Wheat leaf disease classification using vision transformer based deep learning approach

P. Ashwini[1], D. Mythili[2], P. Jyothi[3], K. Swathi[4], G. Rama Krishna[5], M. Sowmya[6]
[1]Department of Computer Science and Engineering, Sreenidhi University, Hyderabad, India
[2]Department of Computer Science and Engineering, Vasavi College of Engineering, Hyderabad, India
[3]Department of Computer Science and Engineering, Teegala Krishna Reddy Engineering College, Hyderabad, India
[4]Department of Computer Science and Engineering, CVR College of Engineering, Hyderabad, India
[5]Department of Computer Science and Engineering, YJR DMS College of Engineering, Machilipatnam, India
[6]Department of Computer Science and Engineering, Stanley College of Engineering and Technology for Women, Hyderabad, India

Article Info

Article history:

Received Jul 27, 2025
Revised Jun 6, 2026
Accepted Jun 17, 2026

Keywords:

Agriculture
Classification
Deep learning
Plant diseases
Preprocessing
Visual transformer

ABSTRACT

Wheat crops can be affected by numerous fungal diseases like leaf rust, stripe rust, and stem rust. These diseases can significantly reduce yield and grain quality and also result in highly economic damage. Detection of diseases at an early stage is the main and most complex task for farmers due to common morphological properties like color, shape, texture, and edges. Deep learning (DL) offers a powerful solution for detecting wheat crop diseases due to its ability to accurately identify diseases, even in complex field conditions, and facilitate early intervention. By analyzing images, DL models can learn complex patterns associated with various diseases, enabling faster and more reliable diagnoses than traditional methods. In this paper use DL models followed by a vision transformer model (ViTM) to extract multilevel features to focus more on the disease-infected area on the plant with enhanced accuracy. DL models extract hierarchical features, whereas transformer models are utilized to capture global dependencies and contextual relationships within the image. In the proposed research evaluated various DL models, InceptionV3, EfficientNet-B0, visual geometry group 16 (VGG-16), and ResNet-50 with visual transformers on the wheat disease detection dataset of 3679 images and achieved accuracy rates of 98.7%, 97.6%, 93.9%, and 97.5%, respectively. The experiment proved that InceptionV3 outperforms the other models.

Corresponding Author:

P. Ashwini
Department of Computer Science and Engineering, Sreenidhi University
Hyderabad, India
Email: ashwini_cse@mvsrec.edu.in or ashwinireddy90@gmail.com.

1. INTRODUCTION

In India, wheat is a vital crop that is used as a staple and greatly enhances the nation's food security. After China, India is the world's second-largest producer of wheat and plays an important role in the global wheat market. Various fungal infections can have a substantial impact on the yield and quality of wheat crops. These diseases include powdery mildew, tan spot, Fusarium head blight, rusts, smuts, and Septoria leaf blotch. Particularly, rusts inflict significant harm that lowers grain quality and reduces yield [1]. These fungal diseases are particularly detrimental because they cause loss in yield and deteriorate the chemical composition of the grain. Nevertheless, the production of wheat is severely constrained by a plethora of diseases like leaf, stem, stripe rusts, and fungi. Most traditional methods of disease detection depend on visuals, which entail manual inspections that are often laced with inconsistencies and subjectivity [2].

In traditional approaches, it is up to the workers to determine whether they can identify the lesions of the specific area through advanced eyesight. Identification becomes a matter of personal judgment, thus leading to the wrong identification of diseases. In some cases, agronomists or plant pathologists mistakenly identify the illness, which leads to ineffective control measures. To prevent damage to crops, there is a need for active intervention and rapid response in diagnosing and prescribing treatment, which makes this issue substantial [3]. Previous work has investigated machine learning (ML) and deep learning (DL) approaches for detecting plant diseases. While traditional ML was based on manually designed features. To resolve these issues, a solution has come in the form of DL, a subdivision of artificial intelligence (AI). With advancements in computer vision technology, DL has been integrated into the field of crop disease detection because it is both easy to use and highly accurate [4], [5]. Even though it is critical for global food security, there is limited research on vision-based disease detection in cereal crops using computer vision due to issues with dataset availability [6].

