SJER Scientific Journal of Engineering Research
futuristech PT. Teknologi Futuristik Indonesia

Date of publication November 25, 2025, date of current version November 25, 2025. *e-ISSN: 3109-1725*
Digital Object Identifier 10.64539/sjer.v1i4.2025.330

Article

Ensemble Learning Framework for Image-Based Crop Disease Detection Using CNN Models

Chidi Ukamaka Betrand[1,*], Mercy Eberechi Benson-Emenike[1], Douglas Allswell Kelechi[1], Chinwe Gilean Onukwugha[1], Nneka Martina Oragba[1]

[1] Department of Computer Science, Federal University of Technology Owerri, Nigeria; chidi.betrand@futo.edu.ng
* Correspondence

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Abstract: Crop diseases pose a significant threat to global food security, causing substantial yield losses estimated at 10-40% annually. Traditional methods of disease identification, reliant on visual inspection by farmers or experts, are often subjective, time-consuming, and limited by the availability of specialists. This study proposes an ensemble learning framework for robust image-based crop disease detection, specifically designed to address the challenges of heterogeneous, non-Independent and Identically Distributed (non-IID) agricultural datasets in decentralized environments. Utilizing the Plant Village dataset, we implement a stacking ensemble model integrating diverse Convolutional Neural Networks (CNNs) such as VGG (Visual Geometry Group), ResNet, and Inception as base learners, with a meta-learner to optimize prediction fusion. The system employs comprehensive data preprocessing, including resizing, normalization, noise removal, segmentation, and augmentation, to enhance robustness against real-world variability. Transfer learning with ResNet50 was adopted as a baseline model. The baseline ResNet50 achieved 59% test accuracy across seven grape and potato disease classes. The ensemble model improved performance, attaining 63% accuracy with average precision, recall, and F1-scores of 56%, 52%, and 52% respectively. Class imbalance remained a limiting factor for certain categories. The ensemble learning approach outperformed individual models, demonstrating improved generalization across diverse datasets. Although computational demands and imbalance challenges persist, the system provides a promising AI-driven pipeline for accurate crop disease diagnosis, supporting sustainable agricultural practices.

Keywords: Crop; Disease Detection; Ensemble; Food; Image; ResNet.

1. Introduction

Agriculture remains crucial for global food security and economic development, especially in developing nations. Despite its importance, this sector faces ongoing challenges from climate shifts, pest invasions, and most importantly, crop diseases. Research indicates that plant diseases cause roughly 10-40% yearly crop yield losses worldwide, threatening both food availability and economic stability [1]. The traditional approach to crop disease identification relies on visual inspection by farmers or agricultural experts. This method suffers from several drawbacks: limited availability of specialists, subjective assessments, time inefficiency, and the challenge of spotting subtle or early-stage symptoms [2]. These limitations have sparked interest in technological solutions, particularly those using artificial intelligence and image analysis.

The field of machine learning, especially deep learning approaches, shows great promise for agricultural disease detection. Convolutional Neural Networks (CNNs) have gained popularity for plant disease identification because they can extract complex visual features directly from images [3]. Various architectures like AlexNet, VGGNet, and ResNet have achieved impressive accuracy on standard plant disease datasets, surpassing conventional methods that depend on manual feature extraction. Yet agricultural datasets present unique challenges—they're typically heterogeneous and non-Independent and Identically Distributed (non-IID). They vary based on geographic location, plant types, disease categories, image quality, environmental factors, and camera equipment. These variations create distributional differences that make it difficult for single models to perform consistently across diverse datasets [4].

Furthermore, modern farming increasingly relies on Internet of Things (IoT) devices and distributed management systems that collect data from multiple sources. This decentralized approach to data gathering, while advantageous for scaling, introduces complications due to data imbalances, noise, and inconsistent standards across collection points [5]. Ensemble learning offers a promising strategy to address these issues. This approach combines multiple base models to enhance predictive performance and adaptability. Techniques such as bagging, boosting, and stacking provide frameworks to overcome single-model limitations. For example, Random Forests minimize variance by averaging results from numerous decision trees, while boosting methods like AdaBoost and XGBoost reduce bias by emphasizing difficult classification cases [6]. Through ensemble learning, we can potentially strengthen crop disease detection systems, particularly in scenarios characterized by data diversity, class imbalance, and multi-crop complexity.

