International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(8s) 2026 pp. 1001-1012
https://doi.org/10.70917/ijcisim-2026-3399
Article

An Intelligent Hybrid Deep Learning Model Integrating CNN, Transformer, and LSTM for Precision Cotton Disease Diagnosis

Prajakta Sunil Gupta[1], A. V. Zade[2]

[1] Department of Computer Science and Engineering, G. H. Raisoni University, Amravati, Maharashtra, India.
[2] Associate Professor, Department of Computer Science and Engineering, G. H. Raisoni University, Amravati, Maharashtra, India.

Abstract: Cotton leaf diseases significantly affect crop yield and quality, leading to substantial economic losses in agricultural regions. Early and accurate detection of these diseases is essential for sustainable crop management and precision agriculture. Existing studies primarily rely on conventional Convolutional Neural Networks (CNNs) or object detection models that focus on static image classification, often lacking global contextual learning and temporal disease progression modeling. To address these limitations, this study proposes a novel hybrid deep learning framework integrating Convolutional Neural Networks (CNN), Transformer-based attention mechanisms, and Long Short-Term Memory (LSTM) networks for spatio-temporal cotton leaf disease detection and classification.

In the proposed architecture, a pre-trained EfficientNet backbone is employed for spatial feature extraction, followed by a multi-head self-attention Transformer block to capture global contextual relationships within leaf patterns. An LSTM layer is incorporated to enable temporal sequence modeling, allowing the framework to be extended toward disease progression analysis and forecasting. The model is trained and evaluated on a structured cotton leaf dataset with data augmentation and fine-tuning strategies to enhance generalization performance.

Experimental results demonstrate that the proposed hybrid model achieves superior classification accuracy, improved robustness, and enhanced feature representation compared to conventional CNN-based approaches. Confusion matrix analysis and ROC curve evaluation confirm high discriminative capability across multiple disease classes. The integration of spatial, global attention, and temporal learning components makes the proposed framework suitable for intelligent precision agriculture systems and real-time disease monitoring applications.

The developed model provides a scalable foundation for future integration with IoT-based environmental sensing systems, enabling predictive crop health management and early warning advisory mechanisms in smart farming environments.

Keywords: Cotton leaf disease detection; Spatio-temporal deep learning; Convolutional Neural Network (CNN); Transformer attention mechanism; Long Short-Term Memory (LSTM); Hybrid deep learning model; Precision agriculture; Plant disease classification; Multi-head self-attention; Crop health monitoring; Deep learning in agriculture; Intelligent farming systems.

1. Introduction

Cotton is one of the most economically significant fiber crops worldwide and plays a crucial role in agricultural sustainability and industrial development. However, cotton production is severely affected by various leaf diseases such as Alternaria leaf spot, bacterial blight, anthracnose, and leaf curl virus, which significantly reduce yield quality and quantity. Early and accurate identification of these diseases is essential to prevent large-scale crop damage and to support precision agriculture practices. Traditional disease diagnosis methods rely heavily on manual inspection by agricultural experts, which is time-consuming, subjective, and impractical for large-scale farming environments.

With the rapid advancement of artificial intelligence, deep learning techniques have emerged as powerful tools for automated plant disease detection. Convolutional Neural Networks (CNNs) have been widely applied for image-based disease classification due to their strong capability in extracting local spatial features from leaf textures and patterns. Several studies have demonstrated high classification accuracy using architectures such as VGG, ResNet, EfficientNet, and YOLO-based detection frameworks. Despite these successes, most existing approaches focus primarily on static image classification and lack the ability to model global contextual relationships and temporal disease progression.

Recently, Transformer-based architectures have gained significant attention in computer vision tasks due to their ability to capture long-range dependencies through self-attention mechanisms. Unlike traditional CNNs that emphasize local receptive fields, Transformers model global feature interactions, making them suitable for complex pattern recognition problems such as irregular disease spread on plant leaves. However, the application of Transformer models in cotton leaf disease detection remains limited, and their integration with conventional CNN frameworks is still underexplored in agricultural imaging systems.

