Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Cotton leaf image dataset for disease classification and health monitoring

Shamim Ripon*, Raiyan Gani, Nazratan Mazumder Niha, Wasimul Bari Rahat, Shafaeat Hasan Toufiq, Mushfida Ferdous Maisha, Jubaer Ahmed

Department of Computer Science and Engineering, East West University, Aftabnagar, Dhaka, Bangladesh

ARTICLE INFO

Article history:
Received 30 June 2025
Revised 29 August 2025
Accepted 30 September 2025
Available online 8 October 2025

Dataset link: Cotton Leaf Image Dataset for Disease Classification (Original data)

Keywords:
Cotton dataset
Cotton leaf disease
Disease identification
Dataset collection
Image analysis
Leaf diseases
Agricultural challenges

ABSTRACT

Cotton, often referred to as “white gold” or the “king of fibers,” is one of the most widely used natural fibers in the global textile industry, supporting approximately 250 million people worldwide. However, cotton plants suffer from a variety of diseases, particularly leaf diseases, which can significantly reduce the yield and fiber quality. To overcome this problem, we propose a carefully curated image dataset that enables research toward early and automated disease detection and health monitoring of cotton plants. The dataset comprises 1373 original and 4963 augmented high-resolution images of cotton leaves with healthy, damaged, and infected samples. The images were captured under different environmental conditions from plants grown at the Sher-e-Bangla Agricultural University in Dhaka, Bangladesh to provide natural variability and realism. The dataset considers four common cotton leaf diseases—Fusarium wilt, Alternaria leaf spot, Verticillium wilt, and bacterial blight—each labeled and classified to support machine learning applications. Captured from different angles and devices, the images have rich visual content that enables the development of strong deep learning models for disease classification. The dataset was designed to advance research relevant to precision agriculture

* Corresponding author.
E-mail address: dshr@ewubd.edu (S. Ripon).

https://doi.org/10.1016/j.dib.2025.112142

by supporting early disease detection studies, crop health monitoring, and sustainable cotton-growing methods.

Specifications Table

Subject	Plant Pathology (Branch of Agriculture)
Specific subject area	Cotton leaf disease identification and classification using deep learning
Type of data	Raw images Dataset, Analysed, Filtered, and Processed
Data collection	Health leaves, infected leaves, and disease leaves of the cotton corresponding to four specific diseases were thoroughly collected for observation. Each leaf was then carefully selected and photographed by means of mobile phone cameras while individualized from other infected leaves in the garden. Once pictures had been taken, images were down-scaled for eliminating photos that are blurred or of poor quality. A total of four different mobile phone cameras was employed in this photographic technique for preserving variability in the data from images. The repository ended up having >1373 images, representing a vivid spectrum of healthy, diseased, and infected leaves of cotton. To support the generation of a large dataset needed for promoting the development of deep learning models essential in the farming industry, a detailed exploration into the diseases involved was undertaken, culminating in the systematic compilation of the repository.
Data source location	1. Sher-e-Bangla Agricultural University cotton plant garden in Dhaka (latitude: 23° 46′ 17.652″, longitude: 90° 22′ 30.2514″)
Data accessibility	Repository name: Mendeley Data Data identification number: 10.17632/t9hgvk2h9p.1 Direct URL to data: https://data.mendeley.com/datasets/t9hgvk2h9p/1

1. Value of the Data

- The database further contains a complete set of high-quality images of healthy and infected leaves of cotton. The database also contains detailed descriptions of corresponding diseases, thus facilitating accurate comparisons and improved diagnostics capabilities.
- This data can be used for building advanced machine learning models for automatic and precise detection of disease in cotton leaves, thereby abolishing human evaluation and thus reducing use of human evaluation for massive monitoring at a vast level, although expert validation is essential at the time of building the dataset and actual deployment.
- Early detection of disease avoids loss and damage by executing timely action. This also saves crops from destruction. This data leads for very large automatic monitoring systems, and it gives authentic detection instruments for the farmers regardless of the location.
- This database can be utilized by scientists for developing enhanced crop disease detection systems. The quantity and quality of images enable more stable machine learning models to be constructed, therefore enabling more advanced and data-driven farming practices.
- Such indications extracted from the dataset can assist in developing effective disease management programs, thereby fostering the health, yield, and farmers' incomes of crops. Even crop protection agencies and policymakers can utilize the data for creating targeted disease-control programs for the crops of the cotton industry.

