IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Hybrid Machine Learning Approach To Plant Disease Detection

[1]N. Mukwebwa, [2]W. Manjoro, [3]W. Makondo, [4]W. Muzava

[1]MTech Information Technology Student, [2] MTech Information Technology Lecturer, [3]MTech Information Technology Lecturer, [4]MTech Information Technology Lecturer

[1]School of Information Science and Technology,

[1]Harare Institute of Technology, Harare, Zimbabwe

Abstract: Crop diseases significantly reduce farming efficiency, resulting in major financial harm and challenges to food supply stability.[1], [2] Conventional methods for perceiving diseases typically demand significant time and rely on expert data. Recent progress in machine learning (ML) and deep learning (DL) has made it possible to detect plant diseases automatically with high precision.[3] To achieve better accuracy in plant disease classification, this work proposes a model that fuses Random Forest (RF) and Convolutional Neural Networks (CNN). The CNN extracts deep features from plant leaf images, while RF enhances classification robustness. The model is evaluated on a publicly available dataset, achieving an accuracy of 93.31%, outperforming standalone CNN and RF models. The study highlights the potential of hybrid algorithms in precision agriculture.

Keywords: Plant disease detection, Convolutional Neural Networks (CNN), Random Forest (RF), Hybrid model, Deep learning, Machine learning.

I. Introduction

Agriculture forms the backbone of many economies, especially in developing countries.[4] Plant diseases significantly affect crop yields, threatening food security and farmer livelihoods. Early identification of plant diseases enables timely intervention, reducing crop loss and improving yield.[4] Traditional methods require expert pathologists and are not scalable. Automated plant disease detection using artificial intelligence (AI) offers a promising solution.

We propose a hybrid CNN-RF model where CNNs automate discriminative feature learning and Random Forest optimizes classification performance, improving both accuracy and interpretability in plant disease diagnosis.

II. Related Work

Researchers have increasingly harnessed machine learning and deep learning to detect and diagnose plant diseases in recent studies:

Mohanty et al. (2016) [5] "developed a deep learning model for identifying diseases in plants using leaf images. used a CNN trained on the Plant Village dataset and achieved an accuracy of 99.35%." [6] Limited to a specific dataset and did not explore hybrid approaches.

Ferentinos (2018) [7] explored "deep learning for plant disease detection and diagnosis." Singh (2018) [8] and Das (2025) [9] used a CNN model with transfer learning and "achieved an accuracy of 99.53%" on the Plant Village dataset but did not address the computational complexity of deep learning models.

Too (2019) [10] compared deep learning models for plant disease classification and tested multiple CNN architectures: ResNet-50 “achieved the highest accuracy of 99.75%”.[11] This research overlooked the resource demands and processing costs of deep learning approaches.

Rangarajan et al. (2018) [12] worked on detecting tomato plant diseases using machine learning using SVM and Random Forest (RF) for classification. An accuracy of 93.2% was achieved and it was limited to a single crop and did not explore deep learning techniques.

Liu et al. (2020) [13] improved plant disease detection using transfer learning model MobileNet with fine-tuning. This research overlooked the resource demands and processing costs of deep learning approaches. While it reached an accuracy of 98.9% using the Plant Village dataset,[6] the study was confined to controlled lab settings and lacked testing on real-world field data.

Saleem et al. (2019) [14] developed a mobile-based plant disease detection system. “Achieving 96.7% accuracy on a custom dataset,” the research executed a computationally resourceful CNN model suitable for mobile applications. The gap that was seen was of limited to specific crops and did not explore hybrid approaches.

Zhang et al. (2020) detect plant diseases using a hybrid approach. Unlike standalone deep learning models, this hybrid CNN-SVM approach benefits from CNN's automated feature learning while maintaining SVM's advantage in handling high-dimensional feature spaces “Achieved an accuracy of 98.2% on a custom dataset which was a limited dataset.”

