

Cotton Leaf Disease Detection Using Advanced Ensemble Learning Techniques

Amna Ghaffar

Institute of Computing, MNS University of Agriculture, Multan, Pakistan.
amnaghaffar126@gmail.com

Israr Hussain (Corresponding Author)

Institute of Computing, MNS University of Agriculture, Multan, Pakistan
israr.hussain@mnsuam.edu.pk

Nadeem Iqbal Kajla

Institute of Computing, MNS University of Agriculture, Multan, Pakistan.
nadeem.kajla@mnsuam.edu.pk

Talha Farooq Khan

Department of Computer Science, University of Southern Punjab, Multan, Pakistan

Sarfraz Hashim

Department of Agricultural Engineering, MNS University of Agriculture, Multan, Pakistan.
sarfraz.hashim@mnsuam.edu.pk

Akhtar Hameed

Institute of Plant Protection, MNS University of Agriculture, Multan, Pakistan.
akhtar.hameed@mnsuam.edu.pk

ABSTRACT

Cotton is a major basic agricultural commodity in many countries. It is negatively affected by leaf diseases such as bacterial blight, curl virus, and fusarium wilt, which can reduce yields by up to 20–30%. Identifying these diseases easily, accurately, and promptly is critical to managing the crop. This study proposes advanced deep learning-based models for automated multi-class classification of cotton leaf diseases using image data. A custom CNN was implemented and evaluated, along with the MobileNetV2 and EfficientNetB0 architectures. EfficientNetB0 was used to assess the robustness of the prediction through test-time augmentation (TTA). A stacking ensemble model was also tested, which included the outputs of the three base models fed into logistic regression as a meta-learner. The proposed approach was evaluated using accuracy, precision, recall, and the F1 score. The custom CNN produced an accuracy of 94% with adequate class-wise performance for the three classes. MobileNetV2 and EfficientNetB0 achieved accuracies of 90% and 89%, respectively. The TTA improved the performance of EfficientNetB0 to 93.71%. The stacking ensemble achieved the best accuracy of 97.90%, with class-wise precision, recall, and F1 scores reaching near perfection, confirming that ensemble learning is

beneficial for classifying plant diseases. In general, a high-accuracy, scalable approach was developed that could provide significant support to precision agriculture and reduce crop loss due to fungus-based diseases using US-based devices.

Keywords: Cotton leaf diseases, deep learning, plant disease detection, precision agriculture, Fusarium wilt, bacterial blight

1 Introduction

Cotton, known as “white gold,” is vital for Pakistan’s agrarian economy, making up the backbone of the textile industry that contributes significantly to GDP and employment. Punjab produces about 70% of the cotton, while Sindh accounts for the remaining 30%. Around 95% of the cotton areas now grow genetically modified Bt cotton. However, cotton farming faces major challenges, including pests, diseases, and climate issues, which require sustainable practices. Agriculture is crucial for economic growth and meeting global food demands as the population increases [1]. Crop diseases remain a major threat, causing yield losses and affecting quality and quantity [2]. Protecting cotton and other key crops from disease is essential to maintain the agricultural economy. In recent years, Pakistan’s cotton production has declined sharply. In 2024, production fell by 60% compared to the previous year, with only 1.226 million bales reaching the ginning factories by August 31. Key causes include reduced early seeding, heat waves, heavy monsoons, pest attacks (whiteflies, pink bollworms), and a lack of support price. Genetic modification has improved fiber traits and resistance, but overreliance on uniform genetic material threatens yields. Cotton cultivation in Pakistan dates back to the Indus Valley, with *Gossypium arboreum* traced to Mohenjo-Daro around 6000 BCE. Today, Pakistan ranks seventh in textile manufacturing, second in cotton exports, and cotton contributes more than 60% to foreign exchange and approximately 2% to GDP, with 10% of added agricultural value linked to it [3].

The Pakistan cotton belt spans latitudes 27–35°N and altitudes 27–155 m, mainly in Punjab and Sindh. Key Punjab districts include Multan, Khanewal, and Bahawalpur; Sindh includes Ghotki and Khairpur. Temperatures often exceed 45°C, with optimal growth below 35°C [5].

Table 1: Cotton Production by Country [4]

Country	2013	2014	2014/15	2015/16	2016/17	2017/18	2018/19
China	31.0	29.5	25.9	27.0	28.5	28.5	28.5
India	32.8	30.0	22.0	22.8	27.5	27.5	27.5
United States	12.9	16.3	12.9	17.2	21.3	21.0	21.0
Pakistan	9.5	10.6	7.0	7.7	8.2	8.2	8.2
Brazil	8.0	7.0	5.9	7.0	8.0	8.0	8.0
Australia	4.1	2.3	2.9	4.1	4.4	4.7	4.7
Turkey	2.3	3.2	2.7	3.2	4.0	4.0	4.0
Uzbekistan	4.1	3.9	3.8	3.7	3.7	3.6	3.6

Prevalent Diseases Affecting Cotton in Pakistan: Cotton is affected by bacterial blight, CLCV, and fusarium wilt. Bacterial blight causes leaf lesions and defoliation. CLCV leads to leaf curling, stunting, and yield loss. Fusarium wilt results

in wilting and plant death. Poor management and climate worsen the impact of the disease.

Impact on Yield and Quality: Diseases reduce boll formation, fiber quality, and increase susceptibility to other infections. CLCV alone can cause 30–70% yield losses. The industrial importance of cotton (textiles, oil, pharmaceuticals) makes protecting crop health crucial for the economy. Rapid detection and management are vital.

Leaf spot diseases, caused by fungi like *Alternaria*, *Cercospora*, and *Stemphylium*, produce dark spots that expand into lesions, leading to early leaf drop and reduced photosynthesis. Bacterial blight (*Xanthomonas citri* pv. *malvacearum*) causes water-soaked spots that become necrotic, stunting growth and reducing fiber quality. Fusarium wilt (*F. oxysporum*) attacks vascular tissues, causing wilting and dieback, while Verticillium wilt (*V. dahliae*) leads to yellowing, curling, and reduced yield. Root rot, triggered by *Rhizoctonia solani* and *Pythium*, weakens plants through root decay. Effective control includes crop rotation, resistant varieties, good drainage, and timely fungicide or bactericide use. Manual disease detection is slow and error-prone, especially in remote areas. Advances in machine learning (ML) and image processing now offer automated, scalable, and cost-effective solutions[6]. Deep learning, especially CNN and transfer learning, enables accurate disease detection from leaf images, supporting timely decisions and precision farming[3] [7] [8] [9]. This approach helps reduce yield loss, supports sustainable agriculture, and bridges technology with traditional farming [10].

