthy leaves.

Random Forest: proposed by Ho (1995), it involves a set of decision trees consisting of a combination of tree classifiers in which each classifier is generated using a random vector sampled independently from the input vector. Each tree casts a unit vote for the most popular class to rank an input vector (Panchal et al., 2019; Qi et al., 2021). The training samples are randomly selected and each tree is trained using these N samples with replacement. The RF classifier used in this article consists in using randomly selected attributes or a combination of attributes at each node to grow a tree.

As these classifiers do not automatically deduce the characteristics to be used, they were deduced at an earlier stage, considering the relevant phenomena a user would consider. Therefore, different characteristics were extracted, using the following descriptors:

- i) Colour – the chlorotic region of the leaf is associated with colour change and the colour characteristics are the most intuitive and evident in the characterisation of the region with disease symptom spots on the plant leaf (Abdu et al., 2020). To identify the infected region, the input image was converted to HSV colour space so that the infected area could be easily segmented across the entire image (Ma et al., 2018; Too et al., 2019).
- ii) Texture – defined as the frequency of a pattern and colour that are visible in an image or object, such as visible disease spots on leaves (Xian and Ngadiran, 2021). Therefore, the texture is used to select more discriminating parameters. For this descriptor, the image was converted to greyscale and, through selected attributes, it is possible to quantify intuitive qualities, such as the roughness and smoothness of the infected area and the healthy area on the leaf (Zaw et al., 2018; Barbedo, 2019).
- iii) Colour Moments – used to differentiate images based on their colour features. This moment is used to measure colour similarity between images. The basis of colour moments lies in the assumption that colour distribution in an image can be interpreted as a probability distribution (Xian and Ngadiran, 2021). For the study, the weighted average of image pixel intensities was used, allowing for pixel colour location and summarising an image's colour values.
- iv) Histogram – the histogram summarises the frequency of the digital values in each colour band. Images of healthy leaves show a similar histogram, while images of infected leaves change their histogram depending on the colour of these spots (Too et al., 2019; Xian and Ngadiran, 2021). Therefore, each channel presented in an image is 8 bits in size and there are 256 possible intensities which can be displayed in the histogram (Xian and Ngadiran, 2021).

After using the morphological operations to extract features from the input images, the different extracted shape features were combined to create global feature vectors and then the classification process was performed. As such, this study proposes an alternative classification method that integrates deep learning convolutional neural networks (CNNs) and machine learning classifiers with the purpose of overcoming the limitations of CNN and machine learning in the classification of real data for plant disease diagnosis, thus improving its accuracy. The new classifier was named CNNRF and comprises two approaches: feature extraction and image classification.

DL methods require a larger amount of data. This is a drawback since, in agriculture, available datasets are usually small and do not contain enough images, which is a necessity for high-quality decisions. A comprehensive dataset must contain images captured in as many different conditions as possible. Currently, available solutions with DL methods for plant disease detection have been somewhat successful; however, there is still much room for improvement. There are several current limitations in this research field. One of them is that currently available datasets do not contain images gathered and labelled from real-life situations. Therefore, training is conducted with images taken in a controlled environment.

Geolocation and mapping - the previous steps describe the disease detection steps, based on RGB images obtained with low-cost mobile phone cameras. Once each image is analysed, and disease symptoms are detected, it is possible to estimate the plant's location based on the information contained in the camera and stored in the EXIF (Exchangeable Image File Format) file. The EXIF information is stored along with the image and was developed to record information about the technical conditions of image capture (such as the focal length of aperture), along with other ancillary tagged metadata. If a mobile phone has a built-in GPS device, the EXIF format has standard tags for location information that can be easily read from the camera. Thus, the collected data not only includes the image, but also the location where the disease was detected. This information will be used to compute the spatial frequency of the symptomatic plants and produce a map to help farmers control the disease. This part of the study is not described here, as it is the central theme of a new paper.

4. Results

In this section, the proposed method's performance is presented and discussed. To demonstrate the proposed method's effectiveness, it was compared to other conventional machine learning methods, using the specific dataset composed of RGB images of healthy/infected leaves. Therefore, to train the model, 100 epochs (a hyperparameter defining the number of times the learning algorithm will train and start over) were defined. The Early Stopping technique was adopted to obtain ideal generalisation performance.