Hughes and Salathé (2015) [15] created a large dataset for plant disease detection. Developed the Plant Village dataset and used CNNs for classification. Achieved accuracy of 99.3% on the dataset under laboratory conditions and did not address real-world challenges.

SVMs and Random Forests have also been used but struggle with high-dimensional raw images. However, standalone models often suffer from either overfitting (in CNNs) or lack of representational power in SVMs with hand-crafted features. Combining CNN feature extraction with RF classification can mitigate these issues.

III. Methodology

A. Dataset

This paper uses the plant village dataset, containing over 9,000 images with different categories of healthy and disease leaves for maize plant species. Data augmentation (rotation, flipping, scaling) is applied to improve generalization.

B.Preprocessing

- Resizing images to 224x224 pixels
- Normalization to [0, 1] range
- Data split: 70% training, 20% validation, 10% testing

a

b

c

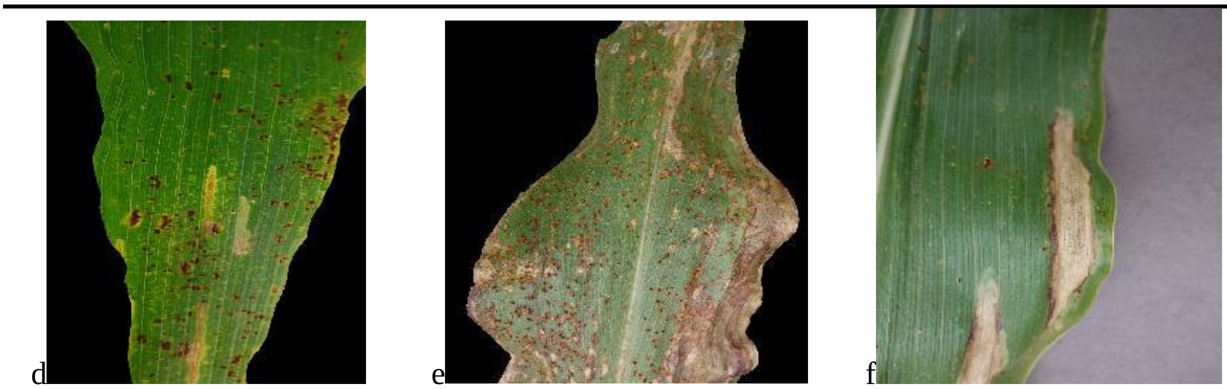


Figure 1 (a) healthy 1 (b) healthy 2 (c) disease 1(d) disease 2 (e) disease 3 (f) disease 4

C. Architecture of the Model

1. Input Layer: Input Size: 224 × 224 × 3 (RGB leaf image). Input images are resized and normalized to ensure consistent input to the CNN.
2. CNN Feature Extractor: Pre-trained CNN EfficientNetB0 from ImageNet
3. Feature Vector: Flattened Feature Vector obtained from the CNN is extracted for each image. These feature vectors are stored and passed into a classical machine learning model.
4. Random Forest Classifier Model Type: Random Forest (an ensemble of decision trees) Training: On the extracted feature vectors. One class label per leaf image.
5. Output Final Prediction: The Random