2. RELATED WORKS

Long *et al.* [7] developed a DL model to classify healthy wheat leaves and four disease categories in the field and glasshouse conditions with an accuracy of 97.05%. Hossen *et al.* [8] proposed a wheat disease detection system based on 2D-convolutional neural networks (CNN) with 4,800 images and reported 98.84% classification accuracy. For cotton leaf disease detection, Ahmad *et al.* [9] evaluated five DL models and two transformer-based models. The accuracy for binary classification was 96.72%, and for multiclass classification, it was 93.39%. Cheng *et al.* [10] incorporated a positional attention block into ResNet to enhance disease localization and achieved an accuracy of 96.4%. Nigam *et al.* [11] obtained an accuracy of 98.70% and an F1-score of 98% by using the EfficientNet-B0-CBAM model on the WheatRust21 dataset. Alharbi *et al.* [12] proposed a few-shot learning framework based on EfficientNet with attention mechanisms for 18 classes of wheat diseases and achieved 93.19% accuracy. Joseph *et al.* [13] developed a multi-crop disease dataset and evaluated several DL models, where the MobileNet variants achieved accuracies above 96% for the classification of wheat disease. Nigus *et al.* [14] developed a DL-based wheat stem rust grading system based on image pre-processing and Softmax classification with 92.01% testing accuracy. Zhang *et al.* [15] applied deep feature fusion for wheat disease diagnosis and obtained better classification performance than single image feature strategies. Fang *et al.* [16] proposed a lightweight multiscale CNN with residual and inception modules, and obtained 98.7% accuracy for wheat disease detection. Wen *et al.* [17] proposed the MnasNet-SimAM model, which combines attention mechanisms and transfer learning to identify wheat diseases with an accuracy of 95.14%. A hybrid MobileNet-DenseNet model for wheat and barley disease classification was proposed by Nagpal *et al.* [18]. The SVM classification achieved an accuracy of 98.89% on crop disease detection. Kalpana *et al.* [19] integrated Swin Transformers and residual networks as an ensemble for crop disease recognition and obtained better performance in feature extraction and classification.

3. PROPOSED METHOD

The proposed architecture shown in Figure 1 uses DL and a vision transformer model (DLViTM) to classify wheat leaf images into three classes: healthy, yellow rust, and brown rust. The method combines the powerful feature extraction of CNNs (EfficientNet-B0, InceptionV3, VGG-16, and ResNet-50) with the global attention of transformers to improve the model's capacity to recognize spatial features. The initial step is carried out by a state-of-the-art transfer learning model, which extracts rich spatial features from input images after pre-training on huge quantities of image data. In the second step, the proposed model uses a 6-layer multihead self-attention (with 8 attention heads each) to discover the relations and dependencies in that space. The sequence data is passed through transformer processing and global average pooling to obtain a single vector. Lastly, the result is passed to a fully connected layer, which calculates class probabilities by projecting these transformed features into the final classes.

3.1. Image dataset

There are 737 testing images and 2,942 training images in the wheat disease detection dataset used in this research, as shown in Table 1. The images shown in Figure 2 are divided into three different groups, where brown rust, as shown in Figure 2(a); healthy, as shown in Figure 2(b); and yellow rust, as shown in Figure 2(c). Two of these classes include pictures of wheat leaves with brown rust and yellow rust disease types, while one class represents the category of healthy wheat leaves. Images from both lab and field situations are included in the dataset. The pictures are 1,024 × 1,024 pixels in size and have a resolution of 96 dots per inch (DPI). Figure 1 illustrates the overall architecture of the proposed model for wheat disease detection and classification.