2. Background

Leaf diseases remain a major issue in cotton production and have a significant effect on farm production and crop yield. Manual checking is usually the norm that farmers undertake to identify the infected leaves. Manual checking is labor-intensive, time-consuming, and prone to errors, particularly when different diseases exhibit the same appearance. Misidentification can lead to the incorrect application of pesticides, both in terms of cost and environmental impact.

The developments in computer vision, deep learning, and machine learning also enabled computerization of plant disease detection. Various research findings have also employed deep learning models in classifying plant diseases [1], while explainable AI (XAI) methods further improved model interpretability for more comprehensible and actionable predictions by agronomists and farmers [2].

Methods of computer vision and image processing have also been generalized for leaf characteristic evaluation and for disease characteristic discrimination [3]. High-technological imaging, such as multispectral and hyperspectral imaging, has also enhanced early disease detection by providing more information for crop status [4]. Precision farming has also benefited from UAVs by enabling disease observation of crops at a massive level [5].

This work contributes to the literature a clean data set of 1373 real images of cotton leaves. The data set has multiple disease classes and healthy classes collected under real-world field conditions for real-world and realistic applications. We aimed at building a strong disease classification model based on convolutional neural networks (CNNs) and explainable AI approaches using this data set. This should give rise to correct and explainable disease diagnosis and consequently allow timely and informed decisions for farmers.

Finally, access to real high-quality data will enhance disease detection, crop loss minimization, and use of AI-based smart-farm methods. They also highly promote the use of conservation farming for sustainable farming and maximization of production for smallholder and commercial cotton farms.

3. Data Description

The dataset [6] comprises a carefully curated collection of 1373 original images and 4963 augmented images of cotton leaves captured under real-world conditions. These high-resolution images are formatted and categorized to facilitate machine learning (ML) and deep learning (DL) research, particularly in the domain of automated plant disease recognition. Each image belongs to one of five well-defined classes: Alternaria Leaf Spot, Bacterial Blight, Verticillium Wilt, Fusarium Wilt, and Healthy Leaf. The organization of this dataset supports efficient navigation and model development, with dedicated subfolders for each disease class (Fig. 1).

3.1. Alternaria Leaf Spot

Alternaria Leaf Spot, illustrated in Fig. 2, is a fungal infection caused by Alternaria macrospora and Alternaria alternata [7]. It tends to appear during the later stages of the growing season under warm and humid conditions. Early symptoms include small round lesions with tan-to-brown centers encircled by purplish margins. These lesions may expand to display concentric rings, giving a characteristic “target spot” appearance. Severe cases can cause extensive defoliation, which significantly reduces the photosynthetic capacity of the plant and may compromise yield. Nutrient-deficient plants, particularly those lacking potassium, are more susceptible to severe infections.

3.2. Verticillium Wilt

As presented in Fig. 3, Verticillium Wilt is a soil-borne fungal disease caused by Verticillium dahliae [8]. The pathogen enters the root system and spreads via xylem vessels, disrupting water and nutrient transport. Symptoms include yellowing of the lower leaves, reddening, necrosis,

