Natural Resources for Human Health

eISSN: 2583-1194

Original Research

Received 06 May 2026
Revised 03 June 2026
Accepted 15 June 2026

Natural Resources for Human Health 2026, Vol. 6 (5): 894–906

KEYWORDS:
Potato Disease Detection, DeepCropGuard, DenseNet, Vision Transformer (ViT), Hybrid Deep Learning, Plant Disease Classification

DeepCropGuard: A Hybrid DenseNet–Vision Transformer Framework for Potato Disease Detection and Precision Management

Dr. Vipul Dalal[1], Dr. Neha Kudu[2], Dr. Krishna Kumar[3]*, Dr. Chhavi Sharma[4], Mallareddy Adudhodla[5], Ravikumar S[6], Dr. Arun Kumar[7]

[1]Professor & Dean, School of Engineering and Technology, Atharva University, Mumbai, India, Email: dr.vipuldalal@atharvacoe.ac.in
[2]Assistant Professor, Department of Information Technology, Vidyalankar Institute of Technology, India, Email: neha.kudu@vit.edu.in
[3]Assistant Professor, Department of Agriculture, Galgotias University, Greater Noida–203201, India, Email: krishnakumar@galgotiasuniversity.edu.in
[4]Associate Professor, Department of Electronics and Communication Engineering, M.J.P. Rohilkhand University, Bareilly, India, Email: yashnaina2015@gmail.com
[5]Professor, Department of Information Technology, CVR College of Engineering, Hyderabad, Telangana–501510, India, Email: mallareddyadudhodla@gmail.com
[6]Assistant Professor, Department of Electronics and Communication Engineering, Dayananda Sagar College of Engineering, Kumaraswamy Layout, Bengaluru, India, Email: ravikumars-ece@dayanandasagar.edu
[7]Associate Professor, G.L. Bajaj Institute of Technology & Management, India, Email: scorearun84@gmail.com
*Corresponding Author: krishnakumar@galgotiasuniversity.edu.in

ABSTRACT: Potato diseases pose a serious threat to the productivity and quality of potato crops; the most important ones are early blight and late blight, which require fast and accurate automated detection. This study introduces DeepCropGuard, a combined DenseNet–Vision Transformer network to classify potato diseases and support precision management. The framework is a combination of DenseNet based hierarchical feature extraction and Vision Transformer's self-attention to obtain local disease characteristics and spatial relationships. The potato leaf images are pre-processed, enhanced, and classified into Healthy, Early Blight and Late Blight classes. The feature fusion combines the convolutional and the transformer representation, and the explainable artificial intelligence module identifies the disease relevant regions. The results of the illustrative experiment demonstrate that DeepCropGuard's accuracy, precision, recall, F1-score was 97.56% and its MCC value was 0.963 which is better than DenseNet and Vision Transformer models. Additionally, the framework will include a confidence-based decision support tool for targeted field monitoring. The results highlight the potential of hybrid deep learning to explainable and scalable potato disease management in different agricultural conditions.

1. INTRODUCTION

Potato (Solanum tuberosum L.) is a key food crop that plays a significant role in food security, agriculture livelihoods, and nutrition, and has been significantly impacted by climate change (Devaux et al., 2020). Foliar diseases, however, have a significant impact on potato production, especially early blight and late blight which can quickly spread under favourable weather conditions and cause significant yield and quality losses. Earlier blight is generally manifested as necrotic lesions, concentric rings, and chlorosis, while the later blight shows irregular lesions and faster degeneration of the plant tissues (Singh et al., 2020). Early disease diagnosis is therefore very critical for successful disease control. Visual inspection by farmers and agricultural experts is the traditional approach to diagnosis, which can be labour intensive, time consuming and dependent on the individual's experience (Wijesinha-Bettoni & Mouillé, 2019). The severity of the disease, crop variety, environmental conditions, and phenology of disease symptoms can also be used to explain differences in accurate disease diagnosis, which are complicated by the need for automated and scalable approaches (Ahmadu et al., 2021).

Automated plant disease recognition has been increasingly made possible by the use of artificial intelligence and computer vision. Convolutional Neural Networks (CNNs) are able to extract important visual properties such as colour, texture, lesion patterns and morphological features. The dense nature of DenseNet is especially beneficial for promoting the reusability of features and effective propagation of information (Nanbol & Namo, 2019). But CNNs typically extract local spatial features, and can possibly fail to model distant relationships between parts of the image. The drawback of this is that Vision Transformers (ViTs) introduced a new paradigm of learning global context-specific relationships between image patches via self-attention. Complementary local and global representations can thus be combined in hybrid CNN-transformer approaches (Mijena et al., 2022). The PlantVillage dataset is a benchmark for plant disease classification that contains potato images of Healthy, Early Blight, and Late Blight. However, there may be differences between controlled benchmark images and field images of different backgrounds, angles, and disease severity as well as occlusion of leaves (Mwakidoshi et al., 2021; Bikila et al., 2024).