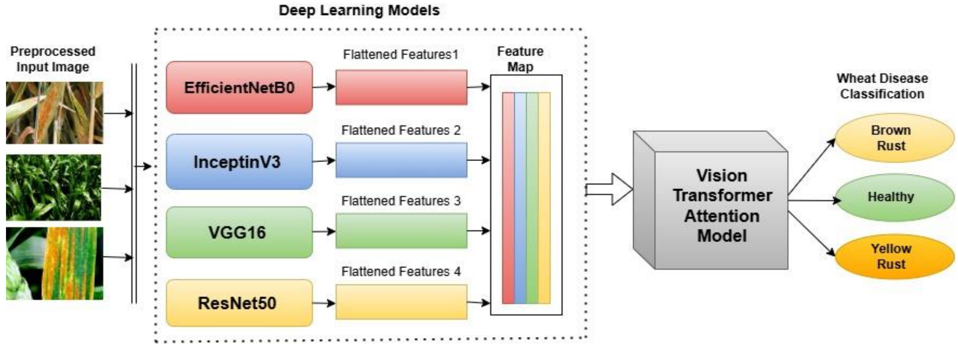


Figure 1. Proposed block diagram for wheat disease classification

(a) (b) (c)

Figure 2. Wheat leaf disease types of (a) brown rust, (b) healthy, and (c) yellow rust

Table 1. Wheat disease dataset description

Class	Training	Validation	Total
Brown rust	902	226	1,128
Healthy	1,116	279	1,395
Yellow rust	924	232	1,156
Total	2,942	737	3,679

3.2. Preprocessing

The wheat disease detection dataset was collected from the Kaggle repository. In the first step, images were resized into two different dimensions, such as 224 × 224 and 299 × 299, to give these images as input to the various transfer learning models. These resized images are normalized to adjust the pixel intensity ranges between 0 and 1. Contrast stretching improves the contrast enhancement of an image, which helps transfer learning models in extracting the most important features for improving classification performance.

3.3. Deep learning models

This study implemented deep transfer learning models that include a vision transformer (ViT) attention module to extract features from preprocessed wheat leaf images. The attention-enhanced features were used to improve representation quality. Various CNN models were then applied for classification.

3.3.1. EfficientNet-B0

EfficientNet-B0, as shown in Figure 3, is a lightweight but strong CNN architecture pretrained on the ImageNet dataset. EfficientNet-B0 is constructed as a series of mobile inverted bottleneck convolutional (MBConv) blocks with squeeze-and-excitation (SE) attention mechanisms [20]. The model employs a compound scaling strategy to maintain depth, width, and resolution balanced, achieving high accuracy with low parameters. The model starts with a stem layer that down samples the input image (224 × 224 × 3) to a 112 × 112 × 32 feature map by a Swish-activated convolutional layer. Seven MBConv blocks are arranged with multiple kernel sizes 3 × 3 or 5 × 5 and with expansion ratios 1 or 6; hence, the no of feature maps is raised from 16 to 320, which captures complex representations of the input. The feature maps are up-sampled to 7 × 7 × 1,280 at the top convolution layer to extract texture, color gradient, and disease patterns from the image. These features are flattened and fed to the transformer block for global context learning [21].

3.3.2. InceptionV3

InceptionV3, as shown in Figure 4, is a high-level architecture for deep convolutional neural networks (DCNNs) and is used for disease classification and feature extraction of wheat leaf diseases. The architecture consists of a series of convolutional, pooling, and inception modules that work towards extracting complex hierarchical features on a wide range of spatial resolutions. The early layers are concerned with the extraction of basic features such as edges and textures. Parallel processing of multi-scale convolutions (1 × 1, 3 × 3, 5 × 5) is enabled in each individual block. As data moves deeper into the network, reduction blocks are utilized to downsample spatial dimensions (from 35 × 35 to 17 × 17, and then to 8 × 8), while increasing the feature depth. The final inception layers produce high-level semantic features that are critical to accurate classification. Then, a global average pooling layer summarizes the spatial information into a 2,048-dimensional feature vector [22]. This feature vector is then passed through a dense layer with 256 units and using rectified linear unit (ReLU) activation, followed by a dropout layer with a standard rate of 0.5 to prevent the risk of overfitting.

