A Powerful Plant Disease Classification based on Ensemble Learning

Noura Ouled Sihamman[1], Assia Ennouni[1], My Abdelouahed Sabri[1] and Abdellah Aarab[2]
[1]Computer Science Department, Faculty of Sciences Dhar el Mahraz, University Sidi Mohamed Ben Abdellah, Fez, Morocco.
[2]Physique Department, Faculty of Sciences Dhar el Mahraz, University Sidi Mohamed Ben Abdellah, Fez, Morocco

Keywords: Ensemble Learning, Deep Learning, Plant Disease Classification, Smart Agriculture, Image Processing.

Abstract: Intelligence in agriculture is becoming more and more necessary. Several tools have been proposed and used to make farmers' tasks automatic. Plant disease detection is a tough and challenging step, especially for inexperienced farmers. In several countries, the production quality is significantly reduced due to various plant diseases, which has a negative impact on the economic sector. To reduce the impact of these diseases, early plant disease diagnosis is seen as a way to treat them in time. In this context, several solutions using artificial intelligence and image processing have been proposed. In this paper, we propose an Ensemble Learning-based approach for the detection and classification of plant diseases. Thus, we propose to combine the performance of 5 Deep Learning architectures to design a robust system for plant disease classification. Simulation results proved the efficiency of the proposed approach, comparing it first with the results obtained for each different DL architecture and with also other approaches from the literature.

1 INTRODUCTION

The economy of several African countries such as Morocco depends heavily on agriculture. In order to meet the huge market demand, countries need to significantly increase their production. The quality of production is essentially linked to the management of plant diseases. Thus, if these diseases are not diagnosed and treated in time, they will affect production. According to the FAO (Food and Agriculture Organization of the United Nations), plant diseases have increased considerably in recent years and this is due to climate change, globalization, ... Generally, plant diseases are identified by visual examination by the farmer and the quality of the diagnosis is strongly linked to the expertise and professional knowledge of the farmer. This expertise is acquired after several years of close experience with plants and the diseases that affect them.

Plant diseases occur when external actors infect the plant and cause changes in physiological and biochemical behaviour. The symptoms of most plant diseases appear on the leaves. Careful examination of the shape, colour and even texture of plant leaves plays a signicant role in the diagnosis of these diseases. In order to overcome this problem, artificial intelligence and image processing have been used to propose systems for plant disease recognition and detection. The aim of such applications is to use machine learning or deep learning algorithms to efficiently detect whether a plant is diseased or not from a plant photo (Ennouni et al., 2021c) (Prakash et al., 2017). The Internet of Things is also used to design efficient systems in smart agriculture (Muhammad et al., 2019) (Gonzalez et al., 2007). Deep and automatic learning have become assets for the detection and classification of medical images, where several approaches have proven to be very efficient (Richard et al., 2005) (Wäldchen et al., 2018).

A large number of DL architectures can be used to design a powerful classification model for plant disease detection. CNN, MobileNet, AlexNet, Inception V3, and VGG16 will be used and tested to identify the behaviour of each architecture. Each architecture allows a relatively different behaviour from the others, but they are complementary. Thus, it has been noticed that some architectures fail to correctly classify some plants that other architectures can classify correctly. Based on this observation, we will propose in this paper a classification approach based on the Learning Set that combines the power of the five architectures. And to evaluate our proposed approach, we will propose a comparative study

Ouled Sihamman, N., Ennouni, A., Sabri, M. and Aarab, A.
A Powerful Plant Disease Classification based on Ensemble Learning.
DOI: 10.5220/0010733500003101
In *Proceedings of the 2nd International Conference on Big Data, Modelling and Machine Learning (BML 2021)*, pages 321-326
ISBN: 978-989-758-559-3

between a SOFT and HARD Ensemble-based classification as well as with other approaches from the literature.

The rest of this paper is as follows: the next section will present plant diseases in detail. Section 3 presents the different DL architectures that we will use to design our classification model. Then we will present our proposed approach. The Dataset used will be detailed as well as the evaluation measures afterwards. The results of experiments will be presented before concluding this paper.