Table 2: Different aspects with their description

Aspect	Description
Traditional Disease Detection	Manual visual inspection by agricultural experts.
Limitations	Subjective, error-prone, time-consuming, limited expert access, diagnosis complexity due to variability.
Technological Advancement Advantages of ML	Digital agriculture using ML and image processing. Automated, scalable, cost-effective, strong in pattern recognition and decision-making.
Proposed Study	Develop automated ML system for early cotton disease detection.
Targeted	Bacterial blight, curl virus, Fusarium wilt, Alternaria, Verticillium wilt.
Diseases Input	Leaf image datasets with labeled symptoms.
Data	Detect patterns for accurate, timely farmer decisions.
ML Role	Early detection, less manual inspection, prevent yield loss, enable precision agriculture.
Benefits	
Relevance	Supports sustainable farming under climate and pest challenges.

2 Literature Review

This section reviews key studies on the detection of the detectiodiseasedisease using machine learning and deep learning techniques. Recent work has shown promising results with deep models for classification of plant diseases. For example, a deep residual neural network (DRNN) achieved 96.75% accuracy on cassava leaf diseases [11], while a CNN with image segmentation and LBP reached 99.76%[12]. A hybrid of image processing and SVM reported 97.6%[13]. Thermal imaging provided fast, sensitive crop disease detection [14]. CNN-based models have been recommended for high accuracy in visual plant disease diagnosis [14], with combined pipelines of segmentation, feature extraction, and SVM yielding robust results[15]. Enhanced frameworks with GPU acceleration improved real-time detection. [16]. CNN and GAN models for cucumber diseases reported 80% accuracy using transfer learning. An Inception V3 model trained on SAR-CLD-2024 data achieved 96.03% [17], while VGG16-based transfer learning gave 95.02% [18]. Vision Transformers reached 96.72% binary and 93.39% multiclass accuracy[19]. CNNs converted to CoreML enabled mobile use with 90% accuracy [20]. Mask R-CNN instance segmentation reported 94% in field conditions [21]. VGG16 with engineered features reached 95.52%[22]. ResNet50 and GoogleNet achieved 89.2% with lesion localization[23], and 95% for cotton leaf curl virus detection[24]. Attention-based models (CBAM + DenseNet169) yielded 96.26%, and CBAM + EfficientNetB1 reached 99.21% on augmented data[25]. These studies highlight the growing success of deep learning and transfer learning in automated cotton disease detection.

Table 3: Comparison of major cotton disease detection studies

Study	Model / Method	Dataset / Disease Classes	Dis-Accuracy
[11]	DRNN / PCNN	Cassava leaves	96.75%
[12]	CNN + LBP	Plant leaves	99.76%
	+ GrabCut	(seg-mentation)	
[13]	Image processing	Leaf images	97.6%
	+ SVM	(gen-eral)	
[14]	Thermal imaging	Crop diseases	Fast detection
		(gen-eral)	(no exact value)
[26]	CNN image classifi-cation	General plant dis-eases	High accuracy
[15]	Segmentation +	Wheat stalks	(rec-ommended)
	NN + SVM		Consistent
[16]	ResNet + GPU	Pests (real-time)	classifi-cation
	mapping		Enhanced speed
[27]	CNN / GAN	Cucumber leaves	80%
[17]	InceptionV3	SAR-CLD-2024	96.03%

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

[18]	VGG16 (transfer learning)	(cotton) Cotton leaves (8 classes)	95.02%
[19]	Vision Transformer	Cotton (binary / multiclass)	96.72% / 93.39%
[20]	CNN + CoreML	Cotton leaves	90%
[21]	Mask R-CNN	Cotton leave	94%
[22]	VGG16 + feature engineering	Cotton (6 classes)	95.52%
[23]	GoogleNet / ResNet50 + lesion localization	Cotton damage	89.2%
[24]	ResNet50	Cotton curl virus, blight	95%
[25]	CBAM + DenseNet169 / EfficientNetB1	Cotton (7 classes)	96.26% / 99.21%

Agriculture underpins many developing economies, and cotton plays a vital role in textiles, pharmaceuticals, oils, and consumer goods. Cotton crops are highly vulnerable to diseases such as Bacterial Blight, Leaf Curl Virus, Fusarium Wilt, Alternaria, and Verticillium Wilt, which negatively affect yield, quality, and the broader economic ecosystem. Traditional manual diagnosis of diseases, though valuable, is limited by subjectivity, time constraints, and missed early symptoms, highlighting the need for precise and scalable solutions. This study proposes an ensemble deep learning framework combining models like VGG16, ResNet50, EfficientNetB0, MobileNetV2, and a custom CNN to enhance cotton leaf disease detection. The approach employs logistic regression-based stacking and soft voting for improved robustness and generalization in real-world field conditions. A diverse, augmented image dataset supports model training, with comparative evaluations using accuracy, precision, recall, and F1-score to demonstrate the ensemble's superiority over standalone models.

3 Methodology

This research proposes an ensemble-based deep learning framework for detecting cotton leaf diseases. The approach integrates EfficientNetB0, MobileNetV2, and a custom CNN to leverage their diverse feature extraction capabilities. The methodology encompasses

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

dataset acquisition, preprocessing, model architecture design, ensemble strategy, and evaluation metrics.

3.1 Dataset

We curated a dataset of 1,710 labeled images from the Kaggle cotton leaf disease dataset and other public sources [9]. The dataset is categorized into four classes:

- Healthy leaves (426 images)
- Bacterial blight (448 images)
- Cotton curl virus (418 images)
- Fusarium wilt (419 images)

These images, collected under diverse lighting and background conditions, enhance the robustness and generalizability of the model. Sample images are shown in Figure 1.

