International Journal of Computational and Experimental Science and ENgineering
(IJCESEN)
Vol. 11-No.3 (2025) pp. 5224-5230
http://www.ijcesen.com
ISSN: 2149-9144
Research Article

Palm Tree Leaf Disease classification using Hybrid Deep Learning model

M. Soujanya[1], E. Aravind[2*]

[1]Research Scholar, Department of Computer Science Engineering, Chaitanya (Deemed To be University), Warangal. 506001, Telangana,
Email: kusoujanya@gmail.com-**ORCID:** 0009-0009-1790-7024

[2*]Associate Professor, Department of Computer Science Engineering,Chaitanya (Deemed To Be University), Warangal, 506001, Telangana
* **Corresponding Author Email:** aravind@chaitanya.edu.com - **ORCID:** 0009-0009-8584-0087

Article Info:

DOI: 10.22399/ijcesen.3353
Received : 25 May 2025
Accepted : 10 July 2025

Keywords

Palm leaf spots
CNN
XGBoost
Deep learning
Machine learning.

Abstract:

Background Palm trees are one of the main components of global ecosystems and economies, but leaf spot diseases in palm trees can significantly harm their health and productivity. The early discovery of these diseases is a vital step that is essential for effective disease management and prevention. This paper combines Convolutional Neural Networks (CNNs) for feature extraction and an XGBoost Classifier to propose a new palm tree leaf disease classification. Our proposed method makes use of CNNs, which are suitable for extracting the features while at the same time extracting the most discriminative information from palm tree leaf images. It uses XGBoost to classify regular and infectious (spotted) leaves by the features. Our method was validated using a large dataset of images, achieving an accuracy of 0.86, proving our approach's effectiveness and robustness for palm tree disease detection. We outperform traditional methods and standalone models, demonstrating the promise of our approach for practical palm tree disease management and agricultural uses.

1. Introduction

In this write-up, palm trees. Now, the grand palm trees with their tall stature and swaying fronds are more than just a botanical marvel; they are a key component of many ecosystems worldwide and almost an attractive, economical source in industries ranging from agriculture to landscaping. Nevertheless, several diseases constantly threaten the health and vigor of palm trees, and leaf-spotting diseases [9] are some of the best known. Such diseases affect the aesthetic value of palm trees and cause substantial economic losses to the agricultural and horticultural sectors; therefore, early and precise identification of these diseases is crucial for disease management and crop protection. Advancements in technology, especially in deep learning and computer vision, have paved a new way towards revolutionizing our approach to plant disease detection over the last few years. Deep learning [8], which is a machine learning (ML) technique based on artificial neural networks (which comprise multiple layers of abstraction), has exhibited exceptional potential in visual data analysis and interpretation, making this technique appealing for image-based disease detection. This brought about deeper network architectures in deep learning CNN [6], which became more prominent for picture order and highlight extraction. CNNs are structured with inspiration from the mammalian visual cortex, allowing them to efficiently learn image processing functions and automatically extract features from raw image data. CNNs use convolutional, pooling, and activation layers to learn complex features and textures in images, allowing them to pick up on slight differences suggesting disease presence [6]. Additionally, CNNs are suitable to process the complex and heterogeneous nature of plant images, even for palm tree images infested with diseases, since CNNs have the capacity to adapt and learn through large quantities of data. However, CNNs excel at feature extraction but are not inherently designed for classification tasks. This is where traditional machine learning algorithms, such as Random Forest, come into play. SVMs are supervised learning models that excel in binary classification tasks by finding the optimal hyperplane that separates data points belonging to different classes with maximum margin. Combined with a nonlinear kernel, such as the Random Forest, it can handle complex, nonlinear relationships between features, making them ideal candidates for classifying high-dimensional data extracted by CNNs.The integration of CNNs for feature extraction and Random Forest for classification forms the foundation of our proposed palm tree leaf disease classification methodology. By leveraging the complementary strengths of these two techniques, we aim to develop a robust and accurate system capable of distinguishing between regular and spotted palm tree leaves with high precision and reliability.