This study explores how ensemble methods can improve image-based crop disease detection, focusing specifically on enhancing generalization with non-IID, decentralized data. Automated crop disease detection represents a sophisticated pattern recognition challenge that requires computer systems to identify subtle visual differences between healthy and diseased plant tissues from digital photographs. This multidisciplinary field integrates computer vision, machine learning, plant pathology, and agricultural engineering to offer comprehensive solutions for modern farming [7]. The complexity of this domain arises from the intricate nature of plant-pathogen interactions, which manifest diverse visual symptoms across varying temporal and spatial scales, ranging from microscopic cellular changes to macroscopic alterations in plant morphology, creating a hierarchical pattern recognition problem that demands sophisticated analytical approaches [8].

Digital image analysis serves as the technological core for automated disease detection, with farmer-captured images containing rich information on plant health encoded in color patterns, texture variations, and morphological changes [9]. However, extracting meaningful diagnostic information from raw pixel data necessitates sophisticated processing techniques to handle variations in image quality, lighting, and camera specifications. A fundamental challenge in agricultural image processing is the inherent variability of natural environments, where unlike controlled laboratory settings, field conditions introduce numerous confounding factors including variations in solar illumination, weather conditions, wind movement causing motion blur, and varying camera angles resulting in perspective distortions that robust processing algorithms must accommodate [10].

Ensemble learning has proven consistently effective in improving both accuracy and robustness, with studies demonstrating typical accuracy improvements of 2-5% plus enhanced robustness to environmental variations [5], [11]. The theoretical foundations for ensemble learning in agriculture are well-established, drawing from statistical learning theory, information theory, and pattern recognition, with the diversity principle being particularly valuable as different models may excel at detecting various visual patterns or perform better under specific environmental conditions, enabling robust disease detection across diverse scenarios. However, substantial gaps remain between laboratory performance and real-world effectiveness, with the non-IID nature of agricultural data creating significant challenges for model generalization, consistently showing degradation when models encounter unfamiliar environmental conditions, crop varieties, or image acquisition parameters [12], [13].

The domain adaptation challenge serves as a major barrier, as systems performing excellently in controlled research often show significant accuracy drops in real agricultural environments due to different lighting, equipment, and geography, while data heterogeneity through geographic, temporal, and equipment variations creates complex distribution shifts difficult to address traditionally.

In this research, we offer an ensemble learning system based on stacking that combines many CNN architectures as base learners, including VGG, ResNet, and Inception's concept with a meta-learner that optimizes the fusing of their outputs. This architecture improves the system's capacity to generalize across many crop disease categories, reduces dataset variability, and increases prediction accuracy.

2. Methods

Ensemble learning is a powerful machine learning paradigm that synergistically combines the predictions of multiple individual models to achieve superior overall predictive performance compared to using a single model [14] - [16]. This strategy is particularly effective in enhancing accuracy, robustness, and generalization capabilities, especially when tackling complex and noisy datasets like those encountered in crop disease detection [17], [18]. The core principle behind ensemble methodology is that a committee of diverse models can collectively make better decisions than any single expert [19], [20]. This improvement stems from reducing errors that individual models might make due to variance, bias, or noise in the data. Ensemble methodologies can be broadly categorized based on how the base classifiers interact during training (sequentially or concurrently) and the strategies employed to combine their outputs. The diversity among ensemble members is crucial, as they should ideally make different errors to maximize the benefit of their combination [21] - [23].

Stacking aims for the highest possible generalization accuracy by learning how to optimally combine the predictions of diverse base learners [24]. It is a powerful technique that can reduce both bias and variance and is particularly effective for combining heterogeneous weak learners [6]. The core idea is to train a "meta-learner" that takes the predictions of several base models as its input and learns to make a final, more accurate prediction.

2.1. Data Acquisition

The dataset utilized for this project is the PlantVillage dataset, a widely recognized and publicly available collection of plant leaf images. This dataset is a cornerstone for research in plant pathology using computer vision due to its comprehensive nature. The PlantVillage dataset is publicly available in kaggle and can be accessed via, https://www.kaggle.com/datasets/sarahgmn/plant-village-dataset) and the original PlantVillage project website [9], [12]. It comprises a large number of high-resolution images of healthy and diseased plant leaves. The dataset is well-structured, typically organized into folders where each folder represents a specific plant species and disease combination (or healthy state). The dataset covers a variety of plant species (e.g., apple, corn, grape, potato, tomato, etc.) and numerous disease categories for each species, as well as healthy leaf images. This diversity is crucial for training a robust model capable of identifying a wide range of plant ailments. The dataset typically contains over 50,000 images, providing sufficient data volume for training deep learning models effectively.

