

A Data-Efficient Image Transformer for Multi-Crop Plant Disease Detection

Muskan Batra ^{1,✉}, Pooja Sharma ²

1. Computer Science and Engineering, I.K. Gujral Punjab Technical University, Kapurthala, IND

2. Computer Science & Engineering, I.K. Gujral Punjab Technical University, Kapurthala, IND

Received: April 20, 2026 | Review began: May 21, 2026 | Review ended: June 04, 2026 | Published: June 08, 2026

© Copyright 2026

This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Accurate plant disease diagnosis is essential for improving agricultural productivity and supporting global food security. Conventional convolutional neural networks may struggle to capture long-range spatial dependencies and subtle visual differences among plant diseases. This study investigates the use of a Data-Efficient Image Transformer for automated recognition of plant leaf diseases across multiple crops. The proposed framework employs transfer learning with pretrained weights, mixed-precision training, and classifier-head fine-tuning to enhance learning efficiency and predictive accuracy. Experiments were conducted on a publicly available dataset containing 30,945 images from eight crop species and 35 disease categories. Performance was evaluated using overall accuracy, macro-averaged precision, recall, and F1-score, together with confidence estimation and disease severity assessment. The model achieved an overall accuracy of 98.50%, with macro-precision, macro-recall, and macro-F1 scores of 0.9843, 0.9826, and 0.9833, respectively. Under identical training settings, DenseNet-121 achieved an accuracy of 97.29%. These findings demonstrate the effectiveness of the proposed framework for multi-crop plant disease classification on the evaluated benchmark dataset.

Categories: Image Processing and Analysis, Computer Vision, Deep Learning

Keywords: plant disease detection, deep learning, vision transformer, image classification, ai in agriculture

Introduction

Agriculture plays an important role in food security, economic development, and employment generation. Nevertheless, plant diseases pose an increasing threat to crop health and are responsible for up to 16% loss in agricultural productivity globally [1]. As the population is likely to grow to 9.10 billion by 2050, food production has to be raised by 70% and it is imperative to focus on effective crop protection [1].

Previous studies have explored deep learning approaches for plant disease analysis using Convolutional Neural Network (CNN)-based architectures, transfer learning methods, transformer models, localization frameworks, and severity estimation techniques [1-15]. CNN-based approaches have demonstrated strong performance but primarily focus on local feature extraction, which may create challenges when distinguishing visually similar disease patterns across multiple crops. More recently, transformer architectures such as Vision Transformers and Data-Efficient Image Transformer (DeiT) have shown improved ability to capture long-range contextual relationships through self-attention mechanisms [7-13]. Despite these advances, many existing studies remain focused on individual crops, limited disease categories, or controlled datasets, motivating further investigation into broader multi-crop disease classification frameworks.

Key contributions include:

How to cite this article:

Batra M, Sharma P (June 08, 2026) A Data-Efficient Image Transformer for Multi-Crop Plant Disease Detection. Cureus J Comput Sci 3 : es44389-026-00108-9. DOI <https://doi.org/10.7759/s44389-026-00108-9>

1. To implement a Data-Efficient Image Transformer (DeiT-Base)-based model for multi-crop and multi-disease leaf classification.
2. Integration of prediction confidence estimation using Softmax probabilities.
3. Addition of an image-based disease severity estimation module using color-space segmentation.

Materials And Methods

Dataset description

The dataset on which this research is conducted is [Plant Leaves Disease Detection](#) on Kaggle that is annotated with the help of VGG Image Annotator (VIA). It has 30,945 RGB images on 35 different disease classes in eight crops: potato (3), tomato (10), bell pepper (2), apple (4), corn (4), grape (4), peach (2), and rice (6), which allows performing multi-crop, multi-disease analysis. The dataset contains variations in lighting conditions, backgrounds, leaf orientation, occlusion, and disease severity. Table 7 presents the class-wise distribution of images for each crop and disease category, providing a detailed overview of the dataset composition and class balance. For experimentation, 80% of the dataset was used for training and 20% for validation using a fixed random seed to ensure reproducibility and maintain class balance and minimize potential data leakage between subsets.

