



Optimization techniques in deep convolutional neuronal networks applied to olive diseases classification

El Mehdi Raouhi ^{a,*}, Mohamed Lachgar ^a, Hamid Hrimech ^b, Ali Kartit ^a

^a LTI Laboratory, ENSA, Chouaib Doukkali University of El Jadida, El Jadida, Morocco

^b LAMSAD Laboratory, ENSA, Hassan First University, Berrechid, Morocco



ARTICLE INFO

Article history:

Received 24 February 2022

Received in revised form 29 May 2022

Accepted 1 June 2022

Available online 7 June 2022

Keywords:

Convolutional neuronal networks (CNN)

Classification

Optimization

Gradient descent

Plant diseases

Olive dataset diseases (ODD)

ABSTRACT

Plants diseases have a detrimental effect on the quality but also on the quantity of agricultural production. However, the prediction of these diseases is proving the effect on crop quality and on reducing the risk of production losses. Indeed, the detection of plant diseases -either with a naked eye or using traditional methods- is largely a cumbersome process in terms of time, availability and results with a high-risk error. The present work introduces a depth study of various CNN architectures with different optimization algorithms carried out for olive disease detection using classification techniques that recommend the best model for constructing an effective disease detector. This study presents a dataset of 5571 olive leaf images collected manually on real conditions from different regions of Morocco, that also includes healthy class to detect olive diseases. Further, one of the goals of this research was to study the correlation effects between CNN architectures and optimization algorithms evaluated by the accuracy and other performance metrics. The highest rate in trained models was 100 %, while the highest rate in experiments without data augmentation was 92.59 %. Another subject of this study is the influence of the optimization algorithms on neuronal network performance. As a result of the experiments carried out, the MobileNet architecture using Rmsprop algorithms outperformed the others combinations in terms of performance and efficiency of disease detector.

© 2022 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

The olive tree is considered as the most cultivated and oldest domesticated fruit trees in the world since old ages (Chliyeh et al., 2014). It is between the scopes 30°–45° in the northern and southern halves of the globe in climatic locales of Mediterranean sort. In this manner, 97% of the world's horticultural region (roughly 10.5 million hectares) is gathered in the Mediterranean basin. Morocco occupies the fourth spot behind Spain, Italy and Greece, the olive-developing region is around 784,000 ha, with a production of 1,483,510 tons of olives each year. Likewise, it serves to contribute effectively in the development of the country's populace by producing in excess of thousand million working days (Vega-Márquez et al., 2020). 5.6% of the region's general surface (PMMA, 2011) spread more than three fundamental regions: The Rif (Taounate, Chefchaoune), the middle (Fes, Meknes, Taza) and the south (Haouz, Tadla and the seaside district among Safi and Essaouira). Lamentably, the olive is exposed to the assault of different micro-organisms, which influences its wellbeing, its yield and the nature of its oil. Either building new olive tree manors, or the effect of water

system on the event of plant infections and the passing of youthful trees have equally expanded, disturbing olive producers (Abade et al., 2021). Otherwise, the exponential increase in the global population is a fact and it tends towards 9.5 billion by 2025, the need of the populations for food will be much higher in the future. Losses of agricultural land (Uguz and Uysal, 2020), climate change and prediction limits for plant diseases mainly explain why the demand for food cannot be met. The challenge on plant pests presents one of the most interesting fields of research in agriculture in terms of difficulty and risks associated with this field. It appears from literature studies that significant increases in yield have been detected in many agricultural products as a result of the control of numerous plant pests (Liu and Wang, 2021). Some of the causes of plant diseases include living agents such as fungi, bacteria and viruses, as well as environmental factors such as bad weather and burning chemicals (Moorthy et al., 2020; Esgario et al., 2019; Uguz and Uysal, 2020; Gavhale and Gawande, 2019). To respond to these challenges. The current review was done in Morocco to add to the identification disease (ID) of the sicknesses which influence the olive tree, through the constitution of a bunch of information gathered under genuine conditions on various districts of the nation and for various illnesses which influence plants and particularly the olive tree, inside and outside learning regions introduce themselves as a basic

* Corresponding author.

E-mail address: raouhi@gmail.com (E.M. Raouhi).

answer to distinguish the discovery and spread of olive tree infections, early ID in the field is a vital initial step (Foyal et al., 2019). This normally prompts moving toward the idea of Deep Learning, which is essential for AI that trains preparing information utilizing network layers with numerous neurons (Hussain et al., 2018), by this CNNs are the best learning calculations for understanding picture content and shown high performance in undertakings identified with picture division, order, recognition and extraction. The accomplishment of CNNs has drawn consideration past scholarly community. In industry, organizations, for example, Google, Microsoft and Facebook have created dynamic examination gatherings to investigate new CNN model structures. Today, a large portion of the pioneers of picture handling and PC vision (CV) contests use CNN-based models (Pantazi et al., 2020). So, the main areas of this work can be summarized as follows. First, the presentation of CNN architectures and plant diseases used mainly in the literature for olive plant diseases classification. Second, this work presents an exploratory analysis with performance analysis simulation of the assembled dataset for olive tree diseases presented previously. In addition, the fundamental aim of this review is to distinguish the side effects of classes of six olive plant leaves illnesses generally seen. The dataset is composed with 5571 images assembled to demonstrate the performance of the system detector as well as m a benchmark for relative works. Research collected have demonstrate that olive plant leaves diseases can be detected with high performance without having to ask for the help of an expert in the field. with the aim of improving classification performance. This work seeks to answer the following questions:

- What's the effects of CNN Model on the performance of olive disease classification? (How performant is the CNN model in classifying olive diseases?)
- What's the effects of optimization on the performance of olive disease classification? (What optimization algorithm gives the best performance in classifying olive diseases?)
- Is there a correlation between CNN Model used for classification and algorithm used to optimize loss function? (Is there a correlation between the CNN model as a classifier and the various loss function optimization algorithms selected?)

The structure of the remainder of this paper is organized as follows. Firstly, Section 2 presents Materials and Methods introduced by the background and related works then describes the simulation workflow processed. Secondly, Sections 3 and 4 covers the experimental setup detailed by the result of simulations and discussions. Finally, Section 4 states some conclusions and some perspectives opening for future work.

2. Materials and methods

2.1. Convolution neuronal network

From 1989 to nowadays, various technological feats in the engineering of neural networks have been demonstrated (Pedrycz, 2020). These innovations can be ordered in the form of improved classification and regularization constraints. In a sense of continuous performance enhancement of its intelligent systems and according to the elaboration of the use cases.

2.1.1. AlexNet

In comparison to conventional techniques this architecture was one of the main deep structures to evolve the performance of ImageNet classification by a crucial step, the AlexNet introduced a new engineering in the works of the works (Alruwaili et al., 2019), using the rectified linear unit ReLu for the nonlinear fraction, instead of a tangent hyperbolic Tanh or Sigmoid function, which before integrated classical neural networks. ReLu is given by:

$$f(x) = \max(0, x) \quad (1)$$

The strong point of the function of ReLu compared to the sigmoid function is the computation time justified by the fact that the function needs to choose $\max(0, x)$ instead of performing exponential operations in the execution time which is more expensive, sigmoid turns out to be tiny in the immersion area and therefore weight updates almost disappear. This is labeled the vanishing gradient problem.

Fig. 1 shows a basic layout of AlexNet architecture showing its five convolution layers and three fully connected layers.

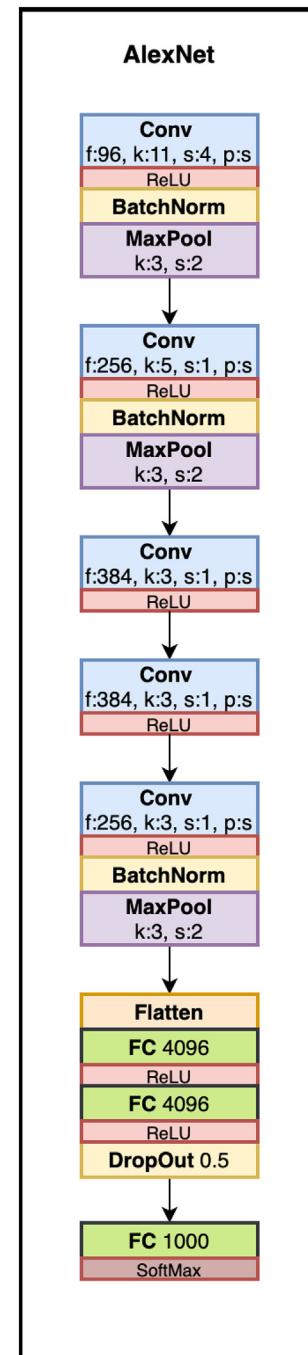


Fig. 1. Classical structure of AlexNet deep neural network.