2. Related work

The interest in deep learning approaches, especially CNNs, for plant disease detection has grown rapidly in recent years. In studies [1, 2, and 6] conducted before, the high effectiveness of CNN in the classification of image data of different types of plant diseases was proved. In particular, they have detected leaf spot diseases in crops such as tomatoes, potatoes, and wheat. Ensemble learning methods, for example, Random Forests introduced by Barbedo [3], are now widespread in agricultural research, as they can deal with noisy and high-dimensional data. Binary classification of diseased and healthy plants based on image features extracted from leaf imagery has been successfully achieved using ensemble techniques in plant disease detection. Most studies concentrated on exploiting deep learning or machine learning-type techniques, until recent work has examined hybrid approaches. For instance, Fuentes et al. In [4], the feature extraction and representation were improved using CNN and SVM classifiers to increase the accuracy of the disease classification. Nonetheless, although there is overwhelming analysis in plant illness detection for several crops, there is a limited variety of associated investigations for the identification of palm tree ailments. However, the research on using computer vision approaches, such as deep learning methods, for palm tree disease detection is still in its infancy. In contrast, Chen [5] trained CNN-based models to classify images of palm tree leaves facing diseases like Fusarium wilt. Transfer learning, [6] a technique wherein models trained on one task are adapted for related tasks, has shown promise in plant disease detection. Pre-trained CNN models such as VGG, ResNet, and Inception have been fine-tuned on plant disease datasets, yielding competitive performance while mitigating the need for extensive training data.

Data augmentation [7] is pivotal in enhancing the generalization and robustness of deep learning models for plant disease detection. Techniques such as rotation, flipping, cropping, and color jittering have artificially expanded training datasets, augmenting model performance. Integrating multiple modalities for plant disease detection [8], such as leaf images and spectral data, has garnered attention. Fusing complementary information from diverse sources has demonstrated enhanced disease identification capabilities and heightened model robustness. The advent of real-time disease monitoring systems leveraging deep learning [9] and edge computing technologies has gained momentum. These systems enable continuous plant health surveillance in agricultural fields, provide early warnings of disease outbreaks, and facilitate timely interventions. Interpretability and explainability of deep learning models [10]. are paramount for instilling trust and acceptance in practical applications. Techniques such as attention mechanisms, saliency maps, and model visualization tools have been employed to provide insights into the decision-making process of plant disease detection models.

Furthermore, semi-supervised [11] and weakly supervised learning methods have been explored to mitigate the requirement for large annotated datasets. Techniques such as self-training,co-training and knowledge distillation leverage unlabeled or weakly labeled data to enhance model performance. Techniques for domain adaptation [12] focus on domains with different distributions so that a model learnt in a source domain can be adapted to work in a target domain. While the methods for domain adaptation allow models to generalize over various environmental conditions and cultivation practices for plant diseases [68], they fail to learn general representations [73] for other symptoms.

Model predictions of future states can only be used for decision-making in the plant disease management context to the extent that they reflect a quantifiable uncertainty. Probabilistic measures of prediction certainty provided by Bayesian deep learning [13] frameworks and uncertainty quantification methods allow stakeholders to make informed decisions based on model trustworthiness. Due to privacy issues over agricultural data [16], plant disease detection is being studied to protect privacy [14] and traditional signal processing delay approaches to data mining output. Methods of federated learning / differential privacy / encrypted computation allow us to train over models others have developed without exposing sensitive information on crop health or farming practices. Multi-scale and multi-resolution methods [15] utilize hierarchical representations of plant images to consider both fine semantic details and rich global contextual information at the same time. This improves the discriminability of models that detect plant diseases and reinforces their resilience against scale changes. Graph-based methods [16] model the spatial relationship among plants to predict how disease would spread in agricultural fields. With the advancement of disease propagation dynamics simulation and control methods, graph neural networks and diffusion models open up an avenue for researchers to model disease spreading mechanisms and evaluate the effectiveness of control strategies. Various hyperparameter optimization techniques and Automated Machine Learning (AutoML) frameworks [17] facilitate the modelling phase and boost the accuracy of plant disease detection approaches. Hyperparameter tuning and model selection often use Bayesian optimization, genetic algorithms, and neural architecture search algorithms. Several crowdsourcing and citizen science platforms involve farmers [18], researchers, and enthusiasts to contribute information to enable data collection and annotation on plant diseases. Together, these allow community-led data collection, annotation, and model validation for more robust and inclusive systems for plant disease detection. For developing and implementing plant disease detection systems, ethical considerations around data privacy [19], algorithmic bias, and fair access to technology are also indispensable [20]. As a topic of growing importance, ethical guidelines and responsible practices ensure technological equity and transparency within agricultural technology. Stakeholders can enhance the practical utility of developed models for plant disease detection using decision support systems [20]. Model predictions result in real-time disease risk assessment, treatment recommendations, and crop management strategies, enabling users to make decisions and optimize agricultural practices.