The proposed model (CNNRF) achieved 96% precision and 97.8% accuracy for leaf disease identification activity. This means that if 100 images are presented to the model, approximately 96 images will be correctly classified. Nevertheless, despite the model presenting consistent scores, some inaccuracies occurred in the process of identifying leaf diseases, which may be related to: i) the size of the training sample; ii) different leaf orientations; iii) lighting, related to the different weather conditions at the time of image collection and; iv) the appearance of shadows on the leaves. These factors may have made it difficult for the proposed model to properly identify the characteristics of interest in the images. In this case, the highest error rate found was 18.3%, a rate that can be explained by the fact that there are images with disease spots that the model confused among the experimental data, as 9 infected leaves were classified as healthy and 5 healthy leaves were considered infected. The last scenario is the case in which the images of the leaves had spots that were not easily distinguishable, increasing the complexity of the classification task. The confusion matrix for a binary classifier shows that from a total of 1240 healthy leaves, 46 leaves were misclassified. In the 1240 leaves with symptoms, 53 leaves were wrongly classified as healthy leaves.

Figure 3 shows the model's results in the prediction stage of the plant leaf-based disease identification process. For example, the first two leaves a) and b) were identified with higher precision scores, while the last two leaves c) and d) demonstrate the imprecision in leaf disease identification.

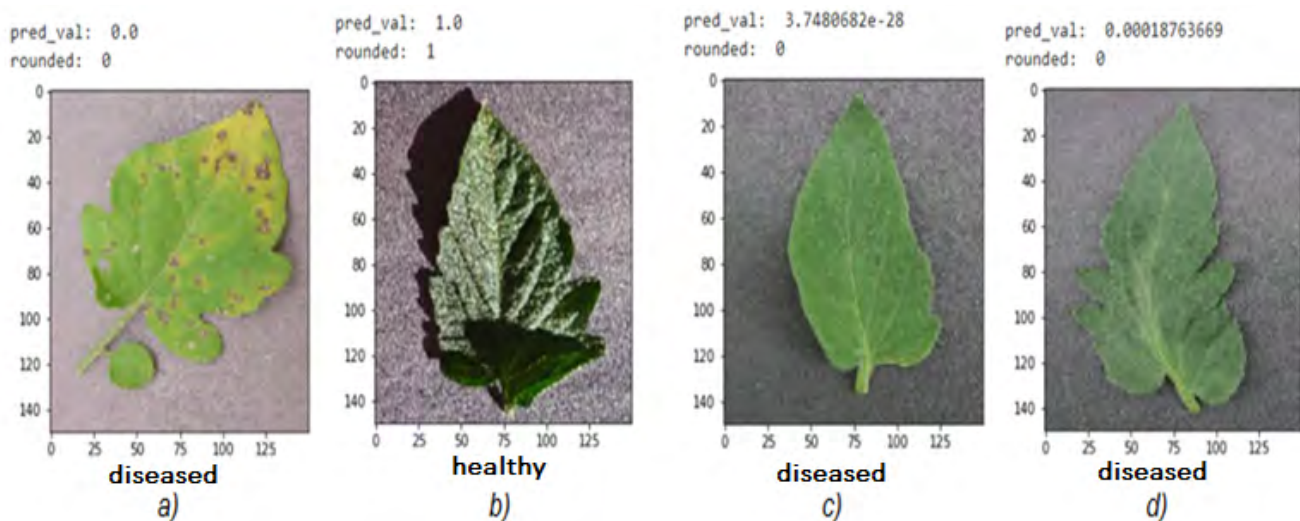


Figure 3: Precision and imprecision of leaf disease predictions: a) and b) were accurately identified and c) and d) imprecision in disease identification.

Figure 4 displays the training accuracy and validation generated by the proposed model. The figure shows a substantial reduction of the model entropy's logarithmic loss after the 60th training iteration. From the 70th epoch onwards, the model tends to converge. Therefore, when the training dataset and epochs are increased, the accuracy also increases.