2.2. Preprocessing

Seven classes of photos of grape and potato leaves, representing both healthy and diseased categories, make up the dataset used in this study as shown in Table 1. The dataset was split into training, validation, and test sets using a roughly 70:20:10 ratio to provide dependable model training and assessment. All photos were preprocessed before being ingested into the model. This involved scaling the pixel values to the range [0,1] by dividing them by 255. During training, data augmentation specifically, random image rotations of up to 20 degrees—was used to improve generalization and decrease overfitting. This allowed the model to acquire more robust and rotation-invariant features.

2.3. Evaluation Metrics

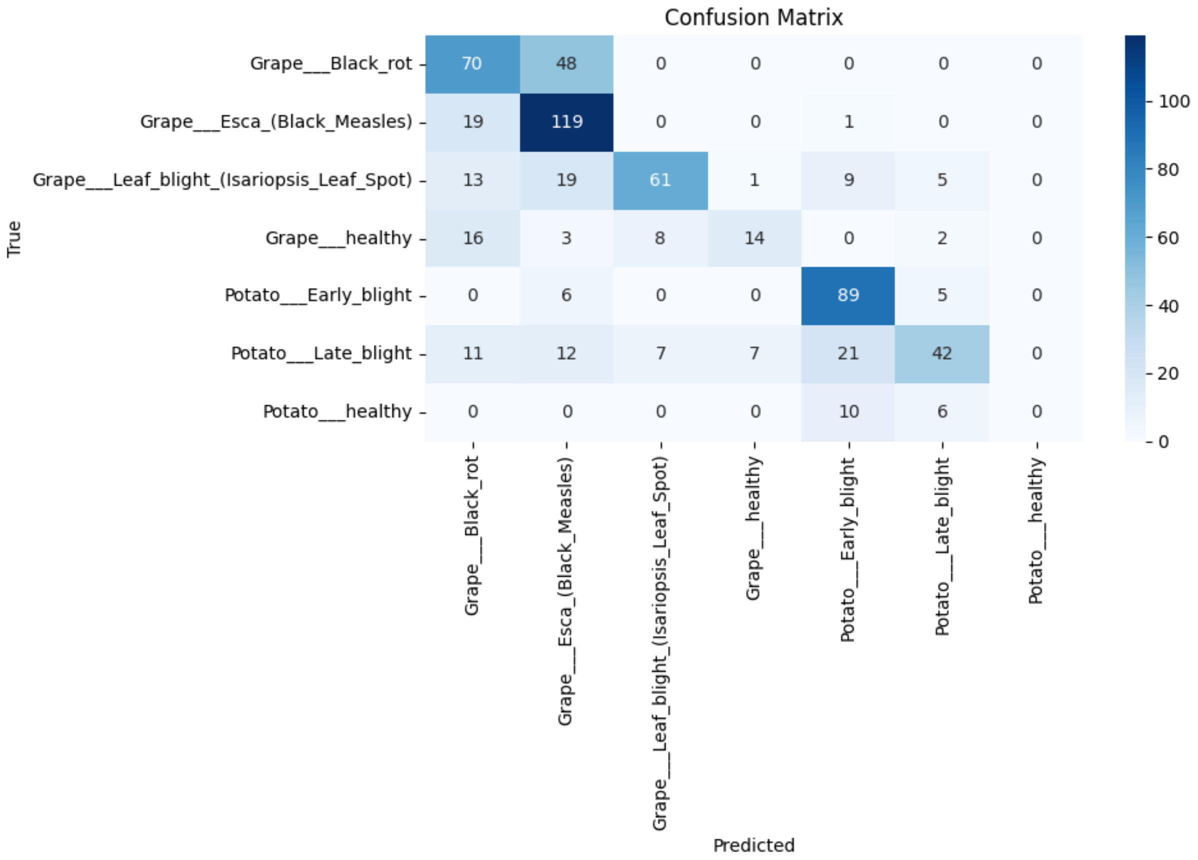
A number of quantitative evaluation metrics were used to compare the performance of the baseline CNN model and the suggested ensemble architecture. In a multi-class environment with varied grape and potato disease categories, these measures were chosen to offer a thorough insight of classification effectiveness. By calculating the percentage of correctly categorized leaf images in relation to the total number of samples, model accuracy served as the main metric of overall performance. However, supplementary metrics including precision, recall, and F1-score were calculated for each disease class because accuracy alone could be deceptive in the case of class imbalance. While recall gauges the model's capacity to find all pertinent cases of a specific disease, precision assesses the accuracy of positive predictions by showing the percentage of genuine disease detections among anticipated positives.

