Advanced Binary Classification for Disease Detection in Trees Using a novel Machine-Deep Learning method

Marouane Kihal[1], Lamia Hamza[1] and Mohammed Charif Kihal[2]

[1]*Laboratory of Medical Informatics (LIMED), Faculty of Exact Sciences, University of Bejaia, 06000 Bejaia, Algeria*
[2]*Laboratory of electrical engineering and industrial electronics (L2EI), Faculty of Science and Technology, Department of Electrical Engineering, University of Jijel, 18000 Jijel, Algeria*

Abstract
Detecting plant health is crucial to prevent losses in the productivity and quality of agricultural products. This study focuses on identifying plant diseases through the visual examination of leaf patterns. Specifically, we aim to efficiently determine the health status (diseased or healthy) of lemon trees by analyzing the condition of their leaves using nine different machine learning algorithms optimized with a deep learning approach. Our experimental results demonstrate that this method achieves a high accuracy rate of 93%, surpassing other machine learning techniques. The integration of multiple machine learning algorithms followed by deep learning proves to be a promising solution for effective detection of tree diseases.

Keywords
Trees diseases detection, Binary detection, Machine learning, Deep Learning, Citrus fruit

1. Introduction

Before the advent of AI-based methods, early detection of disease in trees was often hampered by rudimentary, empirical methods. Observers often had to rely on visible signs such as changes in leaf color or obvious external symptoms, limiting the ability to identify diseases at an early stage of development. In addition, the diversity of diseases and the variability of forest environments made it difficult to implement uniform and reliable detection protocols. These challenges highlighted the urgent need for innovative solutions to improve the efficiency and accuracy of tree disease monitoring. Machine Learning (ML) in the field of tree disease detection involves using algorithms to analyze data such as images or sensory data to identify characteristic signs of disease. This approach enables computer systems to learn from data without being explicitly programmed, thus improving the accuracy and efficiency of diagnosis. On the other hand, Deep Learning (DL), an advanced branch of machine learning, uses artificial neural networks to perform complex recognition and classification tasks. In tree disease detection, deep learning enables in-depth analysis of high-resolution images, facilitating early detection of infection or structural damage thanks to its ability to extract significant features and patterns from large quantities of data. Thus, detecting plant health with ML and DL algorithms is crucial to prevent losses in yield and quality of agricultural products by examining visually observable patterns on plants, such as leaves, stems, and fruits.

In this paper, we aim at efficiently binary detection of the health status of lemon trees (diseased or healthy) from the state of the leaves using an approach based on deep learning optimization of nine machine learning algorithms. The main contribution of this article is to design the following:

- Binary detection of the health status of trees from the state of the leaves using an approach based on deep learning optimization of nine machine learning algorithms.
- Comparison of our approach with nine Machine Learning algorithms.
- Application of four different evaluation metrics to compare results.

Proceedings of the DAAfrica'2024 workshop
*Lamia Hamza
†These authors contributed equally.
marouane.kihal@univ-bejaia.dz (M. Kihal); lamia.hamza@univ-bejaia.dz (L. Hamza); mc.kihal@univ-jijel.dz (M. C. Kihal)
0000-0002-6675-7087 (M. Kihal); 0000-0002-5436-3099 (L. Hamza); 0000-0002-2075-2491 (M. C. Kihal)

The rest of this paper is organized as follows: Section 2 reviews related works. Section 3 introduces our proposed approach for detecting diseased trees. In Section 4, we evaluate our results. Finally, Section 5 concludes the paper and suggests potential directions for future research.

2. Related works

Numerous studies have focused on detecting diseases from the leaves of various plants. For tomato, Prajwala et al.[1] proposed a variation of the convolutional neural network model, LeNet, to detect and identify diseases in tomato leaves. For rice, Kawcher et al.[2] introduced a rice leaf disease detection system utilizing machine learning techniques. Additionally, a study on potatoes [3] suggests a model that employs pre-trained models for fine-tuning to extract relevant features from the dataset, followed by a logistic regression classifier. For lemon tree, Banni and Sksvmacet [4] proposed a model that utilises GLCM (Grey Level Co-Occurrence Matrix) algorithms for the detection of citrus leaf disease. However, this study was unable to obtain appropriate outcomes in order to classify the image data. This study yielded an accuracy rate of approximately 85.71%. Recently, more work has been based on machine learning and deep learning algorithms. Pramanik et al. [5] used Transfer Learning-based Deep Learning models, specifically DenseNet-201, ResNet-50, ResNet-152V2, and Xception, to classify lemon leaf diseases. Xception outperformed all other models in terms of accuracy, with 94.34%. Khattak et al. [6] suggested the use of a CNN model to distinguish between healthy fruits and leaves and those that have prevalent citrus diseases, including black spot, canker, scab, greening, and Melanose. This CNN model has a test accuracy of 94.55%. Hassam et al. [7] proposed a single-stream convolutional neural network architecture to identify illnesses in citrus fruits. The expanded citrus dataset (Citrus Fruits, Leaves, and Hybrid Datasets) were employed in the experiment, and the accuracy was 99.4%, 99.5%, and 99.7% respectively. However, the study reveals little redundant information in the collected deep features. Yuan [8] evaluated and compared two deep learning models, DenseNet and MobileNet, for the case study of lemon leaf image classification. This study indicated that MobileNet is more promising in practice. Islam et al. [9] used InceptionV3 and VGG16 deep learning models to classify diseases in citrus leaves, including melanoses, canker, scab, and black spot. InceptionV3 outperforms VGG16 in terms of accuracy. Despite numerous studies in the field of agronomy, no previous work has employed a comprehensive set of ML algorithms, including DL, to optimize the detection of tree diseases.