Crop species	Class distribution (number of images)
Bell Pepper	Bacterial Spot (997), Healthy (1,478)
Apple	Black Rot (496), Cedar Rust (220), Healthy (1,316), Scab (504)
Corn	Cercospora Leaf Spot (410), Common Rust (953), Healthy (929), Northern Leaf Blight (788)
Grape	Black Rot (944), Esca (Black Measles) (1,107), Healthy (339), Leaf Blight (861)
Peach	Bacterial Spot (1,838), Healthy (288)
Potato	Early Blight (999), Healthy (152), Late Blight (1,000)
Rice	Bacterial Leaf Blight (350), Brown Spot (350), Healthy (350), Leaf Blast (350), Leaf Scald (350), Narrow Brown Spot (350)
Tomato	Bacterial Spot (2,127), Early Blight (1,000), Healthy (1,591), Late Blight (1,909), Leaf Mold (952), Mosaic Virus (373), Septoria Leaf Spot (1,771), Spider Mites (1,676), Target Spot (1,404), Yellow Leaf Curl Virus (422)

TABLE 1: Class-wise distribution of images across crop species.

Data preprocessing and augmentation

Images were resized to 224×224 pixels and normalized using ImageNet mean and standard deviation values to match pretrained weights. Training augmentation included random horizontal flipping and random rotation ($\pm 15^\circ$), while validation images were only resized and normalized. The normalization operation is defined in Equation (1) where I represents the input image, I_{norm} denotes the normalized image, μ is the channel-wise mean, and σ is the channel-wise standard deviation.

$$I_{\text{norm}} = \frac{I - \mu}{\sigma} \quad (1)$$

Hierarchical dataset loading strategy

A custom dataset loader was implemented to traverse the hierarchical crop-disease directory structure for multi-crop and multi-disease classification. Each sample is assigned a unique class index corresponding to its disease label, while preserving crop-disease mapping for visualization and severity analysis. Assume that the dataset D consists of N samples as represented in Equation (2), where x_i denotes the input RGB leaf image and $y_i \in \{1, 2, \dots, C\}$ represents the corresponding disease class label, with $C = 35$ disease categories.

$$D = \{(x_i, y_i)\}_{i=1}^N \quad (2)$$

DeiT-Base vision transformer architecture

The proposed model is based on DeiT-Base, which leverages self-attention to capture long-range contextual relationships. The image patches are flattened and projected into a D -dimensional embedding space. Each input image is divided into 16×16 patches, forming the patch embedding matrix represented in Equation (3) where X denotes the patch embedding matrix, N represents the total number of image patches, D denotes the embedding dimension of each patch token, and $\mathbb{R}$ represents the set of real numbers.

$$X \in \mathbb{R}^{N \times D} \quad (3)$$

Positional embeddings are also used in the patch embeddings to help remember spatial information. The input to the transformer encoder is determined by Equation (4) where Z_0 represents the input sequence to the transformer encoder, X denotes the patch embedding matrix, and P represents the learnable positional embedding matrix used to preserve spatial information.

$$Z_0 = X + P \quad (4)$$

Each transformer encoder block contains a multi-head self-attention (MHSA) module and a feed-forward network (FFN). Multi-head self-attention operation can be defined as in Equation (5) where Q , K , and V represent the query, key, and value matrices, respectively, K^T denotes the transpose of the key matrix, d_k represents the dimensionality of the key vectors, and $\text{softmax}(\cdot)$ computes the attention weights.

$$\text{MHSA}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (5)$$

Use of residual connections and normalization of the layers is applied to improve training stability. Output of the attention is calculated as in Equation (6) where Z'_i denotes the output of the multi-head self-attention block at layer i , Z_{i-1} represents the output from the previous transformer layer, $\text{LN}(\cdot)$ denotes layer normalization, and $\text{MHSA}(\cdot)$ represents the multi-head self-attention operation.

$$Z'_i = \text{MHSA}(\text{LN}(Z_{i-1})) + Z_{i-1} \quad (6)$$

How to cite this article:

The feed forward network then updates the feature representation as in Equation (7) where Z_l denotes the output representation of the transformer encoder at layer l , Z'_l represents the output of the attention block, $\text{LN}(\cdot)$ denotes layer normalization, and $\text{FFN}(\cdot)$ represents the feed-forward network.

