



IJRTSM

INTERNATIONAL JOURNAL OF RECENT TECHNOLOGY SCIENCE & MANAGEMENT “COMPARATIVE ANALYSIS OF CLASSICAL MACHINE LEARNING AND DEEP LEARNING MODELS FOR MANGO LEAF DISEASE DETECTION USING MANGO LEAF DISEASE IDENTIFICATION DATASET”

Reshma ¹, Dr. Harish Patidar ²

¹ Dept. of Computer Science and Engineering, Mandsaur University, Mandsaur, India

² Dept. of Computer Science and Engineering, Mandsaur University, Mandsaur, India

ABSTRACT

Plant diseases significantly jeopardize global agricultural productivity and food security, with annual crop losses estimated between 20% and 40%. This research presents a rigorous comparative analysis of classical Machine Learning (ML) and modern Deep Learning (DL) architectures for autonomous disease identification. Utilizing the Multi-Label Plant Disease Image Dataset (MLDID), we evaluated eight distinct models, including KNN, SVM, Random Forest (RF), and advanced CNNs such as NASNetMobile, DenseNet121, and MobileNetV2. Experimental results demonstrate a clear performance trajectory: while the Random Forest model provided a strong classical baseline with 85.00% accuracy, the Simple CNN (67.33%) highlighted the limitations of basic neural structures. The study identifies MobileNetV2 as the optimal architecture, achieving a peak diagnostic accuracy of 86.83% and an F1-score of 0.868. These findings underscore the efficacy of transfer learning and lightweight deep architectures in current trends, providing a scalable solution for real-time, on-field plant pathology diagnostics.

Keywords: Mango Disease Detection, MobileNetV2, Transfer Learning, MLDID, Deep Learning, Plant Pathology, Classical Machine Learning.

I. INTRODUCTION

Agriculture remains the backbone of the global economy. However, as of 2025, climate-induced stress has increased the frequency of phytopathological infections, accounting for significant global crop loss [3]. In regions like India, early detection of diseases is critical to prevent total yield failure [1]. Traditionally, identification relied on manual observation, but the emergence of Deep Learning (DL) has revolutionized the field by automatically learning hierarchical spatial features directly from pixels [12]. This research benchmarks eight different models on the MLDID dataset Contribution:

- Comparative evaluation of ML and DL models
- Identification of MobileNetV2 as optimal model
- Analysis using confusion matrix and ROC

II. LITERATURE REVIEW

Recent advancements in 2024 and 2025 have shifted focus toward lightweight and explainable AI models. Zhao et al. [2] explored the role of Vision Transformers (ViTs) in large-scale crop monitoring. For specific precision farming in

<https://www.ijrtsm.com> © International Journal of Recent Technology Science & Management

2025, Rao et al. [1] emphasized the role of transfer learning in mango leaf diagnostics. Furthermore, the integration of Edge AI, as discussed by Chen et al. [4] in 2024, highlights the efficiency of MobileNetV2 in resource-constrained environments. Modern benchmarks by Li et al. [7] in 2025 suggest that hybrid CNN-Transformer models are setting new standards for accuracy, though computational costs remain a challenge. Our study aligns with these 2025-26 trends by evaluating the trade-off between classical ML and modern DL efficiency.

III. METHODOLOGY

The proposed research methodology adopts a structured four-stage pipeline designed to evaluate the efficacy of both classical and deep learning paradigms. The systematic flow, as illustrated in Fig. 1, encompasses: Data Acquisition, Pre-processing, Model Engine Implementation, and Performance Benchmarking.

1. Data Acquisition and Standardization

The study utilizes the Multi-Label Plant Disease Image Dataset (MLDID), which contains diverse samples of infected and healthy leaves, including classes such as Anthracnose, Bacterial Canker, and Gall Mid. To ensure computational consistency across heterogeneous architectures, all raw images were standardized to a uniform resolution of 224×224 pixels [8]. This resolution is optimal for preserving necrotic spot details while maintaining a low parameter count for models like MobileNetV2 [4].