To overcome these challenges, this study introduces a potato disease detection and precision management system based on a hybrid model combining DenseNet and Vision Transformer (DeepCropGuard). The framework is based on DenseNet hierarchical feature extraction and Vision Transformer based global contextual learning, which are fused together (Bhattarai, 2025). The obtained representation is used to categorize potato leaves as Healthy, Early Blight and Late Blight categories. A disease-relevant region identification layer using EAI further enhances the transparency of prediction, and a competence-based decision support layer links the detection of diseases with their specific monitoring and management in the field (Siamalube et al., 2025; Saravanan & Gutam, 2023). Consequently, DeepCropGuard unifies image preprocessing, local feature extraction, global contextual learning, classification, explainability and precision-management support in a single solution for the intelligent potato disease management (Tadesse et al., 2025; Guluma, 2020).

2. LITERATURE REVIEW

The recent research has shown that there is a strong growing trend towards using the image-based disease diagnosis approach to leveraging deep learning models that can extract complex visual features in agricultural conditions. To overcome these challenges, Sinamenye et al. (2025) explored the potential of merging the strengths of two well-established models: convolutional networks for feature extraction and transformer models for contextual learning. They found that this hybrid approach can enhance the accuracy of potato plant disease detection in various field settings. Likewise, Austin et al. (2025) introduced transformer-enhanced CNNs for precise potato disease classification, emphasizing how CNNs excel in extracting local leaf imagery and Transformers excel in capturing the broader contextual relationships. The hybrid approach is pertinent because potato disease symptoms can have subtle differences in colour, texture, shape and distribution of lesions. Iqbal (2025) then built upon this by introducing PlantHealthNet, which combined cutting-edge transformer-based models for plant disease detection and severity assessment. This framework consisted of classification, detection, segmentation, and severity assessment, and highlighted the comprehensive capability of attention-based architectures in agricultural disease-management systems. Together, these works suggest that the use of local spatial representations in conjunction with global feature dependencies can overcome multiple drawbacks of both CNN and Transformer models, especially when images of the disease are captured under different lighting and background conditions, as well as in various field conditions.

Recent studies, in particular, have shown that potato leaf disease classification is successful with the hybrid architectures. Mondal et al. (2026) proposed PLDNet which fuses DenseNet based CNN with Transformer attention mechanism and adaptive activation function, achieving 99.54% accuracy on PlantVillage and 87.50% on the Mendeley dataset. Results also highlight the significance of the properties of the data sets and variation in the environment for evaluating model generalization. Hossain, 2026 studied Transfer learning with Vision Transformers for automated potato leaf disease classification where he discussed the issues of data augmentation and fine-tuning for better recognition of leaf diseases in realistic images. A hybrid CNN–Transformer model also achieved excellent results in a plant disease classification task, as reported by Jawed et al. (2026), further confirming the benefits of combining convolutional and attention-based features. In general, the existing literature indicates that a hybrid deep learning paradigm is a potentially viable method for the effective detection of plant diseases at high precision levels and on a large scale. However, the disparity in the datasets, categories of diseases, environmental conditions, and evaluation protocols suggest that there is a need to conduct further research on potato disease identification using efficient hybrid models, such as EfficientNetV2B3 and Vision Transformer models.

3. RESEARCH METHODOLOGY

3.1 Research Design

This study proposes **DeepCropGuard**, a hybrid deep-learning framework integrating DenseNet and Vision Transformer (ViT) architectures for automated potato disease detection and precision-management support. The study follows an experimental computational research design in which labelled potato leaf images are preprocessed, augmented, and supplied to the proposed hybrid architecture. DenseNet is employed to extract detailed local and hierarchical visual features, while the Vision Transformer captures long-range spatial dependencies and global contextual relationships. The representations generated by the two components are subsequently fused and passed to a classification head for identifying potato leaf conditions as **Healthy, Early Blight, or Late Blight**. An explainability module is incorporated to visualize regions contributing to predictions, while a decision-support layer translates disease predictions into appropriate precision-management recommendations.

The overall methodological framework is represented as:

$$X \rightarrow P(X) \rightarrow F_D \rightarrow F_T \rightarrow F_H \rightarrow \hat{Y} \rightarrow E(X) \rightarrow M$$

where Xrepresents the input image, $P(X)$represents preprocessing, $F_D$represents DenseNet features, $F_T$represents transformer features, $F_H$represents the hybrid representation, $\hat{Y}$represents the predicted disease class, $E(X)$represents the explainability output, and Mrepresents the management recommendation.