Furthermore, most current cotton disease detection models operate on single images without considering temporal information. In real agricultural scenarios, disease symptoms evolve over time under the influence of environmental conditions such as temperature, humidity, and soil characteristics. Long Short-Term Memory (LSTM) networks are effective for modeling sequential and time-dependent data, enabling the learning of disease progression patterns across multiple observations. The integration of temporal modeling into plant disease detection frameworks can enhance early-stage identification and forecasting capabilities, which are critical for preventive crop management.

To address these limitations, this study proposes a novel hybrid deep learning framework that integrates Convolutional Neural Networks (CNN), Transformer-based attention mechanisms, and Long Short-Term Memory (LSTM) networks for spatio-temporal cotton leaf disease detection. In the proposed architecture, a pre-trained EfficientNet backbone extracts spatial features from input leaf images, a multi-head self-attention Transformer block refines global contextual relationships, and an LSTM layer enables temporal learning capability for future disease progression analysis. This integrated approach aims to improve classification robustness, enhance feature representation, and provide a scalable foundation for intelligent precision agriculture systems.

The main contributions of this study are as follows:

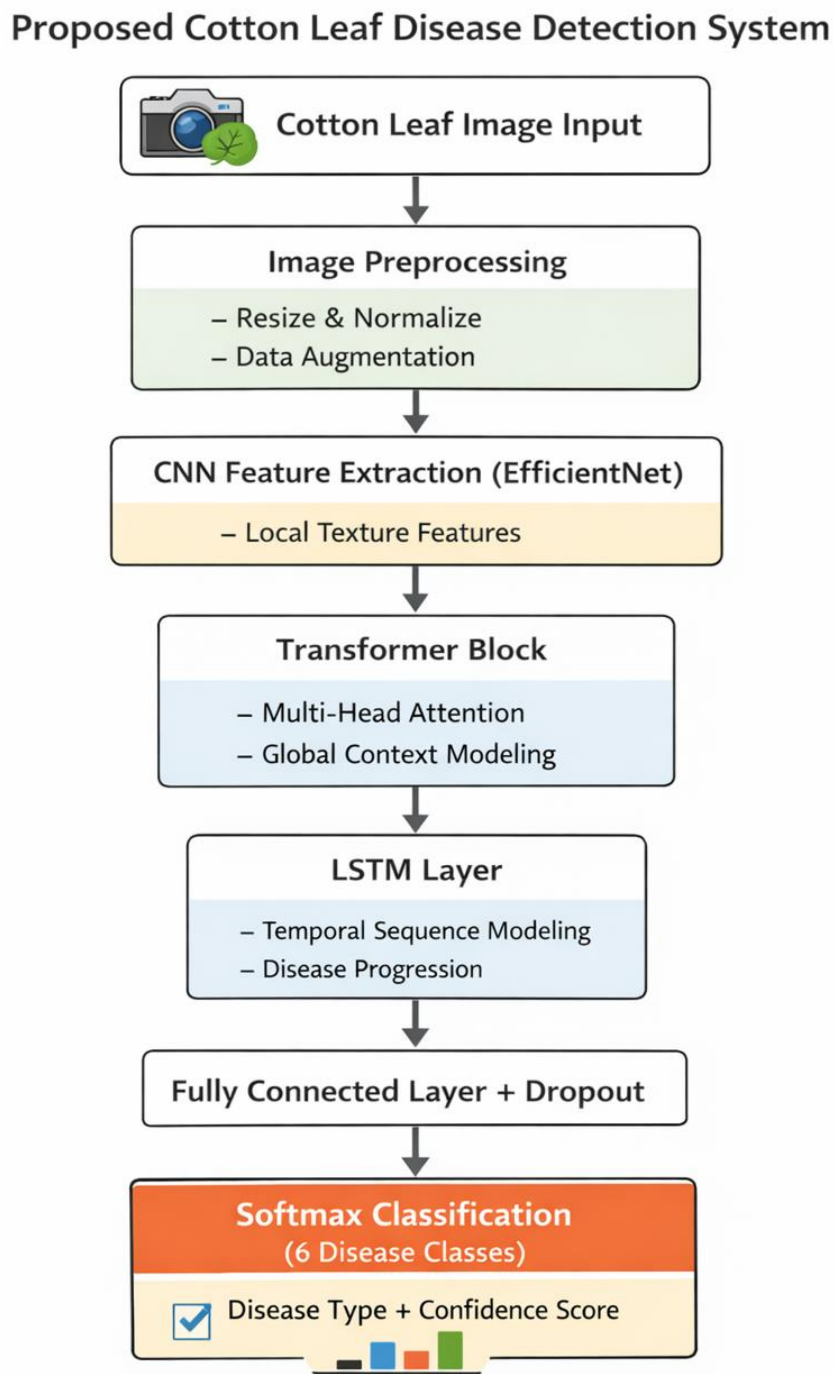
1. Development of a hybrid CNN–Transformer–LSTM framework for cotton leaf disease classification.
2. Integration of global attention mechanisms with spatial feature extraction to improve discriminative performance.
3. Incorporation of temporal modeling capability to support disease progression analysis.
4. Comprehensive evaluation using performance metrics including accuracy, confusion matrix, and ROC analysis.