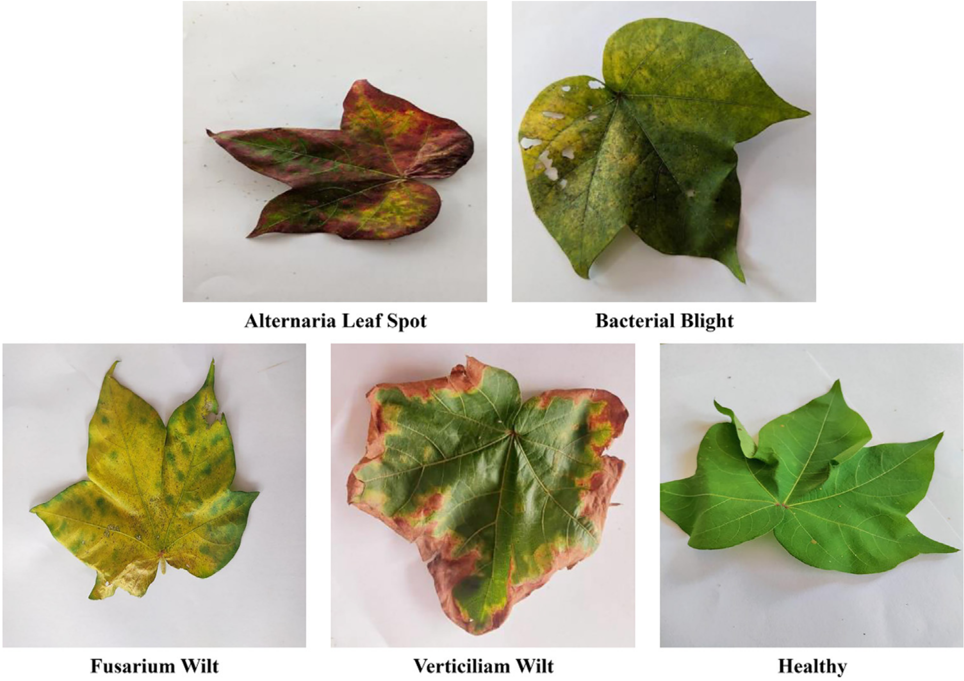

Fig. 1. Overall dataset sample.

Fig. 2. Sample image of Alternaria Leaf Spot.

and eventual wilting, particularly during flowering and boll development phases. In advanced stages, the disease can lead to stunted growth, premature defoliation, and yield loss. Effective management practices include crop rotation with non-host plants and cultivation of resistant cotton varieties.

Fig. 3. Verticillium Wilt.

Fig. 4. Bacterial Blight.

Fig. 5. Fusarium Wilt.

3.3. Bacterial Blight

Fig. 4 shows that bacterial light is a destructive disease caused by Xanthomonas citri subsp. malvacearum [9]. It affects all aerial parts of the cotton plant, leading to reduced fiber quality and overall yield. Key symptoms include angular, water-soaked lesions constrained by veins on the leaves, black lesions on petioles and stems (commonly known as “black arm”), and small, round water-soaked spots on bolls. The infected seeds are the principal vectors of transmission.

3.4. Fusarium Wilt

As shown in Fig. 5, Fusarium Wilt is caused by Fusarium oxysporum f. sp. vasinfectum [10]. This disease commonly occurs in acidic sandy soils and can affect cotton plants at any growth stage. The initial signs included dark green foliage and stunted growth. As the infection pro-

Fig. 6. Healthy leaf.

Table 1
Sample distribution and folder structure of the jackfruit dataset.

Class	Original Dataset		Augmented Dataset	
	Category	No. of Images	Category	No. of Images
1	Healthy Leaf	333	Healthy Leaf	1015
2	Bacterial Blight	218	Bacterial Blight	1027
3	Alternaria Leaf Spot	173	Alternaria Leaf Spot	987
4	Fusarium Wilt	337	Fusarium Wilt	957
5	Verticillium Wilt	312	Verticillium Wilt	977
	Total	**1373**	**Total**	**4963**

gresses, the leaf margins yellow and necrotize inward, leading to complete leaf desiccation. The fungus colonizes vascular tissues, disrupts nutrient flow, and often kills the plant. This pathogen can persist in contaminated seeds, making field sanitation and resistant cultivars essential for control.

3.5. Healthy leaf

Fig. 6 shows a reference image of a healthy cotton leaf. Typically, broad and heart-shaped healthy leaves have three–five distinct lobes and well-defined venation [10]. A uniform green color indicates a sufficient chlorophyll concentration, which is vital for effective photosynthesis. Although the texture can vary by cotton species, a healthy leaf generally reflects optimal plant health. Maintaining such foliage is critical for energy production required for boll development. Proper nutrition and disease prevention are fundamental to preserving leaf health.

3.6. Dataset folder structure

The dataset is organized logically into two broad categories: the Original Dataset and the Augmented Dataset, the summary of which is shown in Table 1. The original dataset contained 1373 raw images of five classes, and the augmented dataset consisted of 4963 images generated through augmentation operations of rotation, flipping, and brightness adjustment. These opera-