2.1.2. VGG

Architectural design research has increased following the success of CNN networks especially in image recognition tasks. In this view the authors of this work presented in [Pedrycz \(2020\)](#), which recommends a simple and powerful plan directive for CNN structures. Their layered presentation, called VGG, was separate. Unlike AlexNet, VGG relies on 16 or 19 deep layers to draw inspiration from the relationship of depth with the capacity to represent the composition of the layers ([Uguz and Uysal, 2020](#)). [Fig. 2](#) shows a basic layout of VGG architecture.

2.1.3. GoogleNet

The GoogleNet is introduced to challenge a simple objective in order to achieve the best performance with the lowest cost ([Pedrycz, 2020](#)). This was made possible through the revolutionary idea of the beginning block in CNN, it joins convolutional modifications at different scales using the concept of division, change and union. In GoogleNet, conventional convolutional layers are exchanged by small squares as underlying to each layer with miniature neural networks as presented in [Fig. 3](#).

2.1.4. Residual network

ResNet redesigned the engineering of CNN by introducing the idea of continuing to develop in CNNs for the preparation of deep organizations. Taking the example of road networks, each of the past

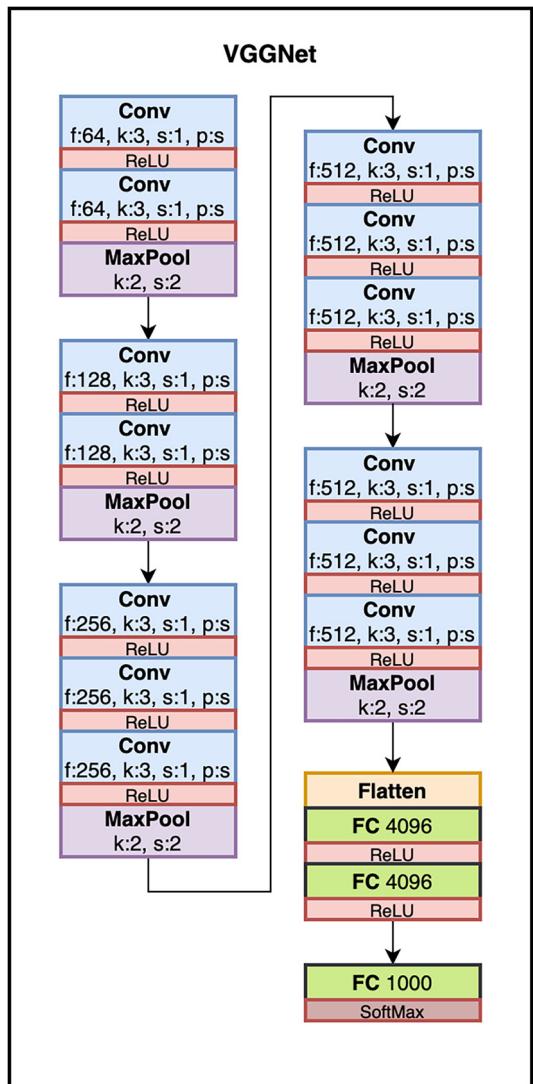


Fig. 2. Classical structure of VGG deep neural network.

models used deep neural configurations in which they coherently superimposed several layers of convolution. The authors in [Pedrycz \(2020\)](#) found that deeper configurations are more efficient. In order to overcome these constraints, the authors of the ResNet architecture presented in [Fig. 4](#) introduced jump associations with the speculation that the deeper layers will allow learning similar to that achieved by the shallower layers.

2.1.5. CNN characteristics

Deep CNN performs well in time series data or in grid type topology. Notwithstanding, there are other challenges, among which a deep CNN architecture has been applied. The fundamental difficulties related with the distinctive CNN models are referenced on [Table 1](#) with a significant strength and prattles.

Deep CNN works well in time series data or in grid type topology. However, there are other challenges, among which a deep CNN architecture has been applied. The main challenges associated with the different CNN architectures are mentioned on [Table 1](#) with major strength and gabs.

2.2. Olive diseases

More than 95% of the 750 million hectares are cultivated in the Mediterranean region. it is important to note that various factors can impact this crop including insects, nematodes and pathogens, the latter causing serious damage to the olive crop production overall the European Union ([Pedrycz, 2020](#)). Commercial operations of goods and people, climate change and changes in agricultural practices have favored the introduction, spread and establishment of certain diseases in olive production ([Sinha and Shekhawat, 2020](#)). However, many pathogenic and pest factors negatively impact the yields of olive trees ([Moorthy et al., 2020](#); [Chliyeh et al., 2014](#)) [Fig. 5](#). Diseases such as:

2.2.1. Black olive disease

The Black Olive disease presented in [Fig. 6](#) is characterized as a fungus spread by mealybugs. The leaves are overcast by black dust and block the tree's breathing process. Laying females have between 1000 and 2000 eggs under their shells. There are two types of impact, direct impacts on the leaves and twigs (by extracting the sap), causing the leaves to drop, weakening the vitality of the trees and the death of the twigs in the event of a severe infestation ([Alves et al., 2019](#)). The main damage is the indirect damage caused by the secretion of a large amount of soot-covered honeydew, which covers the fruits and leaves with a dense black substance. As a consequence, the leaves fall off, the olives quality deteriorates and the shoots become wilted ([Sinha and Shekhawat, 2020](#)) [Fig. 5](#).

2.2.2. Peacock eye disease

The Cycloconium or "peacock's eye" presented in [Fig. 7](#) is a disease in the effects are the most serious because it attacks both the leaves and the fruit, it appears as a spot of an interval between 2 and 10 mm diameter circular yellow or brown and are mainly found on the upper surface of the leaves, stems or even directly on the fruit mainly infecting the plant photosynthetic ability, these complications can be presented by leaf drop or a significant reduction of production in the second year particularly. Newly sprouted leaves are likely to be asymptomatic over a long period of several months, under conditions of high humidity and rain ([Sinha and Shekhawat, 2020](#)).

2.2.3. Verticulosis disease

The *Verticillium dahliae* presented in [Fig. 8](#) occurs as a microscopic fungus that spreads in the soil and infects the tree through a rise of sap. Contagion develops through root wounds or through pruning using infected tools ([Sinha and Shekhawat, 2020](#)). The external pathological symptoms of the disease manifest itself in an acute form in young trees and a specific form in older trees. In addition, the symptoms

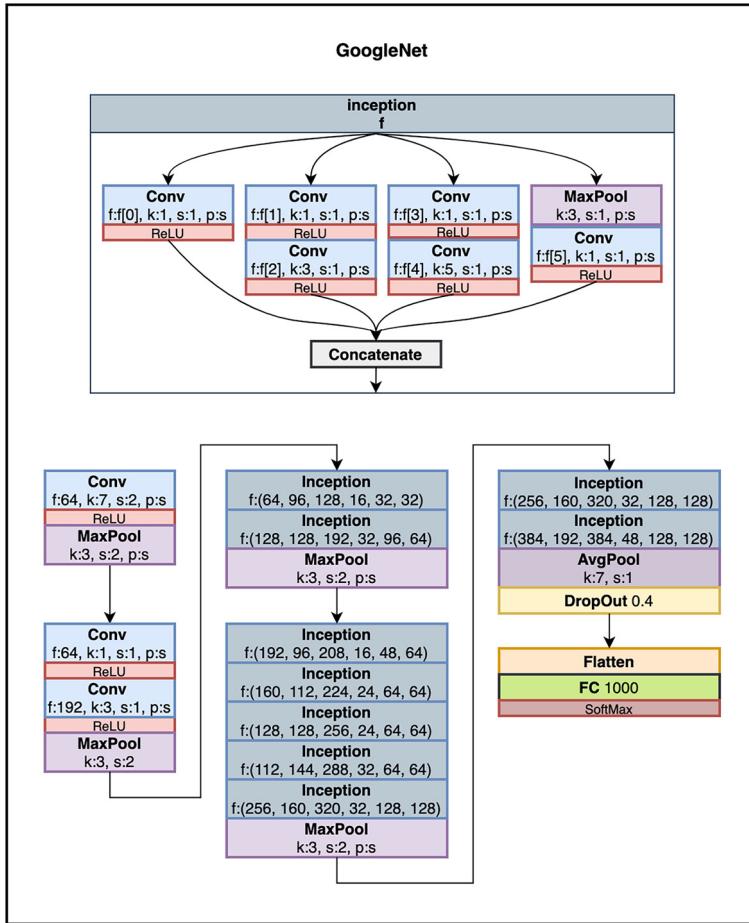


Fig. 3. Classical structure of Google net deep neural network.

of stroke appear in the period of late winter or early spring. It is characterized by a rapid and significant wilting of the twigs, secondary and main branches with a brown bark that tends towards purple developing from tip to base. These leafy organs have leaves that curl towards their underside and whose greenish color degrades to turn light brown, which makes them completely dry (Singh et al., 2020).