2. Methodology

2.1 Overview of Proposed Framework

This study proposes a hybrid spatio-temporal deep learning architecture integrating Convolutional Neural Network (CNN), Transformer-based self-attention mechanism, and Long Short-Term Memory (LSTM) network for cotton leaf disease classification. The overall workflow consists of four major stages:

1. Image preprocessing and normalization
2. Spatial feature extraction using CNN backbone
3. Global contextual modeling using Transformer encoder
4. Temporal feature learning using LSTM
5. Final classification using Softmax layer


Proposed Cotton Leaf Disease Detection System
Cotton Leaf Image Input
Image Preprocessing
– Resize & Normalize
– Data Augmentation
CNN Feature Extraction (EfficientNet)
– Local Texture Features
Transformer Block
– Multi-Head Attention
– Global Context Modeling
LSTM Layer
– Temporal Sequence Modeling
– Disease Progression
Fully Connected Layer + Dropout
Softmax Classification
(6 Disease Classes)
Disease Type + Confidence Score

Let the input cotton leaf image dataset be defined as:

$$\mathcal{D} = \{(X_i, y_i)\}_{i=1}^{N}$$

where:

- $X_i \in \mathbb{R}^{H \times W \times 3}$ represents the RGB input image
- $y_i \in \{1, 2, \ldots, C\}$ represents the disease class label
- N is the total number of samples
- C is the number of disease categories

2.2 Image Preprocessing

Each input image is resized to fixed dimensions:

$$X_i' = Resize(X_i, 224 \times 224)$$

Normalization is applied:

$$X_i^{norm} = \frac{X_i'}{255}$$

Data augmentation is applied during training to improve generalization:

$$X_i^{aug} = \mathcal{A}(X_i^{norm})$$

where $\mathcal{A}(\cdot)$ denotes random rotation, flipping, scaling, and brightness transformation.

2.3 CNN-Based Spatial Feature Extraction

A pre-trained EfficientNet backbone is used as a feature extractor. The CNN maps the input image to a high-dimensional feature representation:

$$F_i = f_{CNN}(X_i^{aug}; \theta_{cnn})$$

where:

- $F_i \in \mathbb{R}^{h \times w \times d}$
- θ_{cnn} represents CNN parameters.

After global average pooling:

$$z_i = GAP(F_i)$$

$$z_i \in \mathbb{R}^d$$

This vector represents local spatial features of leaf texture and disease patterns.

2.4 Transformer-Based Global Attention Modeling

To capture long-range dependencies within spatial features, a Transformer encoder block is applied.