$$Z_l = \text{FFN}(\text{LN}(Z'_l)) + Z'_l \quad (7)$$

Lastly, the resulting output of the class token is fed through a classification head to get the probable predicted classes as represented in Equation (8) where $\hat{y}$ denotes the predicted probability distribution over disease classes, z_c represents the class token embedding, W_c denotes the learnable weight matrix, b_c represents the bias vector, and $\text{softmax}(\cdot)$ converts the output logits into class probabilities.

$$\hat{y} = \text{softmax}(W_c z_c + b_c) \quad (8)$$

Transfer learning strategy

The DeiT-Base model is initialized with ImageNet-1K pretrained weights. The transformer backbone is frozen, and the classification head is refined on the disease data. This reduces computational cost while enabling task-specific adaptation.

Training configuration

Both the models are trained under identical conditions using AdamW optimizer, learning rate of 5×10^{-4} , batch size 32, and 20 epochs. Mixed-precision training was used to improve computational efficiency, and experiments were performed using Kaggle GPU resources (T4 $\times 2$). Cross-entropy loss was used for optimization as defined in Equation (9), where L denotes the cross-entropy loss, C represents the total number of disease classes, y_i denotes the ground-truth label for class i , and $\hat{y}_i$ represents the predicted probability for class i .

$$L = - \sum_{i=1}^C y_i \log(\hat{y}_i) \quad (9)$$

Prediction confidence estimation

For each validation sample, the softmax output is utilized to compute a prediction confidence score, Equation (10) defined as the maximum posterior probability, where $\text{Confidence}(x)$ denotes the prediction confidence score for input sample x , $\hat{y}_k$ represents the predicted probability of class k , and $\max_k(\cdot)$ selects the maximum posterior probability among all disease classes. This confidence score provides an approximate indication of prediction reliability.

$$\text{Confidence}(x) = \max_k (\hat{y}_k) \quad (10)$$

Disease severity estimation

Predicted leaf images were converted from RGB to HSV color space, where green regions were identified using thresholding ($H \in (25, 90)$, $S \in (40, 255)$, $V \in (40, 255)$). Disease severity was calculated using Equation (11), where **Severity** denotes diseased area percentage, A_{total} represents total image pixels, and A_{green} denotes pixels classified as healthy regions. Severity was categorized as low ($\leq 20\%$), medium (21-50%), or high ($> 50\%$). The HSV-based approach provides a lightweight approximation but may be affected by lighting and background variations.

$$\text{Severity} = \left(\frac{A_{\text{total}} - A_{\text{green}}}{A_{\text{total}}} \right) \times 100 \quad (11)$$

Evaluation metrics

Model performance was evaluated using overall accuracy, macro-averaged precision, macro-averaged recall, and macro F1-score to account for class imbalance. Accuracy measures correctly classified samples, while macro-averaged metrics equally weight each disease class. In these equations, C denotes the number of classes, while TP_i , FP_i , and FN_i represent true positives, false positives, and false negatives for class i , respectively.

$$\text{Accuracy} = \frac{1}{N} \sum_{i=1}^N \mathbb{I}(\hat{y}_i = y_i) \quad (12)$$

$$\text{Precision}_{\text{macro}} = \frac{1}{C} \sum_{i=1}^C \frac{TP_i}{TP_i + FP_i} \quad (13)$$

$$\text{Recall}_{\text{macro}} = \frac{1}{C} \sum_{i=1}^C \frac{TP_i}{TP_i + FN_i} \quad (14)$$

$$\text{F1}_{\text{macro}} = \frac{1}{C} \sum_{i=1}^C \frac{2 \cdot \text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \quad (15)$$

DeiT framework flowchart

Figure 7 illustrates the overall workflow of the proposed DeiT-based plant disease detection framework, including dataset preprocessing, transformer-based classification, prediction confidence estimation, disease severity estimation, and performance evaluation.