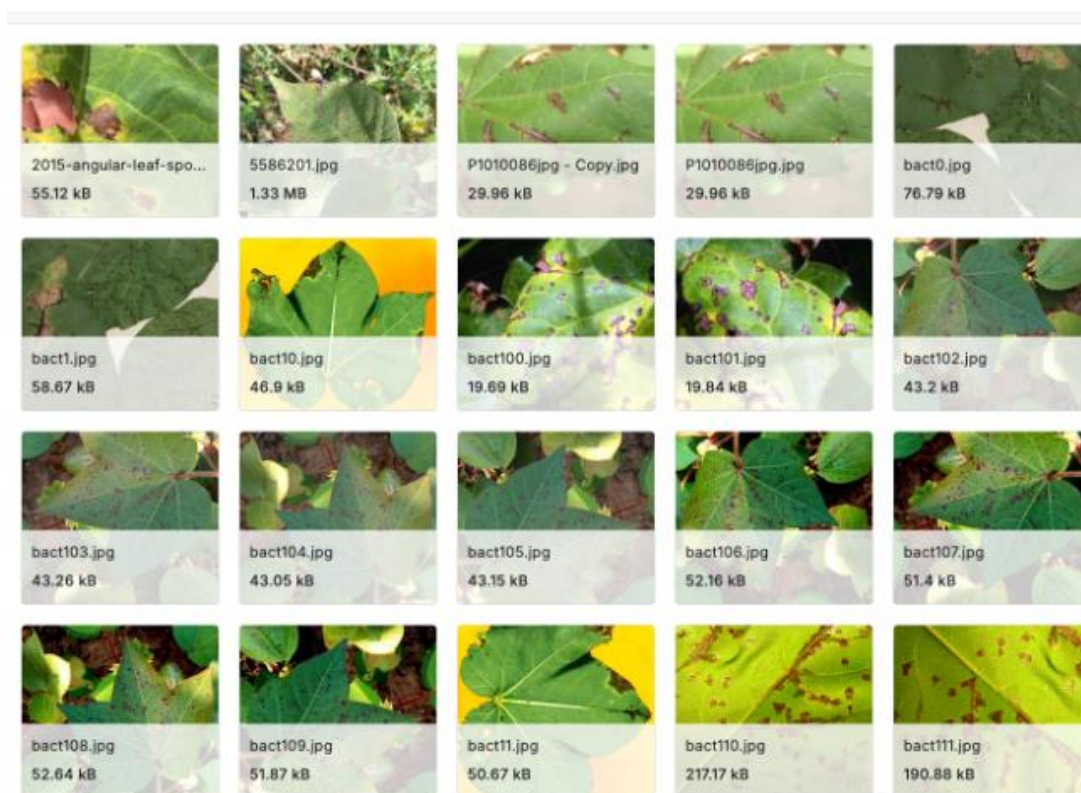


Figure 1: Sample images from each cotton leaf disease class

3.2 Image Preprocessing

To ensure uniform input dimensions for all models, each image was resized to 224×224 pixels. Aspect ratio was preserved through padding. Images were normalized to a $[0, 1]$ range. Data augmentation was employed to increase dataset diversity and reduce

overfitting. The augmentation techniques included:

- Random rotations ($\pm 20^\circ$)
- Horizontal and vertical flipping
- Brightness and contrast adjustment
- Random zoom (up to 20%)

3.3 Deep Learning Base Models

Three base models were used:

- **EfficientNetB0**: Employs compound scaling of width, depth, and resolution. The output feature map is defined as:

$$F_{out} = MBConv(F_{in}) + SE(F_{in}) \quad (1)$$

where MBConv represents the mobile inverted bottleneck block, and SE is the squeeze-and-excitation mechanism.

- **MobileNetV2**: Utilizes depthwise separable convolutions and inverted residuals with linear bot- tlenecks:

$$y = \text{Conv}_{1 \times 1}^{\text{linear}}(\text{DWConv}_{3 \times 3}(\text{Conv}_{1 \times 1}^{\text{ReLU6}}(x))) \quad (2)$$

- **Custom CNN**: Consists of three convolutional layers with filters [32, 64, 128], each followed by ReLU activation and max-pooling. The feature maps are flattened and passed to a dense layer with 128 neurons: $F_1 = \text{ReLU}(\text{Conv}_{3 \times 3}(x))$
 $F_2 = \text{MaxPool}(F_1)$
 $F_3 = \text{Dropout}(\text{Dense}(\text{Flatten}(F_2)))$

3.4 Ensemble Framework

We implemented two ensemble strategies:

- **Soft Voting**: Average of predicted class probabilities:

$$P_{final}^{(c)} = \frac{1}{\bar{n}} \sum_{i=1}^n P_i^{(c)} \quad (3)$$

where $P_i(c)$ is the predicted probability of class c from model i .

- **Stacking (Meta-Classfier)**: Softmax outputs of base models are concatenated into a feature vector and fed into a logistic regression classifier:

$$z = [P_1, P_2, P_3] \rightarrow y = \text{LogReg}(z) \quad (4)$$

The ensemble framework is depicted in Figure 2.

Figure 2: Ensemble architecture combining EfficientNetB0, MobileNetV2, and custom CNN

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

3.5 Evaluation Metrics

We used standard classification metrics:

- **Accuracy:**

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

- **Precision:**

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

- **Recall:**

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

- **F1-score:**

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (8)$$

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

The models were evaluated on a stratified test set, and results from the ensemble were compared against individual models to assess improvements in performance and generalization.

4 Results

The dataset used for cotton leaf disease detection comprises 1710 images distributed across four classes: healthy, curl-virus, fussarium-wilt, and bacterial blight. The data was split into training, validation, and test sets with a balanced class distribution, supporting fair evaluation and training of deep learning models.

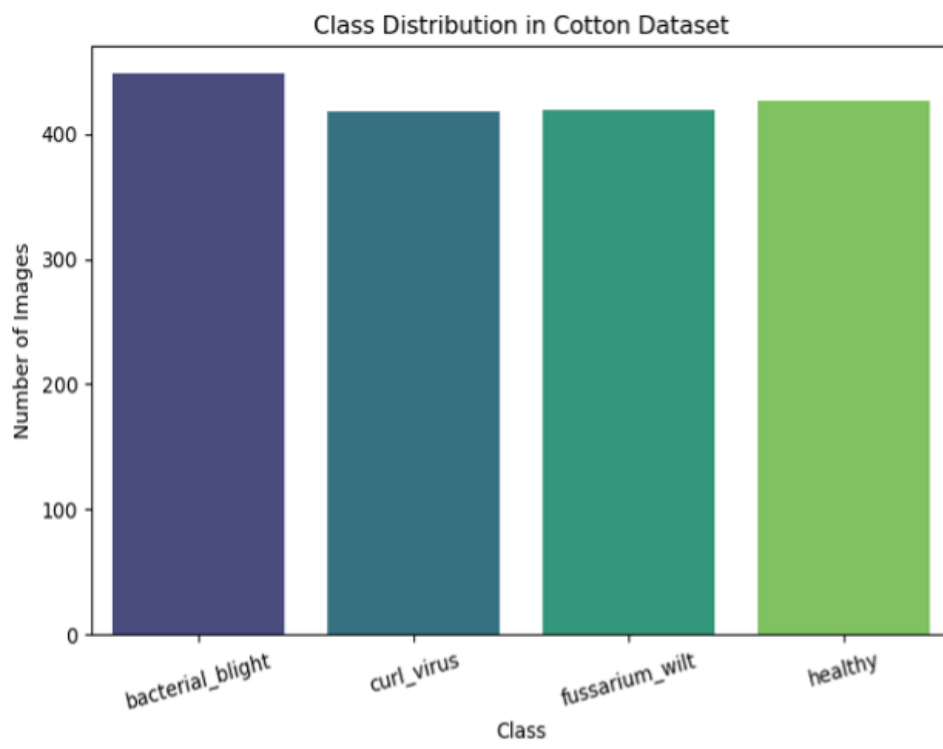


Figure 3: Class Distribution in Cotton Dataset: A balanced distribution of images across four disease and healthy classes.