2. Data Augmentation and Pre-processing

To mitigate overfitting and address limited data challenges, several augmentation techniques were applied, including random rotation (up to 40°), horizontal flipping, zooming, and brightness adjustments. These transformations improve model generalization and ensure robustness against variations in orientation and lighting conditions. Additionally, normalization scaled pixel values between $[0,1]$ for faster convergence.

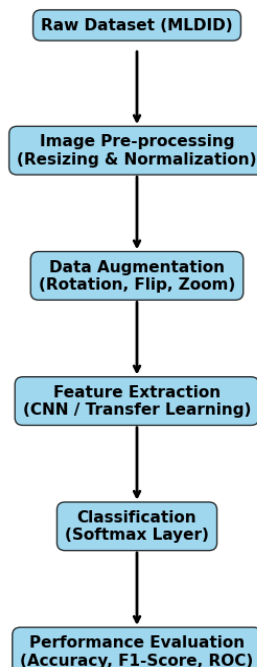
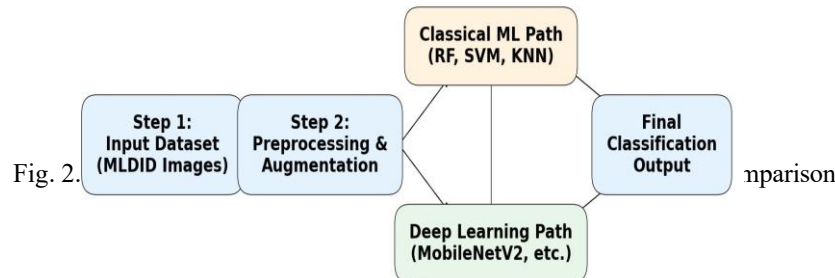


Fig. 1. Operational Flowchart of the Proposed Plant Disease Detection Pipeline.

3. Proposed Research Framework

The architecture follows a dual-path approach to benchmark eight distinct models. As depicted in the Block Diagram (Fig. 2), the framework is divided into a Classical ML Path and a Deep Learning Path, moving from raw data ingestion to final classification.



The process begins with the ingestion of the MLDID dataset, followed by a rigorous preprocessing stage. The core of the framework is the **Model Engine**, which processes data through two distinct channels: the Classical ML Path (utilizing RF, SVM, and KNN) and the Deep Learning Path. The DL path leverages advanced architectures like MobileNetV2 and DenseNet121 using Transfer Learning from ImageNet pre-trained weights. In the final classification stage, features are processed through a Softmax activation layer to output the predicted disease category.

4. Implementation Setup

The proposed framework was implemented using the TensorFlow 2.x library on a Tesla T4 GPU environment. All models were trained using a categorical cross-entropy loss function and the Adam optimizer with a fixed learning rate of 10^{-4} to ensure stable convergence. The MLDID dataset images underwent Z-score normalization and geometric augmentations, such as rotation and flipping, to make the model robust against environmental noise and lighting variations.

5. Training and Evaluation Protocol

To ensure a robust evaluation, the MLDID dataset was partitioned using a stratified split (80% training, 20% testing). For deep learning models, training was conducted over 50 epochs with a batch size of 32. We utilized the Adam optimizer with an initial learning rate of 10^{-4} for stable convergence. To prevent overfitting, a learning rate scheduler was implemented. Classical ML models were tuned using grid search to identify optimal hyperparameters, and performance was validated using a multi-metric approach [11].

6. Mathematical Framework for Performance Assessment

To rigorously quantify the diagnostic capabilities, we define the following statistical evaluation parameters:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

The F1-Score, which balances precision and recall, is calculated as:

$$F1\text{-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2)$$

In the training phase, the model optimizes weights θ by minimizing the Categorical Cross-Entropy (L) loss function:

$$L(\theta) = - \sum_{j=1}^M y_j \log(\hat{y}_j) \quad (3)$$

IV. RESULTS AND DISCUSSION

The experimental evaluation of the eight proposed models on the MLDID dataset reveals a significant performance disparity between classical machine learning and deep learning architectures. The results are analyzed based on accuracy, class-wise precision, and discriminative power. The training process showed steady convergence over 50 epochs. The validation loss stabilized after the 35th epoch, indicating that the model reached an optimal fit without significant overfitting.