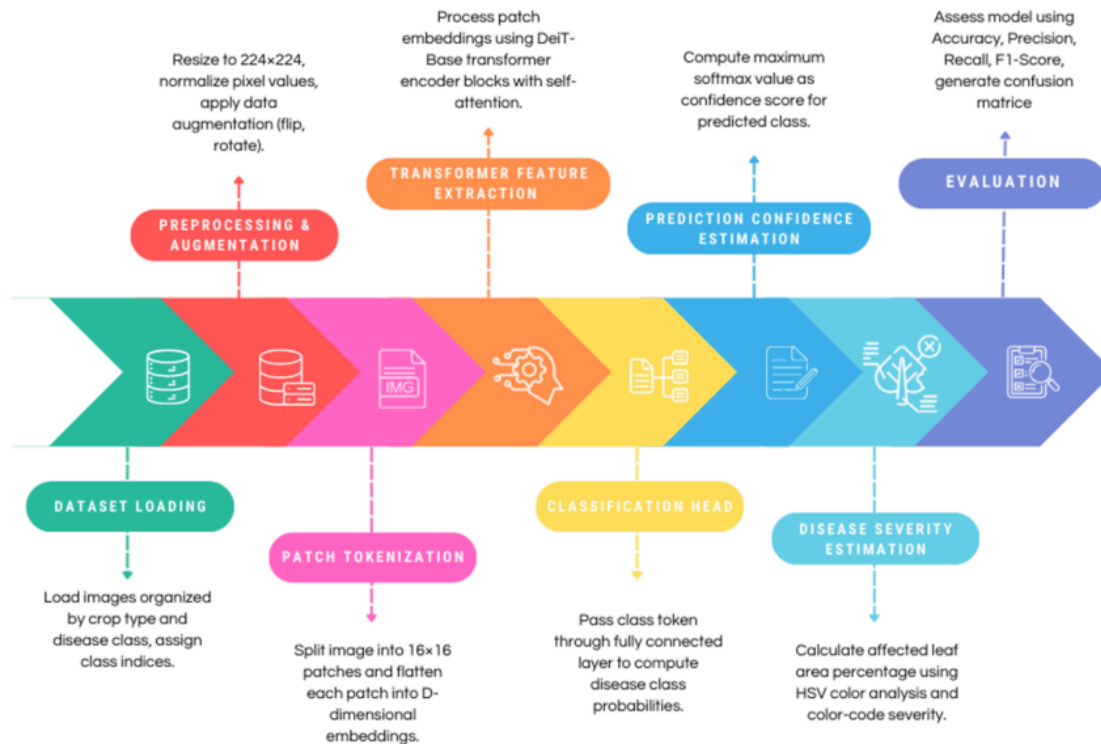


FIGURE 1: Proposed DeiT-based plant disease detection workflow adapted from standard DeiT and Vision Transformer architectures.

DeiT, Data-Efficient Image Transformer

Baseline CNN model for comparative evaluation

DenseNet-121 was selected as the CNN baseline for comparative evaluation against the proposed DeiT-Base model. To ensure fair comparison, both models used identical preprocessing, training, and evaluation settings. The DenseNet layer formulation is specified in Equation (16), where $[x_0, x_1, \dots, x_{l-1}]$ denotes the concatenation of feature maps from all preceding layers, and $H_l(\cdot)$ represents a composite operation consisting of batch normalization, activation function (ReLU), and convolution.

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]) \quad (16)$$

Results And Discussion

Experimental setup

Experiments were conducted using PyTorch on Kaggle GPU resources with NVIDIA Tesla T4 acceleration. The proposed DeiT-Base model was trained using the configuration described in the "Materials and Methods" section, while detailed hyperparameters are summarized in Table 2. Model performance was evaluated using accuracy, macro-precision, macro-recall, macro F1-score, confusion matrix analysis, and prediction visualization.

How to cite this article:

Batra M, Sharma P (June 08, 2026) A Data-Efficient Image Transformer for Multi-Crop Plant Disease Detection. Cureus J Comput Sci 3 : es44389-026-00108-9. DOI <https://doi.org/10.7759/s44389-026-00108-9>

Parameter	Value
Framework	PyTorch
Model	DeiT-Base Patch16 224
Image size	224 × 224
Batch size	32
Epochs	20
Learning rate	5×10^{-4}
Optimizer	AdamW
Loss function	Cross-Entropy loss
Train–Validation split	80:20
Hardware	NVIDIA Tesla T4 GPU
Mixed precision training	Enabled
Data augmentation	Rotation, Horizontal flip

TABLE 2: Experimental setup and hyperparameter configuration of the proposed DeiT-Base model.