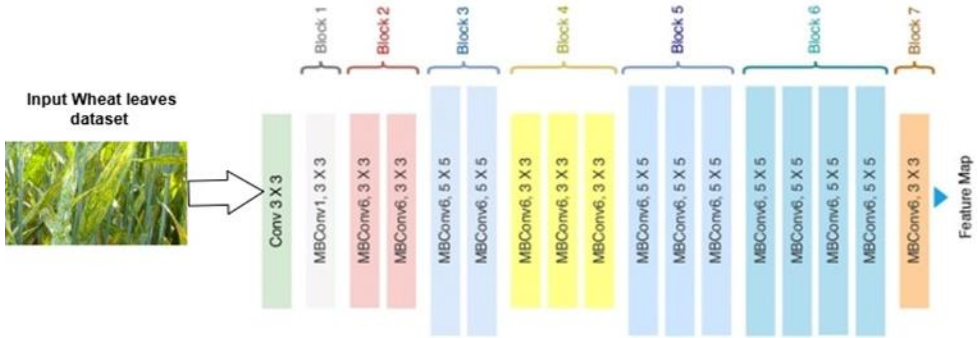


Figure 3. EfficientB0 architecture

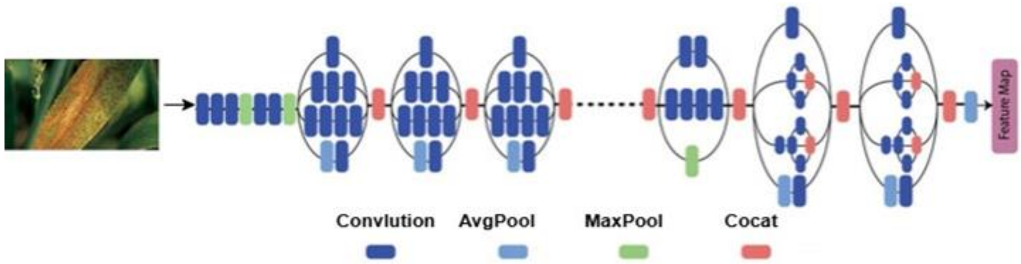


Figure 4. InceptionV3 architecture

3.3.3. VGG-16

The VGG-16 structure has 13 convolutional layers, 5 max-pooling layers, and 3 fully connected layers, as shown in Figure 5. It is designed to work with input images of size 224 × 224 × 3, which is the standard size for red, green, and blue (RGB) leaf images in wheat disease datasets. The Conv1_1 and Conv1_2 apply 64 filters of 3 × 3 size with the same padding, which leads to a feature map of size 224 × 224 × 64. The first 2 × 2 max-pooling operation is followed to reduce the spatial size to 112 × 112 × 64. In block-2, 128 filters are applied to each convolutional layer, which produces an output of 112 × 112 × 128, which is then downsampled to 56 × 56 × 128. The number of filters is doubled to 256 in block-3 and generates an output of 56 × 56 × 256. The second 2 × 2 max-pooling layer reduces it to 28 × 28 × 256. A 512-filter is applied to blocks-4 and 5 in each convolutional layer, which progressively downsample to 14 × 14 × 512, and finally, after the fifth max-pooling layer, the feature map is reduced to a 7 × 7 × 512 dimension [23].

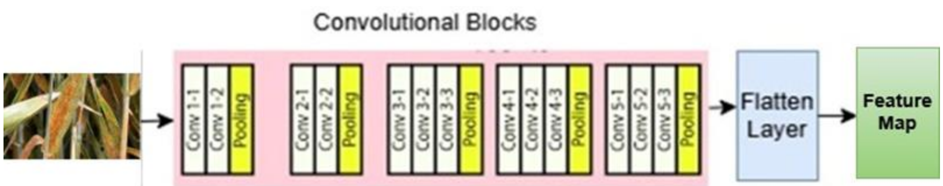


Figure 5. VGG-16 architecture

3.3.4. ResNet-50

ResNet-50, as shown in Figure 6, is a 50-layer CNN architecture with its signature residual learning approach that assists in overcoming the vanishing gradient problem typical of long networks. The model begins by accepting RGB images of wheat leaves with a size of 224 × 224 × 3. A feature map of size 112 × 112 × 64 is generated by applying a 7 × 7 convolutional layers with 64 filters and a stride of 2. Spatial dimensions are further reduced to 56 × 56 × 64 by applying a 3 × 3 max-pooling layers with a stride of 2. ResNet-50 consists of four convolutional layer stages (conv2_x to conv5_x), each of which consists of a series of several bottleneck residual blocks [24].