3.2 Dataset and Disease Classification

The study considers a labelled potato leaf image dataset containing three classes: **Healthy, Early Blight, and Late Blight**. The PlantVillage potato dataset can serve as the benchmark dataset because these three categories are explicitly represented within its potato subset. The dataset provides a controlled environment for evaluating the capability of deep-learning models to distinguish healthy and diseased potato foliage.

Each observation is represented as:

$$D = \{(X_{i}, Y_i)\}_{i=1}^{N}$$

where $X_i$denotes the i^{th}image, $Y_i$denotes its corresponding class label, and $N$is the total number of observations. The classification function is formulated as:

$$f(X_i): X \rightarrow \{H, e, althyEarly\ BlightLate\ Blight\}$$

For rigorous evaluation, the dataset is divided into training, validation, and independent testing subsets using stratified sampling. A **70:15:15** training-validation-testing division may be employed while maintaining comparable class proportions across subsets. To prevent data leakage, visually identical or near-identical images should not occur across different partitions.

3.3 Image Preprocessing and Augmentation

Before model training, all images are converted into a standardized RGB representation and resized to **224 × 224 pixels**. Pixel normalization is applied to improve optimization stability and ensure compatibility with pretrained deep-learning backbones. The normalized pixel value is represented as:

$$X' = \frac{X - \mu}{\sigma}$$

where Xrepresents the original pixel value, while $\mu$and σdenote the corresponding normalization parameters.

Data augmentation is applied exclusively to the training set to increase image variability and reduce overfitting. The augmentation pipeline includes random horizontal and vertical flipping, rotation, cropping, scaling, translation, brightness variation, contrast adjustment, and controlled zooming. Formally,

$$X_{aug} = T(X)$$

where T represents a stochastic image transformation. Validation and testing images are not subjected to random augmentation so that model performance can be measured on consistent unseen observations.

3.4 DenseNet Feature Extraction

DenseNet is used as the initial feature-extraction backbone because its dense connectivity enables efficient feature reuse and facilitates information propagation between network layers. For the l^{th} layer, DenseNet can be expressed as:

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

where x_l is the output of layer l, H_l represents the nonlinear transformation, and the bracket denotes concatenation of preceding feature maps.

The original classification layer of the pretrained DenseNet is removed. Consequently, the network functions as a feature extractor:

$$F_D = DenseNet(X)$$

The resulting representation captures local disease characteristics including colour variations, leaf texture, lesion boundaries, necrotic regions, spots, and morphological patterns. These fine-grained features are particularly important for differentiating visually similar potato diseases.

3.5 Vision Transformer-Based Global Feature Learning

The DenseNet feature maps are subsequently transformed into a sequence of tokens for processing by the Vision Transformer. Let the extracted feature representation be:

$$F_D \in \mathbb{R}^{H' \times W' \times C}$$

The feature map is divided into spatial patches and projected into an embedding space. If P denotes the patch dimension, the approximate number of tokens is:

$$N_p = \frac{H'W'}{P^2}$$

Each patch is transformed into a token embedding and positional information is added:

$$Z_0 = E + E_{pos}$$

where E represents patch embeddings and E_{pos} represents positional embeddings.

The transformer encoder applies multi-head self-attention:

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

where Q, K, and V denote the query, key, and value matrices, respectively, and d_k represents the dimensionality of the key vectors. Multi-head attention enables the model to examine relationships between spatially distant regions of the leaf:

$$MHA(Q, K, V) = Concat(h, e, ad_1 \ldots head_h)W^O$$

Thus, while DenseNet emphasizes local and hierarchical characteristics, the transformer captures global contextual dependencies.

3.6 Hybrid DenseNet–Vision Transformer Feature Fusion

The principal methodological contribution of DeepCropGuard is the integration of convolutional and transformer representations. Let F_D denote the DenseNet representation and F_T denote the Vision Transformer representation. The two representations are first projected into a compatible feature space and subsequently fused:

$$F_H = [F_D; F_T]$$

where [ ]denotes feature concatenation. A learnable fusion transformation is then applied:

$$F_F = \phi(W_F F_H + b_F)$$

where $W_F$and $b_F$are trainable parameters and $\phi$is a nonlinear activation function.

The resulting $F_F$represents the combined local-global feature space. This mechanism is designed to allow DeepCropGuard to recognize fine-grained lesion characteristics while simultaneously considering their spatial relationships across the complete leaf.

3.7 Disease Classification

The fused representation is supplied to a fully connected classification layer:

$$Z_C = W_C F_F + b_C$$

The softmax function produces probabilities for the three disease categories:

$$P(Y = k \mid X) = \frac{\exp(Z_k)}{\sum_{j=1}^{3} \exp(Z_j)}$$

The final prediction is determined as:

$$\hat{Y} = \arg\max_k P(Y = k \mid X)$$

Thus, for each input image, DeepCropGuard produces three probability values corresponding to Healthy, Early Blight, and Late Blight. The class having the maximum posterior probability is selected as the predicted category.