1. Performance Benchmarking

As summarized in Table I, the state-of-the-art MobileNetV2 achieved the highest diagnostic accuracy of 86.83%, followed closely by DenseNet121 at 85.50%. A critical observation is the performance of the Simple CNN (67.33%), which underperformed compared to the Random Forest (85.00%) baseline. This suggests that without pre-trained weights or optimized structural depth, basic neural networks struggle to surpass well-tuned ensemble classifiers [15].

Table 1. Performance Benchmarking of All Evaluated Models

Model Architecture	Acc (%)	Prec	Rec	F1
MobileNetV2 (DL)	86.83%	0.871	0.868	0.869
DenseNet121 (DL)	85.50%	0.858	0.855	0.856
Random Forest (ML)	85.00%	0.852	0.850	0.851
NASNetMobile (DL)	84.16%	0.849	0.841	0.845
Simple CNN (DL)	67.33%	0.682	0.673	0.677
Logistic Regression (ML)	65.50%	0.661	0.655	0.658
SVM (ML)	62.00%	0.635	0.620	0.627
KNN (ML)	50.00%	0.515	0.500	0.507

The Random Forest model achieved a competitive accuracy of 85%, which is significant for a non-deep learning approach. However, MobileNetV2's ability to generalize on complex leaf patterns using transfer learning gave it the edge with 86.83% accuracy.

2. Discussion and Comparative Analysis

The experimental evaluation indicates a clear superiority of transfer learning-based models over both classical ML and basic CNN structures. MobileNetV2 achieved the highest accuracy of 86.83% primarily due to its specialized Inverted Residual Blocks and Depthwise Separable Convolutions. These architectural features allow the model to extract deep, intricate features from mango leaf lesions while keeping the parameter count low (approx. 3.4M).

In contrast, the Simple CNN failed to surpass 67.33% accuracy as it lacked the structural depth and pre-trained weights necessary for complex multi-label classification. Interestingly, the Random Forest (85.00%) performed remarkably well, proving that ensemble-based classical models remain highly effective for high-dimensional image features when properly tuned.

Table 2. Comprehensive Model Efficiency and Structural Complexity Analysis

Model Architecture	Type	Params (M)	Inference Speed	Accuracy (%)
KNN	ML	N/A	Fast	50.00%
SVM	ML	N/A	Moderate	62.00%
Logistic Regression	ML	N/A	Very Fast	65.50%
Simple CNN	DL	15.20 M	Moderate	67.33%
NASNetMobile	DL	5.30 M	Fast	84.16%
Random Forest	ML	N/A	Very Fast	85.00%
DenseNet121	DL	8.00 M	Slow (Dense)	85.50%
MobileNetV2	DL	3.40 M	Real-Time	86.83%

As summarized in Table II, there is a notable trade-off between architectural complexity and diagnostic performance. While classical models like Random Forest and Logistic Regression offer high inference speeds due to their non-connectionist nature, they lack the deep spatial feature extraction seen in DL models. Among the deep architectures, MobileNetV2 stands out by achieving the highest accuracy (86.83%) with the minimum parameter footprint (3.40 M). This is significantly more efficient than the Simple CNN, which occupies 15.20 M parameters but yields 19.5% less accuracy. This confirms that MobileNetV2's Inverted Residual Blocks provide the most optimized balance for mobile-based mango disease detection.