$$Y(i,j) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} X(i+m, j+n).K(m,n) \quad (1)$$

$$F(w * x + b) = (F(X)) \quad (2)$$

$$Y(i,j) = Max_{m=0}^{k-1} Max_{n=0}^{k-1} X(i.s+m, j.s+n) \quad (3)$$

Stage 1 (conv2_x) combines three residual blocks that project the input into feature maps of size 56 × 56 × 256. During Stage 3 (conv4_x) with six bottleneck blocks, the network strengthens feature extraction at a spatial scale of 14 × 14 × 1,024. Stage 4 (conv5_x) downscales the spatial resolution to 7 × 7 × 2,048.

$$s(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}} \quad (4)$$

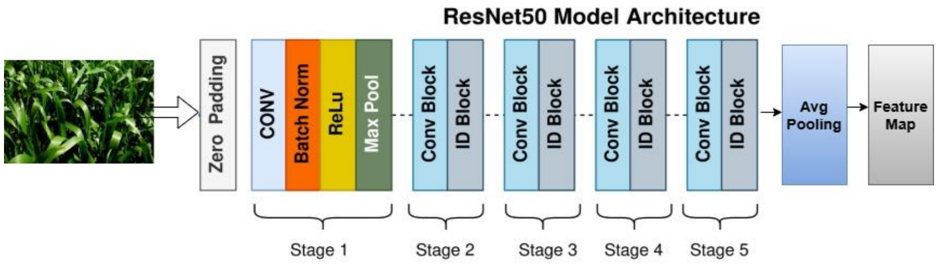


Figure 6. ResNet-50 architecture

3.3.5. Visual transformer attention model

The initial features obtained from the different DL models are passed as input to the ViT attention model to get deep features. ViT, as shown in Figure 7 is a model that utilizes self-attention mechanisms to deal with images. The ViT architecture is made up of a sequence of transformer blocks. Each transformer block has two sub-layers: a multi-head self-attention sub-layer and a feed-forward sub-layer [25].

The self-attention layer computes the attention weights for every pixel of the wheat leaf image in relation to all other pixels, whereas the feed-forward layer performs a non-linear transformation on the output of the self-attention layer. The multi-head attention generalizes this mechanism, which is observed by the model focusing on various regions of the input sequence at the same time. ViT also has a patch embedding layer, which splits the image into patches of fixed size and projects each patch into a high-dimensional vector representation. The patch embeddings are passed through the transformer blocks for processing. The output of the ViT architecture is a class prediction using a fully connected layer.

The first step is to feed the features that are extracted through the transfer learning model to the ViT model, whereby the extracted features IF are mapped into a 2D image $I \in R\ L * W * C$ dimensions, where L is the length, W is the width of the image, and C is the number of channels. The last dimensions of the original features that are extracted by pre-trained models: EfficientNet-B0, InceptionV3, VGG-16, and ResNet-50 are $R\ 7 * 7 * 1{,}280$, $R\ 8 * 8 * 240$, $R\ 7 * 7 * 512$, and $R\ 7 * 7 * 2{,}048$, respectively.

Then these features are passed as input to the ViT model to be processed. To connect to the transformer, the 2D feature map is reshaped into a sequence of image patches. For instance, the feature map is unfolded from (7, 7, 1,280) into (49, 1,280), in which each of the 49 patches is used as a token input to the transformer encoder, and 1,280 denotes the depth of the feature representation (channels) at a given spatial location.