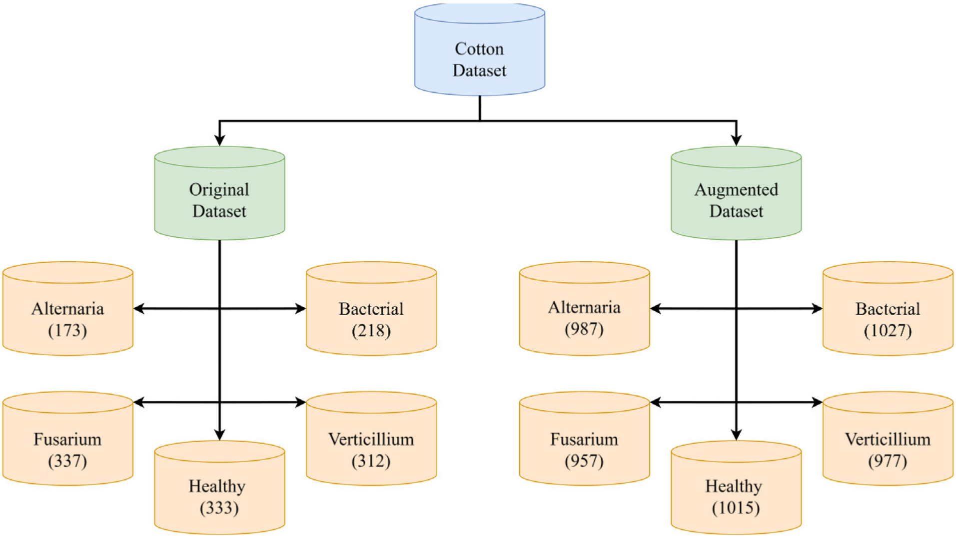

Fig. 7. Dataset folder structure.

tions enrich the dataset by injecting controlled variability, and as a result, improve the overall generalizability of the ML/DL models.

All images were stored in the PNG format to ensure maximum cross-platform compatibility and maintain image quality. The augmentation process is careful not to over distort such that real-world feature validity is maintained while optimizing the dataset diversity.

To facilitate the experimentation, the directory organization was hierarchical, as shown in Fig. 7. The root of the directory has subdirectories labelled by the class and type of dataset (original or augmented). The augmentation methods were calibrated to yield variations without modifying the intrinsic nature of the image. Validation procedures were integrated into the augmentation pipeline to ensure that the images produced were reflective of real-world disease patterns, thus making the model robust and performant.

4. Experimental Design, Materials, and Methods

4.1. Dataset collection

The dataset collection process plays a crucial role in building reliable deep learning models for plant disease detection. In this study, we created a standardized image dataset by capturing high-quality photos of cotton leaves affected by various diseases as well as healthy leaves. All images were collected from the cotton plant garden of Sher-e-Bangla Agricultural University in Dhaka, Bangladesh. Although the dataset originated from a single location, efforts have been made to capture the natural variations that typically occur due to environmental diversity in other regions.

The disease identification process started with a preliminary examination of the gathered leaf images. Visible symptoms were compared with online resources to compile an initial list of probable diseases. For reliability purposes, this list was verified by an agricultural expert, whose affirmation represented the ground truth for proper dataset annotation and was later used for training and testing the machine learning models.

We considered several factors, such as light, leaf orientation, and camera angle, to simulate different scenarios in real farming conditions. Four smartphones with varying camera specifi-

Table 2
Image capturing devices.

Device Name	Camera Resolution
Samsung Galaxy A34 5G	48 MP, f/1.8, 26 mm (wide), PDAF, OIS 8 MP, f/2.2 (ultrawide) 5 MP, f/2.4 (macro) 13 MP, f/2.2 (wide)
OnePlus Nord	48 MP, f/1.8 (wide), OIS 8 MP, f/2.3 (ultrawide) 5 MP, f/2.4 (depth) 2 MP, f/2.4 (macro) 32 MP, f/2.5 (wide) 8 MP, f/2.5 (ultrawide)
Google Pixel 6	50 MP, f/1.85 (wide), laser detect autofocus, OIS 12 MP, f/2.2 (ultrawide) 8 MP, f/2.0 (front camera)
Poco M2 Max Pro	48 MP, f/1.79, ISOCELL Plus (primary) 8 MP, f/2.2 (ultra-wide) 5 MP, f/2.4 (macro) 2 MP, f/2.4 (depth), 16 MP (front camera)

cations were used to improve the quality and diversity of the dataset. All smartphone camera specifications are listed in Table 2.

These mobile phones were selected because of their ability to capture high-resolution images, which are required to detect faint signs of disease on leaves. The use of multiple phones helped to obtain varying image qualities, such as resolution, color quality, and light.

Each leaf was placed against a white background to maintain consistency and prevent background distraction. This also improved the color presentation of disease symptoms. Wherever possible, natural light was employed, although arrangements were made to shoot under altered settings so that issues of oversaturation and shading could be circumvented. Several images were captured systematically during morning and noon sessions from various angles to record natural variation in light and observational perspectives. In order to maintain consistency and quality in the images, environmental conditions such as wind were controlled to avoid leaf movements during photography.