The F1-score is a balanced statistic that is less affected by unequal class distributions because it is the harmonic mean of precision and recall. A more thorough analysis of the model's performance across disease categories with varying sample sizes was made possible by the combination of these metrics. The distribution of true and anticipated class labels was also visualized using a multi-class confusion matrix. This matrix revealed misclassification tendencies, such as the misidentification of potato diseases as grape diseases or the confusion between visually similar grape diseases. These visual insights were essential for deciphering model flaws and comprehending how intra-class similarity affects classification performance.

The study included continuous evaluation measures appropriate for multi-class models in addition to these discrete metrics. A micro-averaging technique was used to calculate the Area Under the Receiver Operating Characteristic Curve (ROC-AUC), which offers a threshold-independent evaluation of the model's capacity to prioritize accurate classes over wrong ones. Strong separability between illness groups is indicated by a high ROC-AUC value. The trade-off between accuracy and recall across different choice thresholds was further captured by the Average accuracy (AP) score, which is derived from the Precision–Recall curve and provides a reliable performance metric under class imbalance.

Table 1. The feature extraction table.

Feature	**Explanation**
Dataset	7 classes of grape and potato leaf images collected
Data split	Train: Validation: Test ≈ 70:20:10
Preprocessing	Pixels value rescale by 1/255
Data Augmentation (training)	Random rotation ≤20°



True \ Predicted	Grape___Black_rot	Grape___Esca_(Black_Measles)	Grape___Leaf_blight_(Isariopsis_Leaf_Spot)	Grape___healthy	Potato___Early_blight	Potato___Late_blight	Potato___healthy
Grape___Black_rot	70	48	0	0	0	0	0
Grape___Esca_(Black_Measles)	19	119	0	0	1	0	0
Grape___Leaf_blight_(Isariopsis_Leaf_Spot)	13	19	61	1	9	5	0
Grape___healthy	16	3	8	14	0	2	0
Potato___Early_blight	0	6	0	0	89	5	0
Potato___Late_blight	11	12	7	7	21	42	0
Potato___healthy	0	0	0	0	10	6	0

Figure 1. Confusion matrix showing crops disease detection.

3. Results and Discussion

The proposed model achieved an accuracy of 59% on the test dataset, which consisted of 624 images across seven classes. This result demonstrates that the model was able to learn meaningful features, as the accuracy is considerably higher than the random baseline of approximately 14%. Nonetheless, the performance also reveals certain challenges, particularly in differentiating between visually similar disease classes and in handling variations caused by inconsistent image capture conditions. Performance improvements were observed through two key modifications: unfreezing the last 30 layers of the network and applying a learning-rate scheduling strategy. These adjustments allowed the model to extract richer feature representations and contributed to the observed accuracy gains.

The confusion matrix in Figure 1 reveals the model's key strengths and weaknesses. It performs very well on Potato_Early_blight (89 correct predictions) and reasonably well on Grape_Black_rot (70 correct). However, significant confusion occurs in two main areas: Cross-Crop Misclassification: The model frequently misclassifies Potato_Late_blight as a grape disease (e.g., 21 times as Grape_Leaf_blight), indicating it is struggling to distinguish between crops themselves, not just their diseases. Also, within crop confusion: There is high confusion between Grape_Esca and Grape_Black_rot, suggesting their visual symptoms (e.g., dark spots) are too similar for the model to tell apart with the current training. The cells represent classified and misclassified samples. Figure 2 shows the trade-off between the model's precision (how often its predictions are correct), and its recall (how well it finds all the relevant cases) across all disease classes. The model achieved an Average Precision (AP) score of 0.68. This means that across all probability thresholds, the model maintains a fairly consistent balance between making accurate predictions and identifying a large portion of the diseased plants in the dataset. The score of 0.68 is a quantitative measure of the model's overall performance in distinguishing between the seven classes of grape and potato diseases.