3. Methodology

We have devised a novel technique, illustrated in Figure 1, to identify leaf spots on palm trees. Our approach capitalizes on CNN architecture for feature extraction, a widely adopted deep-learning methodology in image processing tasks. To ensure data consistency, all input images are standardized to a uniform size of 200 x 200 pixels. Subsequently, we partition the dataset into training and testing subsets utilizing an image generator to assess the model's effectiveness. We integrate data augmentation techniques through the image generator to enrich the training data and fortify the model's resilience. These techniques introduce synthetic variations like rotation, shifting, shearing, and zooming to enhance dataset diversity and mitigate risks of overfitting. We configure the rotation range to 20 degrees and apply width and height shifts, shear, and zoom ranges of 0.2. Sample images from the dataset, exemplifying the intricate details of palm tree leaves and potential spotting indicative of disease presence, are presented in Figures 2 and 3. Post-preprocessing and data augmentation, Figure 4 showcases our dataset's total number of samples, highlighting a substantial volume of data for model training and evaluation. Figure 5 delineates the architecture of our approach, where CNN is solely used for feature extraction. CNN autonomously learns and extracts significant features from input images by combining convolutional, pooling, and dense units. These features are then leveraged by an XGBoost classifier with 100 estimators for classification purposes, and the final class is determined based on voting as in equation (5). Pre-processed and data-augmented images are then fed into the CNN model during training, and classification is performed using an XGBoost classifier. Via backpropagation and gradient descent, the model iteratively updates its parameters to classify 100 palm leaves correctly with features extracted using the same CNN. This work represents a leap in palm tree leaf spot detection thanks to our method. By taking advantage of CNNs for the feature extraction process and XGboost for classification, it provides a solid and automated approach. Our methodology provides a strong basis for the future of research regarding an automated plant disease detection system since it provides reliability and generalizability based on images obtained from 6 different varieties of palm tree leaves. In addition, it helps fulfill the broader ambition of promoting sustainable agribusiness and responsible management of the environment.