2.2.4. Anthracnose disease

Widespread throughout the Mediterranean basin. Anthracnose presented in Fig. 9 is one of the illnesses that most affects the olive tree. Manifests throughout the maturation and growing of the fruit and through further than one cycle throughout the year. appears in the form of a parasite manifested by the phyto pathogenic growth alluding to the enormous types of *Colletotrichum*, this disease is probably the main foliar fungus of the olive tree and the main infection of the natural product. These characteristics in general are rotting and mummification of olives (Sinha and Shekhawat, 2020).

2.2.5. Tuberculosis disease

The disease is mainly manifested by the appearance of protuberances on branches, trunks and roots in spring and summer, usually at the nodes at the base of leaves and fruit stems. Generally, the Tuberculosis disease impacts directly the production (Sinha and Shekhawat, 2020). Olive knots presented in Fig. 10 are visible and very responsive on the tree. Has aerial tumors instead of the typical leaf necrosis and ulcer. A hyperplastic growth (galls or nodules) on the branches and stems of the tree and sometimes on the leaves and fruits presents characteristic symptoms of the disease.

2.2.6. *Saissetia oleae* disease

Part of the coccidae family. The *Saissetia Oleae* is a cochineal presented in Fig. 11. it is spread over a good number of wild and leafy plants. it manifests itself on different crops, in particular fig trees, citrus fruits, apricots and olive trees. It manifests mainly on the branches until they are completely covered in some cases [9]. the disease weakens the plant leading to the destruction of its vital organs. Through the secretion of honeydew on leaves and twigs. As it happens allows the expansion of soot (Sinha and Shekhawat, 2020).

2.3. Related work

Related work can be classified into two categories: Firstly, techniques for detecting plant diseases and secondly, techniques used in the literature particularly for olive plant diseases classification.

2.3.1. Techniques for detecting plant diseases

This section describes the techniques used in the literature for the identification of plant diseases. Firstly, some works (Gavhale and Gawande, 2019; Saleem et al., 2019; Esgario et al., 2019) used datasets based on leaves images, but did not specify whether the datasets were balanced or unbalanced. The authors in Moorthy et al. (2020) used spectroscopy to study the interactions between materials and light, regarding the frequencies that will be joined or mirrored, the review contended that the majority of the early work on leaves pictures utilized as a key information section, additionally when the side effect is shown, it shows that the sickness is as of now at a high-level stage and practically speaking there is no hope to save the tainted plants. Henceforth the requirement for elective means and strategies like spectroscopy to

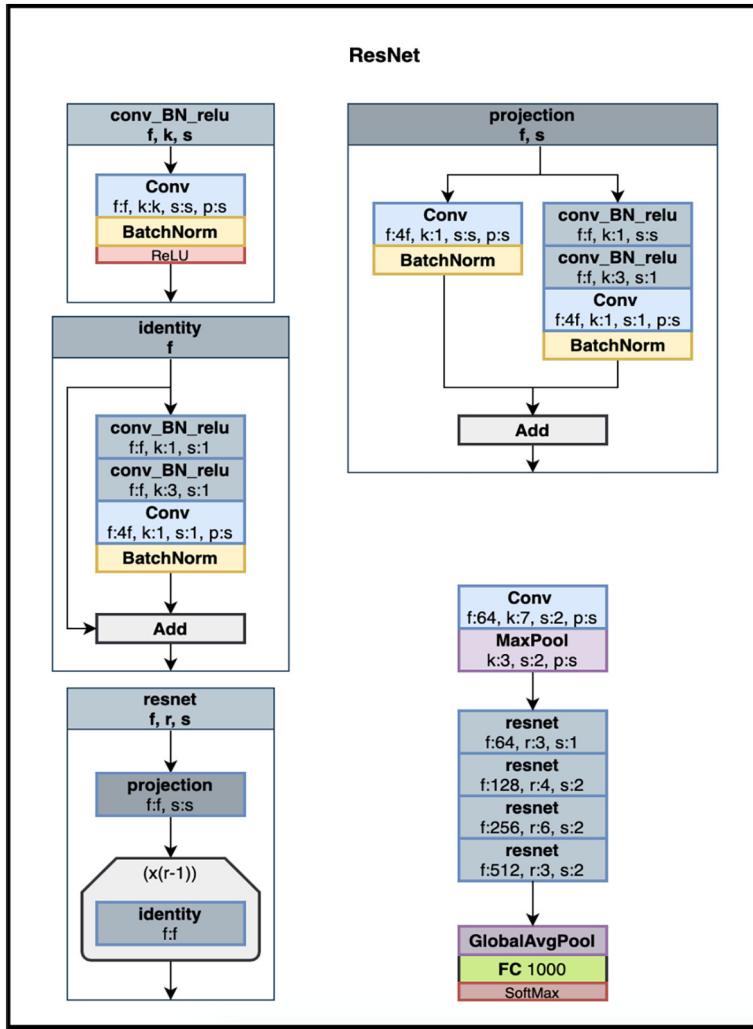


Fig. 4. Classical structure of residual network deep neural network.

recognize and then obstruct the disease at an early phase. Furthermore, the author of [Foyal et al. \(2019\)](#) proposed a useful and appropriate solution for recognizing the classifying areas of sicknesses in tomato plants. The goal is to track down a profound learning engineering to handle units more appropriate than the strategy for of acquiring actual examples (leaves, plants) and to dissect it in the lab as in previous work. Afterwards, at that point, the exploration proposed in [Saleem et al. \(2019\)](#) utilized the Inception v3 engineering to prepare the convolutional neural networks organization on the cassava infection picture dataset collected in Tanzania. To distinguish three sicknesses and two bugs, they demonstrated that figuring out how to relocate is a decent instrument for robotized illness discovery. The model is conveyed on cell phones to recognize disease of cassava plants progressively through a Tensor Flow application. At last, the exploration work introduced ([Abade et al., 2021](#)), treats some compositional models, for example, AlexNet, GoogLeNet and VGG framed on a public dataset containing 87,848 pictures, including 25 unique plants. In a bunch of 58 particular classes of plant sickness types, incorporating sound plants with the best outcomes exhibitions arriving at a triumph pace of 99.53%. Other works address a very important area regarding the limitations of data augmentation techniques, In fact, the authors in [Tassis et al. \(2021\)](#) propose an integrated framework using different convolutional neural networks (CNN) to automate the detection/recognition of lesions from field images collected via a smartphone containing part of the coffee tree. In addition, authors in [Sharma et al. \(2020\)](#) investigate a

potential solution to the model generalization problem on independent data, by using segmented image data to train convolutional neural network (CNN) models. The present experience builds on this previous work, except that it uses a proper dataset with an improved classification model.