The ROC-AUC graph, Figure 3 measures the model's ability to rank predictions correctly across all classes, without being biased by class imbalance. The model achieved an Area Under the Curve (AUC) score of 0.92. An AUC score of 0.92 indicates a high level of performance in distinguishing between the different classes. This means that for a randomly chosen image, the model is very effective at assigning a higher probability to the correct class than to an incorrect one. This result suggests the model has learned strong, separable features for the task. It is important to note that this metric is generally more optimistic than the Precision-Recall curve when a class imbalance exists in the data. It good to note that class imbalance within the dataset reduced the model’s

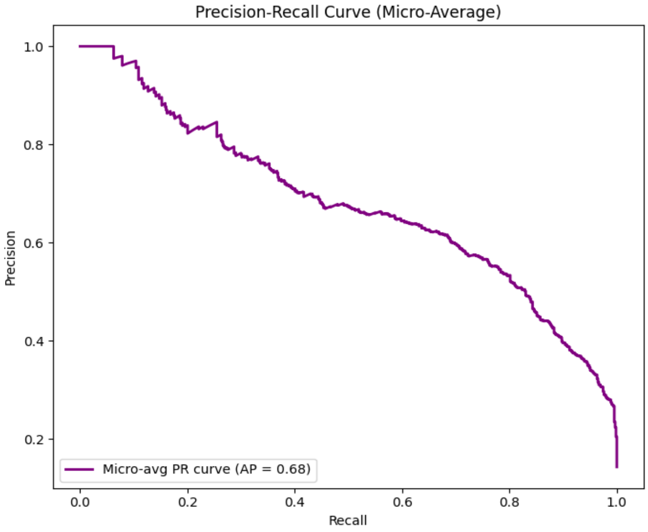


Figure 2. Precision -recall curve.

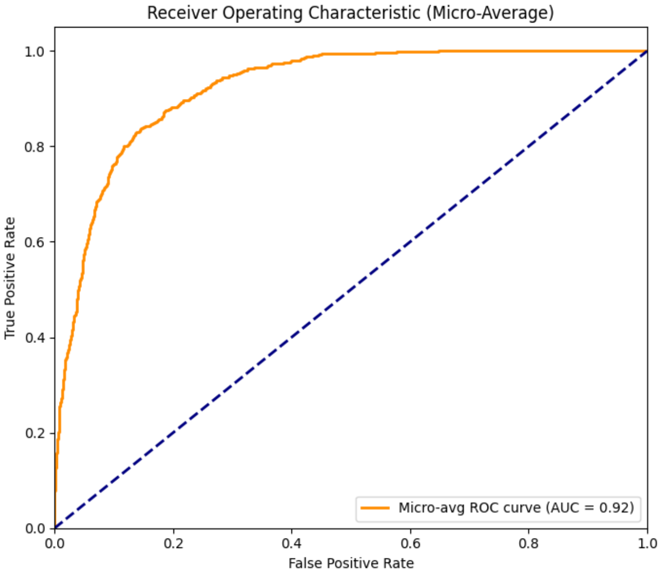


Figure 3. ROC-AUC (micro-average).

ability to generalize equally well across all classes, while the augmentation strategies employed lacked sufficient diversity to simulate real-world variability. In addition, the reliance on CPU-only training restricted both the training duration and the feasibility of utilizing stronger backbone architectures. Finally, the absence of crop-specific domain adaptation limited the model's adaptability to different environmental and agricultural conditions. The model's prediction probabilities for each of the seven target classes are shown in Figure 4.

When the model makes a classification decision, the bar chart shows how its confidence is distributed. The model indicates that the input sample is most likely to fall into the "Grape Black rot" class, which has the highest probability (≈ 37%). Grape Leaf Blight (Isariopsis Leaf Spot) comes next at around 25%, followed by Potato Late Blight and Grape Healthy, all of which have moderate risks of about 12–14%. Other categories, such as Grape Esca (Black Measles), Potato Early Blight, and Potato Healthy, are given significantly lower probability (< 10%), suggesting that the model is less certain of these classifications for the particular case.