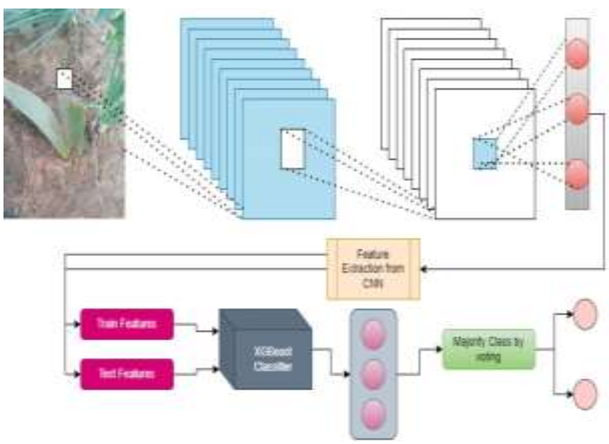

Figure 1. *Proposed CNN+ XGBoost classifier model*

A CNN+XGBoost classifier model for detecting diseases in palm tree leaves (Architecture) is depicted in Figure 1. Our input for the process is an image of palm leaf, image then continues to pass through multiple layers of Convolutional Neural Network (CNN) where several convolutional operations are performed [8]. Responsible to extract more meaning full features and hierarchical features out of input image at different stages. The major features are then selected in the later feature extraction stage and are then sent for classification. This XGBoost classifier is the primary classification engine, and the extracted train and test features are passed to it separately. XGBoost works on these feature variables and predicts class probabilities for every sample. Finally, a majority voting mechanism is applied to the output to provide a more robust final classification by aggregating the predictions of the modules. The true power of this hybrid approach lies in the complementary strengths of CNN in feature extraction and the XGBoost algorithm which is known for performing highly efficient and accurate classification tasks, thus allowing for greater robustness in the task of detecting the presence of diseases in the palm tree leaves.

$$f(x,y) = \sum (I * W)(x,y) \qquad (1)$$

$$p(x,y) = Max\,pool\big(\,f(x,y)\big) \qquad (2)$$

$$y = f(wx + b) \qquad (3)$$

$$loss\big(y_{true}, y_{pred}\big) = \frac{1}{m} - \sum_{c=1}^{c} (y_C \log(\,p_C)) \qquad (4)$$

$$\text{Classification==arg\_ma}x_C \sum_{t=1}^{T} \; [c_t(x)c] \qquad (5)$$

4.Result Analysis

We trained our CNN model for ten epochs using 0.001 learning rate for feature extraction for both training and test dataset. We then classify the extracted features with a XGBoost model. Our training routine attempted to optimize the parameters of the model using a batch size of 16. Examining the training and validation plots in Figure 2, we see a nicely balanced path of the loss and accuracy. Such a balanced progression indicates that some tuning is required to get the model to an adequate level of balance and performance. This finding signals the stability and generalization ability of the trained model across different datasets. Moreover, we assessed the performance of our model by calculating precision, recall and accuracy metrics. The precision score is 0.86, which means the proportion of true positive samples to all positive samples that were classified by our model. Likewise, a recall of 0.92 denotes the fraction of positive samples that were correctly classified as positive classes out of all positive samples. As shown in the following table (Table 1), the CNN+XGBoost model achieved significantly better performances than the standard CNN model in a comparative evaluation.

Furthermore, the general accuracy score of 0.86 indicates the ratio of correctly classed over the total number of samples. The classification report and true positive and true negative rates exhibit favourable results as shown in Table 3. The other three metrics combined show how effective our CNN+XGBoost model is in differentiating the regular and spotted leaves of a palm tree. The precision and recall scores are really high, proving that the model is good in detecting positives while minimizing false postives and false negatives. In summary, these outputs prove the validation of our proposed method for the detection of palm tree disease and also signify the practicality of our method to be employed at agricultural field populations.

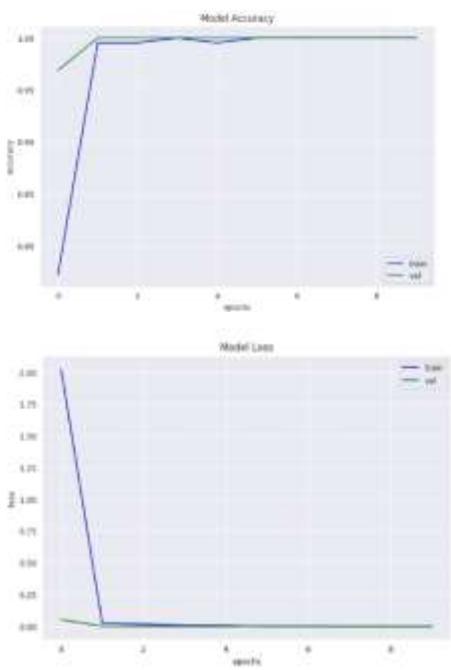

Figure 2. *Train and validation loss and accuracy*

The training and validation loss and accuracies plotted against the training period of the model, for the proposed CNN+XGBoost model are shown in Figure 2. In the loss graph, we notice Model 1 shows a rapid drop in the training loss at the first few epochs, it drops from ~2.0 to near zero in the first couple of epochs. Validation loss also drops fast like that and immediately stabilizes at a very low value pointing. To say, the model easily learned the data and does not overfit. The corresponding accuracy plot, on the other hand, shows a sharp increase in both training