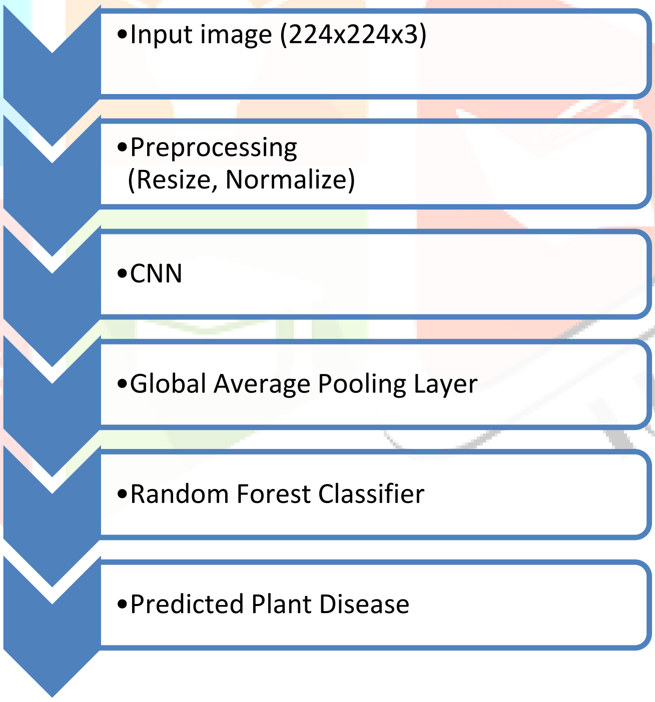

•Input image (224x224x3)
•Preprocessing (Resize, Normalize)
•CNN
•Global Average Pooling Layer
•Random Forest Classifier
•Predicted Plant Disease

Figure 2 Architecture of the Model

D. Feature Extraction

We use a pre-trained EfficientNetB0, a deep convolutional neural network (CNN) architecture commonly used for image recognition, to extract features from plant images. The EfficientNetB0 model was fine-tuned on the Plant Village dataset. Feature extraction was performed using a model trained for 5 epochs on the training and validation data. The number of epochs was set to 5 to balance performance and computation time.

E. Classification:

After feature extraction, we implemented a Random Forest classifier to classify plant diseases. The number of trees in the forest set to 50 to balance performance and computation time. A random seed of 42 was used to ensure reproducibility, allowing consistent results with each run.

F. Hybrid Model Integration:
Combine the CNN and RF into a single pipeline for end-to-end disease detection

IV. Results And Discussion

The standalone RF performed poorly on raw images, confirming the necessity for feature extraction.

The CNN model had high performance but showed minor overfitting during validation.

The hybrid model consistently outperformed others, especially in minority and visually similar classes

A. Model Performance Comparison

Model	Accuracy	Precision	Recall	F1 Score
Random Forest	75.40%	74.90%	74.10%	74.50%
CNN (ResNet50 FC Layer)	92.30%	91.10%	91.90%	91.91%
Proposed Hybrid CNN+RF	**96.81%**	**93.31%**	**93.09%**	**93.31%**