3. Our approach

In this Section, we propose the use of nine ML algorithms and DL techniques to enhance the detection of tree diseases from leaf images. Various machine learning techniques that we employed for this task, including:

1. **Ada Boost :** A technique of grouping a number of individual weak classifiers all together in a single powerful classifier.
2. **Logistic Regression :** A model that maximises the probability for a binary dependent variable.
3. **Decision Trees :** A technique that divides the data into subsets according to specific values of the input dimensions, it can reveal the patterns correlated with plant diseases.
4. **Random Forests :** A learning algorithm that builds many decision trees in the training process.
5. **Support Vector Machines :** A method aims to select the hyperplane that provides the maximum distance between classes of healthy and diseased leaves in the space of features.
6. **k-Nearest Neighbors :** Categorizes a leaf based on the results of a majority vote on the k nearest neighbors using distance.
7. **Naive Bayes :** Utilize the Bayesian model with strong (naive) hypothesis of feature's independence.
8. **Linear Discriminant Analysis :** A method aims to determine the best way of dividing the different classes.

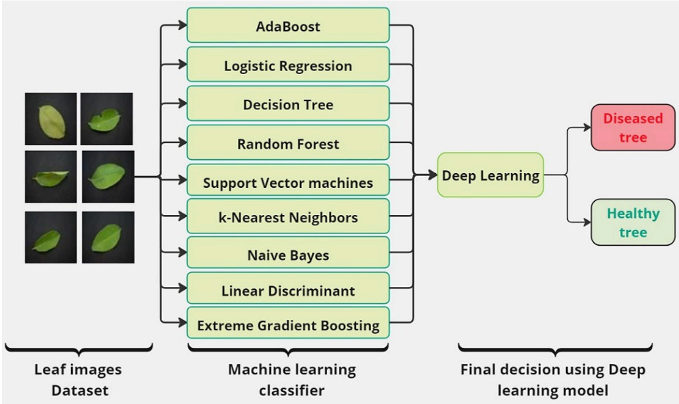

Figure 1: General architecture of our proposed method

9. **Extreme Gradient Boosting :** An optimized gradient boosting.

As mentioned the result of the nine ML algorithms will be passed by a deep learning model to make and optimize the finale decision as shown in Fig. 1.

4. Experimentation and results

In this Section, we will present the details and results of the experiments conducted on images of lemon trees. The images used for this study were obtained from the *Collection of Different Category of Leaf Images*[10]

4.1. Methodology

We employed the nine machine learning algorithms discussed in the previous section, followed by implementing a deep learning model. Specifically, our DL model is a sequential model developed using Keras[11] consisting of five dense layers. The first dense layer has 64 units with a ReLU activation function, taking input data of dimension. The following layers have 128, 256, and 512 units, each also using the ReLU activation. The last dense layer has a single unit with a sigmoid activation function, suitable for binary classification tasks. The model is compiled with the 'adam' optimizer, 'binary crossentropy' loss function. Finally, the model is trained on data for 50 epochs with a batch size of 32.

4.2. Results

The result of experimentation are shown in Table 1. This outcomes demonstrate a clear comparison of various algorithms in terms of Accuracy, Precision, Recall, and F1-Score. Adaboost and SVM both achieve an accuracy of 85%, with Adaboost showing a high recall of 99% and an F1-Score of 90%, indicating strong performance in identifying true positives but a slightly lower precision of 82%. Logistic Regression and XGBoost both achieve an accuracy of 87.5%, with high precision 87% and recall 96%, resulting in an F1-Score of 92%, highlighting their balanced performance. Decision Trees and Random Forests show lower accuracy at 82.5% and 80% respectively, k-NN achieves also 82.5% of accuracy. Linear Discriminant Analysis achieves 85% accuracy, while Naive Bayes has the lowest accuracy at 75%, reflecting its limited effectiveness in this context. Notably, our proposed ML-DL approach outperforms all other algorithms with an accuracy of 93%, precision of 90%, recall of 99%, and an F1-Score of 95%, indicating superior overall performance in terms of both identifying true positives and minimizing false positives. Moreover, we trained the VGG16 model proposed by Islam et al. [9] on the same datasete used to train our model, the VGG16 model obtained about 85% in all evaluation metrics. Furthermore, the loss and accuracy curves show that the loss steadily decreases while the accuracy consistently increases until reaching 93%, indicating good model learning, as shown in Fig. 2.