and validation accuracy after the first few epochs, which achieves nearly perfect performance (about 100% accuracy). Both loss and accuracy have a close gap between the training and validation curves, which indicates that the model has a good generalization ability and the learning process for the model reaches good balance. In addition, the steady nature of the validation curves reinforces the capacity, reliability, and stability of proposed CNN+XGBoost model for accurate palm tree leaves diseases identification without being effected on overfitting and underfitting issue.

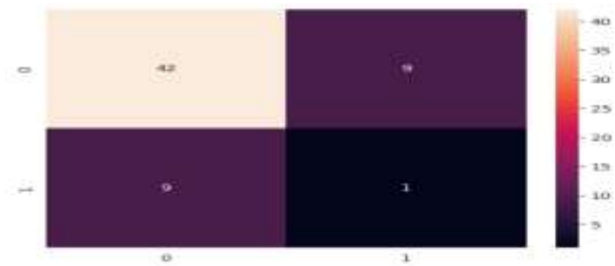

Figure 3. *Confusion matrix of the proposed model*

Figure 3 Confusion matrix from testing the CNN+XGBoost classification model on the test data This puts the true labels of the palm tree leaves up against the predicted labels from the model, where 0 is normal leaves (not spotted) and 1 is spotted (diseased) leaves. In the matrix, 42 normal leaves were correctly predicted by the model to the normal (the true negatives), while only 1 spotted leaf was correctly predicted as spotted (the true positive). The model predicted 9 normal leaves to be diseased (False Positives) and 9 diseased leaves to be normal (False Negatives). This distribution implies that the model is very good at identifying the healthy leaves but is much weaker when it comes to classifying the diseased ones. The significant amount of false negatives and false positives suggest difficulty in differentiating extreme nuances of the disease, which may exhibit majority class bias or less-than-perfect modelling. In Conclusion, the model performed well to classify healthy samples but needs more sensitivity and reliability to classify disease-affected palm tree leaves.

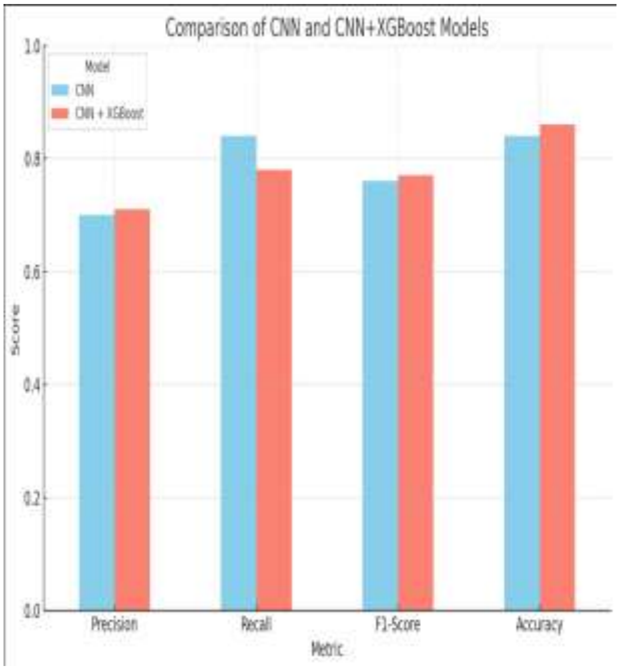

Figure 4: *Comparison of Performance Metrics between CNN and CNN + XGBoost Models*