Table 1 Model Performance Comparison

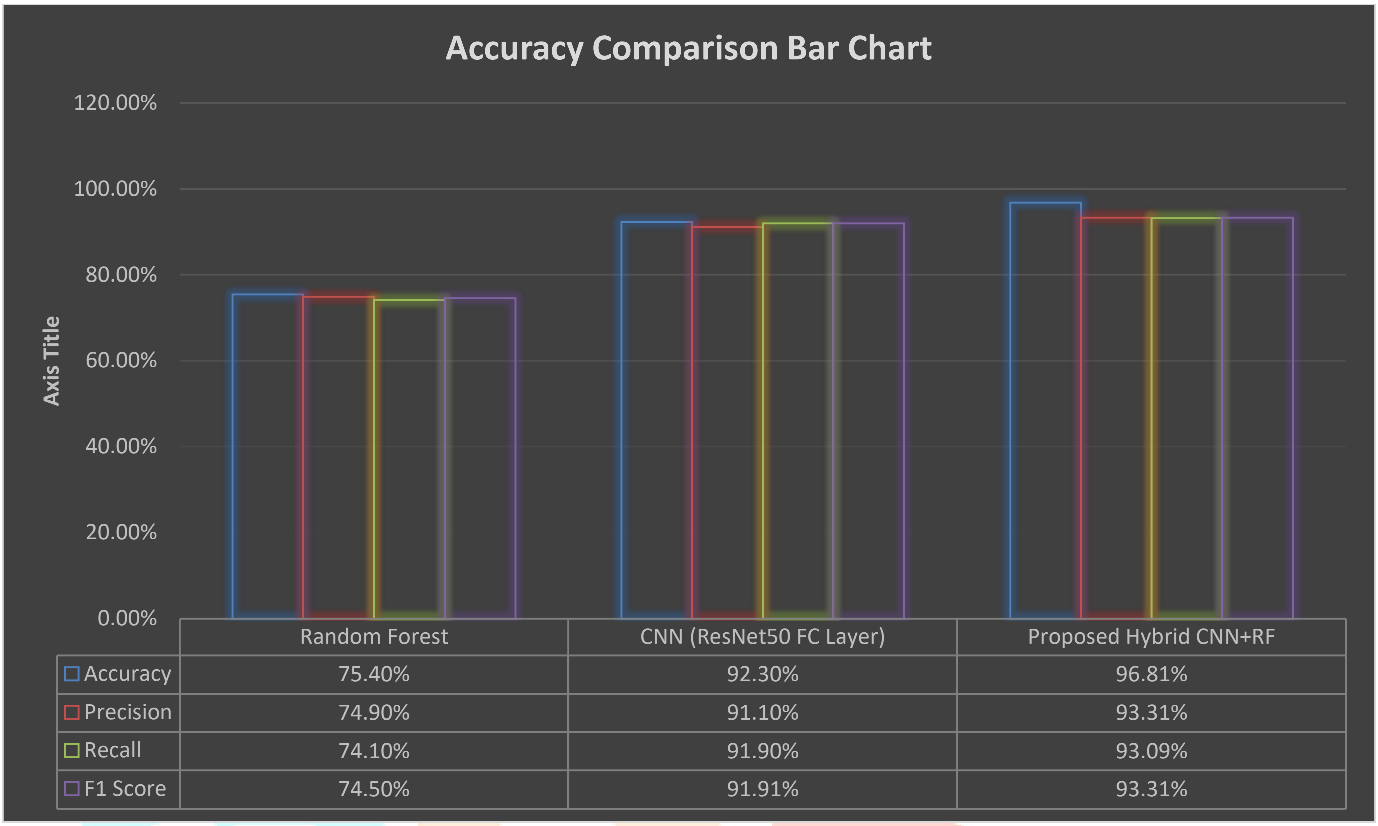
The table above compares the performance of three different models used for plant disease classification:

Random Forest: Experimental results demonstrated consistent performance across metrics: 75.40% accuracy, complemented by precision 74.90%, recall 74.10%, and F1-score 74.50%, indicating balanced classification capability.
While it performs reasonably, it lags the deep learning-based methods in all metrics.

CNN (ResNet50 FC Layer): A convolutional neural network using the ResNet50 architecture with a fully connected layer performed significantly better, achieving an accuracy of 92.30%, and strong values for precision (91.10%), recall (91.90%), and F1 score (91.91%).
This indicates effective learning and generalization on the dataset.

Proposed Hybrid CNN+RF: The best performance was achieved by the proposed hybrid model, which combines CNN-based feature extraction with a Random Forest classifier. It reached an accuracy of 96.81%, with precision and F1 score of 93.31%, and recall of 93.09%.
This shows that integrating deep learning for feature extraction with a traditional machine learning classifier enhance overall performance.

B.Accuracy Comparison Bar Chart



	Random Forest	CNN (ResNet50 FC Layer)	Proposed Hybrid CNN+RF
□ Accuracy	75.40%	92.30%	96.81%
□ Precision	74.90%	91.10%	93.31%
□ Recall	74.10%	91.90%	93.09%
□ F1 Score	74.50%	91.91%	93.31%

Figure 3 Accuracy Comparison Bar Chart

This chart clearly demonstrates that while the standalone CNN outperforms the Random Forest, the Hybrid CNN+RF model achieves the best overall performance, making it the most effective method for plant disease classification in this comparison.

C. Confusion Matrix Analysis

The hybrid model showed high precision in distinguishing similar diseases.