4. Conclusion

The study revealed that a transfer learning approach with ResNet50, enhanced by augmentation, offers a promising baseline for crop disease detection, achieving 59% accuracy on grape and potato datasets. Insights indicate that fine-tuning pre-trained models addresses data scarcity effectively while augmentation mitigates overfitting, aligning with ensemble principles to improve robustness in non- IID settings. Practically, this suggests potential for affordable AI tools to aid farmers in early disease

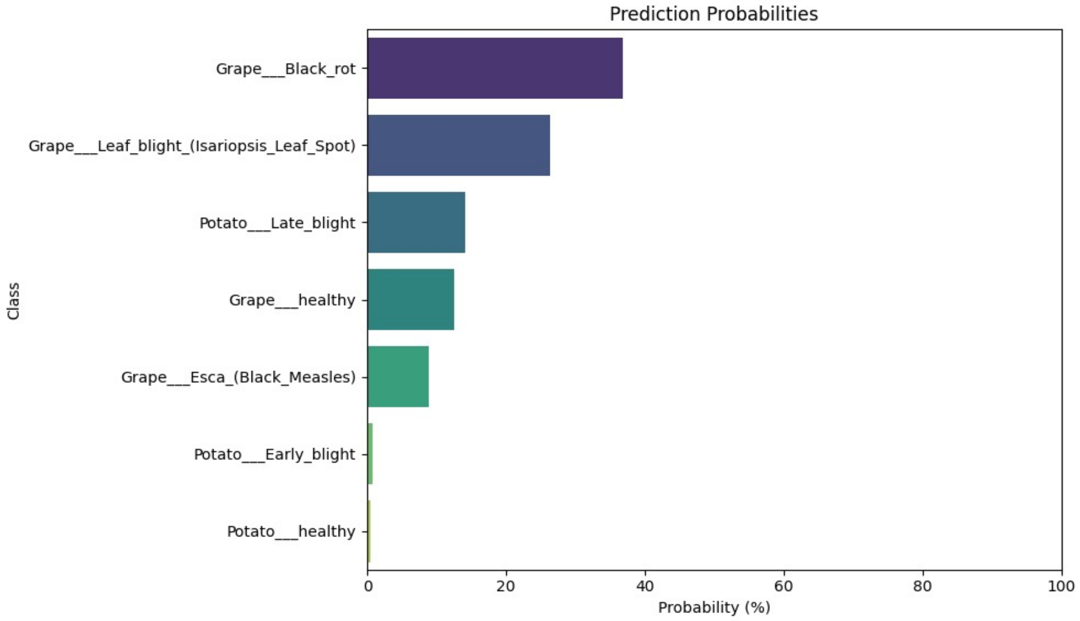


Figure 4. Prediction probabilities for crop diseases.

detection, supporting sustainable agriculture. Class imbalances between grape and potato subclasses, along with insufficient coverage of background/lighting variations by augmentations, led to misclassifications (e.g. potato as grape diseases). The Plant Village dataset's controlled nature may not fully reflect field variability, and time/resource constraints prevented advanced techniques like domain adaptation. These constraints highlight the need for enhanced computational resources and diverse data to achieve production-level accuracy (>90%), reinforcing the study's focus on real-world applicability.

5. Limitations and Future Research

The results of this study demonstrates a promising basis for crop disease identification across grape and potato leaf classes is provided by the suggested ensemble learning framework. Although the model outperformed baseline CNN findings and performed rather well, a number of issues affected its accuracy and generalizability. The limited processing power available during training was one of the biggest limitations. The model's capacity to investigate more intricate ensemble architectures or deeper fine-tuning techniques was limited because the tests were carried out on CPU-only hardware, which prevented the model from being trained for longer epochs or with higher batch sizes. The model's ability to acquire deeper feature representations was unavoidably impacted by this limitation, especially for disease categories that exhibit minor or overlapping visual traits. Another important observation pertains to the underlying dataset. Although the Plant Village dataset is widely used and contains high-quality images, its controlled and homogeneous nature poses challenges for real-world scalability.

Another important observation pertains to the underlying dataset. Although the PlantVillage dataset is widely used and contains high-quality images, its controlled and homogeneous nature poses challenges for real-world scalability.

The uniform backgrounds, consistent lighting, and clean leaf appearances differ substantially from real farm environments where shadows, leaf occlusions, soil particles, variations in weather, and camera noise are present. Models trained on such controlled datasets often struggle when deployed in the field without rigorous domain adaptation techniques. Related to this, class imbalance was evident across the seven disease categories. Some diseases were represented with far more samples than others, resulting in uneven model confidence and recall levels. This imbalance contributed to recurrent misclassifications, especially within the grape diseases that exhibit similar lesion patterns and coloration.