Table 1
Performance of Different Algorithms

Algorithm	Accuracy	Precision	Recall	F1-Score
Adaptative boosting	85%	82%	99%	90%
Logistic Regression	87,5%	87%	96%	92%
Decision Trees	82,5%	86%	89%	88%
Random Forests	80%	78%	99%	88%
Support Vector Machines	85%	82%	99%	90%
k-Nearest Neighbor	82,5%	80%	99%	89%
Linear Discriminant Analysis	85%	87%	93%	90%
Naive Bayes	75%	76%	93%	84%
eXtreme Gradient Boosting	87,5%	87%	96%	92%
Islam et al. [9]	85%	85%	85%	85%
Our approach	**93%**	**90%**	**99%**	**95%**

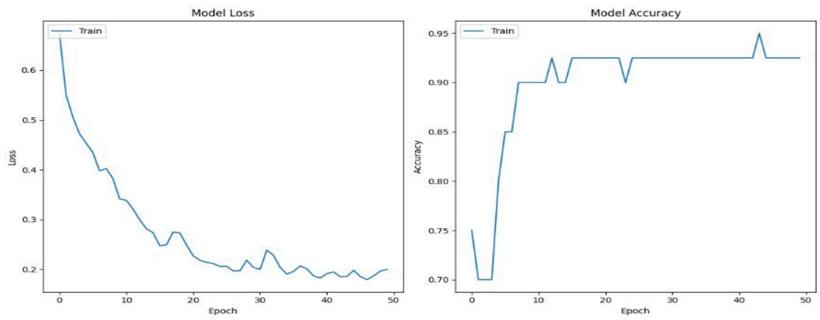

Figure 2: Loss and accuracy model curves

5. Conclusion

In this paper, we have proposed a new approach based on machine learning followed by deep learning to efficiently detect the health status of tree leaves, using nine powerful machine learning algorithms, namely adaboost, logistic regression, decision tree, random forest, support vector machines, k-nearest neighbors, naive bayes, linear discriminant analysis, and extreme gradient boosting. The results presented in the experiment demonstrate that the proposed model outperformed the individual machine learning algorithms on four evaluation measures, achieving accuracy of 93%, precision of 90%, recall of 99%, and an F1 measure of 95%. These results indicate the effectiveness and robustness of the proposed approach, which can be used as an effective solution for tree disease control. Our future research will explore the generalization of this approach to other domains and datasets.

Acknowledgments

This work has been sponsored by the General Directorate for Scientific Research and Technological Development, Ministry of Higher Education and Scientific Research (DGRSDT), Algeria.

References

[1] P. Tm, A. Pranathi, K. SaiAshritha, N. B. Chittaragi, S. G. Koolagudi, Tomato leaf disease detection using convolutional neural networks, in: 2018 eleventh international conference on contemporary computing (IC3), IEEE, 2018, pp. 1–5.
[2] K. Ahmed, T. R. Shahidi, S. M. I. Alam, S. Momen, Rice leaf disease detection using machine

learning techniques, in: 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI), IEEE, 2019, pp. 1–5.

[3] D. Tiwari, M. Ashish, N. Gangwar, A. Sharma, S. Patel, S. Bhardwaj, Potato leaf diseases detection using deep learning, in: 2020 4th international conference on intelligent computing and control systems (ICICCS), IEEE, 2020, pp. 461–466.

[4] R. Banni, L. Sksvmacet, Citrus leaf disease detection using image processing approaches, International Journal of Pure and Applied Mathematics 120 (2018) 727–735.

[5] A. Pramanik, A. Z. Khan, A. A. Biswas, M. Rahman, Lemon leaf disease classification using cnn-based architectures with transfer learning, in: 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), IEEE, 2021, pp. 1–6.

[6] A. Khattak, M. U. Asghar, U. Batool, M. Z. Asghar, H. Ullah, M. Al-Rakhami, A. Gumaei, Automatic detection of citrus fruit and leaves diseases using deep neural network model, IEEE access 9 (2021) 112942–112954.

[7] M. Hassam, M. A. Khan, A. Armghan, S. A. Althubiti, M. Alhaisoni, A. Alqahtani, S. Kadry, Y. Kim, A single stream modified mobilenet v2 and whale controlled entropy based optimization framework for citrus fruit diseases recognition, IEEE access 10 (2022) 91828–91839.

[8] Y. Yuan, Computer vision and deep learning for precise agriculture: A case study of lemon leaf image classification, in: Journal of Physics: Conference Series, volume 2547, IOP Publishing, 2023, p. 012024.

[9] S. F. Islam, N. Chakrabarty, M. S. Uddin, Citrus leaf disease detection through deep learning approach, Bulletin of Electrical Engineering and Informatics 13 (2024) 2643–2649.

[10] M. Nagadia, Collection of different category of leaf images, Kaggle repository, 2023. URL: https://www.kaggle.com/datasets/meetnagadia/collection-of-different-category-of-leaf-images, last accessed 2024-07-25.

[11] F. Chollet, et al., Keras, GitHub repository, 2015. URL: https://github.com/fchollet/keras, last accessed 2024-07-25.