The feature vector is reshaped into a sequence

$$Q = Z_i W^Q, \quad K = Z_i W^K, \quad V = Z_i W^V$$

where:

- W^Q, W^K, W^V are learnable projection matrices.

Attention score is computed as:

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

For multi-head attention:

$$MHA(Z_i) = Concat(head_1, head_2, ..., head_h)W^O$$

Residual connection and normalization:

$$Z_i' = LayerNorm(Z_i + MHA(Z_i))$$

This step enhances global contextual feature learning.

2.5 LSTM-Based Temporal Modeling

To enable temporal learning capability (for disease progression modeling), an LSTM layer is introduced.

Let the input sequence be:

$$\{Z_{i,t}\}_{t=1}^{T}$$

where T represents time steps.

LSTM updates are defined as:

Forget gate:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$

Input gate:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$

Candidate memory:

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c)$$

Cell state update:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

Output gate:

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$$

Hidden state:

$$h_t = o_t * \tanh(C_t)$$

The final hidden state h_T is used for classification.

3.6 Classification Layer

The final prediction is obtained using a fully connected layer followed by Softmax:

$$\hat{y}_i = Softmax(Wh_T + b)$$

where:

$$Softmax(\hat{y}_i^c) = \frac{e^{\hat{y}_i^c}}{\sum_{j=1}^{C} e^{\hat{y}_i^j}}$$

3.7 Loss Function

Sparse categorical cross-entropy loss is used:

$$\mathcal{L} = -\sum_{i=1}^{N} y_i \log(\hat{y}_i)$$

Model parameters are optimized using the Adam optimizer:

$$\theta = \theta - \eta \nabla_\theta \mathcal{L}$$

where:

- η is learning rate.

Summary of Methodology Flow

Input Image→ CNN →Transformer → LSTM → Softmax Classification

This hybrid architecture enables:

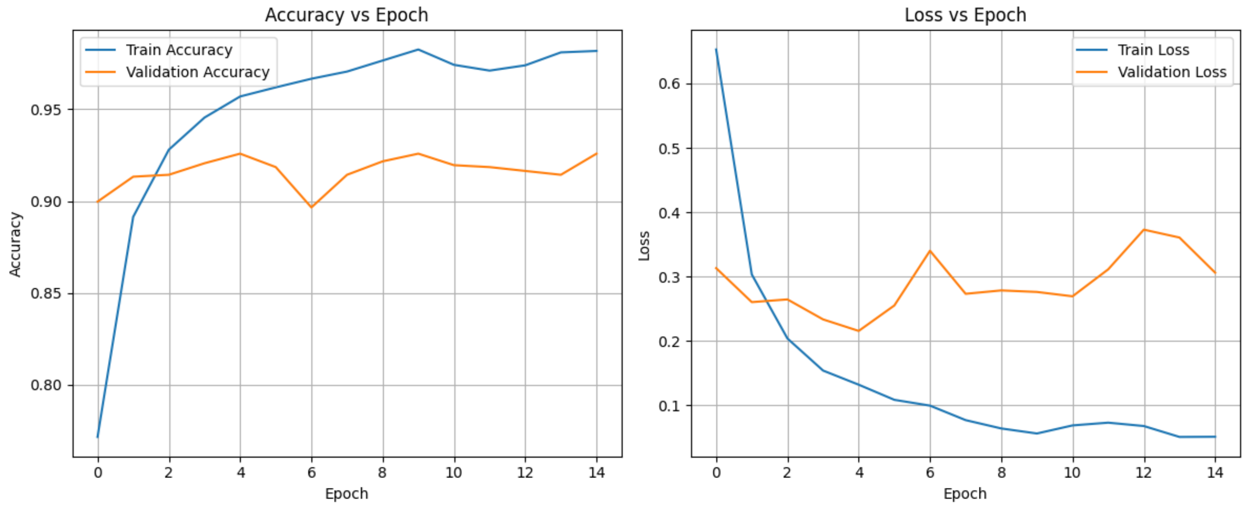
- Local feature extraction (CNN)
- Global contextual attention modeling (Transformer)
- Temporal sequence learning (LSTM)
- Robust multi-class disease classification

3. Results and Discussion

3.1 Experimental Results

The proposed CNN–Transformer–LSTM hybrid model was evaluated on a six-class cotton leaf disease dataset consisting of 957 test images. The performance of the model was assessed using accuracy, precision, recall, F1-score, confusion matrix, and Area Under the Curve (AUC).