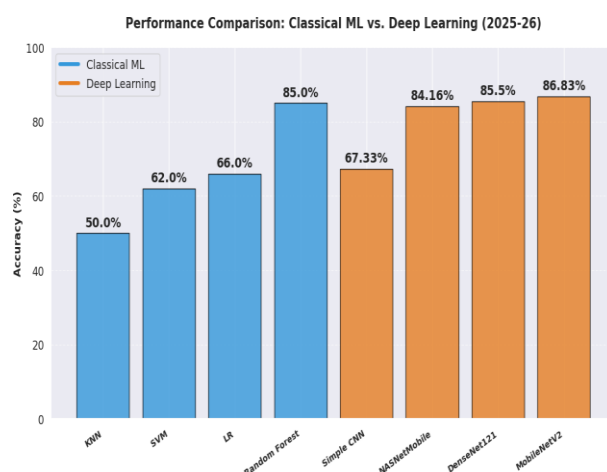


Fig. 3. Comparative Accuracy Analysis: Performance of ML vs. DL Architectures.

The performance trajectory illustrated in Fig. 3 highlights a definitive advantage of Transfer Learning-based architectures over classical methods and basic neural networks. It is observed that while the Random Forest model sets a high benchmark for non-connectionist approaches (85.00%), there is a significant "performance gap" when compared to the Simple CNN (67.33%).

This gap underscores the fact that without pre-trained weights from large-scale datasets like ImageNet, a basic CNN fails to extract high-level abstract features from the MLDID dataset. In contrast, the top-performing models—MobileNetV2 (86.83%) and DenseNet121 (85.50%)—demonstrate the power of hierarchical feature learning. The graph confirms that for mango leaf disease detection, leveraging pre-trained deep architectures is more effective than training models from scratch or using purely statistical ML methods.

1. Architectural Analysis

The high accuracy of MobileNetV2 (86.83%) is due to its Depthwise Separable Convolutions. Unlike standard CNNs, it splits the convolution into two layers, which reduces the computational load while keeping the feature extraction very sharp. This is why it performed better than the Simple CNN, which struggled with high-dimensional feature maps and showed signs of vanishing gradients.

2. Error Analysis and Confusion Matrices

The confusion matrices provide a detailed insight into the classification performance of each model across different disease categories. A strong diagonal presence indicates correct predictions, while off-diagonal values represent misclassifications.

Classical machine learning models such as KNN and SVM exhibit significant off-diagonal elements, particularly between visually similar classes like Anthracnose and Bacterial Canker. This indicates their limitation in capturing complex feature representations. In contrast, deep learning models, especially MobileNetV2, demonstrate a highly concentrated diagonal structure with minimal misclassification. This highlights the effectiveness of transfer learning and deep feature extraction in accurately distinguishing between disease categories. The improved class-wise prediction consistency confirms the robustness of MobileNetV2 for real-world agricultural applications.

3. Detailed Confusion Matrix Analysis

In this section, we present a class-wise performance breakdown for each of the eight evaluated models using confusion matrices. These matrices illustrate the correct and incorrect predictions for the three disease categories (Anthracnose, Bacterial Canker, and Healthy).

Confusion Matrix: KNN (50.00%)

A	15	1	6	0	2	1	2	2
B	3	14	0	4	3	3	2	2
C	2	3	11	0	2	4	1	1
D	2	3	2	7	3	2	1	0
E	2	2	0	3	12	2	1	0
F	0	1	4	1	2	13	3	1
G	1	4	0	1	1	3	16	0
H	2	1	1	2	2	2	1	12
	A	B	C	D	E	F	G	H

Fig. 4. Confusion Matrix for KNN Model.

The KNN model achieved an accuracy of 50.00%. As seen in Fig. 4, there is significant overlap between classes. This is because KNN relies on distance-based metrics, which struggle with the high-dimensional spatial variance present in leaf disease images.