Leaf samples were collected in a structured manner from a cotton field after a randomized sampling strategy. Five to six single leaves were selected from each plant to ensure adequate representativeness of the population. To completely capture visual disease characteristics, each sample was photographed from multiple angles under stable and controllable light conditions. This enabled morphological symptoms, including discoloration, necrotic spots, and surface unevenness, to be visually documented with clarity. The compiled image dataset consequently reflects the basic dataset for subsequent-stage visual analysis and precise disease recognition.

To minimize potential bias and ensure representative coverage of the research area, a randomized sampling method was employed. Leaves were sampled from multiple plants in the field to obtain variations in disease expression. By doing this, it was possible to ensure that a wide range of morphological characteristics was represented, thereby enhancing the accuracy and stability of future disease analysis and detection.

After taking the pictures, they were inspected for quality and relevance. Dark, obstructed, and blurred images were excluded. The remaining photos were categorized into neatly labeled folders based on a particular disease or health condition. These five categories included healthy leaves, Alternaria leaf spot, bacterial light, Verticillium wilt, and Fusarium wilt. During editing, slight adjustments were made to enhance clarity and standardize brightness and alignment across the dataset. The result is a high-quality, neatly organized dataset that is ready for use in machine learning. Fig. 8 shows the overall dataset collection and pre-processing steps.

4.2. Dataset preparation

To enhance the performance of deep learning models and generalize them more effectively, data augmentation was performed during the data preparation stage [11]. This boosts the amount and diversity of the dataset, making it a stronger and better representative of actual conditions.

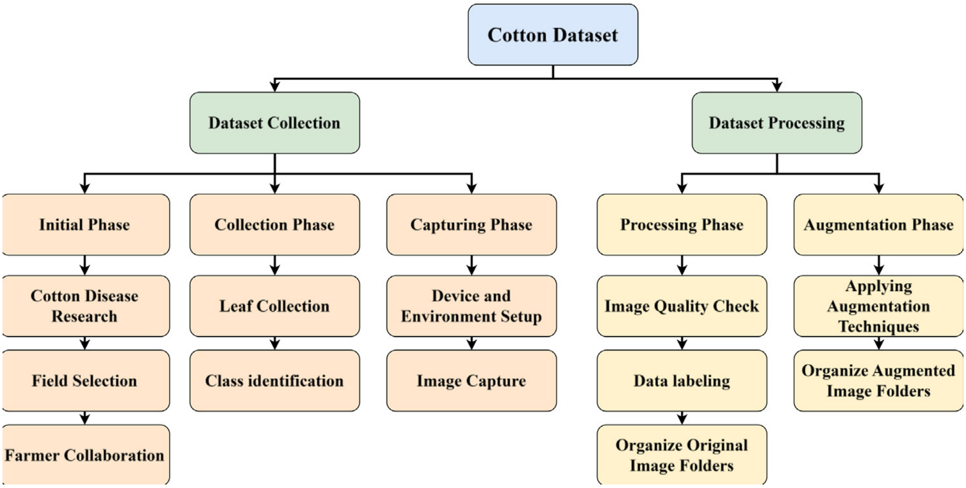

Fig. 8. Overall dataset collection and pre-process procedure.

Unlike synthetically generated data, images enhanced in this study were derived from real photographs of cotton leaves, preserving the integrity of the disease characteristics [12]. Controlled variability was created using different methods of augmentation without changing the most essential visual features of the leaves.

The transformations included both geometric and photometric operations as follows:

- Horizontal and vertical flipping
- Random rotation
- Cropping and zooming
- Brightness and contrast adjustment

These operations produce different representations of the same image of a leaf, simulating the actual changes in orientation, lighting, and background. This type of methodology allows the model to learn invariant and discriminative features under a broad range of conditions [13].

The most important goal of the augmentation process was two-fold.

1. To increase the dataset size, a deep learning model with a large training sample size was developed.
2. To reduce overfitting, we need to ensure that the model does not just learn to memorize patterns in the training data but also learns features that can generalize well to unseen, new data.

To obtain a realistic and well-balanced augmented dataset, a systematic augmentation pipeline was used. This included tweaking the augmentation parameters and checking the impact on the performance of the model training. We retained only reasonable transformations: those that preserved the biological feasibility and boosted the model's generalization ability.

In particular, rotation, flipping, and cropping are basic geometric transformations required for simulating the potential varied appearances of leaves under various field conditions [14]. These techniques were employed to simulate the orientation and position changes of infected leaves encountered in real farm fields.

Through this guided and biologically aware improvement mechanism, we constructed a dataset that was not only large in quantity but also diverse, and in the process, enhanced the stability and performance of disease detection models for cotton leaves.