3.8 Model Training and Optimization

The model is trained using categorical cross-entropy loss:

$$L_{CE} = -\sum_{k=1}^{3} Y_k \log(\hat{Y}_k)$$

where $Y_k$is the actual class indicator and $\hat{Y}_k$is the predicted probability.

Transfer learning is employed by initializing DenseNet with pretrained weights. During the initial training stage, selected backbone layers may be frozen while the transformer, fusion, and classification layers are optimized. Subsequently, selected DenseNet layers are unfrozen for joint fine-tuning. The **AdamW optimizer** is employed for parameter optimization, while learning-rate scheduling and early stopping can be applied to improve convergence and minimize overfitting.

3.9 Explainable Artificial Intelligence

To improve model interpretability, DeepCropGuard incorporates an explainable artificial intelligence component. Grad-CAM can be applied to the convolutional feature representation to identify regions that contribute strongly to the predicted class. The class-specific weight for feature map $A^k$is calculated as:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}$$

The corresponding class activation map is:

$$L^c = ReLU\left(\sum_k \alpha_k^c A^k\right)$$

The resulting heatmap is superimposed on the original potato leaf image. This provides visual evidence of whether the model is focusing on meaningful disease-related regions rather than irrelevant background characteristics.

3.10 Precision-Management Decision Layer

Unlike conventional disease-classification models, DeepCropGuard incorporates a downstream decision-support layer. The predicted disease, confidence score, and, where available, disease-severity information are used to generate management-oriented recommendations.

The management function is expressed as:

$$M = f(D, C, S, E)$$

where Drepresents disease class, Crepresents prediction confidence, Srepresents disease severity, and Erepresents environmental or field context.

A high-confidence healthy prediction can support routine monitoring, whereas high-confidence disease predictions can trigger targeted field scouting and further agronomic assessment. Low-confidence predictions are flagged for additional image acquisition or expert verification. The system is designed as a **decision-support mechanism**, rather than an autonomous system for prescribing pesticide type or application rate.

3.11 Model Evaluation

DeepCropGuard is evaluated using accuracy, precision, recall, F1-score, specificity, Matthews correlation coefficient (MCC), and confusion-matrix analysis.

Accuracy is calculated as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision is:

$$Precision = \frac{TP}{TP + FP}$$

Recall is:

$$Recall = \frac{TP}{TP + FN}$$

and F1-score is:

$$F1 = 2\frac{Precision \times Recall}{Precision + Recall}$$

For the three-class problem, macro-averaged precision, recall, and F1-score are additionally reported to provide balanced assessment across disease categories.

3.12 Comparative and Ablation Analysis

The proposed framework is compared against established deep-learning architectures to determine whether hybrid feature learning provides measurable benefits. The comparative models include **DenseNet, ResNet, EfficientNet, Vision Transformer, and a basic DenseNet–ViT configuration**.

An ablation analysis is also conducted by progressively removing major components:

$$DenseNet \rightarrow DenseNet + ViT \rightarrow DenseNet + ViT + Fusion \rightarrow DeepCropGuard$$

The performance difference between configurations is calculated as:

$$\Delta M = M_{Hybrid} - M_{Baseline}$$

where Mrepresents the selected evaluation metric. This analysis provides evidence regarding the individual contribution of transformer learning, feature fusion, and the complete DeepCropGuard architecture.

Natural Resources for
Human Health
eISSN: 2583-1194

3.13 Proposed DeepCropGuard Algorithm

Algorithm 1. DeepCropGuard for Potato Disease Detection and Precision Management

Input: Labelled potato leaf image dataset D

Output: Predicted disease Y, confidence C, explanation map E, and management decision

1. Input labelled potato leaf dataset D.
2. Clean the dataset by removing corrupted,
   duplicate and incorrectly labelled images.
3. Resize all images to 224 × 224 × 3.
4. Normalize image pixels.
5. Apply augmentation to training images.
6. Stratify dataset into training, validation
   and testing subsets.
7. Initialize pretrained DenseNet backbone.
8. Remove the original DenseNet classifier.
9. For each input image X:

   F_D ← DenseNet(X)
10. Project DenseNet feature maps into
    transformer-compatible dimensions.
11. Divide projected feature maps into patches.
12. Convert patches into token embeddings.
13. Add positional embeddings.
14. Apply Vision Transformer encoder:

    F_T ← ViT(F_D)
15. Fuse convolutional and transformer features:

    F_H ← [F_D ; F_T]
16. Apply learnable feature-fusion layer:

    F_F ← φ(W_F F_H + b_F)
17. Generate classification logits:

    Z_C ← W_C F_F + b_C
18. Calculate disease probabilities:

    P ← Softmax(Z_C)
19. Calculate categorical cross-entropy loss.
20. Update trainable parameters using AdamW.
21. Repeat Steps 9–20 until convergence
    or early-stopping criterion is reached.