2 PLANT DISEASES

Plants are living beings, the majority of which are attached to the earth by their roots. This characteristic makes that the behaviour of these plants is strictly related to their environment. Plants are subject to various diseases and these diseases can be grouped by regions (El-Sayed et al., 2020). Moreover, as for humans, diseases that can affect plants disrupt and, or modify their vital functions. The different diseases that can affect plants can be categorized into three classes: fungal, bacterial, or viral (El-Sayed et al., 2020) (Vijai, 2020).

- Viral diseases: Viral diseases are transmitted to plants mainly by insects or worms. Typical symptoms of viral diseases include mosaic patterns, yellowing, stripes, and leaf rolling (Fig. 1).

Figure 1: Necrotic spot virus on chilli leaves.

- Fungal diseases: Fungal diseases are diseases that are harmful to plants but not with great risk. Fungi, mould, mildew, and others cause these diseases. The symptoms of these diseases are leaf rust, especially for corn, Stem rust, white mould (Sclerotinia). Figure 2 shows an example of a plant affected by Sclerotinia fungus.

Figure 2: Sclerotinia Infected by Soybeans.

- Bacterial diseases: Plant infections of bacterial origin have almost the same symptoms as fungal diseases. They can be identified by the presence of rot, scab, scorch, wilt and leaf spots. These types of diseases, if not treated in time, can cause serious and disastrous diseases (Sujeet et al., 2016). Two examples of bacterial diseases are shown in figure 3.

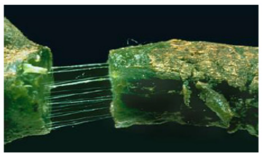

Figure 3: Bacterial strands on stems and Bacterial blight of wheat leaves.

The three different categories of diseases presented below are identifiable based on specifications and characterizations extracted after visualization of plants, stems, and leaves. Therefore, one can use image processing algorithms in combination with artificial intelligence to develop a robust system for automatic detection of plant diseases.

3 DEEP LEARNING BASED APPROACHES FOR PLANT DISEASES CLASSIFICATION

Automatic images classification is an area of research that is very focused in recent years. A distinction is made between supervised classification, unsupervised classification and reinforcement classification. In supervised image classification, machine learning, deep learning and image processing have been widely used in recent years. The main objective is to design a powerful classification model capable of predicting the class of a new image. In machine learning we proceed to a primordial step to extract relevant features before using them as input for a classification algorithm (Ennouni et al., 2021a). In contrast, in deep learning, the image is generally used directly as input to design a classification model. Several DL architectures have been proposed for the case of supervised image classification. Most of these approaches are based on convolutional neural networks (CNN).

The deep learning architecture CNN, also called ConvNet, is generally composed of a sequence of convolution layers, pooling layers (Max, min, … pooling), and at the end one or more fully connected layers. The difference between the Neural Network Deep Learning architectures is essentially in the prearrangement of the neural network components. Also, for a given architecture we should fixe a number of parameters (Kafi et al., 2015) (Shin et al., 2016):

- The number of layers
- The number of neurons in each layer
- The filters mask size
- The neurons weight
- The activation function
- The learning rate
- ....

The following table lists the most used architectures from the literature the authors, the year, the number of parameters and the depth.

Table 1: The most successful CNN architectures.

	Reference	Parameters	depth
LeNet-5	(LeCun, 1998)	60,000	5
AlexNet	(Krizhevsky, 2012)	60 M	8
VGG	(Simonyan, 2016)	138 M	19
GoogleNet	(Szegedy, 2015)	4 M	22
Inception V3	(Szegedy, 2015)	23 M	159
ResNet	(He, 2016)	25 M	152
MobileNet V2	(Sandler, 2019)	3.47 M	53

4 ENSEMBLE LEARNING-BASED APPROACHES FOR PLANT DISEASE CLASSIFICATION

Several classification algorithms can be used for plant disease classification. Each classification algorithm has strengths and weaknesses. Also, it is often noticed that diseases are misclassified by some classifiers while another has correctly classified them and vice versa. Based on this principle, it is possible to combine the performance of a set of classification algorithms to obtain a single robust classification model (Yawen et al., 2018). The Ensemble Learning is a meta-classifier allowing to combine similar or conceptually different machine learning classifiers (by majority or plurality voting) (Sabri et al., 2020). Ensemble Learning is often used in professional applications and has helped win several competitions.

Several Ensemble Learning techniques can be used (Mao et al., 2019):

1- Hard voting: The predicted label is defined as the class label most frequently predicted by the classification models.