DeiT, Data-Efficient Image Transformer

Performance comparison

Table 3 compares the performance of DeiT-Base and DenseNet-121 under identical training conditions. DeiT-Base achieved higher overall accuracy (98.50%), macro-precision (0.9843), macro-recall (0.9826), and macro F1-score (0.9833), indicating improved classification performance on the evaluated dataset.

Metric	DeiT-Base	DenseNet-121
Overall accuracy (%)	98.50	97.29
Macro precision	0.9843	0.9728
Macro recall	0.9826	0.9702
Macro F1-Score	0.9833	0.9713

TABLE 3: Comparative performance analysis of the DeiT-Base model and the DenseNet-121 CNN baseline under identical training and evaluation conditions.

CNN, Convolutional Neural Network; DeiT, Data-Efficient Image Transformer

Confusion matrix analysis

Figure 2 presents the normalized confusion matrix across 35 disease classes. Most predictions are concentrated along the diagonal, indicating strong classification performance. Misclassifications primarily occur among visually similar disease categories, while healthy classes are generally distinguished effectively.

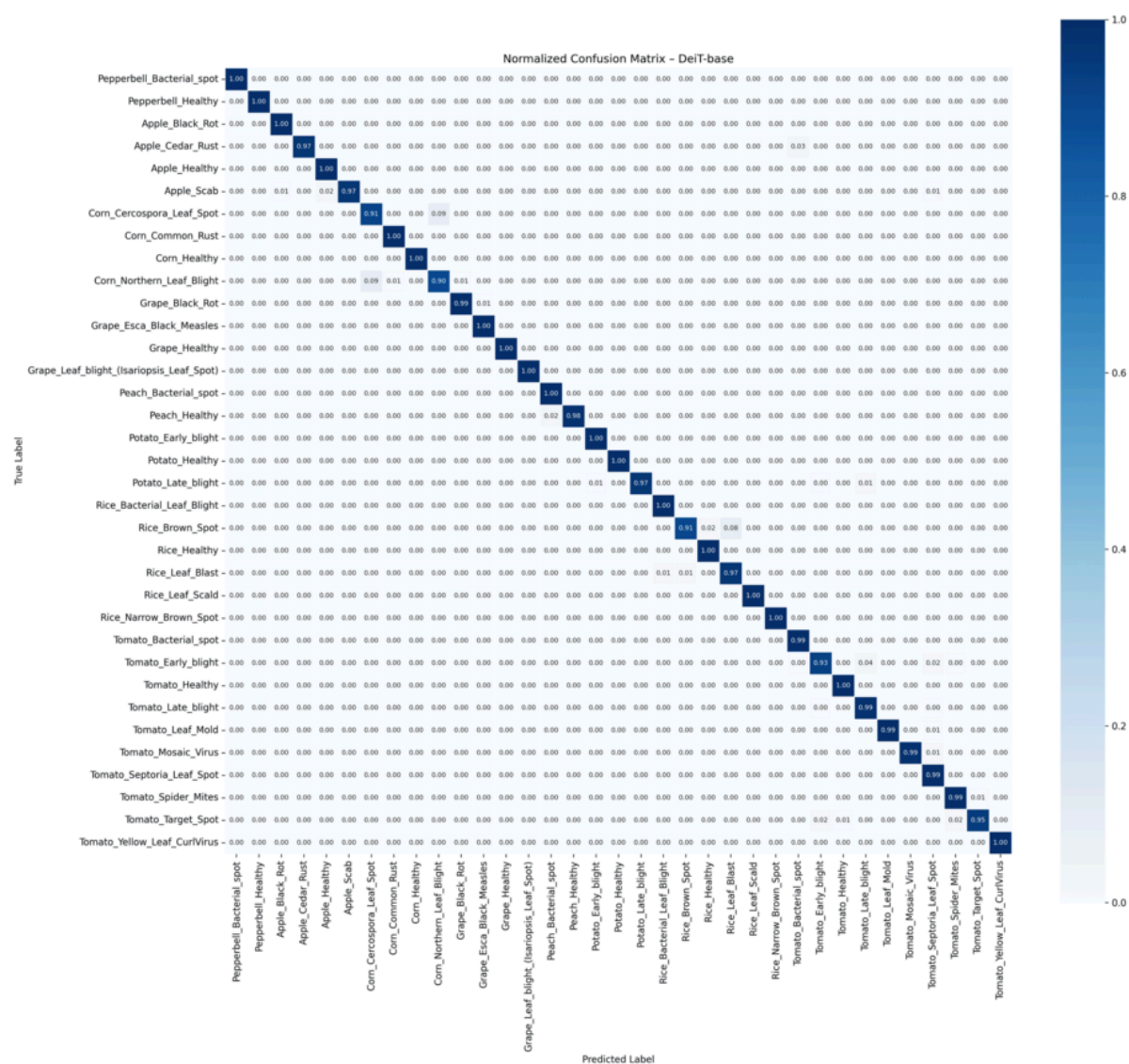


FIGURE 2: Normalized confusion matrix of DeiT-Base model across 35 disease classes.

DeiT, Data-Efficient Image Transformer

Accuracy and loss curve

Figure 3 illustrates training and validation accuracy and loss across 20 epochs. Training accuracy increased from 86.73% to 99.89%, while validation accuracy stabilized at 98.50%. Both training and validation losses decreased progressively, suggesting stable convergence with limited signs of overfitting.

How to cite this article:

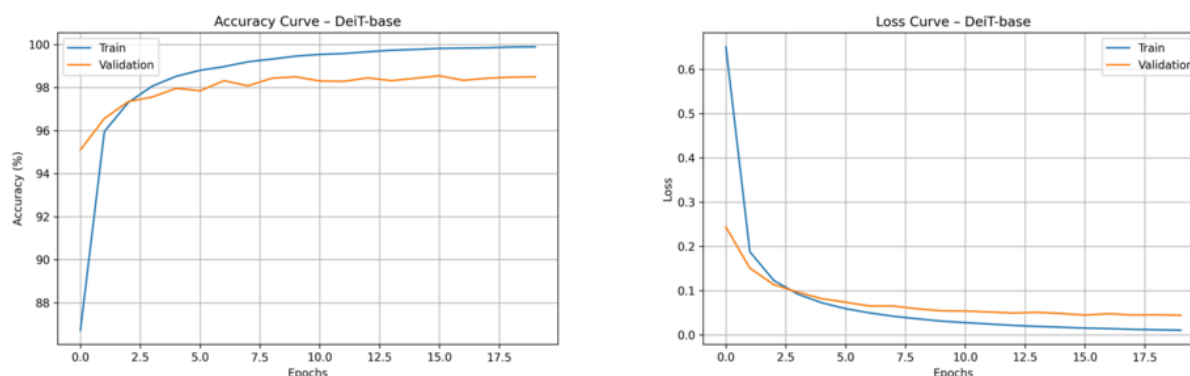
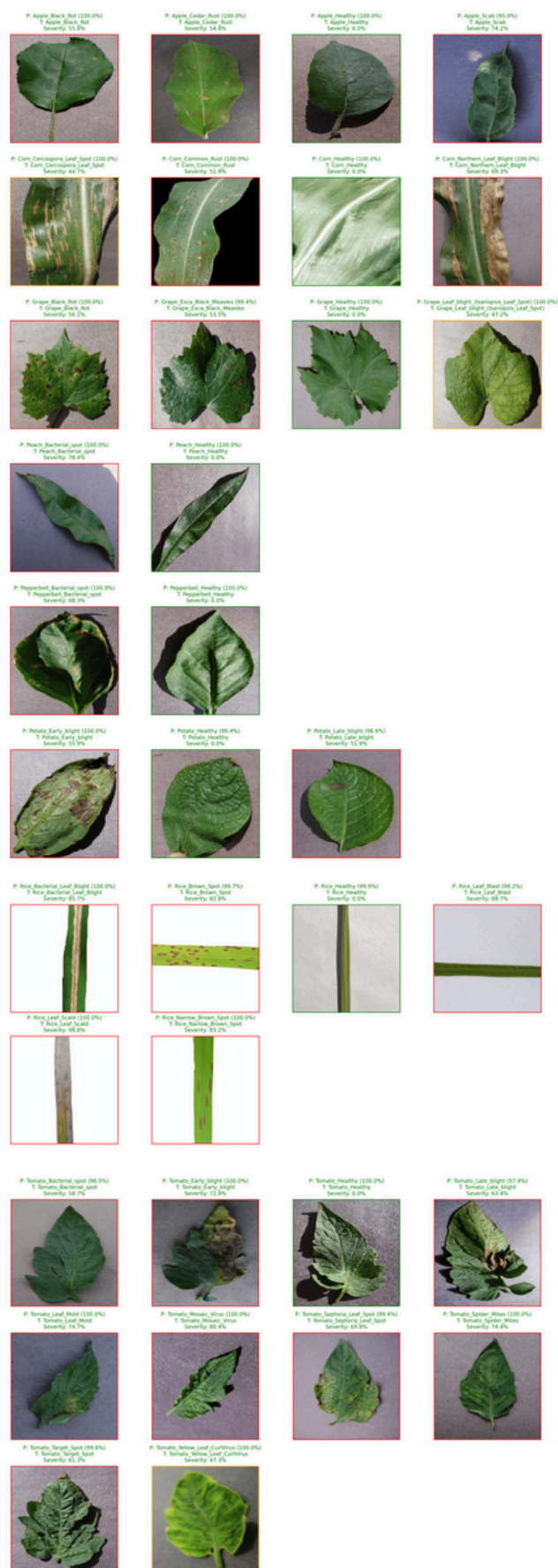