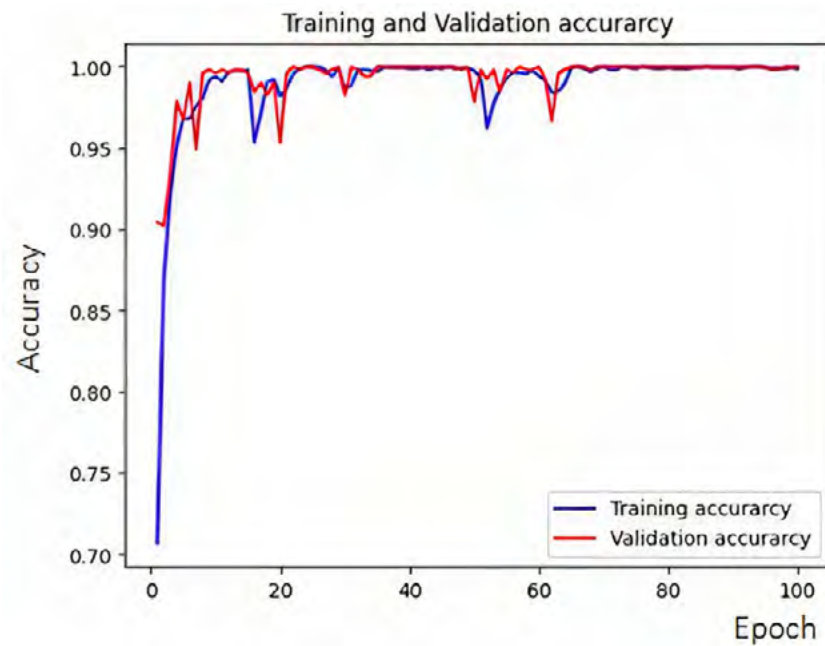


Figure 4: Training and Validation Accuracy.

The training and validation losses are illustrated in Figure 5. These losses are calculated based on the errors computed in the training dataset, that is, the computation is performed based on evaluating the trained model exclusively with the training dataset, in contrast with the precision metric. The loss is the sum of the discrepancies between the model's predictions and the real values for each example in the training and validation sets. Thus, the loss value implies how well or poorly a given model performs after each optimisation iteration. If the training loss is less than the validation loss, this suggests the model is overfitting, as observed by Jaiganesh et al (2020).

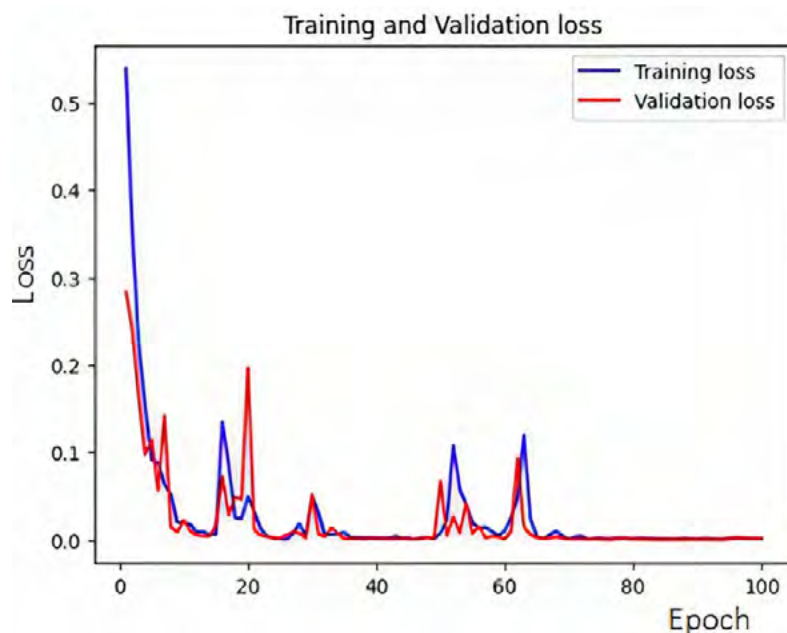


Figure 5: Training and validation Loss.

To access the proposed model's efficiency, the same dataset was used to train machine-learning algorithms commonly used for digital image processing, going through the attribute extraction process for further classification. Figure 6 shows the result of the texture segmentation process for visible disease spots on the tomato crop leaf using K-means clustering applied to the input image. For image segmentation, which is a process for extracting attributes, global descriptors such as colour, texture, colour moments and colour histogram were used. These descriptors permitted the conservation of healthy leaf information in segmented disease points.