2- Soft voting: This type of technique can be divided into two sub-elements:
   a. Progressive voting: we predict the class labels by averaging the class probabilities (recommended only if the classifiers are well calibrated).
   b. Weighted progressive voting: this is an extension of progressive voting where we assign different weights to the classifiers used. These weights are defined according to the importance and quality of the classifier.

3- Bagging: Bagging consists of creating random samples of the training dataset with substitution (subsets of the training dataset) and then applying the same classifier (Used usually with Decision Tree) for each sample and applying a hard vote.

4- Boosting: Is considered as progressive voting with adaptation of the weights of the classifiers in an iterative way.

4.1 Majority and Hard Voting

Majority voting is a widely used approach that combines many classification algorithms where each algorithm makes its prediction and the predicted class is the one obtained by majority voting

$$\tilde{y} = mode\ \{C_1(x), C_2(x), \dots, C_m(x)\}$$

where ỹ is the predicted class and Cj (i= 1 to m) is the j classifier

4.2 Weighted Majority Voting

Weighted voting is a special case of the majority voting where each prediction algorithm is used with a predefined weight.

$$\tilde{y} = argmax \sum_{j=0}^{m} w_j C_j(x)$$

4.3 Soft Voting

Soft voting is also a special case of majority voting where each prediction algorithm is used with a probability p. This approach is only effective if the classification algorithms are calibrated.

$$\tilde{y} = \sum_{j=0}^{m} w_j p_j$$

where wj is the weight assigned to the jth classifier.

5 THE PROPOSED APPROACH

The objective of this work is to propose a powerful approach for plant disease classification based on Ensemble Learning. Indeed, we have opted in a first step to the design of a classification model of plant diseases using 5 Deep Learning architectures, namely: VGG16, AlexNet, Inception V3, CNN and MobileNet.

When we design classification models using the five architectures distinctively. Each of the 5 classification models have different behaviour. We noticed after a judicious study of the classification errors for each of the five architectures that they are complementary. Thus, we noticed that for each image in the database at least one of the five architectures manages to classify it correctly. From there, we thought of combining the performances of the five architectures in an Ensemble Learning architecture to propose a powerful approach for the classification of plant diseases.

Figure 4 presents the proposed approach

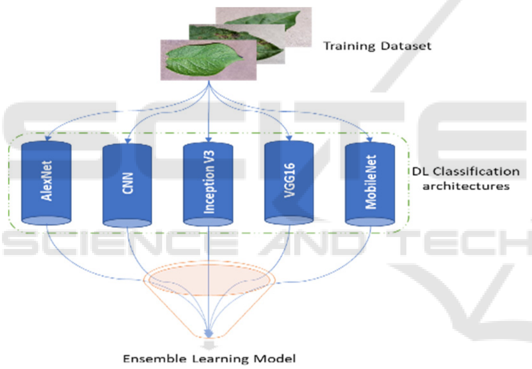

Figure 4: The proposed approach.

We will implement the two Ensemble Learning techniques; the Hard Voting approach and the Soft Voting approach. We will compare these two approaches and also compare them with recent approaches in the literature.

6 DATASET AND EVALUATION

6.1 Plant Village Dataset Description

The Plant Village (PV) is a well-known dataset used to evaluate the plant disease classification algorithms. It contains infected and healthy crop leaves images. The dataset contains 87K RGB images classified into 38 subsets. The dataset is provided with training dataset containing 70295 images and validation dataset with 17572 images (Ennouni et al., 2021b).

6.2 Performance Measures

In images classification, precision, recall and accuracy are the most evaluations measures used. We evaluating multiclass classification algorithm, their values are the mean of all the values for each class:

$$precision = \frac{\sum_{i=1}^{n} precision_i}{n} \tag{1}$$

$$recall = \frac{\sum_{i=1}^{n} recall_i}{n} \tag{2}$$

$$accuracy = \frac{\sum_{i=1}^{n} accuracy_i}{n} \tag{3}$$

$$precision_i = \frac{TP_i}{TP_i+FP_i} \tag{4}$$

$$recall_i = \frac{TP_i}{TP_i+FN_i} \tag{5}$$

$$accuracy_i = \frac{TP_i+FP_i}{TP_i+TN_i+FP_i+FN_i} \tag{6}$$

where n is the number of classes, i the current class, TP is The True Positive, FN is the False Negatives, TN is the True Negatives and FP is the False Positives numbers.