Table 3
Previous work comparison.

Data Source	No. of Class	Sample Size	
		Original Dataset	Augmented Dataset
Bhoi, J. [15]	3	1951	–
Our dataset [6]	5	1373	4963

4.3. Dataset comparison

This section compares our dataset to existing datasets for cotton leaf disease. As shown in Table 3, the most frequently cited publicly available Kaggle dataset by Bhoi [15] has 1951 images of only three types of cotton leaf conditions. However, our dataset contains a diversified set of five different classes: healthy leaves, bacterial light, Alternaria leaf spot, Fusarium wilt, and Verticillium wilt.

Our dataset included 1373 original images and an additional 4963 augmented images, significantly expanding its diversity. The augmented dataset was constructed using established image transformation techniques such as flipping, rotation, cropping, and color adjustments, which are widely used to improve model generalization and performance in deep learning applications. These augmentations were derived from real images, preserving the authenticity of disease symptoms while introducing meaningful variations for robust training. By offering a wider range of disease categories and an enriched image pool, our dataset supports more accurate, scalable, and generalizable disease-classification models. Class-wise labeling and well-organized folder structures further enhance their utility in practical machine learning tasks.

In summary, the increased class diversity, greater volume of samples, and systematic augmentation approach make our dataset a valuable resource for researchers working on cotton disease diagnosis and plant health monitoring using deep learning.

4.4. Dataset use case

To display the practical use of the proposed dataset, we also tested the performance of two popular deep learning architectures, DenseNet121 and MobileNetV2, for classifying cotton leaf disease. Each leaf from the dataset constitutes a unique sample, and care was taken not to keep more than one image of a single leaf in both the testing as well as the training set in order to maintain dataset independence. Also, all the images were randomized prior to splitting into training, validation, and testing partitions in order to avoid data leakage and also for carrying out unbiased model evaluation. The dataset was divided into 70 %, 20 %, and 10 % for training, validation, and testing, respectively, ensuring balanced class representation and allowing for unbiased performance evaluation.

The model evaluation focused on three aspects.

- Training and validation loss behavior.
- Classification performance metrics.
- Class-level prediction accuracy using confusion matrices.

The convergence plots in Fig. 9 reveal that DenseNet121 has smoother and more stable convergence and a smaller training vs. validation loss gap. This is an indication of improved generalization and a reduced risk of overfitting. MobileNetV2, on the other hand, had a slow decrease in validation loss and a larger training vs. validation gap, indicating a higher susceptibility to class imbalance or data noise.

These trends can be accounted for by differences in the model structure. DenseNet121, owing to its densely connected layers, can detect fine-grained structures and intraclass variations in

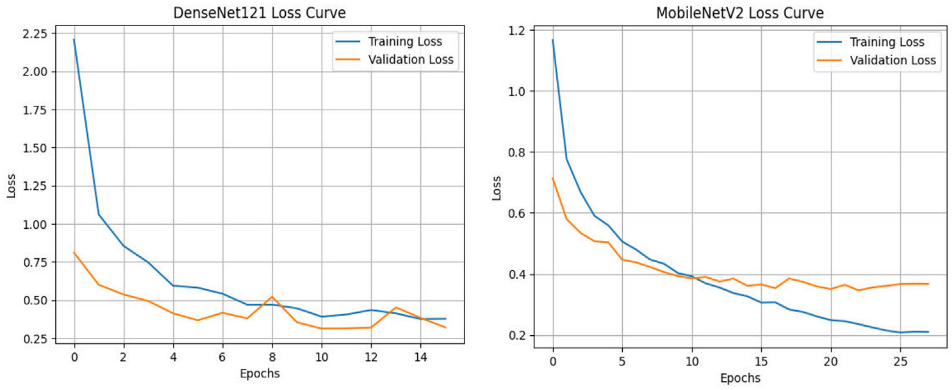

Fig. 9. Training and validation loss curves for DenseNet121 and MobileNetV2.