2.3.2. Techniques for detecting olive leaf diseases

In the methodology proposed in [Waleed et al. \(2020\)](#), using segmentation of the k-mean algorithm which offers greater precision, texture analysis is applied using first to fourth order moments, this helped to identify the relation of infection with one or more textures using first to fourth order moments, in this case this allowed us to identify the infection relationship with a high correlation between the area of infection and texture features acting as homogeneity, entropy, which also helped to classify the two homogeneous diseases of neofabrie and leaf spot of peacock. Subsequently in [Chandra and Matthias \(2017\)](#), the authors present a proficient model utilizing the idea of move learning applied to distinguish olive tree diseases, a savvy increment of information with a weighted number of pictures in each class, and it works in more composite conditions with a broadened and further developed informational index. In the simulation results authors exhibit that the optimized model proposed accomplishes higher estimations, as far as exactness, accuracy, recall and F1 estimation with a general precision of 99.11%.

Then in [Waleed et al. \(2020\)](#), the authors introduced an automatic olive tree detection system using an improved K-Means algorithm,

Table 1

Major challenges associated with implementation of depth based CNN architectures.

Architecture	Strength	Shortcomings
AlexNet	<ul style="list-style-type: none"> - No more convultive layers and parameters used to adapt to the high volume data set. - The introduction of the rectified unit and the preprocessing represent an important advance in computer vision tasks. 	<ul style="list-style-type: none"> - Limited efficiency compared to new CNN architectures.
ResNet	<ul style="list-style-type: none"> - Simplification of the process of knowledge as an extential layer in neuronal networks as an identity function. - Inputs can propagate faster via residual connections between layers. - The convolutional and sequential characteristic of the network. 	<ul style="list-style-type: none"> - Overfitting due to stacking the same modules.
VGG	<ul style="list-style-type: none"> - Based on a compact and efficient design of complex networks. - Multitudes of layers of deep and reduced convolutions performed better than fewer layers of wider convolutions. 	<ul style="list-style-type: none"> - The cost of calculation using fully associated layers.
GoogLeNet	<ul style="list-style-type: none"> - Convolutions 1×1 minimize the dimensionality of the channel on the pixel pane. Thus the maximum pooling reduces the resolution. - Offers a validation performance close to similar architectures with a complexity (it can't be considered as an advantage I guess) and a more optimal calculation time. 	<ul style="list-style-type: none"> - Risk of data loss due to a bottleneck

they propose an automated detection approach of olive trees disease including an implementation model based on the classification in several steps. Therefore, Random Forest outperformed the others with an overall accuracy of 97.5% with a 70–30 ratio between training and testing.

In addition to the research presented previously, the experimental studies of [Uguz and Uysal \(2020\)](#) using data augmentation



Fig. 6. Sample of black olive disease.

techniques in the trained models obtained a performance rate of 95%, while in the experiments without this techniques, the higher valuer was 88%. One more subject of this work is the impact and connection of the Adam, AdaGrad, SGD and RMSProp streamlining calculations on network performance. Because of the performed analyses, the authors presumed that the Adam and SGD by and large have unrivaled outcomes. Eventually, when implementing the Random Forest algorithm to Olive Anthracnose, the results were acceptable but could be improved if real data had been collected. The [Table 2](#) summarizes the results obtained of the related work carried out in the literature of olive plant diseases classification and parameters used in terms of augmentation techniques, number of classes, transfer learning, CNN architecture and data-set volume.

The [Table 2](#) summarize the results obtained of the related work carried out in the literature for the classification of olive tree diseases and parameters used in terms of augmentation techniques, number of classes, transfer learning, CNN architecture and data-set volume.

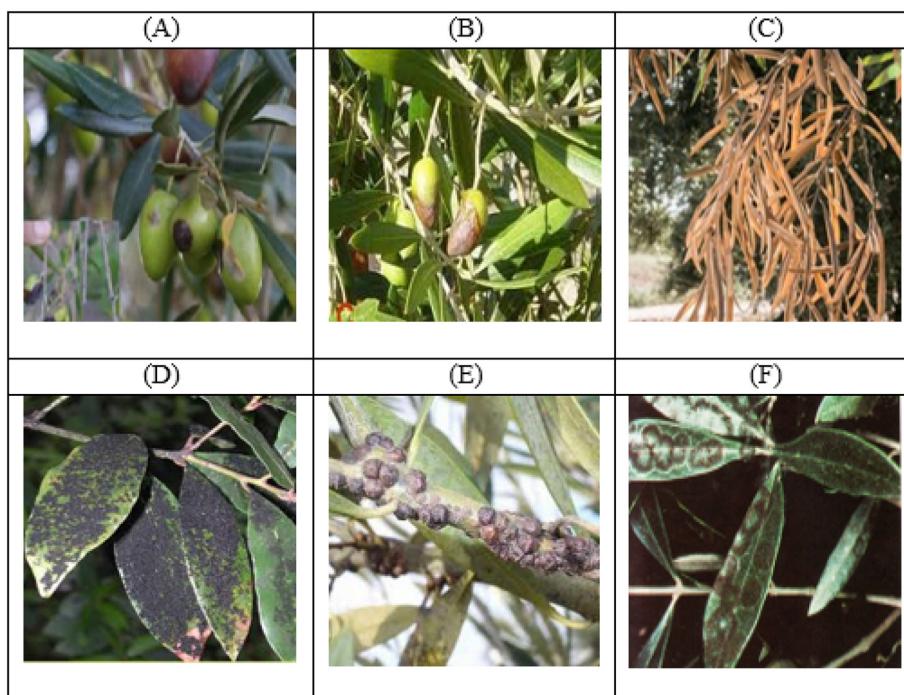


Fig. 5. Sample of olive dataset diseases used in experimentation (A,B) Anthracnose (C) Verticilliose, (D) Fumigina, (E) Saissetia oleae, (F) Cyclonium OP.



Fig. 7. Sample of peacock eye disease.

2.4. Simulations

The simulation process presented in the Fig. 12 below gives a general idea of the process adopted for the implementation of the optimization techniques studied and the approach followed to obtain the best results.

2.4.1. Simulation workflow

The unknown samples in the training dataset are used to train the classification algorithm to determine the CNN models that are most suitable for detecting a specific disease. Finally, suitable model associated with best optimizer were obtained, and the results were evaluated using different performance indicators. The analysis shows the optimal methodology of feature selection employing suitable methods for



Fig. 9. Sample of anthracnose disease.

classification. The most important tasks associated in the image classification techniques detailed in the schema above are the designation of a suitable classifier, the extraction of features, the choice of a training sample, the preprocessing of the images and determining the optimal classification model, optimization algorithm, processing after classification and lastly analyzing the accuracy and performance of the evaluation (Dhingra et al., 2018).

2.4.2. Context of simulations

Below is the description of the configuration of the computer hardware used for the simulation.

- CPU: Processeur i5-8250U @ 1.6 GHz (8 cpu)
- RAM: 16384MB
- Langage de programmation: Python Version 3.7
- Software: Anaconda 3 / Spyder Version 3.3.6

2.4.3. Data collection

The dataset contains 5571 images of different plant leaves classes (health and diseases) taken in real conditions from the crop olive fields. It's composed with seven class diseases with an unbalanced distribution described in Fig. 13 including healthy one.

It is important to note that the dataset contains a variety of images taken at different stages of the evolution of the olive diseases presented and this makes it possible to enrich the learning set and thus allow the classifier to obtain best result on the validation set (Yousuf and Khan, 2021). At the beginning of this work, many human and material resources were devoted to the collection of diseased olive because few suitable datasets were available for the real-time detection of olive



Fig. 8. Sample of verticulosis disease.



Fig. 10. Sample of tuberculosis disease.



Fig. 11. Sample of saissetia oleae disease.

diseases. The disease patterns of olive vary with the season and with other factors such as the humidity, temperature and radiance. For example, rainy weather is conducive to the generation and spread of germs, thereby resulting in the expansion and diffusion of the disease spots on affected leaves. Taking that into consideration, images are collected under various weather conditions for more comprehensive applications. Finally, all diseased images in the dataset are annotated manually by experts.

2.4.4. Data preprocessing

The data preprocessing tasks begins by spanning the data points. Secondly, an achievement of a split on the data using 80 of the images for training and 20 for testing. In addition improvement approaches were applied for the upgrade of the distribution of pixels over an expansive range of intensities, direct discrepancy stretching was applied on the images.