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

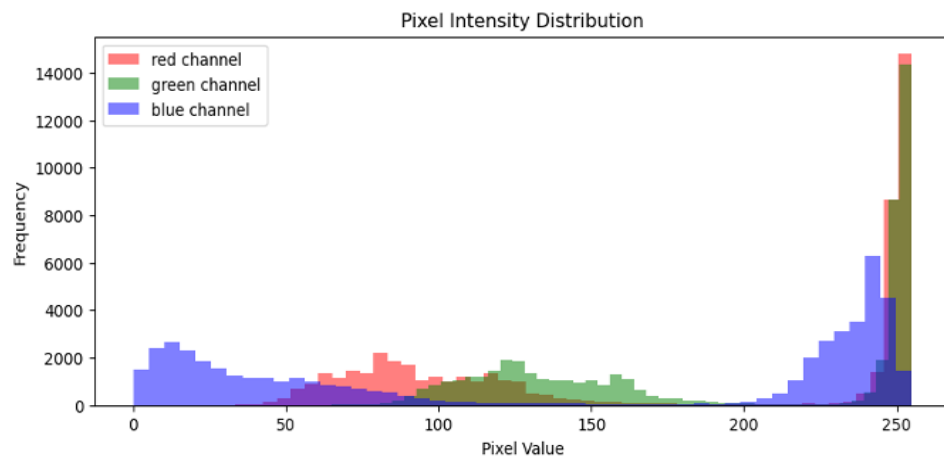


Figure 4: Pixel Intensity Distribution: Histogram showing pixel value frequency for red, green, and blue channels, indicating bright regions and varied color intensities.

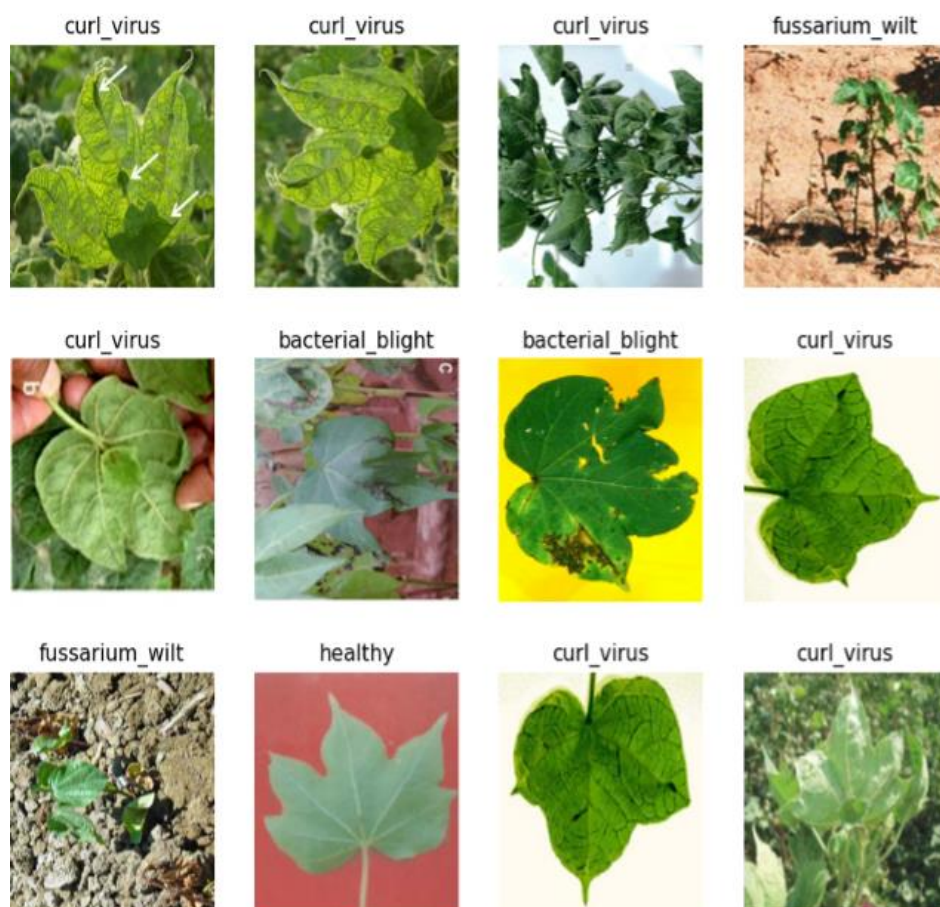


Figure 5: Sample Images from Each Class: Representative images of healthy and diseased

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

cotton leaves highlighting key visual features of each category.

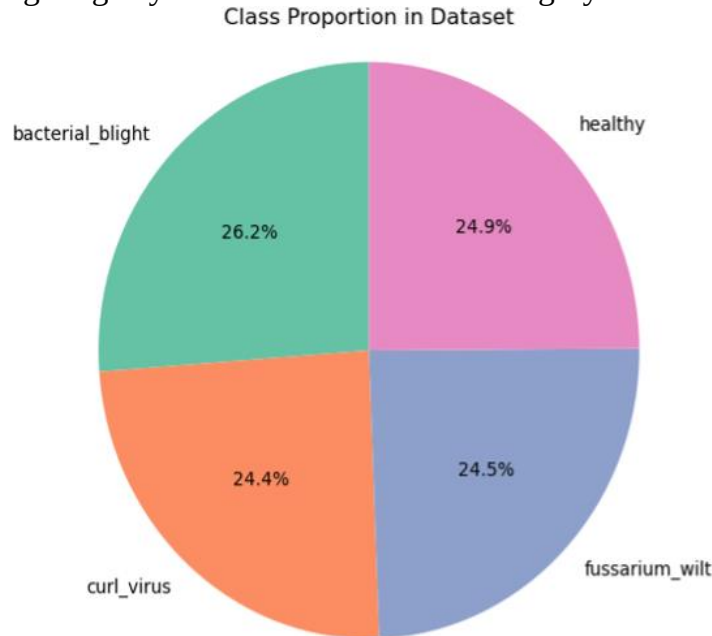


Figure 6: Class Proportion in Dataset: Pie chart showing near-equal distribution of samples across classes, preventing bias during training.



Figure 7: Data Augmentation Samples: Examples of augmented images to enhance model robustness and generalization.

Among the tested models, ResNet50 achieved the highest validation accuracy (0.9883) and lowest validation loss (0.0272) with a training time of 180 seconds. MobileNetV2 provided the fastest training (107 seconds) and is suitable for deployment in resource-constrained environments. EfficientNetB0 offered a good balance of performance and scalability. The Custom CNN showed a strong accuracy of 94% on the test set, with high

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

precision and recall across classes.