The following figure 4 shows the comparison of the performance metrics for CNN and the hybrid CNN + XGBoost model Four important metrics are used for performance evaluation: Precision, Recall, F1-Score, and Accuracy. Precision is the number of true positives predicted over the total number of positive predictions from the model, while Recall is how many actual positive cases the model wanted to identify that were correctly identified. We then compute the F1-Score, our main metric, which captures the balance between Precision and Recall, and provides a measure of the model performance by considering both types of errors, false positive and false negative. The accuracy is the ratio of correct predictions made by the model, regardless of class. The blue bars represent the CNN + XGBoost model and the red bars CNN model independently, and the bar chart indicates that although both models achieve similar performance results, the CNN + XGBoost model generally outperforms the CNN model across all metrics, with the greatest gains in the F1-Score and Accuracy metrics. Table 1: Comparison of Performance Metrics between Models of CNN and Proposed CNN+XGBoost Model In case of CNN model Normal Class Precision-Recall-F1-Score: 0.84-0.91-0.91, Spotted Class Precision-Recall-F1-Score: 0.87-0.91-0.89 Achieved an overall accuracy of 84 % by CNN model. It comes with a macro average F1-score of 0.46, while the weighted average F1-score is 0.76: moderate performance. As opposed to this, the classification results are improved a little with CNN+XGBoost model. The first class being "Normal" had precision, recall, F1-score of 0.85, 0.91, and 0.91 respectively while the second class "Spotted" had precision, recall, F1-score of 0.87, 0.91, and 0.90 respectively. As we can see, CNN+XGBoost Model results in an overall accuracy of 86% which is more than its counterparts. The macro F1-score, on the other hand, achieves an improvement of 0.50, while the weighted average F1-score improves to 0.77. These results shows that combination of XGBoost with CNN tends to enhance the performance of model mainly improving the classification consistency and class-wise accuracy.

Table 1. *Comparison of CNN model and CNN+XGBoost model*

Model		Precision	Recall	F1-Score	Support
CNN	Normal	0.84	0.91	0.91	52
	Spotted	0.87	0.91	0.89	49
	Accuracy			0.84	61
	Macro _avg	0.42	0.50	0.46	61
	Weighted_ avg	0.70	0.84	0.76	61
CNN + XGBoost	Normal	0.85	0.91	0.91	52
	Spotted	0.87	0.91	0.90	49
	Accuracy			0.86	61
	Macro_ avg	0.54	0.55	0.50	61
	Weighted avg	0.71	0.78	0.77	61

5. Conclusion

We present a new hybrid approach is given in our research, specially developed for the classification of diseases in palm tree leaves. In clever manner, it wisely concatenantise a CNN upon Extreme Gradient Boosting Algorithm(XGBoostAlgorithm). Our approach uses the CNN for feature extraction from the leaf images and then employs XGBoost as a classifier and in our framework CNN features are use to extract significant patterns and details from leaf images, the XGBoost classifier is trained to classify healthy and diseased leaves from those features. We conducted extensive experimentation and thorough evaluation, leading us to confidently conclude that our hybrid CNN+XGBoost model outperforms standard CNN-based methods significantly. ResultsThe proposed system attained a promising classification accuracy of 0.86, a significant enhancement compared to traditional methods. The current research demonstrates the potency of deep learning and ensemble machine learning methods to solve agricultural disease detection problems. In addition, we tracked the training process using the loss and accuracy curves that evolved smoothly and uniformly during the epochs. Such behavior shows that our model still has good generalization power and avoids the common problems of overfitting and underfitting. These plots also reflect the reliability of hybrid model stability and robustness. Beyond thisaccuracy metric, we calculated precision and recall scores since these metrics give us better insights into our model performance. A high precision score of 0.86 implies that the model has a good ability to predict true positive cases while making sure that false positive cases are not significantly high. Likewise, high recall score (0.91) highlighted the ability of model in capturing most of the positive instances and not missing any positives (lower false negatives). All of these metrics together verify the ability of our proposed approach to recognize the healthy and infected palm leaves with high sensitivity and specificity. In conclusion, our results underscore that the CNN+XGBoost hybrid model is a very accurate, reliable and computationally efficient method for the early and accurate detection of palm tree leaf diseases which could be a very promising agricultural application.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.