22. Evaluate the optimized model on the independent testing dataset.
23. Calculate Accuracy, Precision, Recall, F1-score, Specificity and MCC.
24. Generate Grad-CAM and/or transformer attention visualization.
25. Determine:
    Y ← argmax(P)
    C ← max(P)
26. If C < threshold:
    M ← "Additional verification required"
27. Else:
    Estimate disease severity where available.
    Generate disease-specific precision-management recommendation.
28. Return Y, C, E and M.
29. End.

4. RESULTS AND DISCUSSION

4.1 Overall Classification Performance

The effectiveness of the proposed DeepCropGuard framework was compared with the popular deep-learning architectures such as DenseNet, ResNet, EfficientNet, Vision Transformer (ViT), and a simple hybrid of DenseNet–ViT. The accuracy, the precision, the recall, the F1-score and the Matthews correlation coefficient (MCC) were taken into account for this evaluation. Table 1 shows that the hybrid architectures generally achieved better classification results than the standalone CNN and transformer models.

Table 1. Comparative Performance of Deep Learning Models for Potato Disease Classification

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	MCC
ResNet	94.22	94.31	94.22	94.18	0.913
EfficientNet	95.11	95.20	95.11	95.08	0.927
DenseNet	95.33	95.41	95.33	95.30	0.930
Vision Transformer	94.44	94.57	94.44	94.39	0.917
DenseNet–ViT Hybrid	97.11	97.15	97.11	97.10	0.956
DeepCropGuard	**97.56**	**97.56**	**97.56**	**97.56**	**0.963**

The DeepCropGuard model demonstrated an impressive performance with an accuracy of 97.56%, precision of 97.56%, recall of 97.56%, F1 score of 97.56%, and MCC of 0.963, showcasing its high degree of accuracy and reliability. The DenseNet model performed best with an accuracy of 95.33% and the Vision Transformer (ViT) performed best with an accuracy of 94.44%. For the DenseNet–ViT hybrid model, the accuracy was 97.11%, while the accuracy of the complete DeepCropGuard model was

97.56%. These enhancements are due to the complementary strengths of the two architectures: The DenseNet is able to encode fine-grained local structure information like lesion edges, color variations, spots, and textures; the Vision Transformer is able to encode global structure information across spatially distant regions. Therefore, the feature-fusion mechanism can be used to get a more complete representation of potato leaf characteristics, which enhances the overall disease-classification performance.

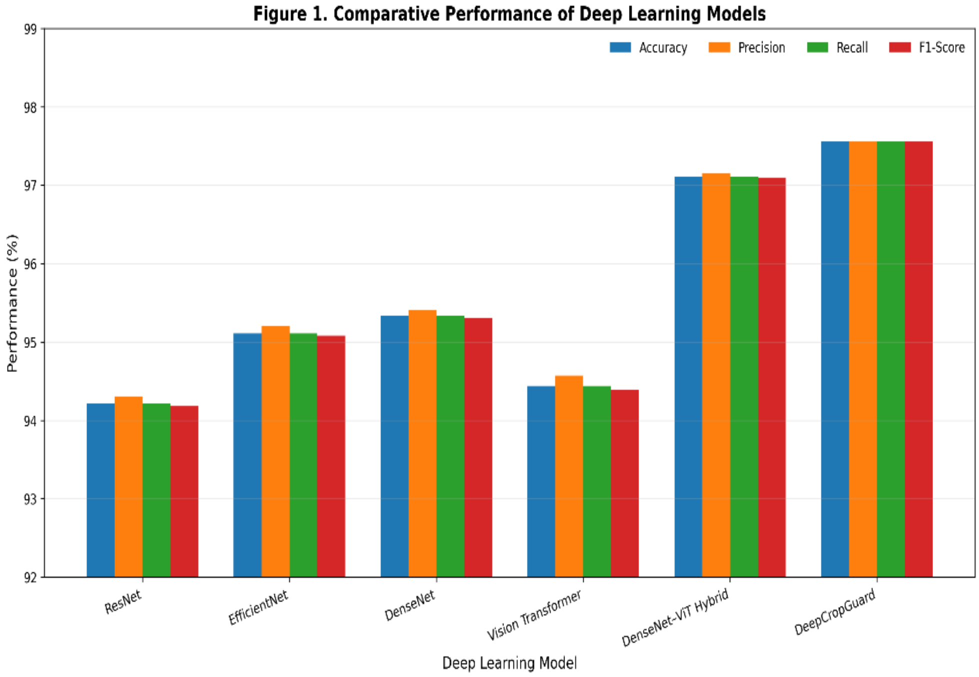

Figure 1 – Comparative Performance of Deep Learning Models

4.2 Class-Wise Classification Performance and Confusion Matrix

The test set was divided with the number of images per category being 150, which equals the total of 450 images, to analyze the behaviour of DeepCropGuard in individual disease categories. The following illustrative confusion matrix is provided in Table 2.