In light of these difficulties, a number of avenues for improving future research become apparent. To improve generalization, the dataset would need to be expanded with large-scale field photos taken in a variety of environmental settings. The model's resilience in uncontrolled environments may be greatly increased by implementing domain adaptation and domain generalization techniques, such as adversarial learning, feature alignment techniques, or transformation-invariant architectures. Furthermore, investigating more complex ensemble fusion techniques, such as weighted decision fusion, uncertainty-aware ensembling, and hybrid CNN–Transformer combinations, will be beneficial for future research. These improvements may result in more reliable forecasts for a variety of illness categories. Lastly, using explainable AI methods like Grad-CAM or SHAP visualizations could improve transparency and help agronomists and farmers make meaningful decisions. Combining reliable deep learning models with interpretability and deployment-friendly architecture will be essential for developing workable disease detection solutions as agriculture increasingly uses smart technologies for crop monitoring.

6. Declarations

6.1. Author Contributions

Chidi Ukamaka Betrand: Conceptualization, Methodology, Software, Validation, Supervision**; Mercy Eberechi Benson-Emenike:** Formal analysis, Investigation; **Douglas Allswell Kelechi:** Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft; **Chinwe Gilean Onukwugha:** Writing - Review & Editing, Visualization, Project administration; **Nneka Martina Oragba:** Supervision, Project administration.

6.2. Institutional Review Board Statement

Not applicable.

6.3. Informed Consent Statement

Not applicable.

6.4. Data Availability Statement

Available on Kaggle: https://www.kaggle.com/datasets/sarahgmn/plant-village-dataset.

6.5. Acknowledgment

The author would like to thank all parties who assisted in this research.

6.6. Conflicts of Interest

The authors declare no conflicts of interest.

7. References

[1] S. Savary, L. Willocquet, S. J. Pethybridge, P. Esker, N. McRoberts, and A. Nelson, "The global burden of pathogens and pests on major food crops," *Nat Ecol Evol*, vol. 3, no. 3, pp. 430–439, Feb. 2019, https://doi.org/10.1038/s41559-018-0793-y.

[2] G. Mohyuddin, M. A. Khan, A. Haseeb, S. Mahpara, M. Waseem, and A. M. Saleh, "Evaluation of Machine Learning Approaches for Precision Farming in Smart Agriculture System: A Comprehensive Review," *IEEE Access*, vol. 12, pp. 60155–60184, 2024, https://doi.org/10.1109/ACCESS.2024.3390581.

[3] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Comput Electron Agric*, vol. 145, pp. 311–318, Feb. 2018, https://doi.org/10.1016/j.compag.2018.01.009.

[4] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using Deep Learning for Image-Based Plant Disease Detection," *Front Plant Sci*, vol. 7, Sep. 2016, https://doi.org/10.3389/fpls.2016.01419.

[5] S. Ji *et al.*, "Emerging trends in federated learning: from model fusion to federated X learning," *International Journal of Machine Learning and Cybernetics*, vol. 15, no. 9, pp. 3769–3790, Sep. 2024, https://doi.org/10.1007/s13042-024-02119-1.

[6] Z.-H. Zhou, *Ensemble Methods: Foundations and Algorithms*, 1st ed. CRC Press, 2012. Accessed: Nov. 25, 2025. [Online]. Available: https://books.google.co.id/books?id=hm0-EQAAQBAJ.

[7] J. A. Wani, S. Sharma, M. Muzamil, S. Ahmed, S. Sharma, and S. Singh, "Machine Learning and Deep Learning Based Computational Techniques in Automatic Agricultural Diseases Detection: Methodologies, Applications, and Challenges," *Archives of Computational Methods in Engineering*, vol. 29, no. 1, pp. 641–677, Jan. 2022, https://doi.org/10.1007/s11831-021-09588-5.

[8] F. Tanner, S. Tonn, J. de Wit, G. Van den Ackerveken, B. Berger, and D. Plett, "Sensor-based phenotyping of above-ground plant-pathogen interactions," *Plant Methods*, vol. 18, no. 1, p. 35, Dec. 2022, https://doi.org/10.1186/s13007-022-00853-7.