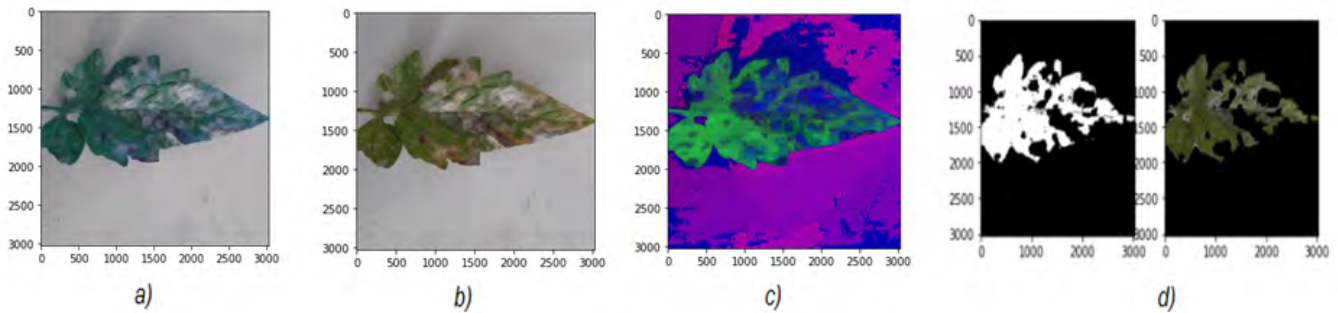


Figure 6: Segmentation results - a) Original image of tomato leaf; b) Enhanced RGB image; c) Image converted to HSV color; d) Segmented image - infected areas.

To facilitate data analysis, a box diagram (Figure 7) was constructed, which permitted a comparison of the different machine learning algorithms concerning position, dispersion, symmetry and outliers.

In the diagram, the LR and RF algorithms indicate a low variability and standard deviation, unlike the LDA, KNN, NB and SVM algorithms, which present relatively higher variability and standard deviation. Algorithms with higher variability indicate that the predictability of disease identification and classification based on plant leaves is relatively lower compared to those with low variability and lower standard deviation. Some classifiers have outliers, with values that are distant from the mean value, which is relatively low, with a higher emphasis on the SVM classifier, which influences the prediction process.

From the accuracy point of view, the best results were obtained with the RF classifier and the worst results were found with the SVM method. This justified the use of the RF method for classification based on characteristics derived from the convolutional network.

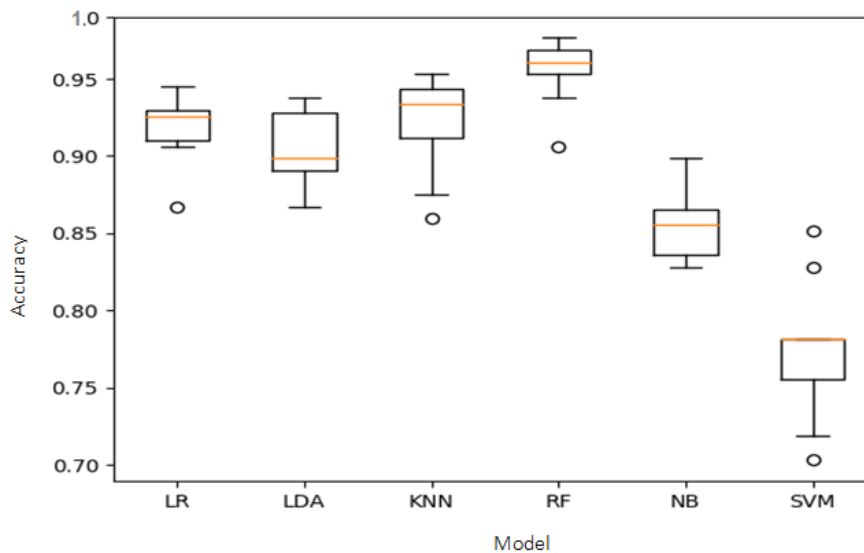


Figure 7: Boxplot of classifiers.

Table 2 shows the values of the performance parameters computed from the classifications using all six methods, including the proposed hybrid model, based on the collected dataset.

Table 2: Comparison of the proposed model with other classifiers.

Model	Precision	Recall	F-Score	Accuracy
CNNRF	0.96	0.93	0.93	0.98
RF	0.94	0.92	0.93	0.93
LR	0.93	0.85	0.89	0.89
CNN	0.86	0.81	0.81	0.89
LDA	0.90	0.84	0.87	0.88
KNN	0.90	0.86	0.88	0.88
NB	0.91	0.63	0.75	0.78
SVM	0.87	0.52	0.65	0.71

The proposed model achieved an accuracy score of 98%, higher than the RF method, which obtained a value of 93%. Other methods (LR, CNN, LDA and KNN) produced similar results, around 89%. The worst results are attributed to the NB and SVM methods (78% and 71%, respectively). By comparison, the proposed model, i.e. the combination of the CNN deep learning network as a feature extractor and the RF machine learning classifier, improves the accuracy score considerably. A similar tendency was found when analysing and comparing the F-score values. However, in this case, the proposed method produces equivalent results to the Random Forest classification (93%). Again, the worst results were produced with NB and SVM classifiers. The differences can be attributed to how these parameters are calculated. The recall and precision values are above 93%.