7 SIMULATION RESULTS

All simulations in this article were performed on an Intel Core i7 3.6 GHz processor, 16 GB of RAM and a Windows 10 operating system.

In our study, we will test each of the five DL architectures to evaluate the behaviour of each one. Then, we will test the two Ensemble learning proposed approaches. For the CNN architecture, we will use three layers depth. And for the VGG16, AlexNet, Inception V3 and MobileNet the number of epochs is 25.

Table 2 presents the accuracy of the five DL architectures, the accuracy of the Soft Ensemble Learning and the accuracy of the hard ensemble learning

Table 2: Classification results of the proposed approach.

	Precision	Recall	Accuracy
VGG16	0.89	0.96	0.92
AlexNet	0.86	0.96	0.95
Inception V3	0.89	0.95	0.91
CNN	0.95	0.95	0.95
MobileNet	0.94	0.97	0.95
Soft EL	0.95	0.96	0.96
Hard EL	0.99	0.99	0.99

We notice from the classification results presented in Table 1 that each of the five architectures has a different behaviour and with different values of precision, recall and accuracy. We notice that the Inception V3 architecture had the lowest values in terms of accuracy followed by AlexNet. The AlexNet, CNN and MobileNet architectures had the same accuracy score. But in terms of Recall we can say that the MobileNet architecture is the best. However, the proposed Ensemble Learning approaches have significantly improved the classification results and especially using the HARD technique with a value equal to 0.99. We can say that the use of the vote of the 5 architectures allowed us to have almost 100% of classification rate.

To evaluate our proposed approach in regards to the most relevant approaches from literature, we conducted a comparative study. Table 3 presents the classification results of our proposed approach and the approaches proposed by (Srdjan, 2016) and (Sibiya, 2019). in (Srdjan, 2016) authors proposed to use Deep CNN. Authors in (Sibiya, 2019) used also CNN with 50 hidden layers.

Table 3: A comparison results of the classification of the approaches we propose with recent approaches.

	(Srdjan, 2016)	(Sibiya, 2019).	Hard EL
Accuracy (%)	94.60	95.81	99.21

As stated at the outset, the Ensemble Learning has a higher classification accuracy than the other approaches. Although the work presented in [25] uses a large number of layers, it is still less efficient than our proposed approach.

8 CONCLUSION

Ensemble learning consists in combining a number of classification models in order to make them complementary in terms of classification accuracy. Thus, assuming that at least one of the classification models can correctly classify an image that the other classification models have misclassified, the objective of this work is to propose an ensemble learning classification approach based on 5 Deep Learning architectures namely; VGG16, AlexNet, Inception V3, CNN and MobileNet. We have implemented two techniques of ensemble learning; SOFT and HARD. The ensemble learning using HARD allowed us to significantly improve the classification rate of plant diseases. With a rate of almost 100% we can say that our proposed approach can be effectively used in smart agriculture to automatically classify plant diseases.

REFERENCES

El-Sayed, A.M.A., Rida, S.Z., Gaber,Y.A. : Dynamical of curative and preventive treat-ments in a two-stage plant disease model of fractional order, Chaos, Solitons & Fractals, Volume 137, 109879, ISSN 0960-0779, (2020).

Ennouni A., Sabri M.A., Aarab A. (2021) Plant Diseases Detection and Classification Based on Image Processing and Machine Learning. In: Gherabi N., Kacprzyk J. (eds) Intelligent Systems in Big Data, Semantic Web and Machine Learning. Advances in Intelligent Systems and Computing, vol 1344. Springer, Cham. https://doi.org/10.1007/978-3-030-72588-4_20

Ennouni A., Sihamman N.O., Sabri M.A., Aarab A. (2021) Analysis and Classification of Plant Diseases Based on Deep Learning. In: Motahhir S., Bossoufi B. (eds) Digital Technologies and Applications. ICDTA 2021. Lecture Notes in Networks and Systems, vol 211. Springer, Cham. https://doi.org/10.1007/978-3-030-73882-2_12.

Ennouni, A., Sihamman, N. O., Sabri, M. A. & Aarab, A. (2021). Early Detection and Classification Approach for Plant Diseases based on MultiScale Image Decomposition. Journal of Computer Science, 17(3), 284-295. https://doi.org/10.3844/jcssp.2021.284.295

Gonzalez RC, Woods RE (2007) Digital image processing, 3rd edn. Prentice Hall.