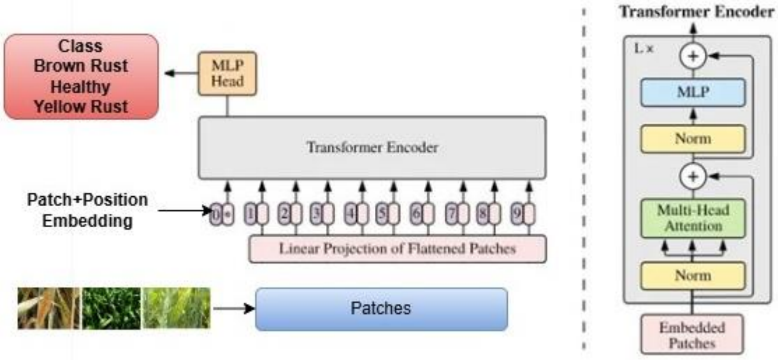


Figure 7. Visual transformer attention model for wheat disease classification

4. RESULTS AND DISCUSSION

Different DL models with a visual transformer are experimented with and evaluated with respect to their performance, as shown in Table 2. The various performance metrics, accuracy, precision, recall, and F1-score, are evaluated and compared for four different DL models: EfficientNet-B0, InceptionV3, VGG-16, and ResNet-50 with a hybrid DLViTM. Table 2 presents the performance evaluation of the DL models integrated with ViT using precision, recall, F1-score, and accuracy metrics. Figure 8 illustrates the precision, recall, and F1-score achieved by the InceptionV3, EfficientNet-B0, VGG-16, and ResNet-50 models for brown rust, healthy, and yellow rust wheat leaf classes. The results indicate that InceptionV3 and ResNet-50 provide superior classification performance across all classes.

Table 2. Performance measures of DL models with ViT

Model	Class	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
InceptionV3 + ViT	Brown rust	98.7	98.2	98.4	98.7
	Healthy	99.3	99.6	99.4	
	Yellow rust	98.1	98.7	98.4	
EfficientNet-B0 + ViT	Brown rust	96.8	97.2	97	97.6
	Healthy	97.9	97.5	97.7	
	Yellow rust	97.3	97.6	97.4	
VGG-16 + ViT	Brown rust	93.4	92.8	93.1	93.9
	Healthy	94.7	94.2	94.4	
	Yellow rust	93.9	94.1	94	
ResNet-50 + ViT	Brown rust	97	96.8	96.9	97.5
	Healthy	97.4	97.7	97.6	
	Yellow rust	97.1	97.3	97.2	

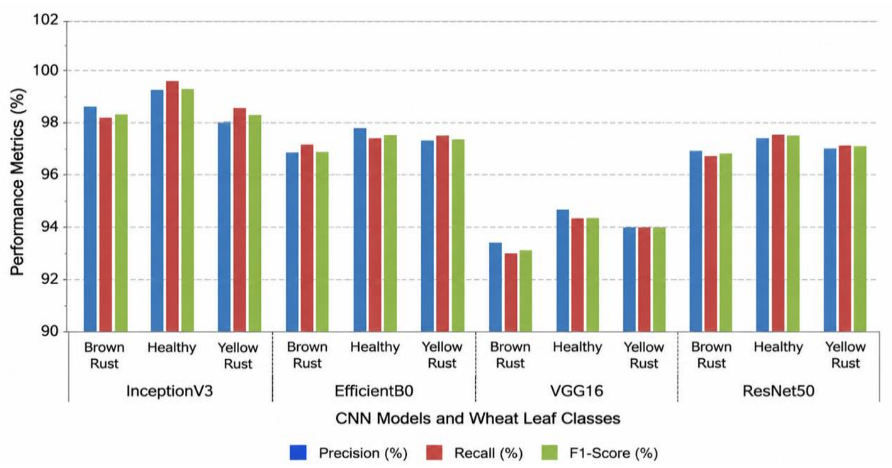
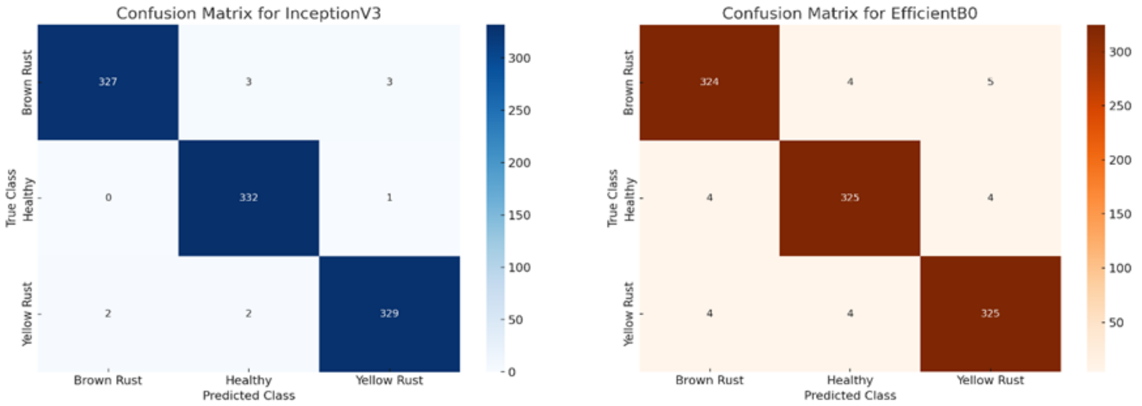