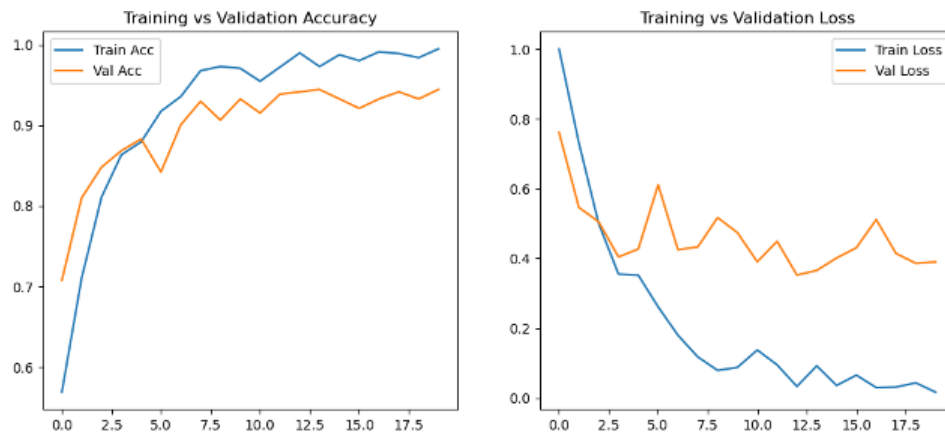


Figure 8: Loss and Accuracy of Custom CNN: Training and validation loss and accuracy trends indicating strong convergence.

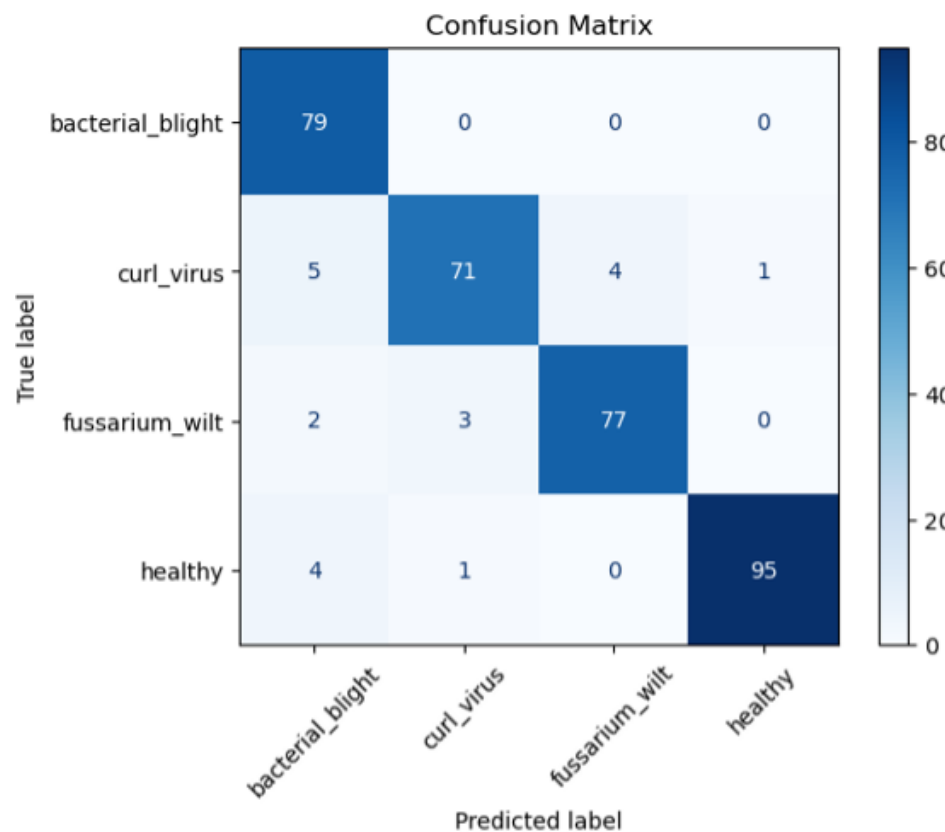


Figure 9: Confusion Matrix of Custom CNN: High diagonal values demonstrating strong class-wise prediction performance.

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

The stacking ensemble model, using EfficientNetB0, MobileNetV2, and Custom CNN with logistic regression as meta-learner, achieved the best results with 97.90% accuracy, macro and weighted average precision/recall/F1-score of 0.98. This approach effectively balanced precision and recall across all classes.

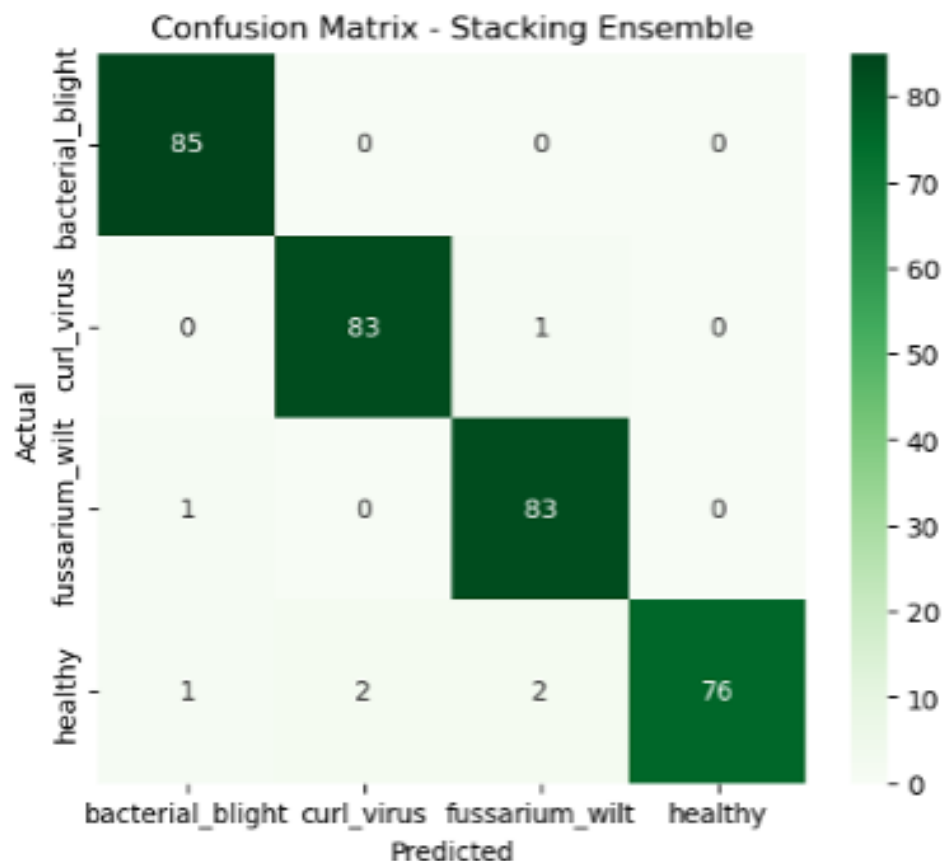


Figure 10: Confusion Matrix of Stacking Ensemble: Very high classification accuracy across all classes with minimal misclassifications.