[9] A. Ganguly, B. Tiwari, G. P. K. Reddy, and M. Chauhan, "Ensemble Learning for Plant Leaf Disease Detection: A Novel Approach for Improved Classification Accuracy," *Research Square Preprint*, Aug. 2023, https://doi.org/10.21203/rs.3.rs-3257323/v1.

[10] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Comput Electron Agric*, vol. 147, pp. 70–90, Apr. 2018, https://doi.org/10.1016/j.compag.2018.02.016.

[11] M. Brahimi, K. Boukhalfa, and A. Moussaoui, "Deep Learning for Tomato Diseases: Classification and Symptoms Visualization," *Applied Artificial Intelligence*, vol. 31, no. 4, pp. 299–315, Apr. 2017, https://doi.org/10.1080/08839514.2017.1315516.

[12] A. Oad *et al.*, "Plant Leaf Disease Detection Using Ensemble Learning and Explainable AI," *IEEE Access*, vol. 12, pp. 156038–156049, 2024, https://doi.org/10.1109/ACCESS.2024.3484574.

[13] N. Rane, S. P. Choudhary, and J. Rane, "Ensemble deep learning and machine learning: applications, opportunities, challenges, and future directions," *Studies in Medical and Health Sciences*, vol. 1, no. 2, pp. 18–41, Jul. 2024, https://doi.org/10.48185/smhs.v1i2.1225.

[14] B. Mohammed, "A Comprehensive Overview of Federated Learning for NextGeneration Smart Agriculture: Current Trends, Challenges, and Future Directions," *Informatica*, vol. 49, no. 1, Mar. 2025, https://doi.org/10.31449/inf.v49i1.6764.

[15] F. Wang, Y. Li, F. Liao, and H. Yan, "An ensemble learning based prediction strategy for dynamic multi-objective optimization," *Appl Soft Comput*, vol. 96, p. 106592, Nov. 2020, https://doi.org/10.1016/j.asoc.2020.106592.

[16] S. Abimannan, E.-S. M. El-Alfy, Y.-S. Chang, S. Hussain, S. Shukla, and D. Satheesh, "Ensemble Multifeatured Deep Learning Models and Applications: A Survey," *IEEE Access*, vol. 11, pp. 107194–107217, 2023, https://doi.org/10.1109/ACCESS.2023.3320042.

[17] I. D. Mienye and Y. Sun, "A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects," *IEEE Access*, vol. 10, pp. 99129–99149, 2022, https://doi.org/10.1109/ACCESS.2022.3207287.

[18] A. F. Jadama and M. K. Toray, “Ensemble Learning: Methods, Techniques, Application,” *ResearchGate Preprint*, pp. 1–7, 2024, https://doi.org/10.13140/RG.2.2.28017.08802.

[19] T. A. Shaikh, T. Rasool, P. Verma, and W. A. Mir, “A fundamental overview of ensemble deep learning models and applications: systematic literature and state of the art,” *Ann Oper Res*, Dec. 2024, https://doi.org/10.1007/s10479-024-06444-0.

[20] R. Polikar, “Ensemble based systems in decision making,” *IEEE Circuits and Systems Magazine*, vol. 6, no. 3, pp. 21–45, 2006, https://doi.org/10.1109/MCAS.2006.1688199.

[21] M. Sakib, S. Mustajab, and M. Alam, “Ensemble deep learning techniques for time series analysis: a comprehensive review, applications, open issues, challenges, and future directions,” *Cluster Comput*, vol. 28, no. 1, p. 73, Feb. 2025, https://doi.org/10.1007/s10586-024-04684-0.

[22] S. Wang and X. Yao, “Diversity analysis on imbalanced data sets by using ensemble models,” in *2009 IEEE Symposium on Computational Intelligence and Data Mining*, IEEE, Mar. 2009, pp. 324–331. https://doi.org/10.1109/CIDM.2009.4938667.

[23] Z. Fan *et al.*, “Diverse Models, United Goal: A Comprehensive Survey of Ensemble Learning,” *CAAI Trans Intell Technol*, vol. 10, no. 4, pp. 959–982, Aug. 2025, https://doi.org/10.1049/cit2.70030.

[24] D. H. Wolpert, “Stacked generalization,” *Neural Networks*, vol. 5, no. 2, pp. 241–259, Jan. 1992, https://doi.org/10.1016/S0893-6080(05)80023-1.