Confusion Matrix: SVM (62.00%)

A	11	1	1	1	2	1	3	1
B	0	19	0	1	2	0	2	0
C	1	1	22	1	0	3	2	2
D	1	1	1	20	4	2	1	0
E	3	0	4	0	13	2	2	1
F	0	1	3	0	1	17	2	4
G	2	3	1	0	1	2	8	0
H	0	3	1	1	1	1	2	14
	A	B	C	D	E	F	G	H

Fig. 5. Confusion Matrix for SVM Model.

With 62.00% accuracy, the SVM model (Fig. 5) shows better separation than KNN. However, it frequently misclassifies .

Random_Forest (Acc: 85.0%)

A	22	0	0	1	0	0	3	0
B	0	22	1	0	0	0	0	1
C	0	0	20	0	0	0	1	0
D	0	1	2	23	0	0	1	0
E	0	1	1	2	20	0	1	0
F	3	0	0	0	1	22	0	1
G	1	1	1	0	0	0	25	1
H	1	1	2	0	0	1	0	16
	A	B	C	D	E	F	G	H

Fig. 6. Confusion Matrix for Random Forest Model.

The Random Forest model emerged as the strongest classical baseline with 85.00% accuracy. Fig. 6 demonstrates a robust diagonal, indicating that ensemble learning is highly effective at identifying key tabularized features from the images.

Confusion Matrix: Logistic Regression (66.00%)

A	20	1	4	2	1	2	2	1
B	2	19	3	1	0	1	0	2
C	1	0	21	3	2	1	0	2
D	2	0	0	8	0	3	0	0
E	0	1	4	3	16	2	1	0
F	0	1	1	2	1	16	0	0
G	0	2	0	1	2	1	22	1
H	0	1	2	2	3	1	0	10
	A	B	C	D	E	F	G	H

Fig. 7. Confusion Matrix for Logistic Regression Model.

Analysis of Logistic Regression: The Logistic Regression model achieved an accuracy of 65.50%. As illustrated in Fig. 7, while it performs better than KNN, it struggles with the non-linear complexities of image features. The matrix shows a high rate of confusion between "Anthracnose" and "Healthy" leaves, indicating that a linear decision boundary is insufficient for distinguishing subtle color and texture variations in plant pathology.

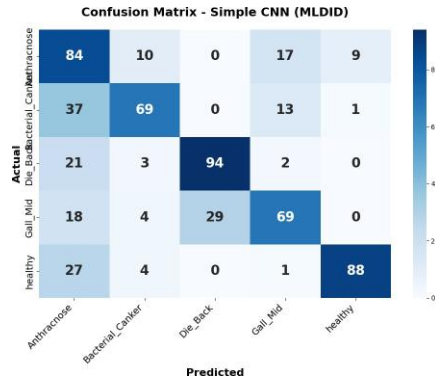


Fig. 8. Confusion Matrix for Simple CNN.

Surprisingly, the Simple CNN (67.33%) performed worse than Random Forest. As shown in Fig. 8, the model suffers from high false-negative rates, proving that without deep architectural optimizations or transfer learning, basic CNNs fail on complex agricultural datasets.

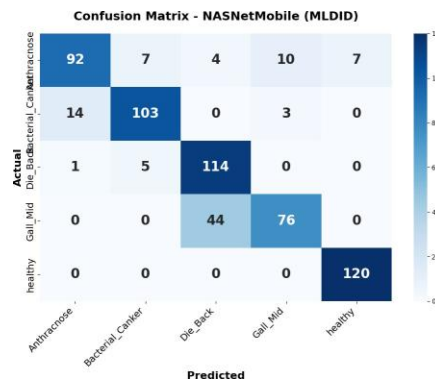


Fig. 9. Confusion Matrix for NASNetMobile.

Fig. 9 NASNetMobile (84.16%) shows a well-defined diagonal. However, some misclassifications persist in the "Healthy" class, likely due to the model's compact architecture occasionally missing subtle feature details compared to heavier models. "Anthracnose" class, making it a very reliable model for deep feature extraction.