2.4.5. Data annotation

Image annotation is a vital step in which the objective is to label the positions and classes of object spots in the diseased images. In this stage, an algorithm that provides a frame selection function is developed in Python. With this algorithm and together with the knowledge provided by experts in agriculture, the diseased areas of an image can be selected and assigned to the corresponding classes (Pantazi et al., 2020). Annotations are provided as images in the same size as the originals, stored in JPG or JIFF format. A black colored pixel indicates the background, while all other colors are used to uniquely identify the leaves of the plants in the scene. Across the time-lapse footage, the approach consistently used the same color code to label occurrences on the same sheet. Then, within the binary mask of each plant, by the delimitation of the leaves, branches and fruits individually, following an approach based solely on manual labeling (Garcia and Barbedo, 2018).

2.4.6. Data augmentation

The overfitting problem in the training stage of CNNs can be overcome via data augmentation. The overfitting problem occurs when it is random noise or errors are described, rather than the underlying relationship, see Cap et al. (2020). With more images after expansion via data augmentation techniques, the model can learn as many irrelevant patterns as possible during the training process, thereby avoiding overfitting and realizing higher performance. For this end, several

techniques are used for data augmentation operations, including rotation transformations, horizontal and vertical flips, and intensity disturbance, which include disturbances of brightness, sharpness and contrast. A Gaussian noise processing operation is also applied. Via the above operations, new diseased images are generated from each image (Cap et al., 2020). The authors use cross-validation before the approach of over-sampling the olive dataset classes (Vega-Márquez et al., 2020), just as how feature selection should be implemented. Only by resampling the data repeatedly, randomness can be introduced into the olive disease dataset to make sure that there won't be an overfitting problem.

The Fig. 14 present the distribution of olive disease dataset after applying the augmentation techniques detailed on the above paragraphs. The data augmentation process is carried out by fine-tuning the following parameters includes rotation_range = 15, width_shift_range = 0.2, height_shift_range = 0.2, shear_range = 0.2, zoom_range = 0.2, horizontal_flip = True, fill_mode = 'nearest'.

2.4.7. Feature extraction

Some variables or characteristics are very important in forming the different models. In this case, and in order to keep only the most relevant variables and eliminate the harmful characteristics that can disturb the learning of the proposed system, a selection process of variables is then applied. To this extent, the methods of variable selection used in this study are the sequential backward selection of variable selection.

2.4.8. Classification

Classification is a function that need the use of machine learning algorithms that learn how to assign a marker to different classes from the problem sphere. An easy illustration example to understand is classifying emails spam. Multi-class classification cites to those classification functions and characteristics that have further than two class markers.

2.4.9. Evaluation

Performance metrics provide intelligence for decision makers to support evaluation. The results are evaluated using distinctive performance metrics like Accuracy, Error Rate, Kappa, Precision, Recall, F1 Score, Mean Absolute Error, and Log Loss to recommend the appropriate feature selection method for prediction [5]

2.4.9.1. Accuracy.

Accuracy is calculated according to the formula below.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (2)$$

Where, TP- True Positive; TN- True Negative; FP- False Positive; FN- False Negative.

2.4.9.2. Precision. The present of predicted positive that is in fact positive. It also called as Positive Predicted Value (PPV).

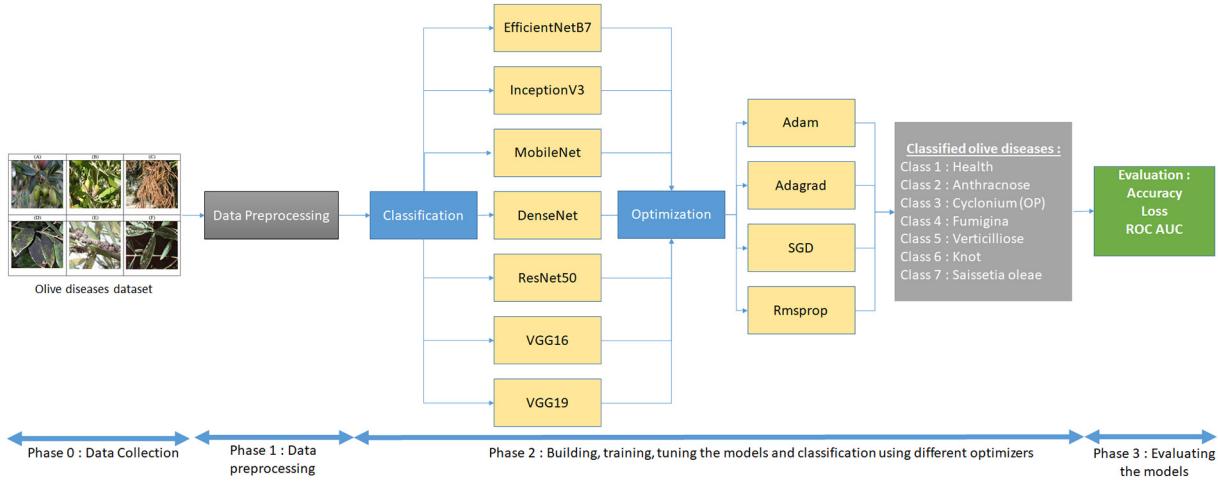
$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FD}} \quad (3)$$

2.4.9.3. Recall. The proportion of positive results out of the number of samples which were in fact positive. It is recognized as Sensitivity.

Table 2

Related works carried out of olive disease classification.

Paper	Validation accuracy	Augmentation	Nb. of classes	Transfer learning	Architecture	Dataset	Country source of images
Sinan Uguz, 2020	95%	Yes	(2 + 1 healthy)	Yes	VGG16 and VGG19 architectures	3400 olive leaves images	Turkey
Sinan Uguz, 2020	96%	Yes	(1 + 1 healthy)	No	SSD architecture	1460 olive leaves images	Turkey
Mario Milicevic, 2020	97.20% ± 0.57%	Yes	(1 + 1 healthy)	No	VGG-inspired network	1000 images	Croatia
Madallah Alruwaili et al, 2019	99.11% ± 0.75%	Yes	(6 + 1 healthy)	No	Alexnet architecture	2287 olive leaves images	-

**Fig. 12.** General schema for the experimental study.

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

2.4.9.4. F1 score. F1 Score is designed as the geometric mean of Precision or PPV and Recall or True Positive Rate (TPR).

$$\text{F1Score} = 2 * \frac{\text{PPV} \cdot \text{TPR}}{\text{PPV} + \text{TPR}} \quad (5)$$

2.4.9.5. Loss. The paired cross entropy misfortune work ascertains the d_{scale}90%eficiency of a model by figuring the accompanying normal:

$$\text{Loss} = - \frac{1}{\text{output}} \sum_{i=1}^{\text{output}} y_i \cdot \log \hat{y}_i + (1 - y_i) \cdot \log (1 - \hat{y}_i) \quad (6)$$

where \hat{y}_i is the i-th scalar worth in the model yield, y_i is the relating objective value, and yield size is the quantity of scalar qualities in the model output/result.

3. Results

Models are implemented, using the appropriate optimizer, to find accurate olive characteristics for predicting specific diseases. Classification methods are optimized with augmentation techniques to find the most suitable model for a particular disease. The techniques are

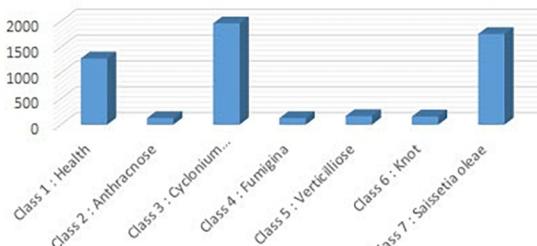
evaluated thereafter, using metrics such as Accuracy, Loss and Area under the ROC Curve (AUC).

Fig. 15 displays a chart analyzing the results of models using the algorithm of optimisation.