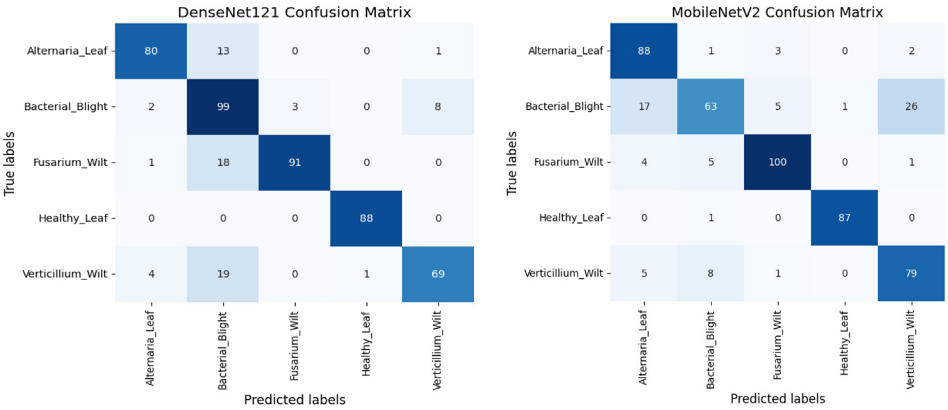

Fig. 10. Confusion matrix for DenseNet121 and MobileNetV2.

Table 4
Performance metrics comparison of DenseNet121 and MobileNetV2.

Model	Accuracy	Precision	Recall	F1-Score
DenseNet121	**0.8591**	**0.8850**	**0.8608**	**0.8672**
MobileNetV2	0.8390	0.8434	0.8491	0.8394

disease symptomatology. MobileNetV2, being computationally lightweight, may not handle fine intraclass variations, especially if the disease features are either subtle or visually overlapping.

Table 4 indicates that DenseNet121 outperformed MobileNetV2 in all performance metrics. It had higher precision, that is, lower false positives, and higher recall, that is, higher sensitivity to correctly classify true cases of disease. The F1-score also supported the improved trade-off between recall and precision, which is a measure of better consistency in classification among all disease classes.

The confusion matrices in Fig. 10 provide a detailed class-level performance analysis for each model. DenseNet121 exhibited a balanced performance with relatively fewer misclassifications between disease types. This implies their ability to distinguish between classes that are visually similar. MobileNetV2, on the other hand, experienced more confusion between similar disease types, possibly because of its lightweight structure, which overlooks fine-grained visual information while extracting features.

Incorporating real-world variation in the dataset, that is, variation in leaf angle, lighting, and severity of symptoms, is a challenge for simpler models such as MobileNetV2. Although the variation makes the dataset more realistic and practical, it necessitates a more complex architecture, for example, DenseNet121, to maintain classification accuracy.

In short, this case demonstrates the applicability of the dataset for training and testing deep learning models for plant disease detection. It also indicates the need to choose a suitable model architecture based on the complexity and diversity of image data.

Limitations

Although the dataset and methodology presented in this study offer significant value for cotton leaf disease classification, several limitations should be acknowledged to guide future improvements and applications.

- **Environmental Influence:** Although the dataset contained diverse examples of healthy and diseased cotton leaves, all the images were captured under controlled conditions. As a result, the dataset may not fully reflect the challenges encountered in natural field environments, such as extreme lighting, shadow interference, background clutter, or inconsistent soil surfaces. These uncontrolled variables could affect the model performance in real-world deployment.
- **Device Specificity:** The dataset was created using four smartphone models: Samsung Galaxy A34, OnePlus Nord, Google Pixel 6, and Poco M2 Max Pro. Although this multidevice approach introduces some variation in image quality and characteristics, the generalization of trained models may still be limited when applied to images captured using different camera types, especially those with differing color reproduction, resolution, or image processing algorithms.
- **Limited Disease Spectrum:** Despite including five important categories, the dataset might not cover the full range of cotton leaf diseases encountered in agricultural settings. Certain rare diseases, mixed infections, or early-stage symptoms may be underrepresented. This limits the ability of the model to be generalized across all potential leaf disease scenarios in the field.
- **Regional Constraints:** All images were collected from a single geographic location: Sher-e-Bangla Agricultural University in Dhaka, Bangladesh. As a result, the dataset may not capture variations in disease expression owing to regional differences in climate, soil conditions, plant varieties, or pest populations. This geographic limitation may restrict the effectiveness of the models when applied to other regions with distinct environmental factors.
- **Leaf-level Independence:** Complete independence at the leaf level cannot be guaranteed, as multiple samples may originate from the same plant or from plants exposed to similar environmental conditions. This introduces potential data correlation, which could lead to an overestimation of model performance. To mitigate this, future work should enforce stricter sampling protocols—such as plant-level separation or randomized sampling across distinct individuals—to ensure greater independence during dataset collection and evaluation.
- **Domain Gap:** While the dataset was gathered in natural field conditions, some real-world phenomena, like changing lighting, wind motion, and cluttered backgrounds, were not completely captured. Such discrepancies could lead to a domain gap between controlled data and large-scale autonomous agricultural setups. Thus, additional domain adaptation and field testing are crucial for robust model generalization.