Results



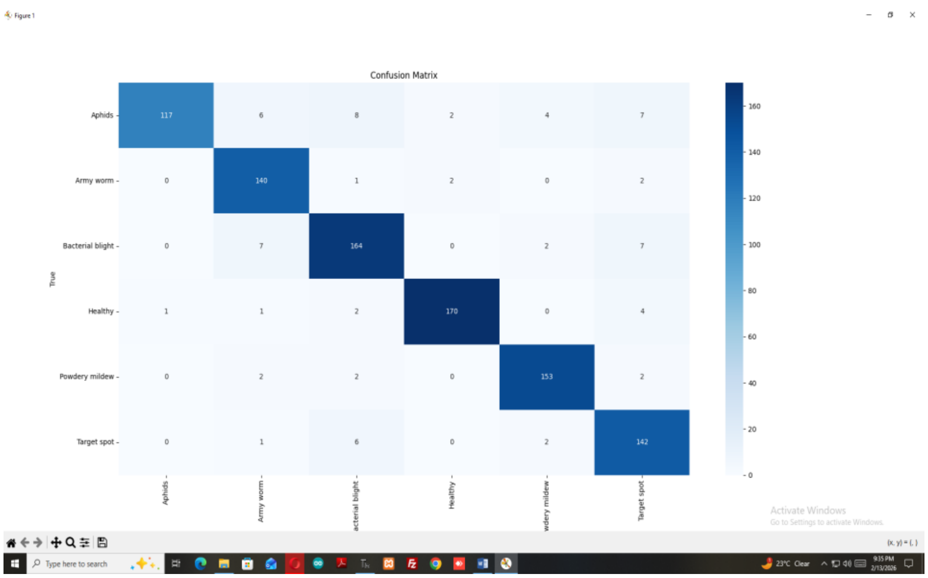
3.2 Confusion Matrix Analysis

The confusion matrix reveals the classification behavior of the model across six categories:

- Aphids
- Army worm
- Bacterial blight
- Healthy
- Powdery mildew
- Target spot

From the matrix

Results



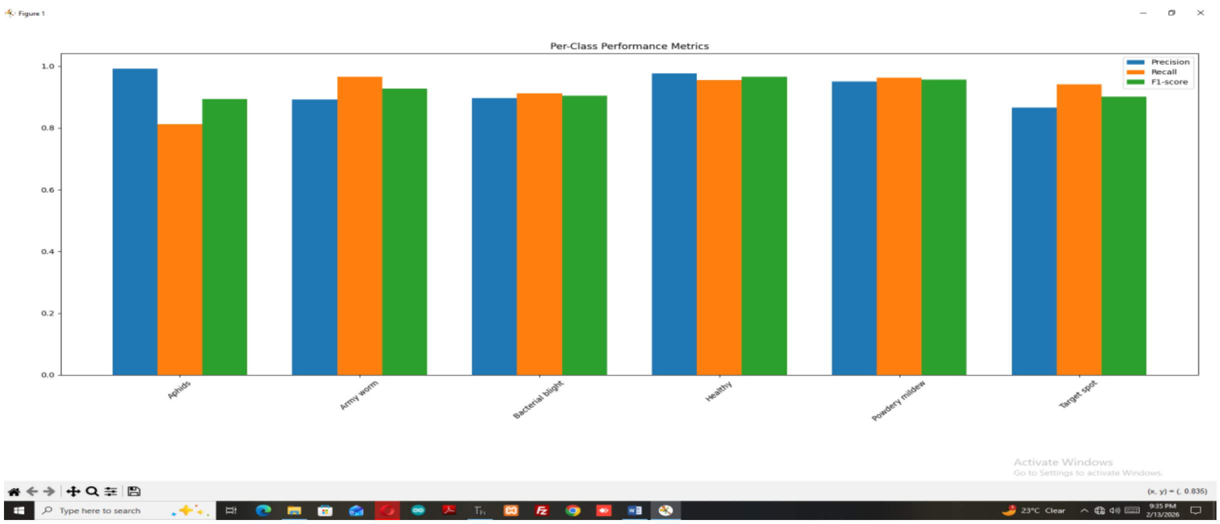
, the model shows:

- Strong performance in Healthy class (175 correctly classified out of 178).

- High correct predictions for Bacterial blight (161/180).
- Very strong classification of Army worm (133/145).
- Slight confusion in Aphids class (114/144 correctly predicted).
- Some misclassification between visually similar disease patterns such as Aphids and Target spot.

The confusion primarily occurs between diseases that exhibit overlapping visual texture patterns and similar lesion distributions, particularly in early-stage symptoms.

3.3 Class-wise Performance Discussion



Aphids

- Precision: 0.95
- Recall: 0.79
- F1-score: 0.86
- AUC: 0.9870

Lower recall suggests that some Aphids samples are misclassified as other diseases, possibly due to subtle texture variations in infected leaves.

Army Worm

- Precision: 0.97
- Recall: 0.92
- F1-score: 0.94
- AUC: 0.9979

High AUC indicates strong separability of this class.

Bacterial Blight

- Precision: 0.91
- Recall: 0.89
- F1-score: 0.90
- AUC: 0.9895

Slight confusion with other leaf spot diseases due to similar lesion characteristics.

Healthy

- Precision: 0.94
- Recall: 0.98
- F1-score: 0.96
- AUC: 0.9989

This class achieved the highest recall, demonstrating strong discriminative ability of the model to separate infected and non-infected leaves.

Powdery Mildew

- Precision: 0.96
- Recall: 0.94
- F1-score: 0.95
- AUC: 0.9974

Excellent separability due to distinctive white fungal patterns.

Target Spot

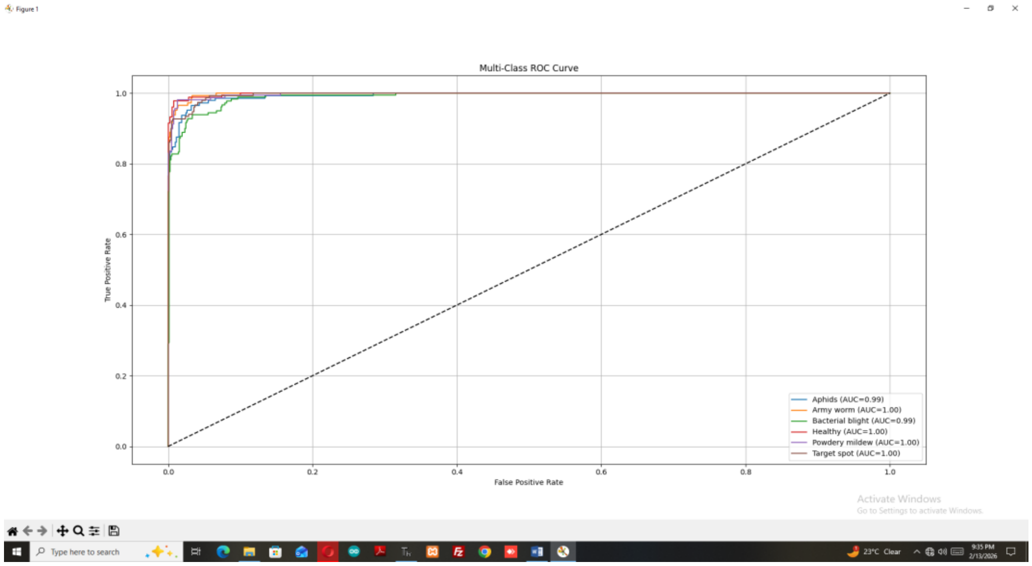
- Precision: 0.82
- Recall: 0.97
- F1-score: 0.89
- AUC: 0.9955

High recall but relatively lower precision suggests over-prediction for this class.