FIGURE 3: Training and validation accuracy and loss curves of the DeiT-Base model over 20 epochs.

DeiT, Data-Efficient Image Transformer

Visual prediction analysis

Figure 4 presents prediction examples across eight crop species and multiple disease categories using the proposed DeiT-Base model. The model produced high-confidence predictions while estimating disease severity from infected leaf regions. Green borders indicate correct predictions, while severity levels are visualized using orange and red indicators. The predictions suggest that the model captures visual characteristics such as discoloration and texture variations across disease categories. These results indicate that the model can effectively distinguish disease symptoms across multiple crop categories while maintaining high classification accuracy.



How to cite this article:

Batra M, Sharma P (June 08, 2026) A Data-Efficient Image Transformer for Multi-Crop Plant Disease Detection. Cureus J Comput Sci 3 : es44389-026-00108-9. DOI <https://doi.org/10.7759/s44389-026-00108-9>

FIGURE 4: Experimental results of the proposed DeiT-Base model for plant disease classification, prediction confidence, and disease severity estimation across eight crop species and 35 disease classes.

DeiT, Data-Efficient Image Transformer

Conclusions

This study presented a DeiT-Base-based framework for automated plant disease recognition across multiple crop species. By leveraging self-attention mechanisms, the model effectively captured global contextual information and achieved superior classification performance compared with DenseNet-121 under the same experimental conditions. Transfer learning enabled efficient adaptation of pretrained representations while requiring only classifier-head fine-tuning. Analysis of prediction results showed that most errors occurred among visually similar disease categories and early-stage infections. In addition to disease classification, the framework incorporated confidence estimation through softmax probabilities and an HSV-based approach for approximate severity assessment. Evaluation on a dataset of 30,945 images spanning 35 disease classes yielded an overall accuracy of 98.50%, demonstrating the potential of transformer-based architectures for large-scale plant disease recognition tasks.

Limitations and future work

Although the proposed DeiT-Base framework achieved high classification accuracy on the selected dataset, several limitations remain. The experiments were conducted using a single 80:20 train-validation split without external field validation or cross-validation. Therefore, additional evaluation under real agricultural conditions is necessary to further assess generalization capability. Furthermore, the HSV-based disease severity estimation method may be sensitive to illumination, background variation, and camera conditions. The confidence scores were derived from softmax probabilities and may not fully represent calibrated uncertainty. Future work will include additional transformer and CNN baselines, statistical significance analysis, ablation studies, and real-time deployment validation under field environments.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Muskan Batra, Pooja Sharma

Acquisition, analysis, or interpretation of data: Muskan Batra, Pooja Sharma

Drafting of the manuscript: Muskan Batra, Pooja Sharma

Critical review of the manuscript for important intellectual content: Muskan Batra, Pooja Sharma

Supervision: Pooja Sharma

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

How to cite this article:

Batra M, Sharma P (June 08, 2026) A Data-Efficient Image Transformer for Multi-Crop Plant Disease Detection. Cureus J Comput Sci 3 : es44389-026-00108-9. DOI <https://doi.org/10.7759/s44389-026-00108-9>

Data Availability Statements

The data supporting this study are openly available in Kaggle at <https://www.kaggle.com/datasets/naimur006/plant-leaves-disease-detection>. The code used in this study is available from the corresponding author upon reasonable request.