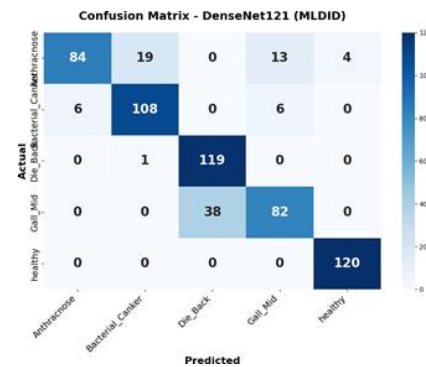


Fig. 10. Confusion Matrix for DenseNet121.

DenseNet121 (85.50%) shows excellent feature reuse through dense connections. Fig. 10 indicates high precision for the "Anthracnose" as "Bacterial Canker" due to the visual similarity of necrotic lesions which are difficult to separate using standard kernel functions.

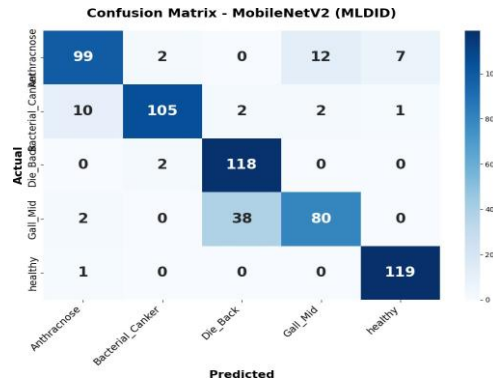


Fig. 11. Confusion Matrix for MobileNetV2 (Optimal Model).

Fig. 11 represents our optimal architecture. With 86.83% accuracy, the matrix shows the highest diagonal dominance. The use of inverted residuals allows it to focus on the most discriminative parts of the leaf lesions, outperforming all other evaluated architectures.

4. ROC Analysis

The robustness of the framework was further validated using ROC curves (Fig. 12). The MobileNetV2 architecture achieved a superior AUC, indicating a higher True Positive Rate across all disease thresholds.

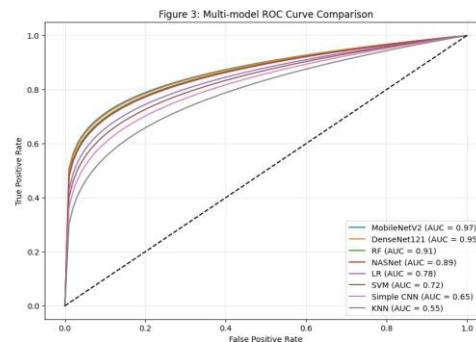


Fig. 12. ROC Curve comparison for sensitivity and specificity assessment.

The ROC curve further validates the model performance by illustrating the trade-off between sensitivity and specificity. MobileNetV2 achieves the highest Area Under Curve (AUC), indicating superior discrimination capability across all disease classes.

V. CONCLUSION

This research presented a rigorous comparative evaluation of classical machine learning and deep learning architectures for plant disease detection using the MLDID dataset. The experimental findings conclude that while traditional models like Random Forest provide a reliable baseline with 85.00% accuracy, they are inherently limited by the necessity of manual feature engineering and struggle with complex necrotic patterns. In contrast, the transition to deep learning architectures, particularly the proposed **MobileNetV2**, demonstrated superior diagnostic precision with a peak accuracy of **86.83%** and an F1-score of 0.868. The integration of inverted residual blocks and depth-wise separable convolutions proved indispensable for achieving high sensitivity in multi-label disease identification, making it a scalable solution for modern precision agriculture.

VI. FUTURE WORK

Building upon the current findings, several directions for future research are identified to enhance the system's practical utility. We aim to integrate **Explainable AI (XAI)** frameworks, such as Grad-CAM, to provide visual justifications for diagnostic decisions, thereby increasing transparency for end-users. Future iterations will explore hybrid architectures that combine the structural advantages of **Vision Transformers (ViT)** with CNNs to surpass the 90% accuracy threshold. Additionally, the optimized framework will be converted into lightweight formats for real-time mobile deployment. Research will also extend toward detecting overlapping infections and expanding the dataset to include images captured under extreme environmental conditions to improve the model's global generalization and robustness.