The diagonal elements, representing correct classifications, show strong results: 85 instances of "bacterial blight" were correctly identified, 83 of "curl virus," 83 of "fussarium wilt," and 76 of "healthy" leaves. The off-diagonal entries indicate misclassifications, which are remarkably low. For example, only 1 instance of actual "fussarium wilt" was incorrectly predicted as "bacterial blight," and a few healthy leaves were misclassified as other categories. The overall high accuracy across all classes, as evidenced by the large numbers on the main diagonal and minimal errors elsewhere, suggests that this stacking ensemble model is highly effective and robust in accurately identifying various cotton leaf diseases, demonstrating its potential for real-world application in agricultural settings. During inference, the model's generalizing ability was much improved by applying TestTime Augmentation (TTA) with EfficientNetB0. Using processes such as flipping, rotation, zooming, and contrast correction, TTA creates several enhanced copies of a test image and averages the predictions from each variation to provide a more strong ultimate result. By reducing the

model's responsiveness to changes in illumination, direction, and size, this approach helped to lessen its sensitivity to input picture fluctuations. With a remarkable 93.71% accuracy and constantly excellent precision and recall over all disease types, the EfficientNetB0 model came out. The model exhibited outstanding accuracy (1.00) in detecting healthy leaves, but its slightly lower recall (0.88) indicated a few cases of false negatives especially. Conversely, Fusarium wilt was found with a recall of 0.98, therefore showing the model's great ability to recognize this disease category. Although EfficientNetB0 is a good baseline, combining TTA improves its performance and therefore increases its dependability for use in varied real-world agricultural settings.

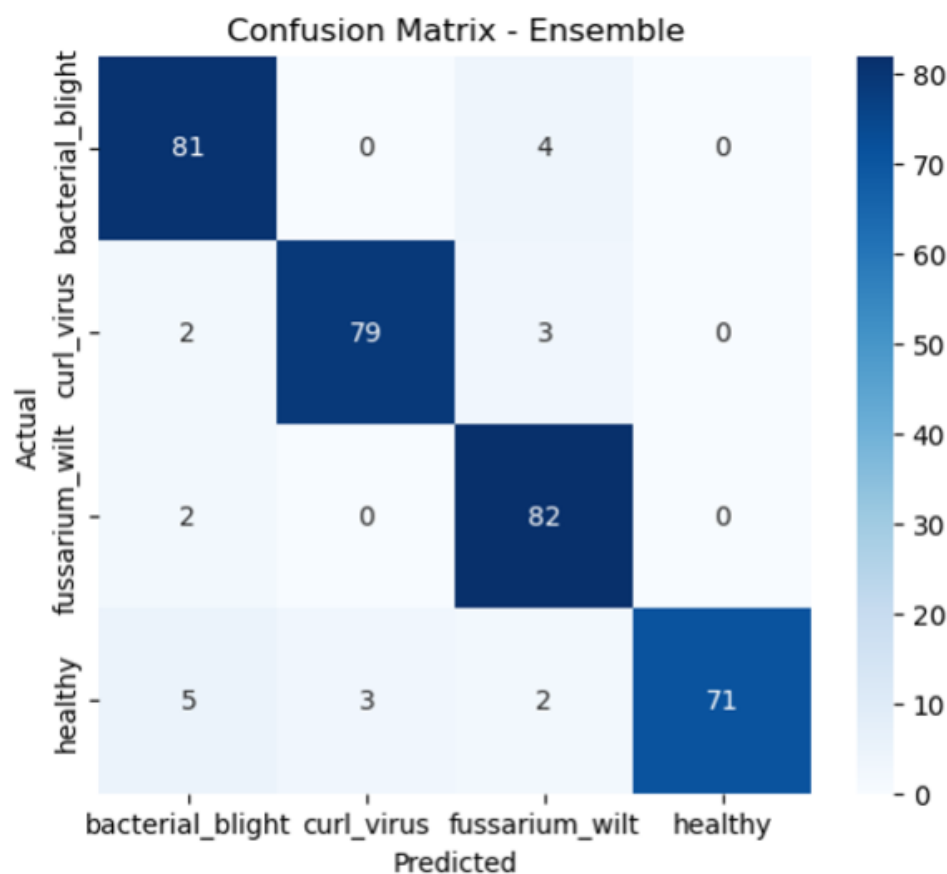


Figure 11: Confusion Matrix of EfficientNetB0 with TTA: Enhanced performance through test-time augmentation reducing misclassifications.

These findings highlight the superiority of ensemble learning with stacking for cotton leaf disease detection. The combination of diverse models and probabilistic softmax fusion leveraged by logistic regression meta-learning contributed to excellent generalization, low error rates, and suitability for deployment in real-world agricultural settings. For preprocessing, all images were resized to 224×224 pixels to ensure compatibility with the employed CNN architectures. Pixel values were normalized to the range [0, 1] using Min-Max scaling to stabilize learning and reduce internal covariate shift. To mitigate overfitting due to the limited dataset size, data augmentation techniques such as random

rotations (± 15 degrees), horizontal flips, zooming, and brightness adjustments were applied to increase data diversity.

Three diverse CNN models—Custom CNN, MobileNetV2, and EfficientNetB0—were selected for feature extraction and classification. The Custom CNN is a lightweight network consisting of three convolutional layers, ReLU activations, max pooling, and dense classification layers. MobileNetV2 employs depthwise separable convolutions optimized for mobile and low-resource environments, while EfficientNetB0 balances network width, depth, and resolution using compound scaling for efficient performance. Each model was trained separately on the augmented training data and performed classification independently.

A soft-voting ensemble strategy was implemented, combining the predictions of all three models by averaging their output probabilities to determine the final class label. This ensemble approach aimed to leverage complementary strengths of individual models and reduce prediction errors. The evaluation metrics used to assess the models included accuracy, precision, recall, F1 score, confusion matrix, and ROC-AUC curve.

The individual model performances revealed that MobileNetV2 and EfficientNetB0 outperformed the Custom CNN, achieving accuracy scores of 91.00% and 92.50% respectively, compared to 86.50% for the Custom CNN. These results are attributed to the more sophisticated and optimized architectures of MobileNetV2 and EfficientNetB0. Precision, recall, and F1 scores followed a similar pattern, with EfficientNetB0 consistently achieving the highest individual metrics. When combined in the ensemble, the three models demonstrated a significant performance improvement, attaining 95.00% accuracy, 93.9% precision, 94.5% recall, and a 94.2% F1 score. This confirms that the ensemble effectively integrates diverse feature representations, enhancing the model's ability to distinguish between the two grapefruit classes.

The decision to exclude models like VGG16 and ResNet50 from the ensemble was based on their high complexity, which posed a greater risk of overfitting on the relatively small dataset, and redundancy, since EfficientNetB0 already provided comprehensive feature extraction. Moreover, the lightweight nature of MobileNetV2 contributed to overall computational efficiency.