Ethics Statement

The authors adhere to the journal's ethical guidelines and confirm that this study does not involve humans, animals, or data obtained from social media. The datasets utilized in this study are publicly accessible, and appropriate citation protocols should be followed when utilizing these datasets.

CRediT Author Statement

Shamim Ripon: Conceptualization, Methodology, Supervision, Visualization, Project administration, Validation; **Raiyan Gani:** Investigation, Methodology, Supervision, Project administration, Writing – original draft, Writing – review & editing; **Nazratan Mazumder Niha:** Investigation, Data collection, Writing – original draft, Writing – review & editing; **Wasimul Bari Rahat:** Investigation, Data collection, Writing – original draft, Writing – review & editing; **Shafaeat Hasan Toufiq:** Investigation, Data collection, Writing – original draft, Writing – review & editing; **Mushfida Ferdous Maisha:** Investigation, Data collection, Writing – original draft, Writing – review & editing; **Jubaer Ahmed:** Investigation, Methodology, Writing – original draft, Writing – review & editing;

Data Availability

Cotton Leaf Image Dataset for Disease Classification (Original data) (Mendeley Data)

Acknowledgements

This research did not receive any specific grants from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this study.

References

[1] S. Abbas, R. Jain, V. Mangat, R. Sharma, A comprehensive survey on plant leaf disease detection using deep learning, Front. Plant Sci. 12 (2021) 708223, doi:10.3389/fpls.2021.708223.
[2] Ü. Atila, M. Uçar, K. Akyol, E. Uçar, Plant disease detection using deep learning: a comparative study of CNN architectures, Data Brief. 35 (2021) 106970, doi:10.1016/j.dib.2021.106970.
[3] J. Ma, H. Zhang, J. Zhang, Z. Liang, Deep learning for agricultural applications: a review on plant disease detection techniques, ISPRS J. Photogrammet. Remote Sens. 163 (2020) 71–87, doi:10.1016/j.isprsjprs.2020.01.006.
[4] W. Zhang, Y. Zhang, Review on plant disease detection using hyperspectral imaging and deep learning techniques, Comput. Electron. Agric. 188 (2021) 106367, doi:10.1016/j.compag.2021.106367.
[5] V. Singh, A.K. Misra, N. Sharma, Coconut plantation health monitoring using UAV-based multispectral imaging and deep learning models, IEEE Access. 8 (2020) 9102–9114, doi:10.1109/ACCESS.2020.9102764.
[6] S. Ripon, R. Gani, N.M. Niha, W.B. Rahat, S.H. Toufiq, M.F. Maisha, J. Ahmed, Cotton leaf image dataset for disease classification, Mendeley Data V1 (2025), doi:10.17632/t9hgvk2h9p.1.
[7] C. Li, J. Wang, Q. Wang, Y. Feng, A survey on deep learning approaches for plant disease classification, Expert. Syst. Appl. 207 (2022) 117163.
[8] N. Dhingra, S. Kumar, Automated papaya leaf disease detection using convolutional neural networks, IEEE Access. (2019), doi:10.1109/ACCESS.2019.8664333.

[9] Crop Protection Network, "Bacterial blight of cotton," [Online]. Available: https://cropprotectionnetwork.org/encyclopedia/bacterial-blight-of-cotton. [Accessed: June 28, 2025].
[10] Cotton Australia, "The cotton plant," [Online]. Available: https://cottonaustralia.com.au/the-cotton-plant. [Accessed: June 28, 2025].
[11] Shorten, Khoshgoftaar, A Survey on Image Data Augmentation for Deep Learning, 2019.
[12] S.P. Mohanty, D.P. Hughes, M. Salathé, Using deep learning for image-based plant disease detection, Front. Plant Sci. 7 (2016) 1419, doi:10.3389/fpls.2016.01419.
[13] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in: Proc. Adv. Neural Inf. Process. Syst., 2012.
[14] TensorFlow, "Image data augmentation," TensorFlow, [Online]. Available: https://www.tensorflow.org/tutorials/images/data_augmentation. [Accessed: April 23, 2025].
[15] J. Bhoi, "Cotton disease dataset," Kaggle, [Online]. Available: https://www.kaggle.com/datasets/janmejaybhoi/cotton-disease-dataset. [Accessed: March 25, 2025].