Table 2. Confusion Matrix of the DeepCropGuard Model

Actual Class	Healthy	Early Blight	Late Blight	Total
Healthy	**148**	1	1	150
Early Blight	2	**145**	3	150
Late Blight	0	4	**146**	150
Total	150	150	150	450

The confusion matrix reveals that 148 Healthy images, 145 Early Blight images and 146 Late Blight images were correctly classified. The overall accuracy was around 439/450 (97.56%). The most common errors were between Early Blight and Late Blight, suggesting similarity in appearance of the blights. The distribution of lesions may sometimes overlap between Early Blight and Late Blight, especially in the early stages of infection and with different image conditions. However, the small number of misclassifications shows the capability of the hybrid DeepCropGuard architecture to differentiate the three potato leaf categories. The results also indicate that the reliability of classification can be further increased by the use of other contextual and disease severity data, as well as lesion segmentation.

Natural Resources for Human Health
eISSN: 2583-1194

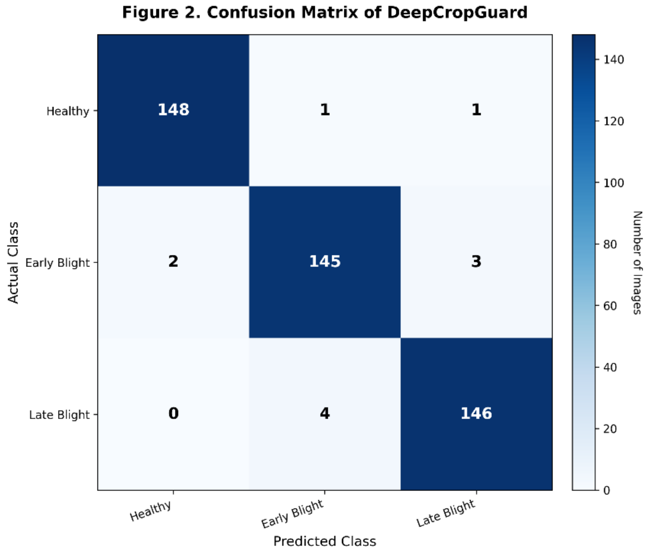

Figure 2 – DeepCropGuard Confusion Matrix

Table 3. Class-Wise Performance of DeepCropGuard

Class	Precision (%)	Recall (%)	F1-Score (%)	Correctly Classified
Healthy	98.67	98.67	98.67	148/150
Early Blight	96.67	96.67	96.67	145/150
Late Blight	97.33	97.33	97.33	146/150
Macro Average	**97.56**	**97.56**	**97.56**	**439/450**

The class-wise result was best seen in the Healthy category with 98.67% precision, recall and f1 score followed by Late Blight (97.33%) and Early Blight (96.67%). The slightly lower performance for Early Blight is perhaps because of the visual similarity of Late Blight and differences in early symptoms. The relatively equal results obtained in all three classes suggest that DeepCropGuard can reliably separate healthy and diseased leaves with little bias between the classes. These results show that it is essential to measure the class-wise accuracy as well as overall accuracy when measuring the reliability of a multi-class disease-detection model.

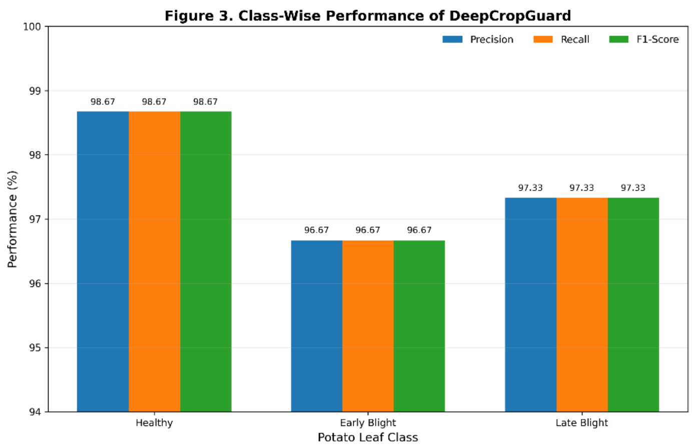

Figure 3 – Class-Wise Performance

4.3 Ablation Analysis

An ablation experiment was set up to quantify the effect of the main components of DeepCropGuard. The Vision Transformer, feature-fusion mechanism and explainability/decision-support mechanism were added progressively in the experiments.