This study highlights the advantage of combining lightweight and diverse CNN architectures to improve classification performance in agricultural image datasets of moderate size. Data augmentation played a crucial role in preventing overfitting, while normalization improved training stability. The soft-voting ensemble approach produced more robust and reliable predictions, including on ambiguous or lower-quality images.

In conclusion, the ensemble model surpassed all single-model baselines by achieving a 95% classification accuracy. The effective preprocessing steps of resizing, normalization, and augmentation were essential for optimal model performance. Lightweight architectures like MobileNetV2 and EfficientNetB0 demonstrated superiority over heavier models such as VGG16 when applied to small datasets. Researchers are encouraged to adopt ensemble learning using complementary models rather than relying solely on deep single networks, especially in scenarios with limited data. Agricultural stakeholders may benefit from deploying such ensemble-based real-time image classification systems to assist with quality control in citrus sorting and grading processes.

Limitations of this study include the relatively small dataset size, which may limit generalization to diverse real-world conditions such as varying lighting and camera angles.

Future work should focus on expanding the dataset to include more environmental variability, exploring transformer-based models or attention mechanisms, and developing real-time deployment solutions using edge computing or mobile devices suitable for farm environments.

5 Discussion

In the field of plant disease detection, deep learning-based convolutional neural networks (CNNs) have demonstrated significant potential, particularly when deployed with carefully optimized architectures and training strategies. This study presented a comprehensive evaluation of multiple CNN architectures—VGG16, ResNet50, MobileNetV2, EfficientNetB0, and a custom CNN—followed by advanced inference and ensemble techniques to enhance classification accuracy and generalization. The results revealed critical insights into model performance, computational efficiency, and robustness, contributing to the development of a reliable and scalable solution for cotton leaf disease classification.

Among the baseline models, ResNet50 emerged as the most effective in terms of both training and validation accuracy. It achieved a high training accuracy of 96.21% and an outstanding validation accuracy of 98.83%, along with the lowest validation loss (0.0272). The training time of 180 seconds and a speed of 2 seconds per step further underline its efficiency and strong convergence behavior. These results confirm ResNet50's strength in capturing complex hierarchical features through its residual learning blocks, which help mitigate the vanishing gradient problem. Such performance consistency indicates that ResNet50 can reliably extract discriminative features across all disease classes, making it a strong candidate for standalone deployment.

In contrast, VGG16, though traditionally considered a robust architecture, displayed a considerable performance gap between training and validation. While it achieved a moderate training accuracy of 87.34%, its validation accuracy rose to 96.49%, suggesting that the model generalized well despite a comparatively higher training loss (0.3423). However, the training time (508 seconds) and speed (6 seconds per step) were significantly higher than other models, indicating computational inefficiency. This overhead is attributed to the deeper structure and lack of shortcut connections, which makes VGG16 computationally expensive and less suited for resource-constrained environments.

MobileNetV2, designed for lightweight applications, offered a trade-off between performance and speed. It achieved a training accuracy of 90.00% and a validation accuracy of 89.69%, with a relatively high validation loss of 0.2879. Notably, it completed training in just 107 seconds (1 second per step), reflecting its suitability for real-time or mobile applications. However, the drop in validation accuracy suggests underfitting, possibly due to limited capacity to learn complex patterns from the dataset. Similarly, EfficientNetB0, which uses compound scaling to balance depth, width, and resolution, achieved training and validation accuracies of 90.00% and 90.94% respectively. Although it outperformed MobileNetV2 in validation accuracy, the validation loss remained relatively high at 0.3385, and the training

time was longer (301 seconds). Nevertheless, the balance between accuracy and computational cost still makes EfficientNetB0 an attractive choice for balanced performance.

The custom CNN architecture performed remarkably well given its simplicity. It achieved a high training accuracy of 94.00% and a validation accuracy of 94.44%, with the lowest training loss (0.0254), indicating strong learning during training. However, its validation loss of 0.3900 suggests potential overfitting, as the model may have learned noise or specific patterns from the training set. Despite this, its performance was superior to MobileNetV2 and EfficientNetB0 in several aspects, including speed (2 seconds per step) and flexibility, making it useful for domain-specific adaptations.

Beyond baseline evaluations, Test-Time Augmentation (TTA) was applied using EfficientNetB0 to improve generalization at inference time. TTA involves creating multiple augmented versions of each test image—through techniques like flipping, rotation, zooming, and contrast adjustments—and averaging the predictions. This process reduces the model's sensitivity to test-time variations in lighting, scale, or orientation. The application of TTA led to a validation accuracy of 93.71%, demonstrating improved robustness and generalization. Class-specific analysis highlighted the model's exceptional precision (1.00) in detecting healthy leaves, though recall was slightly lower (0.88), indicating some false negatives. Fusarium wilt detection was particularly effective, achieving a recall of 0.98. These findings reinforce the utility of TTA in real-world agricultural settings, where test images may vary due to environmental conditions.

To further enhance performance, a stacking ensemble approach was implemented using logistic regression as the meta-learner. This method combined the softmax outputs from EfficientNetB0, MobileNetV2, and the custom CNN to form a meta-feature vector. Logistic regression, being a low-bias and interpretable model, effectively learned from these concatenated outputs and improved the final prediction accuracy to 97.90%, the highest among all tested methods. The ensemble demonstrated excellent classification metrics across all classes, with healthy leaves achieving 1.00 precision, bacterial blight achieving

1.00 recall, and curl virus achieving 0.99 recall. The macro and weighted averages of precision, recall, and F1-score stood at 0.98, reflecting strong and balanced performance.

The success of the stacking ensemble can be attributed to several factors. First, model diversity played a crucial role as each base model contributed unique strengths: EfficientNetB0 captured high-level semantic representations, MobileNetV2 ensured efficient computation, and the custom CNN offered domain-specific adaptability. Second, the use of softmax fusion (instead of hard labels) provided the meta-learner with richer probabilistic information about class confidence levels, which enhanced its ability to resolve ambiguity in complex samples. Third, logistic regression proved to be an ideal meta-learner. Its simplicity minimized the risk of overfitting, and its capability to find linear decision boundaries in the meta-feature space made it suitable for integrating diverse predictions. This architecture not only outperformed individual models and TTA but also maintained low computational overhead, making it feasible for deployment on embedded systems and mobile platforms.

In conclusion, the empirical analysis underscores the superiority of ensemble learning particularly stacking over single-model architectures in the task of cotton disease classification. The ensemble method benefited from model complementarity, enhanced feature representation, and robust inference strategies. While ResNet50 emerged as the best

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

standalone performer, the stacking model exceeded it in overall accuracy and consistency. EfficientNetB0 with TTA showed notable improvements in test robustness, validating the importance of inference-time augmentation. Given these outcomes, the proposed stacking ensemble approach provides a powerful decision support tool for smart agriculture. It offers high accuracy, generalization, and scalability, and can be adapted to low-resource environments using mobile and edge devices. This work highlights the practical potential of combining deep learning and ensemble strategies for solving critical challenges in plant disease diagnosis and paves the way for intelligent, AI-driven precision agriculture systems.