He K, Zhang X, Ren S, Sun J : Deep Residual Learning for Image Recognition. Multimed Tools Appl 77:10437–10453, 2015.

Kafi M, Maleki M, Davoodian N : Functional histology of the ovarian follicles as deter-mined by follicular fluid concentrations of steroids and IGF-1 in Camelus dromedarius. Res Vet Sci 99:37–40. (2015).

Krizhevsky A, Sutskever I, Hinton GE. : ImageNet Classification with Deep Convolutional Neural Networks. Adv Neural Inf Process Syst 1–9, (2012).

LeCun Y., Bottou L., Bengio Y, : Haffner P Gradient-based learning applied to document recognition. Proc IEEE 86:2278–2324, (1998).

Mao, S., Lin, W., Jiao, L., Gou, S., & Chen, J.-W. End-to-End Ensemble Learning by Exploiting the Correlation Between Individuals and Weights. IEEE Transactions on Cybernetics, 1–12. doi:10.1109/tcyb.2019.2931071. (2019).

Muhammad Hammad Saleem, Johan Potgieter and Khalid Mahmood Arif. "Plant Disease Detection and Classification by Deep Learning". Plants 2019, Vol. 8, Issue 11, 468. doi:10.3390/plants8110468.

Richard N Strange and Peter RScott. 2005. Plant disease:a threat to global food security. Annu. Rev. Phytopathol.43(2005),83–116.

Prakash, R. M. Saraswathy, P. G. Ramalakshmi, K. H. Mangaleswari and T. Kaviya.: Detection of leaf diseases and classification using digital image processing, 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), pp. 1-4, Coimbatore (2017).

Sabri, M. A., Filali Y., Khoukhi , H. and Aarab, A.:Skin Cancer Diagnosis Using an Im-proved Ensemble Machine Learning model, 2020 International Conference on Intelligent Systems and Computer Vision (ISCV),pp. 1-5, Fez, Morocco (2020).

Sandler Mark, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, Liang-Chieh Chen. MobileNetV2: Inverted Residuals and Linear Bottlenecks. arXiv:1801.04381,(2019).

Shin H-CC, Roth HR, Gao M, et al Deep Convolutional Neural Networks for Comput-erAided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. IEEE Trans Med Imaging 35:1285–1298, (2016).

Sibiya M, Sumbwanyambe M (2019) A computational procedure for the recognition and classification of maize leaf disease out of health leaves using convolutional neural networks. Agric Eng 1:119–131

Simonyan K, Zisserman: A Very deep convolutional networks for large-scale image recog-nition. ICLR 75:398–406,(2015).

Srdjan Sladojevic, Marko Arsenovic, Andras Anderla, Dubravko Culibrk, Darko Stefanovic, "Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification", Computational Intelligence and Neuroscience, vol. 2016, Article ID 3289801, 11 pages, 2016. https://doi.org/10.1155/2016/3289801.

Sujeet V and Tarun D. : A Novel Approach for the Detection of Plant Diseases International Journal of Computer Science and Mobile Computing Vol. 5 Iss. 7 p. 44 - 54 ISSN: 2320–088X, (2016).

Szegedy C, Vanhoucke V, Ioffe S, et al : Rethinking the Inception Architecture for Comput-er Vision. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, pp 2818–2826, (2016).

Szegedy C, Wei Liu, Yangqing Jia, et al : Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp 1–9, (2015).

Vijai S,. Namita S., Shikha S.: A review of imaging techniques for plant disease detection, Artificial Intelligence in Agriculture, Volume 4, Pages 229-242, ISSN 2589-7217, (2020).

Wäldchen, J., Mäder, P. Plant Species Identification Using Computer Vision Techniques: A Systematic Literature Review. Arch Computat Methods 25, 507-543 (2018). https://doi.org/10.1007/s11831-016-9206-z

Yawen Xiao, Jun Wu, Zongli Lin, Xiaodong Zhao. A deep learning-based multi-model ensemble method for cancer prediction. Computer Methods and Programs in Biomedicine. Volume 153, Pages 1-9. ISSN 0169-2607. https://doi.org/10.1016/j.cmpb.2017.09.005, (2018).