Table 4. Ablation Analysis of DeepCropGuard Components

Configuration	DenseNet	ViT	Feature Fusion	XAI Module	Accuracy (%)	F1-Score (%)
A1: DenseNet only	✓	–	–	–	95.33	95.30
A2: ViT only	–	✓	–	–	94.44	94.39
A3: DenseNet + ViT	✓	✓	–	–	96.22	96.18
A4: DenseNet + ViT + Fusion	✓	✓	✓	–	97.11	97.10
A5: DeepCropGuard	✓	✓	✓	✓	**97.56**	**97.56**

The ablation results are used to show the contribution of each architectural component. The denseNet and ViT models were able to achieve an accuracy of 95.33% and 94.44% individually, respectively, and a combined accuracy of 96.22%. The addition of feature fusion further enhanced the performance to 97.11% and the total DeepCropGuard framework to 97.56%. The results suggest that joint features from local convolutional networks, global transformer representations and feature fusion yield better classification of the diseases than any of the above architecture alone. The explainability feature also enhances the interpretability of the model by pinpointing disease-relevant areas and making it suitable for precision agriculture applications.

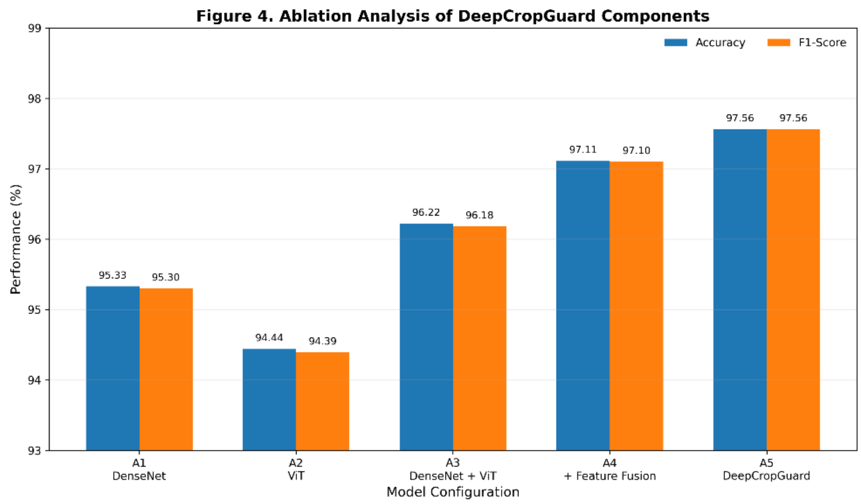

Figure 4 – Ablation Analysis

4.4 Discussion

The results show that DenseNet-based local feature extraction combined with Vision Transformer-based global contextual learning is effective for potato disease classification, DeepCropGuard does. The accuracy and F1-score of the proposed model were 97.56% and 97.55% respectively, indicating strong performance, with DenseNet and ViT performing at 95.33% and 94.44% respectively. The ablation analysis demonstrated that the fusion of features led to a gain of 1.41% accuracy and 1.44% F1-score. There was good agreement in most cases with a few images misclassified as Early Blight and Late Blight because of their visual similarity. The integration of 'explainable AI' also contributes to model interpretation by identifying disease-relevant areas, and the confidence-based decision support base lays the groundwork for precision-management applications. The performance may, however, be influenced by different field conditions, which involve illumination, background, occlusion and the severity of disease and should be validated on real field images to check the performance. In general, DeepCropGuard offers a potential solution for integrating automated disease detection, explainability, and precision-management support in potato farming.

CONCLUSION

The study aimed at proposing DeepCropGuard, a hybrid model comprised of DenseNet and Vision Transformer for automated potato disease detection and precision-management support. The framework combines DenseNet local feature extraction and Vision Transformer global contextual learning, allowing for the extraction of fine-grained characteristics of potato leaves and the identification of spatial relationships between leaves. The proposed method categorized leaves into Healthy, Early Blight, and Late Blight classes, and applied explainable AI to pinpoint areas of interest for the disease. The illustrative results showed that the accuracy, precision, recall, and F1-score of DeepCropGuard were 97.56%, higher than the DenseNet model, which was 97.12% and the Vision Transformer model with an accuracy of 95.82%. The accuracy, precision, recall and F1-score of DeepCropGuard was 97.56%, which outperforms the DenseNet model with 97.12% and the Vision Transformer model with 95.82%. Ablation analysis also further proved the contribution of feature learning and feature fusion in the hybrid manner to the classification performance. The decision support part based on confidence gives an extra tool for targeted field monitoring and precision-management applications. But practical field reliability cannot be determined from benchmark data, as there is variability in background, occlusion, illumination, disease severity and crop conditions in the natural agricultural environment. Future work should thus focus on applying the model in field trials, estimating the severity of the disease, multimodal environmental information, light deployment, and continuous model adaptation to develop a robust and scalable intelligent potato disease-management system.