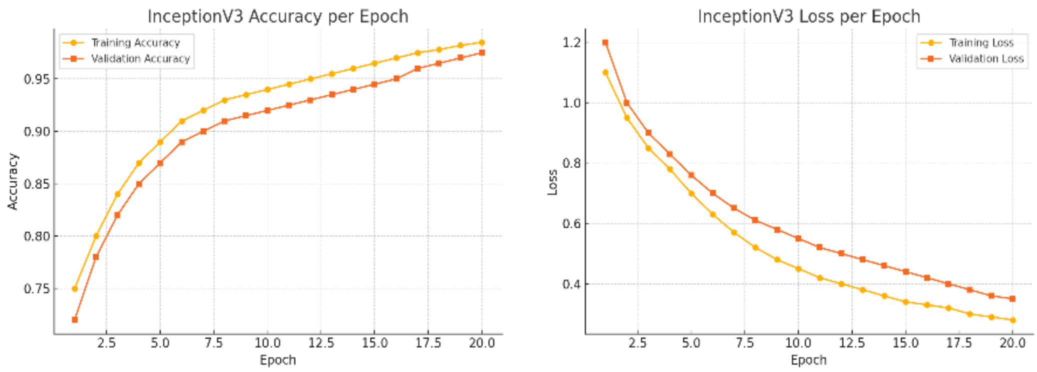
Figure 8. Graphical performance of proposed models for wheat leaf disease classification

Figure 9 presents the confusion matrix obtained for the proposed wheat rust disease classification using Inceptionv3 (Figure 9(a)) and EfficientNet-B0 (Figure 9(b)) models. The rows represent the actual class labels, while the columns denote the predicted class labels. Figure 10 depicts the training and validation accuracy (Figure 10(a)) and loss (Figure 10(b)) curves of the proposed InceptionV3 model over successive epochs. Figure 11 illustrates the training and validation accuracy (Figure 11(a)) and loss (Figure 11(b)) curves of the EfficientNet-B0 model across multiple training epochs. Table 3 compares the performance of the proposed InceptionV3 + ViT model with several recently reported DL approaches for plant disease classification. The proposed model achieves an accuracy of 98.7%, outperforming MnasNet-SimAM, DenseNet, CNN, and EfficientNet-based methods.



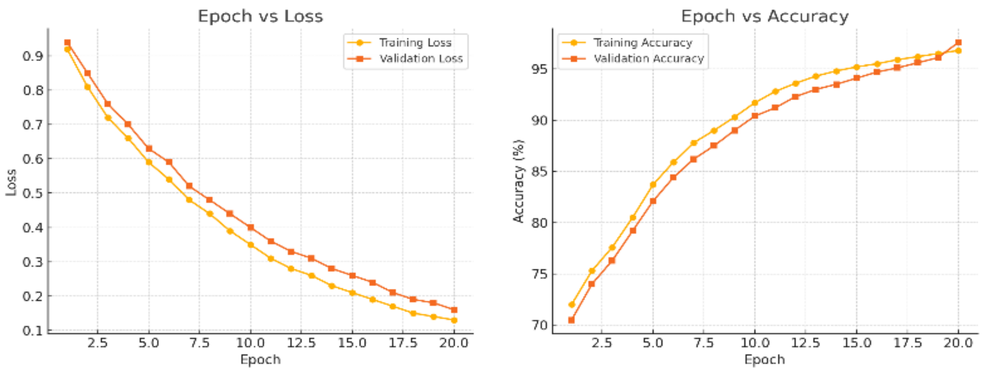
(a) (b)

Figure 9. Confusion matrix for wheat disease classification using (a) Inceptionv3 and (b) EfficientNet-B0



(a) (b)