Acknowledgements

This article was previously presented at the Scope Global Skills University for the International Conference on Emerging Technologies in Artificial Intelligence, held on March 12–13, 2026.

References

1. Rahman KN, Banik SC, Islam R, Al Fahim A: [A real-time monitoring system for accurate plant leaves disease detection using deep learning](#). Science Direct. 2025, 4:1-16. [10.1016/j.crope.2024.100092](#)
2. Azath M, Zekiwo M, Bruck A: [Deep learning-based image processing for cotton leaf disease and pest diagnosis](#). Journal of Electrical and Computer Engineering. 2021, 2021:1-10. [10.1155/2021/9981437](#)
3. Islam MM, Adil MAA, Talukder MA, et al.: [DeepCrop: deep learning-based crop disease prediction with web application](#). Journal of Agriculture and Food Research. 2023, 14:1-11. [10.1016/j.jafr.2023.100764](#)
4. Jung M, Song JS, Shin AY, et al.: [Construction of deep learning-based disease detection model in plants](#). Scientific Reports. 2023, 13:1-13. [10.1038/s41598-023-34549-2](#)
5. Bhargava A, Shukla A, Goswami OP: [Plant leaf disease detection, classification, and diagnosis using computer vision and artificial intelligence: A review](#). IEEE. 2024, 12:37443-37469. [10.1109/ACCESS.2024.3373001](#)
6. Prasad KMA, Pushpa BR: [WMC-Leafset: a dataset of wax gourd and Mangalore cucumber plants for leaf miner and pest infestation diseased object detection](#). Data in Brief. 2025, 63:1-14. [10.1016/j.dib.2025.112231](#)
7. Murugavalli S, Gopi R: [Plant leaf disease detection using vision transformers for precision agriculture](#). Scientific Reports. 2025, 15:1-17.
8. Hemalatha S, Jayachandran JJB: [A multitask learning-based vision transformer for plant disease localization and classification](#). International Journal of Computational Intelligence Systems. 2024, 17:1-21. [10.1007/s44196-024-00597-3](#)
9. Salman Z, Muhammad A, Han D: [Plant disease classification in the wild using vision transformers and mixture of experts](#). Frontiers in Plant Science. 2025, 16:1-19. [10.3389/fpls.2025.1522985](#)
10. Günde M, Keskin F, Işık G: [Empowering few-shot plant disease diagnosis with DeiT and MAML++](#). Cluster Computing. 2025, 29:1-13. [10.1007/s10586-025-05810-2](#)
11. Sun C, Li Y, Song Z, Liu Q, Si H, Yang Y, Cao Q: [Research on tomato disease image recognition method based on DeiT](#). ScienceDirect. 2025, 162:127400. [10.1016/j.eja.2024.127400](#)
12. Roy PS, Kukreja V: [Vision transformers for rice leaf disease detection and severity estimation: a precision agriculture approach](#). Journal of the Saudi Society of Agricultural Sciences. 2025, 24:1-15. [10.1007/s44447-025-00007-w](#)
13. Rezaei M, Diepeveen D, Laga H, Soheli F: [Automatic pixel-level annotation for plant disease severity estimation](#). Computers and Electronics in Agriculture. 2026, 241:1-17. [10.1016/j.compag.2025.111316](#)
14. Ji M, Wu Z: [Automatic detection and severity analysis of grape black measles disease based on deep learning and fuzzy logic](#). Computers and Electronics in Agriculture. 2022, 193:106718. [10.1016/j.compag.2022.106718](#)
15. Nguyen DK, Choi YS, Lee JH, Tran MT, Xin X: [An effective deep learning model for classifying diseases on strawberry leaves and estimating their severity based on the multi-task U-Net](#). Multimedia Tools and Applications. 2025, 84:31329-31350. [10.1007/s11042-024-20413-6](#)

How to cite this article:

Batra M, Sharma P (June 08, 2026) A Data-Efficient Image Transformer for Multi-Crop Plant Disease Detection. Cureus J Comput Sci 3 : es44389-026-00108-9. DOI <https://doi.org/10.7759/s44389-026-00108-9>