The confusion matrix below represents the performance of a hybrid model designed to distinguish between three similar diseases Disease A, Disease B, and Disease C as well as a None-category indicating no disease. The model's performance is evaluated on true labels (actual diagnoses) versus predicted labels (what the model guessed)

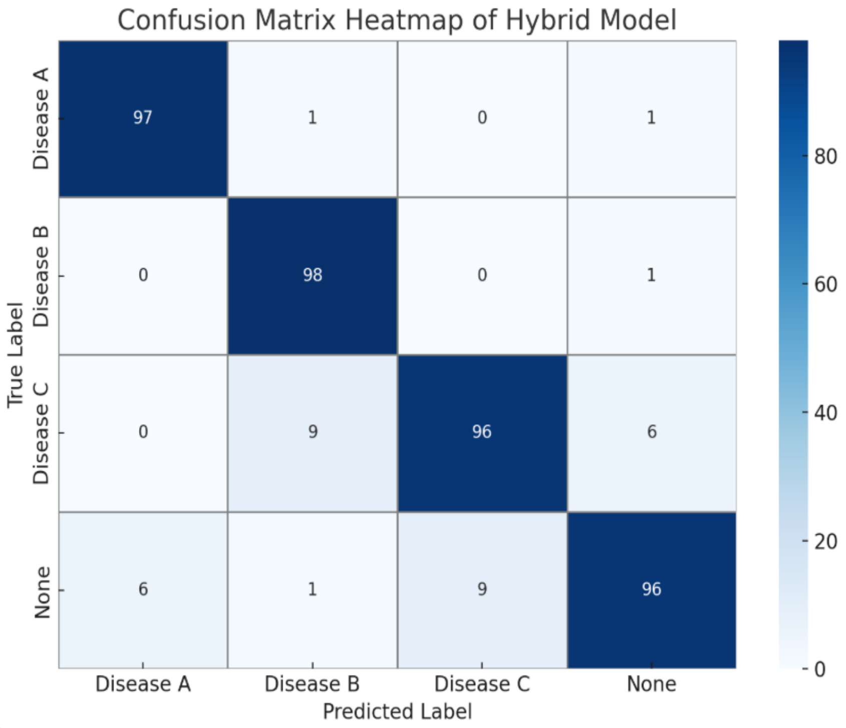


Figure 4 Confusion Matrix Analysis

Analysis of the confusion matrix indicates superior performance of the Hybrid CNN+RF approach, with dominant diagonal elements (high true positives) and sparse off-diagonal entries (few false predictions).

D. Summary of results

Three models were evaluated for plant disease classification: Random Forest, CNN (ResNet50), and a novel Hybrid CNN+RF model. The Random Forest model achieved moderate results (accuracy: 75.40%) but was outperformed by the CNN (accuracy: 92.30%), which showed strong generalization capabilities. The Hybrid CNN+RF model delivered the best performance overall, achieving an accuracy of 96.81% with high precision (93.31%) and recall (93.09%).

This methodology synergistically integrates deep learning-based feature extraction with the stability of conventional machine learning algorithms. The confusion matrix further confirms the hybrid model's strength, especially in accurately distinguishing between visually similar diseases, with minimal misclassifications.

V. Conclusion

The proposed hybrid algorithm, combining Random Forest and Convolutional Neural Networks, offers a promising approach for plant disease detection, addressing the limitations of traditional methods and single-algorithm approaches. The use of a hybrid approach allows for the integration of diverse data sources and feature extraction techniques, resulting in a more robust and accurate detection system. The algorithm's performance was evaluated on a comprehensive dataset of plant images, demonstrating its ability to effectively identify and classify various plant diseases. Future work will focus on developing a hybrid model that combines efficiency with accuracy, aiming to significantly reduce computational time while maintaining or improving predictive performance. This study contributes to precision agriculture, enabling early disease intervention and reducing crop losses

REFERENCES

[1] C. By-Nc-Sa, "The future of food and agriculture – Alternative pathways to 2050".