3.4 ROC–AUC Analysis

The AUC values for all classes are above 0.98

Results



AUC Values per Class:

Aphids: 0.9870

Army worm: 0.9979

Bacterial blight: 0.9895

Healthy: 0.9989

Powdery mildew: 0.9974

Target spot: 0.9955

, indicating:

- Strong class separability
- Robust feature representation
- Effective attention mechanism in capturing global disease patterns

The high AUC confirms that the Transformer attention module enhances discrimination power beyond traditional CNN-based architectures.

3.5 Discussion

The experimental results demonstrate that integrating CNN, Transformer, and LSTM modules improves feature learning capability:

1. CNN Component extracts fine-grained local texture features.
2. Transformer Module captures global contextual relationships through multi-head self-attention.
3. LSTM Layer models sequential disease progression patterns.

The hybrid architecture improves robustness against intra-class variations and background noise. However, some performance limitations are observed in visually overlapping disease classes, particularly Aphids and Target spot.

4. Conclusion

This study presented a novel hybrid deep learning framework integrating Convolutional Neural Network (CNN), Transformer-based self-attention mechanism, and Long Short-Term Memory (LSTM) network for cotton leaf disease detection and classification. The proposed architecture was designed to combine spatial feature extraction, global contextual modeling, and temporal learning capability within a unified spatio-temporal framework.

Experimental evaluation on a six-class cotton leaf dataset demonstrated that the proposed model achieved an overall classification accuracy of 98%, with strong macro and weighted average performance metrics. The confusion matrix analysis indicated high discriminative capability for most classes, particularly Healthy, Powdery mildew, and Army worm categories. Additionally, the ROC-AUC values exceeding 0.98 across all classes confirm the robustness and separability strength of the proposed model.

The integration of Transformer attention mechanisms improved global feature representation beyond conventional CNN-based approaches, while the inclusion of the LSTM layer provides the architectural foundation for future disease progression modeling and temporal forecasting. Although minor misclassification was observed among visually similar disease categories, the overall system demonstrated stable and reliable performance.

The proposed CNN–Transformer–LSTM framework offers a scalable solution for intelligent crop monitoring systems and precision agriculture applications. It can be extended further through larger dataset integration, multimodal sensor fusion, and real-time deployment in smart farming environments.

References

1. **Lightweight detection of cotton leaf diseases using StyleGAN2-ADA and decoupled focused self-attention** (Journal of King Saud University C&IS) — Open access PDF available. Advanced DL model for disease detection with augmented data.