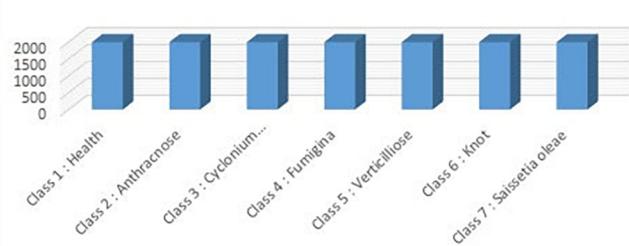
Looking at **Fig. 15**, the CNN MobileNet model offers superior results. The optimal number of epochs using Early Stopping call back function where fixed to 100 iterations. So the augmentation of the total of iterations will further minimize the loss graph oscillations. Also, the augmentation would extend the duration of the experiments significantly. The model should be trained for an optimal number of epochs to reduce overfitting and increase the neural network's generalization capacity. Training dataset part is set aside for model validation, which involves evaluating the model's performance after each epoch of training. Loss and accuracy on both the training and validation sets are monitored to determine the epoch number at which the model begins over-fitting. The best performances in terms of precision values were obtained on validation dataset by the MobileNet model at 0.98 for the six disease classes using the Rmsprop optimization algorithm, **Fig. 15** show precision and loss value graphs relating only to models developed with the Adam, Adagrad, Rmsprop and SGD algorithms. The precision and loss value results of the training and validation datasets with the application of data augmentation techniques can be seen on **Table 3**. **Fig. 16** show the ROC curve of the best models using Rmsprop algorithm, that is plotted with TPR (true positive rate) against FPR (the false positive rate) of each the seven olive plant category including health class.

The results mentioned in the classification report mention that the classes 2 and 4 outperforms the other classes in terms of precision.

Unbalanced distribution of Olive Diseases Dataset

**Fig. 13.** Unbalanced distribution of olive disease dataset.

Balanced distribution of Olive Diseases Dataset

**Fig. 14.** Balanced distribution of olive disease dataset.

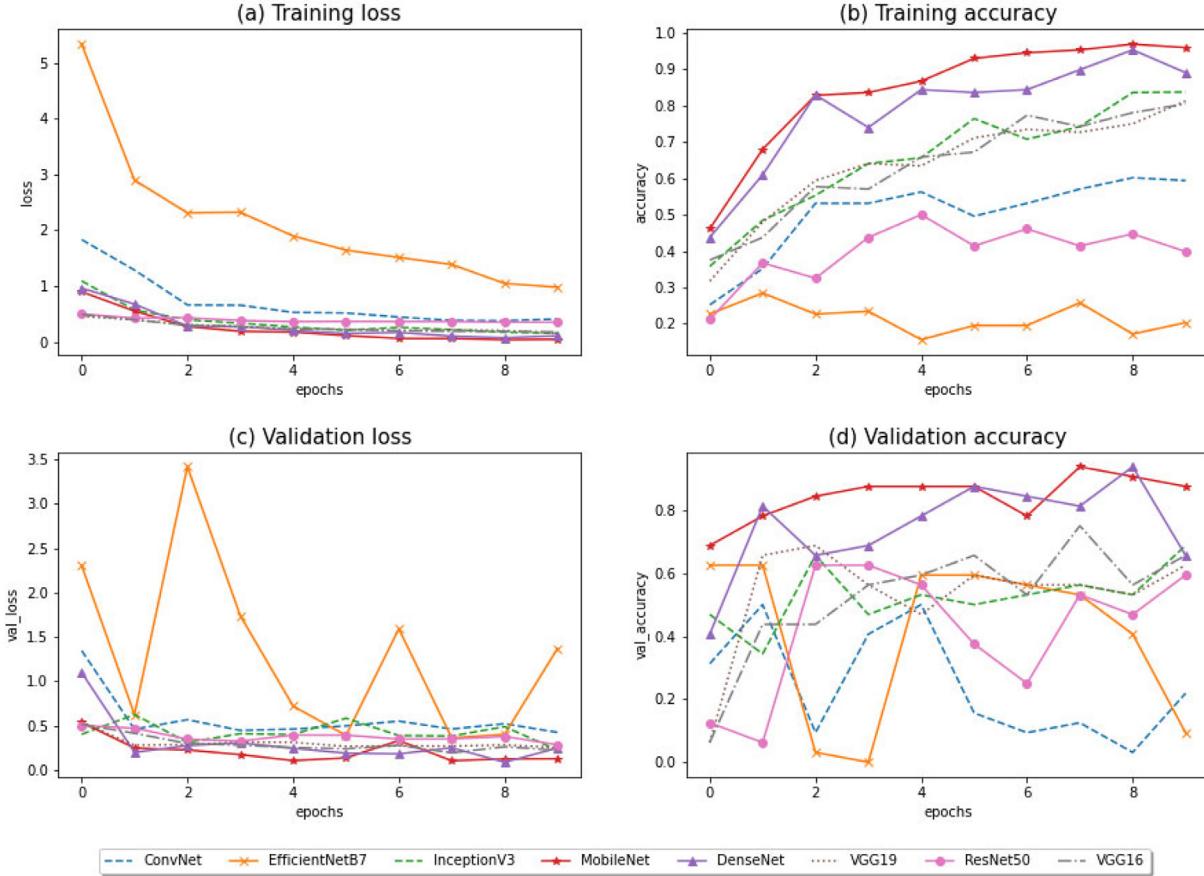


Fig. 15. Comparison of models performances: Rmsprop.

Table 3

Performance evaluation results of CNN models based on the applied optimization algorithms (10 epochs).

Optimization	CNN model	Training loss	Training accuracy	Validation loss	Validation accuracy
Adam	ConvNet	0.5394	0.9479	0.5399	0.8958
	EfficientNetB7	0.8908	0.6838	0.1418	0.8531
	InceptionV3	0.1991	0.9051	0.3465	0.9187
	MobileNet	0.0327	1.000	0.1385	0.9792
	DenseNet	0.2038	0.957	0.2404	0.9187
	VGG19	0.2048	0.8695	0.3391	0.9187
	ResNet50	0.4667	0.834	0.3907	0.8969
	VGG16	0.2796	0.8941	0.4325	0.8531
Adagrad	ConvNet	0.1809	0.9242	0.305	0.8749
	EfficientNetB7	0.9578	1.000	0.1899	0.1968
	InceptionV3	0.4050	1.000	0.1885	0.9625
	MobileNet	0.0664	1.000	0.0159	0.9792
	DenseNet	0.2370	0.9843	0.0946	0.9583
	VGG19	0.4084	0.9686	0.124	0.8333
	ResNet50	0.5161	0.8340	0.5074	0.7875
	VGG16	0.4037	1.000	0.1483	0.0875
Rmsprop	ConvNet	0.2928	0.9598	0.2799	0.8969
	EfficientNetB7	0.5352	0.8872	0.5948	0.8749
	InceptionV3	0.2192	0.9652	0.1963	0.9375
	MobileNet	0.0257	1.000	0.0160	0.9843
	DenseNet	0.2267	1.000	0.1621	0.9792
	VGG19	0.0269	1.000	0.1370	0.9583
	ResNet50	0.4752	0.8494	0.3795	0.8749
	VGG16	0.1908	0.9543	0.2718	0.8969
SGD	ConvNet	0.1479	0.9982	0.2661	0.8969
	EfficientNetB7	0.1601	0.9351	0.2714	0.8749
	InceptionV3	0.1534	0.9952	0.1244	0.9792
	MobileNet	0.0241	1.000	0.0435	0.9801
	DenseNet	0.1228	1.000	0.1937	0.9625
	VGG19	0.3500	0.8831	0.2936	0.8969
	ResNet50	0.1855	0.9707	0.2777	0.8312
	VGG16	0.2581	0.9460	0.2488	0.9405

classification of olive diseases. So this leads us to deduce that except from the quality and the quantity of image also serves to ameliorate the performance as well as the types of classes used. Also the ROC curves clearly show that the MobileNet model associated with the Rmsprop optimization algorithm offered the best results. In addition, the AUC obtained confirm the performance of this combination and more details will be provided in the following discussion.