4.1 Discussion

Based on the study, it is possible to say that the proposed method performed relatively better in the task of identifying plant diseases compared to other methods. Competing results can be found in the models developed by Kawasaki et al. (2015), Fujita et al. (2016), DeChant et al. (2017), Ramcharan et al. (2017), Sharif et al. (2018), Selvaraj et al. (2019), Hu et al. (2019), Ma et al. (2018), Zhang et al. (2019), and Picon et al. (2019). It is important to emphasise that the deep models developed by these authors also used specific datasets.

Therefore, compared to state-of-the-art-methods, the proposed model brings about a slight improvement, as the results it produced are, on average, 4.8% more accurate. The smallest difference is 2.9% and corresponds to the method used by Kawasaki et al. (2015), and the highest difference of 14.6% is connected to the method used by Fujita et al. (2016). Both studies developed deep models for disease identification in cucumber cultures.

Also based on the literature review, a marked difference was found in terms of precision and accuracy between studies that used public datasets and specific datasets, with higher scores found by those using public data and lower scores being obtained by the ones based on specific data. This trend was not confirmed in the present study, in which the results surpassed those found by Sladojevic et al. (2016) and Sharif et al. (2018), who obtained lower scores than those achieved with the present methodology.

However, some of the differences may be associated with the training sample size, image orientation, the difficulties of the method used for extracting features, and factors that increase the classification process's complexity. Sladojevic et al. (2016) implemented a deep model for the recognition of 13 types of plant diseases based on leaf image classification, using deep convolutional neural networks, and the experiment presented in the present study is restricted to leaf disease identification. Sharif et al. (2018), on the other hand, proposed a hybrid method for the detection and classification of diseases in citrus plants, considering fruits and leaves, and in the present study, only the leaf part was analysed. However, the study presented by Saberi Anari (2022), proposed a hybrid model for foliar disease classification based on modified deep transfer learning and ensemble approach for agriculture-based monitoring. This model used deep neural networks as well as the PlantVillage and UCI databases with approximately 90,000 images, obtaining the best model performance.

The results from the methodology adopted in this pepper, which includes the combination of the CNN and RF models to improve the detection of diseases in plants through the analysis of leaf spots, present promising perspectives for their extension to the identification of diseases in other crops, when compared with other studies.

5. Conclusions

In the study, a hybrid model was proposed, combining a CNN as a feature extractor and the Random Forest decision tree classifier to identify powdery mildew fungus, which attacks the tomato leaf. The results show that the proposed hybrid model outperforms conventional CNN and RF by 8.8% and 4.8%, respectively. The hybrid model achieved an accuracy of 98% in our two-class balanced dataset collected in real field conditions, under different levels of light exposure. The results indicate an improvement in accuracy for automatic plant disease identification using low-cost optical images, contributing to the process of plant disease identification. The hybrid model helps to increase performance in terms of accuracy for plant disease identification problems. Thus, in the two classes of interest within this dataset, the proposed hybrid model produced a better (i.e., yielding higher classification accuracy) performance, than using one of the models separately.

The methodology performance in terms of overall accuracy was compared with six state-of-the-art methods, i.e., Naive Bayes, Random Forest, Artificial Neural Networks, Support Vector Machines and decision trees. It was

demonstrated that the proposed methodology outperforms all the compared methods and that even with a reduction of the training set, the average accuracy of the presented method is above 93%.

Conventional artificial intelligence classifiers, commonly used in digital image processing, allowed the conservation of healthy leaf information in segmented disease points, some with higher relative predictability for the identification and classification of diseases visible in the plant leaf. Therefore, compared to other disease identification methods based on artificial intelligence, the proposed method is very close to the one based on Random Forest, even while using different characteristics as input. This indicates that convolutional networks were able to detect useful classification features and can compensate for the proposed feature-based analysis based on knowledge. On the other hand, the proposed method showed, in some situations, the ability to overcome other similar approaches, especially if precision and recall rates are considered.