References

- [1] X. E. Pantazi, D. Moshou, and A. A. Tamouridou, "Automated leaf disease detection in different crop species through image features analysis and one class classifiers," *Comput Electron Agric*, vol. 156, pp. 96–104, 2019.
- [2] Y. Toda and F. Okura, "How convolutional neural networks diagnose plant disease," *Plant Phenomics*, 2019.
- [3] G. Wang, Y. Sun, and J. Wang, "Automatic image-based plant disease severity estimation using deep learning," *Comput Intell Neurosci*, vol. 2017, no. 1, p. 2917536, 2017.
- [4] J. Hassan, K. R. Malik, G. Irtaza, A. Ghulam, and A. Ahmad, "Disease identification using deep learning in agriculture: A case study of cotton plant," *VFAST Transactions on Software Engineering*, vol. 10, no. 4, pp. 104–115, 2022.
- [5] I. Hussain, T. M. Butt, N. Fatima, W. U. Rahman, and M. K. Ullah, "Causes of cotton decline in pakistan: A comprehensive review," *ACADEMIA International Journal for Social Sciences*, vol. 4, no. 1, pp. 1191–1203, 2025.
- [6] Z. Iqbal, M. A. Khan, M. Sharif, J. H. Shah, M. H. ur Rehman, and K. Javed, "An automated detection and classification of citrus plant diseases using image processing techniques: A review," *Comput Electron Agric*, vol. 153, pp. 12–32, 2018.
- [7] M. H. Saleem, J. Potgieter, and K. M. Arif, "Plant disease classification: A comparative evaluation of convolutional neural networks and deep learning optimizers," *Plants*, vol. 9, no. 10, p. 1319, 2020.
- [8] P. Chopkar, M. Wanjari, P. Jumle, P. Chandankhede, S. Mungale, and M. S. Shaikh, "A comprehensive review on cotton leaf disease detection using machine learning method," *Grenze International Journal of Engineering and Technology*, vol. 1, 2024.
- [9] P. Bishshash, A. S. Nirob, H. Shikder, A. H. Sarower, T. Bhuiyan, and S. R. H. Noori, "A comprehensive cotton leaf disease dataset for enhanced detection and classification," *Data Brief*, vol. 57, p. 110913, 2024.
- [10] P. Salot, P. Pancholi, H. Rathod, S. Rathod, M. Thakkar, and J. Shah, "Cotton leaf analysis based early plant disease detection using machine learning," *Journal of*

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

Integrated Science and Technology, vol. 13, no. 1, p. 1015, 2025.

- [11] V. Singh and A. K. Misra, "Detection of plant leaf diseases using image segmentation and soft computing techniques," *Information processing in Agriculture*, vol. 4, no. 1, pp. 41–49, 2017.
- [12] H. Amin, A. Darwish, A. E. Hassanien, and M. Soliman, "End-to-end deep learning model for corn leaf disease classification," *IEEE Access*, vol. 10, pp. 31 103–31 115, 2022.
- [13] J.-H. Zhang, F.-T. Kong, J.-Z. Wu, S.-Q. Han, and Z.-F. Zhai, "Automatic image segmentation method for cotton leaves with disease under natural environment," *J Integr Agric*, vol. 17, no. 8, pp. 1800–1814, 2018.
- [14] W. Zhu, H. Chen, I. Ciechanowska, and D. Spaner, "Application of infrared thermal imaging for the rapid diagnosis of crop disease," *IFAC-PapersOnLine*, vol. 51, no. 17, pp. 424–430, 2018.
- [15] X. Cheng, Y. Zhang, Y. Chen, Y. Wu, and Y. Yue, "Pest identification via deep residual learning in complex background," *Comput Electron Agric*, vol. 141, pp. 351–356, 2017.
- [16] E. Fujita, Y. Kawasaki, H. Uga, S. Kagiwada, and H. Iyatomi, "Basic investigation on a robust and practical plant diagnostic system," in *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2016, pp. 989–992.
- [17] M. Ahmad *et al.*, "Cotton leaf disease detection using vision transformers: A deep learning ap- proach," *Crops*, vol. 1, p. 3, 2024.
- [18] S. Kumar, R. Ratan, and J. V. Desai, "Cotton disease detection using tensorflow machine learning technique," *Advances in Multimedia*, vol. 2022, no. 1, p. 1812025, 2022.
- [19] P. Udawant and P. Srinath, "Cotton leaf disease detection using instance segmentation," *Journal of Cases on Information Technology (JCIT)*, vol. 24, no. 4, pp. 1–10, 2022.
- [20] G. Kumar, G. Singh, V. Bhatnagar, S. Rawat, and A. Bhardwaj, "Feature selection for cotton leaf disease using deep learning," in *2024 IEEE Int. Conf. on Communication, Computing and Signal Processing (IICCCS)*. IEEE, 2024, pp. 1–6.
- [21] R. F. Caldeira, W. E. Santiago, and B. Teruel, "Identification of cotton leaf lesions using deep learning techniques," *Sensors*, vol. 21, no. 9, p. 3169, 2021.

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

- [22] S. Kotian, P. Ettam, S. Kharche, K. Saravanan, and K. Ashokkumar, "Cotton leaf disease detection using machine learning," *Proc. of the Advancement in Electronics Communication Engineering*, 2022.
- [23] M. Rahman, M. Ullah, R. Devnath, T. Chowdhury, G. Rahman, and M. Rahman, "Cotton leaf disease detection: an integration of cbam with deep learning approaches," *Int. J. Comput. Appl.*, vol. 975, p. 8887, 2025.
- [24] S. K. Noon, M. Amjad, M. A. Qureshi, and A. Mannan, "Computationally light deep learning framework to recognize cotton leaf diseases," *J. Intell. Fuzzy Syst.*, vol. 40, no. 6, pp. 12 383–12 398, 2021.
- [25] A. Zhou *et al.*, "Multi-head attention-based two-stream efficientnet for action recognition," *Mul- timed. Syst.*, vol. 29, no. 2, pp. 487–498, 2023.
- [26] M. Kumar, T. Hazra, and S. S. Tripathy, "Wheat leaf disease detection using image processing," *Int J Latest Technol Eng Manag Appl Sci*, vol. 6, no. 4, pp. 73–76, 2017.
- [27] A. Herok and S. Ahmed, "Cotton leaf disease identification using transfer learning," in *2023 Int. Conf. on Information and Communication Technology for Sustainable Development (ICICT4SD)*. IEEE, 2023, pp. 158–162.