4. Discussions

Firsly, the results obtained detailed in the Table 4 show that the Adam and Adagrad algorithms, present the weakest performances with the largest average absolute errors, respectively 0.3346 and 0.2682, as well as the greatest number of iterations 190 for the Rmsprop algorithm (the number of iterations is fixed at 75 for the stochastic gradient). The CNN MobileNet model represents superior performance for mean absolute error as well as for accuracy. Also the results indicated a significant difference between the solvers, for example Adam's algorithm converged in only 70 iterations, however Adagrad's and Rmsprop's algorithms converged in 143, 190 iterations respectively. Furthermore, the RMSProp and Adagrad algorithms represents the best MAEs which are of the order of 0.2071 and 0.2682 respectively, with an increase in the number of iterations, 190 and 143 iterations for RMSProp and Adagrad respectively. Among the algorithms studied, the Stochastic algorithm is of great interest for use on a mobile phone given its good performance in terms of convergence speed and MAE. Secondly, to the best of our knowledge this research paper has unprecedently addressed the problem of the choice of optimization techniques at the level of convolutional neural networks and its impact on the performance of the latter. Thud, CNN models used on simulation had the lowest performance compared to the MobileNet model using

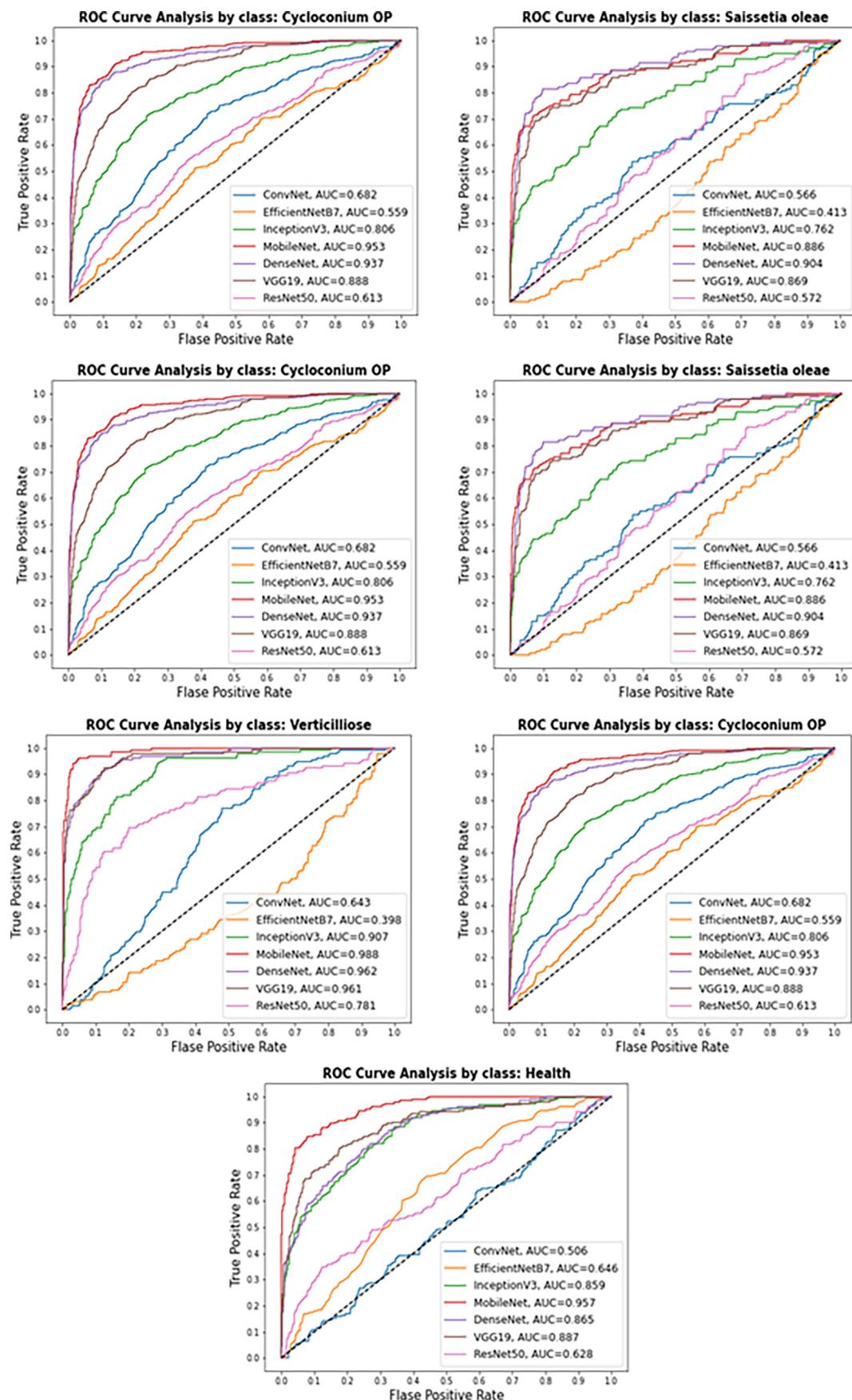
**Fig. 16.** ROC curve by class using Rmsprop.

Table 4

Comparison of the performances of studied algorithms.

Algorithm	Number of iterations	MAE	Tensorflow
Adam	70	0,3346	AdamOptimizer
Adagrad	143	0,2682	AdagradOptimizer
Rmsprop	190	0,2071	RMSPropOptimizer
Stochastic	75 (fixed)	0,2263	SGDRegressor

the most optimal optimization algorithm namely Rmsprop. The batch norm and ReLU are the primary features that characterize MobileNet architecture from another CNN. The MobileNet was built to support classification, identification, and segmentation. Due to the number of parameters and layers, the capability to run deep networks on terminal mobile devices promotes user experience, providing the advantages and characteristics in terms of security and energy consumption of mobile applications. Thirdly, the enhancement of a model that performs better than the pre-trained model definitely depends on more simulations with parameter values. Again, if augmenting the number of simulations and periods has a positive impact on performance, it will also take more time in terms of model compilation. Furthermore, The performance of the same solver will change if a different set of hyperparameters and starting conditions are applied. So Larger learning rate values tend to exceed the gradient value, making it difficult for the weights to converge to the global minimum. Also, small values of the learning rate make the progress of the global minimum very slow, which can be seen in the validation and training losses. But the optimal value of the learning rate leads to a global minimum, which can be visualized by a steady reduction in the loss. So if it is to develop real time disease identification research via a mobile application, it would be more appropriate to converge quickly, since different objects will bring about some effects in the background of the objects and make the learning phase slower. Then, the exploration related to olive trees bandied in this exploration is truly limited. So the present work differs from researches (Foyal et al., 2019) at many levels. Some of the differences are the types of diseases identified, the CNN architecture used, and the increase in data. One of the infections that the authors of Altarawneh (2015) is trying to identify is olive peacock spot, which was also studied. Using the CNN model proposed in this research, the disease accuracy is 95%. The author's work obtained an accuracy of 90.2%. The comparison of the loss functions employed in this study, in addition to compilation time, were not addressed not only in these investigations, but also on other plant species studies. Various diseases found on fruits, leaves, and branches can also be investigated in this subject. The challenge is to be able to collect data of these plant diseases over different regions and particularly when they occur. For the reason that some infections are only seen in particular areas, then obtaining these photos may be difficult. As a result, each study on plant disease detection and particularly in this original dataset is regarded as a valuable case of the linked research provided (Kurmi et al., 2020). After, the classification of different diseases classes was conducted in this study. The challenge to increase the number of class diseases without impacting the success rate. Because diseases of the same type of plant can have similar symptoms. The optimization performed in this field are considered promising because of the difficulty of the success rate. When the ROC curve presented by olive tree disease class provides more relevant information on the quality of learning than the simple error rate. This is concretely present in this case because the classes are very unbalanced in terms of the number and quality of the images appearing at the level of each class, in addition to the moment that the cost of bad prediction or assignment is likely to change. Finally, some researchers around the world are carrying out studies using deep learning techniques for identifying plant diseases. One of the biggest challenges is collecting enough data. For this reason, sharing datasets in papers has become quite important. As a consequence the development and public availability of datasets related to plant diseases is very important to advance detection

systems based on artificial intelligence, in particular Machine Learning, the enrichment of existing datasets goes without saying. to improve and optimize applications intended for the farmer with the objective of increasing agricultural productivity as well as the social development of farmers whose crops are low yielding but also for adaptation to future changes that may impact agricultural crops with a view to parasitic diversity and direct or indirect environmental factors.