The results presented in this study can be improved either by improving the proposed model by increasing the number of training sets, increasing the number of layers in the model or merging with more than one classifier, or by testing the same model on leaves from different plants in a different stage of infection and/or with a different leaf size. Another improvement could occur in the pre-processing phase, based on the choice of better spectral and spatial descriptors, which would facilitate the discrimination of disease spots in the leaf image. However, the results prove that the success rate is considerable. It is also expected that better results can be obtained if the images are acquired with better lighting conditions and less shadow; however, this would diminish the method's applicability under normal conditions, in the field.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ACKNOWLEDGEMENT

We extend our thanks to the National Council for Scientific and Technological Development (CNPq) and The World Academy of Sciences (TWAS) for the scholarship [Process: 167085/2018-2]. We would also like to thank the *Postgraduate* Program in Geodetic Sciences at the Federal University of Paraná for their support.

AUTHOR'S CONTRIBUTION

Author 1: designed and implemented the algorithm, data collection, conducted experimental study and wrote the paper; Author 2: study conception, formal analysis, development of research methods and critical review; Author 3: Substantial contributions to the conception and design of the manuscript, analysis, and interpretation of the results; Author 4: implemented the algorithm and critical review; Author 5: data collection, analysis and final revision.

REFERENCES

- Abdu, A. M., Mokji, M. M., & Sheikh, U. U. (2020). Automatic vegetable disease identification approach using individual lesion features. *Computers and Electronics in Agriculture*, *176*, 105660.
- Abdu, A. M., Mokji, M. M., Sheikh, U. U., & Khalil, K. (2019, March). Automatic disease symptoms segmentation optimized for dissimilarity feature extraction in digital photographs of plant leaves. In *2019 IEEE 15th International Colloquium on Signal Processing & Its Applications (CSPA)* (pp. 60-64). IEEE.
- Afifi, A., Alhumam, A., & Abdelwahab, A. (2020). Convolutional neural network for automatic identification of plant diseases with limited data. *Plants*, *10*(1), 28.
- Amara, J., Bouaziz, B., & Algergawy, A. (2017). A deep learning-based approach for banana leaf diseases classification. *Datenbanksysteme für Business, Technologie und Web (BTW 2017)-Workshopband*.
- Barbedo, J. G. A. (2019). Plant disease identification from individual lesions and spots using deep learning. *Biosystems engineering*, *180*, 96-107.
- Bedi, P., & Gole, P. (2021). Plant disease detection using hybrid model based on convolutional autoencoder and convolutional neural network. *Artificial Intelligence in Agriculture*, *5*, 90-101.
- Boulent, J., Foucher, S., Théau, J., & St-Charles, P. L. (2019). Convolutional neural networks for the automatic identification of plant diseases. *Frontiers in plant science*, *10*, 941.
- Brahimi, M., Boukhalifa, K., & Moussaoui, A. (2017). Deep learning for tomato diseases: classification and symptoms visualization. *Applied Artificial Intelligence*, *31*(4), 299-315.
- Chen, R., Qi, H., Liang, Y., & Yang, M. (2022). Identification of plant leaf diseases by deep learning based on channel attention and channel pruning. *Frontiers in Plant Science*, *13*, 1023515.
- DeChant, C., Wiesner-Hanks, T., Chen, S., Stewart, E. L., Yosinski, J., Gore, M. A., ... & Lipson, H. (2017). Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning. *Phytopathology*, *107*(11), 1426-1432.
- Dhaka, V. S., Meena, S. V., Rani, G., Sinwar, D., Ijaz, M. F., & Woźniak, M. (2021). A survey of deep convolutional neural networks applied for prediction of plant leaf diseases. *Sensors*, *21*(14), 4749.
- Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and electronics in agriculture*, *145*, 311-318.
- Fujita, E., Kawasaki, Y., Uga, H., Kagiwada, S., & Iyatomi, H. (2016, December). Basic investigation on a robust and practical plant diagnostic system. In *2016 15th IEEE international conference on machine learning and applications (ICMLA)* (pp. 989-992). IEEE.
- Golhani, K., Balasundram, S. K., Vadamalai, G., & Pradhan, B. (2018). A review of neural networks in plant disease detection using hyperspectral data. *Information Processing in Agriculture*, *5*(3), 354-371.
- Ho, T. K. (1995, August). Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition* (Vol. 1, pp. 278-282). IEEE.
- Hu, F., Xia, G. S., Hu, J., & Zhang, L. (2015). Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sensing*, *7*(11), 14680-14707.
- Hu, G., Wu, H., Zhang, Y., & Wan, M. (2019). A low shot learning method for tea leaf's disease identification. *Computers and Electronics in Agriculture*, *163*, 104852.
- Hu, L. Y., Huang, M. W., Ke, S. W., & Tsai, C. F. (2016). The distance function effect on k-nearest neighbor classification for medical datasets. *SpringerPlus*, *5*(1), 1-9.