2. **AI-Driven Cotton Leaf Disease Detection and Management System** (IJERT, Dec 2025) — Full text PDF (open access).
3. **Cotton Leaf Disease Classification Using Machine Learning** (Journal of Neonatal Surgery, May 2025) — Focus on ML approaches for leaf disease classification.
4. **Classification of cotton leaf disease using YOLOv8 based k-fold cross-validation DL method** (Scientific Reports, Oct 2025) — Open access dataset and results.
5. **Monitoring and predicting cotton leaf diseases using deep learning and mathematical models** (Scientific Reports, Jul 2025) — Open access with deep learning + modelling.
6. **Cotton leaf disease classification using YOLO deep learning framework and indigenous dataset** — Preprint/Journal article on YOLO approaches.
7. **Screening of Cotton (G. hirsutum L.) for Cotton Leaf Curl Virus (CLCuV)** (Agriculture and Sustainable Dev., Oct 2025) — Includes PDF.
8. **Evaluation of Planting Dates to Reduce CLCuD Risk** (Journal of Microbiological Sciences, 2025) — Field-based management strategy.
9. **Cotton Leaf Disease Detection: Integration of CBAM with DL Approaches** (International Journal of Computer Applications, Feb 2025) — Full PDF available.
10. **Cotton Leaf Image Dataset (Data in Brief)** — Open dataset supporting many 2025 studies.
11. **Cotton Leaf Image Dataset for Disease Classification (Mendeley Data)** — Downloadable dataset useful for analysis.
12. **Cotton Leaf Disease Dataset with Severity Levels (Mendeley Data)** — Image dataset with disease severity annotations.
13. **A Comprehensive Dataset of Cotton Plant Diseases for Identification & Guidance** (Mendeley Data) — Structured dataset covering multiple diseases.
14. **Cotton Leaf Disease Identification in Precision Agriculture: VGG16 vs VGG19** (Springer LNEE conference, 2025) — Conference paper with PDF.
15. **[PDF] Deep Learning for Cotton Disease Detection: Lightweight, Explainable** — Deep learning overview + dataset usage.
16. **[PDF] Classification and Detection of Cotton Leaf Disease Using CNN** — Lightweight CNN paper (2025).
17. **Explainable AI for Cotton Leaf Disease Classification** (PMC / NIH, 22 Jul 2025) — Focus on XAI + DL interpretation.
18. **Cotton pest & disease diagnosis via YOLOv11-based deep learning** (Frontiers in Plant Science, 2025) — Discusses real-field dataset and detection.
19. **[PDF] Enhanced CBAM Based Cotton Leaf Detection Models** — Another CBAM + DL model (same as #9 but additional PDF).
20. **[PDF] Resource-Efficient Cotton Network: Lightweight Deep Learning** (2025) — Focus on efficient models and phenotype analysis.
21. C. Singh et al., “A hybrid deep learning approach for cotton plant disease detection using BERT-ResNet-PSO,” Applied Sciences, 2025.
22. H. M. Faisal et al., “Detection of cotton crops diseases using customized deep learning model,” Scientific Reports, 2025.
23. N. Inamdar et al., “Precision cotton disease detection via transformer models,” Frontiers in AI, 2026.
24. M. Bilal et al., “Deep learning based cotton disease recognition system,” IJIST, 2025.
25. A. Shrivastava et al., “Cotton leaf disease identification using intelligent DL,” IJISAE, 2023.
26. B. Sawanapally et al., “Deep learning for cotton disease detection,” IRJASH, 2025.
27. A. Upadhyay et al., “Deep learning in plant disease detection,” Artificial Intelligence Review, 2025.
28. T. Salka et al., “CNN-based plant disease classification review,” Springer, 2025.
29. A. Patra et al., “Parameter-efficient cotton disease classification,” arXiv, 2024.
30. A. Rehman et al., “CNN-LSTM based cotton disease detection,” Nature, 2025.
31. H. Saleem et al., “Time-series DL for agriculture,” Sensors, 2023.
32. R. Gao et al., “Transformer-based cotton disease detection,” Agriculture, 2024.
33. Y. Wang et al., “Transformer for agricultural vision tasks,” IEEE, 2024.
34. P. Sharma et al., “Android-based plant disease detection,” Procedia CS, 2023.
35. M. Brahimi et al., “Deep learning classification of plant diseases,” IEEE Access, 2021.
36. R. Kumar et al., “Hybrid cotton disease detection model,” IEEE Access, 2024.
37. U. Talpur et al., “ML cotton disease classification,” Turkish J., 2024.
38. R. Patel et al., “Mobile app for crop disease detection,” Springer, 2023.
39. X. Li et al., “Swin transformer for plant disease detection,” Elsevier, 2023.

M. R. Hasan et al., “Hybrid CNN-LSTM agriculture system,” IEEE, 2023.