5. Conclusions

It is crucial to diagnose diseases of growing agricultural products at an early stage, consequently, it helps farmers take the required preventive measures, thereby reducing costs. In addition, the extracted oil's quality and quantity depend both on the fruits' health process. Therefore, disease detection at different growth stages would play a considerable role in the olive food industry. The current article aims to an Olive disease dataset (ODD) collected from different regions of Morocco at different growth stages based on seven classes, six of them being ill and the seventh healthy. The novelty in the classification approach here lies in the fine-tuning based on the MobileNet model architecture which outperformed eight other contenders among standard CNNs, besides being tested with four optimizers, where Rmsprop turned out to be most performant. Admittedly, the suggested model significantly distinguished all the seven olive diseases classes with a high precision, taking into account the fruit texture characteristics and the factors that impact its effectiveness and its efficiency, with consideration of the challenge of classifying the disease spreading in foliage, fruit and branches. Although the proposed method for automatic identification of olive diseases has obtained satisfactory results, further work is needed in the future to improve its accuracy and reliability. Therefore, future work will focus on:

- Further develop the olive tree disease dataset and set up an intelligent olive tree disease diagnosis system based on the use of drones.
- Integrate a layer for image segmentation to increase system accuracy.
- Investigate other architectures of deep neural networks, in order to improve the classification accuracy, reliability and robustness of diagnostic systems for olive diseases.

Credit authorship contribution statement

EI Mehdi Raouhi: Conceptualization, Formal analysis, Software, Data curation, Implementation, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Mohamed Lachgar:** Conceptualization, Methodology, Study design, Formal analysis, Software, Implementation, Investigation, Validation, Writing – original draft, Writing – review & editing, Visualization. **Hamid Hrimech:** Methodology, Study design, Supervision, Formal analysis, Software, Resources, Implementation, Validation, Investigation. **Ali Kartit** Supervision, Project administration, Funding acquisition

Declaration of Competing Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

- Abade, A., Ferreira, P.A., Vidal, F.B., 2021. Plant diseases recognition on images using convolutional neural networks: A systematic review. Computers and Electronics in Agriculture. Elsevier, <https://doi.org/10.1016/j.compag.2021.106125>.
- Alruwaili, M., Alanazi, S., El-Ghany, S.A., Shehab, A., 2019. An efficient deep learning model for olive diseases detection. Int. J. Adv. Comput. Sci. Appl. 10, 486–492. <https://doi.org/10.14569/ijacs.2019.0100863>.
- Altarawneh, M., 2015. An empirical investigation of olive leave spot disease using auto-cropping segmentation and fuzzy c-means classification. World Appl. Sci. Emp. Invest. Olive Leav. Spot Diseases. <https://doi.org/10.5829/idosi.wasj.2013.23.09.1000>.

- Alves, L., Silva, R.R., Bernardino, J., 2019. System to predict diseases in vineyards and olive groves using data mining and geolocation ICSOFT 2018. Proceedings of the 13th International Conference on Software Technologies, pp. 679–687 <https://doi.org/10.5220/0006914306790687>.
- Cap, Q.H., Uga, H., Kagiwada, S., Iyatomi, H., 2020. LeafGAN: an effective data augmentation method for practical plant disease diagnosis. *IEEE Trans. Autom. Sci. Eng.* <https://doi.org/10.1109/TASE.2020.3041499>.
- Chandra, M., Matthias, M., 2017. SC-Adagrad and SC-RMSProp. *arXiv:1706.05507v2*.
- Chliyeh, M., Selmaoui, K., Touhami, A.O., Abdelkarim, F., 2014. Survey of the fungal species associated to olive-tree (*Olea europaea* L.). *IJRBS Survey of the Fungal Species Associated to Olive-tree (*Olea europaea* L.) in Morocco*.
- Dhingra, G., Kumar, V., Joshi, H.D., 2018. Study of digital image processing techniques for leaf disease detection and classification. *Multimed. Tools Appl.* 77, 19951–20000. <https://doi.org/10.1007/s11042-017-5445-8>.
- Esgario, J.G., Krohling, R.A., Ventura, J.A., 2019. Deep Learning for Classification and Severity Estimation of Coffee Leaf Biotic Stress. <http://arxiv.org/abs/1907.11561>.
- Foysal, F.A., Islam, M.S., Abujar, S., 2019. A Novel Approach for Tomato Diseases Classification Based on Deep Convolutional Neural Networks A Novel Approach for Tomato Diseases Classification Based on Deep Convolutional Neural Networks. July. Springer, Singapore <https://doi.org/10.1007/978-981-13-7564-4>.
- Garcia, J., Barbedo, A., 2018. Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. *Comput. Electron. Agricult.* 153, 46–53. <https://doi.org/10.1016/j.compag.2018.08.013>.
- Gavhale, M.K.R., Gawande, P.U., 2019. An Overview of the research on plant leaves disease detection using image an overview of the research on plant leaves. *Diseas. Detect. Image Proess. Techn.* <https://doi.org/10.9790/0661-16151016>.
- Hussain, S.A., Hasan, R., Hussain, S.J., 2018. Classification and Detection of Plant Disease using Feature Extraction Methods13 pp. 4219–4226.
- Kurmi, Y., Gangwar, S., Agrawal, D., Kumar, S., Srivastava, H.S., 2020. Leaf Image Analysis-Based Crop Diseases Classification Orre Cted Unc Pro Of. Signal, Image and Video Processing. Springer-Verlag London Ltd., part of Springer Nature 2020 <https://doi.org/10.1007/s11760-020-01780-7>.
- Liu, J., Wang, X., 2021. Plant diseases and pests detection based on deep learning: a review. *Plant Meth.*, 1–18 <https://doi.org/10.1186/s13007-021-00722-9>.
- Moorthy, S.G., Meenakshi, K., Nithya, M., 2020. Plant leaf disease classification and detection system using machine learning plant leaf. *Diseas. Classific. Detect. Sys. Mach. Learn.* <https://doi.org/10.1088/1742-6596/1712/1/012012>.
- Pantazi, X.E., Moshou, D., Bochtis, D., 2020. Artificial Intelligence in Agriculture. *Intelligent Data Mining and Fusion Systems in Agriculture*. Springer, pp. 17–101 <https://doi.org/10.1016/b978-0-12-814391-9.00002-9>.
- Pedrycz, W., Chen, S.m., 2020. Deep Learning: Algorithms and Applications. *Studies in Computational Intelligence865*. Springer. <https://doi.org/10.1007/978-3-030-31760-7>.
- Saleem, M.H., Potgieter, J., Arif, K.M., 2019. Plant disease detection and classification by deep learning. *Plants* 8, 32–34. <https://doi.org/10.3390/plants8110468>.
- Sharma, P., Berwal, Y.P.S., Ghai, W., 2020. Performance analysis of deep learning CNN models for disease detection in plants using image segmentation. *Inf. Process. Agricult.* 7, 566–574. <https://doi.org/10.1016/j.ipa.2019.11.001>.
- Singh, D., Jain, N., Jain, P., Kayal, P., Kumawat, S., Batra, N., 2020. PlantDoc: a dataset for visual plant disease detection. *ACM Int. Conf. Proceed. Ser.*, 249–253 <https://doi.org/10.1145/3371158.3371196>.
- Sinha, A., Shekhawat, R.S., 2020. Olive spot disease detection and classification using analysis of leaf image textures. *Proc. Comput. Sci.* 167, 2328–2336. <https://doi.org/10.1016/j.procs.2020.03.285>.
- Tassis, L.M., Tozzi de Souza, J.E., Krohling, R.A., 2021. A deep learning approach combining instance and semantic segmentation to identify diseases and pests of coffee leaves from in-field images. *Comput. Electron. Agricult.* 186. <https://doi.org/10.1016/j.compag.2021.106191>.
- Uguz, S., Uysal, N., 2020. Classification of olive leaf diseases using deep convolutional neural networks. *Neur. Comput. Appl.* 5. <https://doi.org/10.1007/s00521-020-05235-5>.
- Vega-Márquez, B., Nepomuceno-Chamorro, I., Jurado-Campos, N., Rubio-Escudero, C., 2020. Deep learning techniques to improve the performance of olive oil classification. *Front. Chem.* 7, 1–10. <https://doi.org/10.3389/fchem.2019.00929>.
- Waleed, M., Um, T.W., Khan, A., Khan, U., 2020. Automatic detection system of olive trees using improved K-means algorithm. *Rem. Sens.* 12, 1–16. <https://doi.org/10.3390/rs12050760>.
- Yousuf, A., Khan, U., 2021. Ensemble Classifier for Plant Disease Detection. 1st. 10. *IJCSMC*, pp. 14–22. <https://doi.org/10.47760/ijcsmc.2021.v10i01.003>.