- Jaiganesh, M., Sathyadevi, M., Chakravarthy, K. S., & Sarada, C. (2020). Identification of plant species using CNN-classifier. *Journal Of Critical Reviews*, 7(3), 923-931.
- Kamal, K. C., Yin, Z., Wu, M., & Wu, Z. (2019). Depthwise separable convolution architectures for plant disease classification. *Computers and electronics in agriculture*, 165, 104948.
- Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). A review of the use of convolutional neural networks in agriculture. *The Journal of Agricultural Science*, 156(3), 312-322.
- Karthik, R., Hariharan, M., Anand, S., Mathikshara, P., Johnson, A., & Menaka, R. (2020). Attention embedded residual CNN for disease detection in tomato leaves. *Applied Soft Computing*, 86, 105933.
- Kaur, P., Harnal, S., Tiwari, R., Upadhyay, S., Bhatia, S., Mashat, A., & Alabdali, A. M. (2022). Recognition of leaf disease using hybrid convolutional neural network by applying feature reduction. *Sensors*, 22(2), 575.
- Kawasaki, Y., Uga, H., Kagiwada, S., & Iyatomi, H. (2015). Basic study of automated diagnosis of viral plant diseases using convolutional neural networks. In *Advances in Visual Computing: 11th International Symposium, ISVC 2015, Las Vegas, NV, USA, December 14-16, 2015, Proceedings, Part II 11* (pp. 638-645). Springer International Publishing.
- Khan, M. A., Akram, T., Sharif, M., Awais, M., Javed, K., Ali, H., & Saba, T. (2018). CCDF: Automatic system for segmentation and recognition of fruit crops diseases based on correlation coefficient and deep CNN features. *Computers and electronics in agriculture*, 155, 220-236.
- Kuswidiyanto, L. W., Noh, H. H., & Han, X. (2022). Plant Disease Diagnosis Using Deep Learning Based on Aerial Hyperspectral Images: A Review. *Remote Sensing*, 14(23), 6031.
- Lasisi, A., & Attoh-Okine, N. (2018). Principal components analysis and track quality index: A machine learning approach. *Transportation Research Part C: Emerging Technologies*, 91, 230-248.
- Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, 18(8), 2674.
- Liu, J., & Wang, X. (2021). Plant diseases and pests detection based on deep learning: a review. *Plant Methods*, 17, 1-18.
- Lu, J., Hu, J., Zhao, G., Mei, F., & Zhang, C. (2017). An in-field automatic wheat disease diagnosis system. *Computers and electronics in agriculture*, 142, 369-379.
- Ma, J., Du, K., Zheng, F., Zhang, L., Gong, Z., & Sun, Z. (2018). A recognition method for cucumber diseases using leaf symptom images based on deep convolutional neural network. *Computers and electronics in agriculture*, 154, 18-24.
- Magsi, A., Mahar, J. A., Razzaq, M. A., & Gill, S. H. (2020, November). Date palm disease identification using features extraction and deep learning approach. In *2020 IEEE 23rd International Multitopic Conference (INMIC)* (pp. 1-6). IEEE.
- Mahlein, A. K. (2016). Plant disease detection by imaging sensors—parallels and specific demands for precision agriculture and plant phenotyping. *Plant disease*, 100(2), 241-251.
- Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in plant science*, 7, 1419.
- Moore, K., & Bradley, L. K. (Eds.). (2018). *North Carolina Extension gardener handbook*. NC State Extension, College of Agriculture and Life Sciences, NC State University.
- Panchal, P., Raman, V. C., & Mantri, S. (2019, December). Plant diseases detection and classification using machine learning models. In *2019 4th international conference on computational systems and information Technology for Sustainable Solution (CSITSS)* (pp. 1-6). IEEE.
- Picon, A., Alvarez-Gila, A., Seitz, M., Ortiz-Barredo, A., Echazarra, J., & Johannes, A. (2019). Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild. *Computers and Electronics in Agriculture*, 161, 280-290.
- Qi, H., Liang, Y., Ding, Q., & Zou, J. (2021). Automatic identification of peanut-leaf diseases based on stack ensemble. *Applied Sciences*, 11(4), 1950.

- Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J., & Hughes, D. P. (2017). Deep learning for image-based cassava disease detection. *Frontiers in plant science*, 8, 1852.
- Rehman, T. U., Mahmud, M. S., Chang, Y. K., Jin, J., & Shin, J. (2019). Current and future applications of statistical machine learning algorithms for agricultural machine vision systems. *Computers and electronics in agriculture*, 156, 585-605.
- Rezk, N. G., Attia, A. F., El-Rashidy, M. A., El-Sayed, A., & Hemdan, E. E. D. (2022). An Efficient Plant Disease Recognition System Using Hybrid Convolutional Neural Networks (CNNs) and Conditional Random Fields (CRFs) for Smart IoT Applications in Agriculture. *International Journal of Computational Intelligence Systems*, 15(1), 65.
- Saberi Anari, M. (2022). A hybrid model for leaf diseases classification based on the modified deep transfer learning and ensemble approach for agricultural aiot-based monitoring. *Computational Intelligence and Neuroscience*, 2022.
- Selvaraj, M. G., Vergara, A., Ruiz, H., Safari, N., Elayabalan, S., Ocimati, W., & Blomme, G. (2019). AI-powered banana diseases and pest detection. *Plant methods*, 15, 1-11.
- Şengür, A., Akılotu, B. N., Tuncer, S. A., Kadiroğlu, Z., Yavuzkılıç, S., Budak, Ü., & Deniz, E. (2018, May). Optic disc determination in retinal images with deep features. In *2018 26th Signal Processing and Communications Applications Conference (SIU)* (pp. 1-4). IEEE.
- Sharif, M., Khan, M. A., Iqbal, Z., Azam, M. F., Lali, M. I. U., & Javed, M. Y. (2018). Detection and classification of citrus diseases in agriculture based on optimized weighted segmentation and feature selection. *Computers and electronics in agriculture*, 150, 220-234.
- Singh, A. K., Sreenivasu, S. V. N., Mahalaxmi, U. S. B. K., Sharma, H., Patil, D. D., & Asenso, E. (2022). Hybrid feature-based disease detection in plant leaf using convolutional neural network, bayesian optimized SVM, and random forest classifier. *Journal of Food Quality*, 2022, 1-16.
- Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., & Stefanovic, D. (2016). Deep neural networks-based recognition of plant diseases by leaf image classification. *Computational intelligence and neuroscience*, 2016.
- Tharwat, A. (2020). Classification assessment methods. *Applied computing and informatics*, 17(1), 168-192.
- Too, E. C., Yujian, L., Njuki, S., & Yingchun, L. (2019). A comparative study of fine-tuning deep learning models for plant disease identification. *Computers and Electronics in Agriculture*, 161, 272-279.
- Tuncer, A. (2021). Cost-optimized hybrid convolutional neural networks for detection of plant leaf diseases. *Journal of Ambient Intelligence and Humanized Computing*, 12(8), 8625-8636.
- Türkoğlu, M., & Hanbay, D. (2019). Plant disease and pest detection using deep learning-based features. *Turkish Journal of Electrical Engineering and Computer Sciences*, 27(3), 1636-1651.
- Xian, T. S., & Ngadiran, R. (2021, July). Plant diseases classification using machine learning. In *Journal of Physics: Conference Series* (Vol. 1962, No. 1, p. 012024). IOP Publishing.
- Xu, S. (2018). Bayesian Naïve Bayes classifiers to text classification. *Journal of Information Science*, 44(1), 48-59.
- Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, 9, 611-629.
- Zaw, K. K., Myo, Z. M. M., & Thoung, D. T. H. (2018). Support vector machine based classification of leaf diseases. *Int. J. Sci. Eng. Appl.*, 7, 143-147.
- Zhang, S., Zhang, S., Zhang, C., Wang, X., & Shi, Y. (2019). Cucumber leaf disease identification with global pooling dilated convolutional neural network. *Computers and Electronics in Agriculture*, 162, 422-430.
- Zhang, X., Yao, L., Wang, X., Monaghan, J., Mcalpine, D., & Zhang, Y. (2021). A survey on deep learning-based non-invasive brain signals: recent advances and new frontiers. *Journal of neural engineering*, 18(3), 031002.