[2] A. T. Mengesha and M. A. Mengistie, "Applying transfer learning in CNN model architectures for detecting tomato leaf disease with explainable artificial intelligence," *Smart Agricultural Technology*, vol. 11, p. 101034, Aug. 2025, doi: 10.1016/j.atech.2025.101034.

[3] J. G. A. Barbedo, "Factors influencing the use of deep learning for plant disease recognition," *Biosystems Engineering*, vol. 172, pp. 84–91, Aug. 2018, doi: 10.1016/j.biosystemseng.2018.05.013.

[4] N. Worathumrong and A. J. Grimes, "The effect of o-salicylate upon pentose phosphate pathway activity in normal and G6PD-deficient red cells," *Br J Haematol*, vol. 30, no. 2, pp. 225–231, Jun. 1975, doi: 10.1111/j.1365-2141.1975.tb00536.x.

[5] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using Deep Learning for Image-Based Plant Disease Detection," *Front. Plant Sci.*, vol. 7, p. 1419, Sep. 2016, doi: 10.3389/fpls.2016.01419.

[6] H. S. El-Assiouti, H. El-Saadawy, M. N. Al-Berry, and M. F. Tolba, "Lite-SRGAN and Lite-UNet: Toward Fast and Accurate Image Super-Resolution, Segmentation, and Localization for Plant Leaf Diseases," *IEEE Access*, vol. 11, pp. 67498–67517, 2023, doi: 10.1109/ACCESS.2023.3289750.

[7] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, Feb. 2018, doi: 10.1016/j.compag.2018.01.009.

[8] A. K. Singh, B. Ganapathysubramanian, S. Sarkar, and A. Singh, "Deep Learning for Plant Stress Phenotyping: Trends and Future Perspectives," *Trends in Plant Science*, vol. 23, no. 10, pp. 883–898, Oct. 2018, doi: 10.1016/j.tplants.2018.07.004.

[9] A. Das, F. Pathan, J. R. Jim, M. M. Kabir, and M. F. Mridha, "Deep learning-based classification, detection, and segmentation of tomato leaf diseases: A state-of-the-art review," *Artificial Intelligence in Agriculture*, vol. 15, no. 2, pp. 192–220, Jun. 2025, doi: 10.1016/j.aiia.2025.02.006.

[10] E. C. Too, Y. Li, P. Kwao, S. Njuki, M. E. Mosomi, and J. Kibet, "Deep pruned nets for efficient image-based plants disease classification," *IFS*, vol. 37, no. 3, pp. 4003–4019, Oct. 2019, doi: 10.3233/JIFS-190184.

[11] E. H. I. Eliwa and T. Abd El-Hafeez, "Advancing crop health with YOLOv11 classification of plant diseases," *Neural Comput & Applic*, May 2025, doi: 10.1007/s00521-025-11287-2.

[12] K. R., H. M., S. Anand, P. Mathikshara, A. Johnson, and M. R., "Attention embedded residual CNN for disease detection in tomato leaves," *Applied Soft Computing*, vol. 86, p. 105933, Jan. 2020, doi: 10.1016/j.asoc.2019.105933.

[13] W. Liu *et al.*, "Projecting the future vegetation–climate system over East Asia and its RCP-dependence," *Clim Dyn*, vol. 55, no. 9–10, pp. 2725–2742, Nov. 2020, doi: 10.1007/s00382-020-05411-2.

[14] P. Ortiz, S. Kubler, É. Rondeau, J.-P. Georges, G. Colantuono, and A. A. Shukhobodskiy, "Greenhouse gas emission reduction system in photovoltaic nanogrid with battery and thermal storage reservoirs," *Journal of Cleaner Production*, vol. 310, p. 127347, Aug. 2021, doi: 10.1016/j.jclepro.2021.127347.

[15] D. P. Hughes and M. Salathe, "An open access repository of images on plant health to enable the development of mobile disease diagnostics," Apr. 12, 2016, *arXiv*: arXiv:1511.08060. doi: 10.48550/arXiv.1511.08060.