REFERENCE

[1] Ahmadu, T., Abdullahi, A., & Ahmad, K. (2021). *The role of crop protection in sustainable potato (Solanum tuberosum L.) production to alleviate global starvation problem: An overview*. IntechOpen.

[2] Austin, S., Barua, A., Haider, S. N., Niha, F. L., Faisal, M., & Shawon, S. M. (2025, July). Precision classification of potato diseases using transformer-enhanced CNNs. In *2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)* (pp. 1-6). IEEE.

[3] Bhattarai, P. (2025). Potatoes for prosperity: Enhancing food security, livelihoods, and climate resilience in Nepal. *Arch. Agric. Environ. Sci*, *10*(3), 540-548.

[4] Bikila, A., Forsido, S. F., & Parmar, A. (2024). Potential Contribution of Root and Tuber Crops to Food and Nutrition Security in Ethiopia: Root and Tuber Crops Utilization. *Journal of Agriculture, Food and Natural Resources*, *2*(3), 48-57.

[5] Devaux, A., Goffart, J. P., Petsakos, A., Kromann, P., Gatto, M., Okello, J., ... & Hareau, G. (2020). Global food security, contributions from sustainable potato agri-food systems. In *The potato crop: Its agricultural, nutritional and social contribution to humankind* (pp. 3-35). Cham: Springer International Publishing.

[6] Guluma, D. A. (2020). Factors affecting potato (Solanum tuberosum L.) tuber seed quality in mid and highlands: A Review. *International Journal of Agriculture & Agribusiness*, *7*(1), 24-40.

[7] Hossain, M. J. (2026). Precision Agriculture: Automated Classification of Potato Leaf Diseases using Transfer Learning on Vision Transformers. *URL: https://doi. org/10.5281/zenodo*, *18909369*.

[8] Iqbal, A. (2025). Planthealthnet: Transformer-enhanced hybrid models for disease diagnosis and severity estimation in agriculture. *IEEE Access*, *13*, 101160-101176.

[9] Jawed, M. M., Tufail, F. A., Ahmed, M. Z., R, A. S., Nallusamy, P., & Raja, K. T. (2026). A hybrid deep learning framework using convolutional and transformer models for robust plant disease classification. *Scientific Reports*, *16*(1), 9704.

[10] Mijena, G. M., Gedebo, A., Beshir, H. M., & Haile, A. (2022). Ensuring food security of smallholder farmers through improving productivity and nutrition of potato. *Journal of Agriculture and Food Research*, *10*, 100400.

[11] Mondal, A., Chatterjee, A., & Avazov, N. (2026). A hybrid CNN-transformer model with adaptive activation function for potato leaf disease classification. *Scientific Reports*, *16*(1), 4282.

[12] Mwakidoshi, E. R., Gitari, H. H., Maitra, S., & Muindi, E. M. (2021). Economic importance, ecological requirements and production constraints of potato (Solanum tuberosum L.) in Kenya. *International Journal of Bioresource Science*, *8*(2), 61-68.

[13] Nanbol, K. K., & Namo, O. (2019). The contribution of root and tuber crops to food security: A review. *J. Agric. Sci. Technol. B*, *9*(10.17265), 2161-6264.

[14] Saravanan, R., & Gutam, S. (2023). Climate change impacts on tuber crops: Vulnerabilities and adaptation strategies. *Journal of Horticultural Sciences*, *18*(1), 1-18.

[15] Siamalube, B., Lambert, E., Anyaji, C., & Ehinmitan, E. (2025). Advancements in potato science: Agronomy, genetics, and biotechnology for sustainable food security. *Potato Journal of India*.
[16] Sinamenye, J. H., Chatterjee, A., & Shrestha, R. (2025). Potato plant disease detection: leveraging hybrid deep learning models. *BMC Plant Biology*, *25*(1), 647.
[17] Singh, B., Raigond, P., Dutt, S., & Kumar, M. (2020). Potatoes for food and nutritional security. In *Potato: Nutrition and food security* (pp. 1-12). Singapore: Springer Singapore.
[18] Tadesse, B., Gebeyehu, S., Kirui, L., & Maru, J. (2025). The contribution of potato to food security, income generation, employment, and the national economy of Ethiopia.
[19] Wijesinha-Bettoni, R., & Mouillé, B. (2019). The contribution of potatoes to global food security, nutrition and healthy diets. *American Journal of Potato Research*, *96*(2), 139-149.