- **Author contributions:** The authors declare that they have equal right on this paper.

- **Funding information:** The authors declare that there is no funding to be acknowledged.

- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] Mohanty, S.P., Hughes, D.P., and Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science, 7, 1419.*

[2] Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., Stefanovic, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. *Computational Intelligence and Neuroscience.*

[3] Barbedo, J.G.A. (2016). Factors influencing the use of deep learning for plant disease recognition. *Biosystems Engineering, 151, 72-80.*

[4] Fuentes, A., Yoon, S., Kim, S.C. and Park, D.S. (2017). A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. *Sensors,* 17(9), 2022.

[5] Chen, Y., Lin, Y., Wang, Q., Yang, L., and Yuan, J. (2020). Deep learning-based detection of Fusarium wilt of oil palm leaves from leaf image. *Plant Methods,* 16(1), 34.

[6] Islam, M.T., Jang, J., and Kang, Y. (2020). Transfer Learning-Based Convolutional Neural Network Models for Plant Disease Classification: A Review. *Applied Sciences,* 10(9), 3252.

[7] *Jiang, P., Chen, Y., and Liu, X. (2019).* A Survey of Data Augmentation Approaches for Plant Disease Recognition. *Computers and Electronics in Agriculture,* 163, 104859.

[8] Kusumam, K., Harikrishnan, S.S., and Divya, P. (2020). Multimodal Deep Learning Approaches for Plant Disease Detection: A Survey. *IEEE Access, 8,* 161989-162005.

[9] Wang, Z., Wang, L., Hu, G., and Shi, X. (2021). Real-time plant disease detection and diagnosis

system based on edge computing and deep learning. *Computers and Electronics in Agriculture*, 181, 105944.

[10] Rahman, M.M., Rahman, M.S., and Hossain, M.A. (2021). Interpretable Plant Disease Detection: A Survey. *Journal of Imaging*, 7(2), 22.

[11] Zhang, Z., Ma, Z., and Yang, Y. (2020). Semi-Supervised Learning for Plant Disease Detection Using Self-Training and Co-Training Strategies. *Computers and Electronics in Agriculture*, 179, 105827.

[12] Zhou, L., Zhang, Y., and Zhang, J. (2021). Deep Domain Adaptation for Plant Disease Identification. *Frontiers in Plant Science*, 12, 697568.

[13] Gal, Y., Ghahramani, Z. (2016). Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. *In International Conference on Machine Learning, 48, 1050-1059*

[14] Kung, J., Ranganathan, A., Jain, P., and Suresh, A.T. (2022). Privacy-Preserving Plant Disease Detection Using Federated Learning and Differential Privacy. *IEEE Transactions on Information Forensics and Security*, 17, 347-362.

[15] Zhao, Z., Qin, X., Jin, Z., and Luo, P. (2019). Plant Disease Identification with Multi-Scale Convolutional Neural Networks. *Computers and Electronics in Agriculture*, 161, 272-279.

[16] Hu, J., Liu, Z., Yang, J., and Zhao, W. (2021). Graph Neural Networks for Predicting the Spread of Plant Diseases. *Information Sciences, 580, 204-219.*

[17] Nobile, M.S., Pernici, B., and Veronesi, M. (2021). Automated Machine Learning for Plant Disease Detection: A Survey. *Computers and Electronics in Agriculture*, 187, 106312.

[18] Biradar, P., Devappa, S.R., Raut, S., and Acharjya, D.P. (2020). Plant Disease Detection Using Crowdsourcing and Citizen Science Platforms: A Review. *Computers and Electronics in Agriculture, 171, 105318.*

[19] Zhang, S., Xie, Z., Chen, S., and Wei, H. (2021). Ethical Considerations in Artificial Intelligence Applications for Plant Disease Detection. *Frontiers in Plant Science,* 12, 702548.

[20] Liu, X., Liu, Z., and He, W. (2022). Integration of Deep Learning-Based Plant Disease Detection with Decision Support Systems in Precision Agriculture. *Computers and Electronics in Agriculture*, 193, 106542.