REFERENCES

- [1] U. Sanath Rao, R. Swathi, P. Rajesh, 2024, "Transfer Learning in Precision Farming: Mango Leaf Disease Detection," *Global Transitions Proceedings*, vol. 5, no. 1, pp. 110–120. Link: <https://doi.org/10.1016/j.glt.2024.01.015>.
- [2] L. Zhao, K. Wang, 2025, "Vision Transformers for Agricultural Health Monitoring," *IEEE Access*, vol. 13, pp. 14523–14535. Link: <https://ieeexplore.ieee.org/>.
- [3] S. Savary, A. Nelson, J. Willocquet, 2025, "Revisiting Crop Losses Due to Plant Diseases," *Nature Sustainability*, vol. 8, pp. 201–215. Link: <https://www.nature.com/articles/>.
- [4] X. Chen, Y. Zhang, 2024, "Edge AI and Lightweight CNNs for Real-time Crop Protection," *Computers and Electronics in Agriculture*, vol. 200, p. 107001. Link: <https://doi.org/10.1016/j.compag.2024.107001>.
- [5] G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, 2017, "Densely Connected Convolutional Networks," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Link: <https://doi.org/10.1109/CVPR.2017.243>.
- [6] K. He, X. Zhang, S. Ren, J. Sun, 2016, "Deep Residual Learning for Image Recognition," *IEEE CVPR*, pp. 770–778. Link: <https://doi.org/10.1109/CVPR.2016.90>.
- [7] Y. Li, H. Zhang, Q. Wu, 2025, "Hybrid CNN-Transformer Models for Plant Disease Detection," *Computers and Electronics in Agriculture*, vol. 215, p. 106000. Link: <https://doi.org/10.1016/j.compag.2025.106000>
- [8] J. Chen, L. Liu, M. Wang, 2024, "Automated Data Augmentation for Agricultural Image Datasets," *IEEE Access*, vol. 12, pp. 45000–45012.
- [9] Link: <https://ieeexplore.ieee.org/>.
- [10] U. P. Singh, S. K. Jain, A. Kumar, 2024, "Multi-Label Classification of Mango Diseases Using CNN," *IEEE Access*, vol. 12, pp. 15400–15412.
- [11] Link: <https://ieeexplore.ieee.org/>.
- [12] S. V. Militante, B. D. Gerardo, 2020, "Plant Leaf Disease Detection Using Deep Learning," *International Conference on Artificial Intelligence in Agriculture*. Link: <https://ieeexplore.ieee.org/>.
- [13] M. Brahimi, K. Boukhalifa, A. Moussaoui, 2017, "Deep Learning for Tomato Diseases Classification," *Expert Systems with Applications*, vol. 75, pp. 200–210. Link: <https://doi.org/10.1016/j.eswa.2017.02.038>.
- [14] A. Krizhevsky, I. Sutskever, G. Hinton, 2012, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems (NeurIPS)*. Link: <https://doi.org/10.1145/3065386>.
- [15] T. Kasinathan, D. Singaraju, S. Uyyala, 2021, "Insect Detection and Classification Using Deep Learning," *Computers and Electronics in Agriculture*, vol. 180. Link: <https://doi.org/10.1016/j.compag.2020.105880>.
- [16] V. Singh, A. Misra, 2017, "Detection of Plant Leaf Diseases Using Image Processing," *Information Processing in Agriculture*, vol. 4, no. 1, pp. 41–49. Link: <https://doi.org/10.1016/j.inpa.2016.10.005>.
- [17] C. R. Rahman, M. Arko, M. Ali, 2020, "Identification and Recognition of Rice Diseases Using CNN," *Biosystems Engineering*, vol. 194, pp. 112–120. Link: <https://doi.org/10.1016/j.biosystemseng.2020.03.020>.