Figure 10. InceptionV3 curve of (a) accuracy and (b) loss



(a) (b)

Figure 11. Accuracy and loss curve of EfficientNet-B0

Table 3. Comparison of proposed models with existing methods

Reference	Dataset	Class	Model	Accuracy (%)
Wen *et al.* [17]	Wheat fungi diseases	6	MnasNet-SimAM	95.14
Ahmad *et al.* [9]	Cotton leaf images	3	DenseNet	91.66
Khamparia *et al.* [20]	Multi-crop leaf	5	CNN	97.5
Alharbi *et al.* [12]	Wheat disease	18	EfficientNet	93.19
Proposed model	Wheat disease detection	3	InceptionV3 + ViT	98.7

5. CONCLUSION

In this work handled the fundamental issue of early and precise detection of wheat crop disease by employing DL and visual transformer models. Through the use of the power of CNNs in local feature extraction and ViTs in capturing global contextual relationships, our method significantly enhances performance in disease detection in challenging field scenarios. Experimental results on a wheat disease detection image dataset show the superior performance of InceptionV3, with a maximum accuracy of 98.7%, followed by EfficientNet-B0, ResNet-50, and VGG-16. These results show the future of hybrid DL with the ViT model to enhance accuracy in the diagnosis of wheat leaf diseases. As a future scope, the proposed models can be extended to different crop disease datasets for comprehensive performance evaluation and integrated into real-time agricultural applications to facilitate early disease detection and support farmers.

FUNDING INFORMATION

This research received no specific grant from any funding agency.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
P. Ashwini	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
D. Mythili		✓				✓			✓					
P. Jyothi	✓	✓	✓	✓				✓		✓				
K. Swathi		✓				✓		✓	✓	✓	✓	✓		
G. Rama Krishna	✓	✓	✓		✓				✓	✓				
M. Sowmya	✓		✓	✓			✓			✓	✓		✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare. All co-authors have seen and agree with the contents of the manuscript.

DATA AVAILABILITY

The authors confirm that the data supporting the findings of this study are available within the article.

REFERENCES

[1] Z. Chen, X. Zhang, T. Lv, and G. Wang, "Using a hybrid convolutional neural network with a transformer model for tomato leaf disease detection," *Agronomy*, vol. 14, no. 4, 2024, doi: 10.3390/agronomy14040673.

[2] P. Kaur, A. M. Mishra, N. Goyal, S. K. Gupta, A. Shankar, and W. Viriyasitavat, "A novel hybrid CNN methodology for automated leaf disease detection and classification," *Expert Systems*, vol. 41, no. 7, 2024, doi: 10.1111/exsy.13543.

[3] S. Nigam *et al.*, "Deep transfer learning model for disease identification in wheat crop," *Ecological Informatics*, vol. 75, 2023, doi: 10.1016/j.ecoinf.2023.102068.

[4] O. Jouini, M. O. E. Aoueileyine, K. Sethom, and A. Yazidi, "Wheat leaf disease detection: a lightweight approach with shallow CNN based feature refinement," *AgriEngineering*, vol. 6, no. 3, pp. 2001–2022, 2024, doi: 10.3390/agriengineering6030117.

[5] R. Mao *et al.*, "DAE-mask: a novel deep-learning-based automatic detection model for in-field wheat diseases," *Precision Agriculture*, vol. 25, no. 2, pp. 785–810, 2024, doi: 10.1007/s11119-023-10093-x.

[6] S. M. N. Raja and A. S. Nargunam, "An optimal feature selection-based deep learning approach for wheat disease identification," *Multimedia Tools and Applications*, vol. 84, no. 14, pp. 12855–12876, 2025, doi: 10.1007/s11042-024-19453-9.

[7] M. Long, M. Hartley, R. J. Morris, and J. K. M. Brown, "Classification of wheat diseases using deep learning networks with field and glasshouse images," *Plant Pathology*, vol. 72, no. 3, pp. 536–547, 2023, doi: 10.1111/ppa.13684.

[8] M. H. Hossen, M. Mohibullah, C. S. Muzammel, T. Ahmed, S. Acharjee, and M. B. Panna, "Wheat diseases detection and classification using convolutional neural network (CNN)," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 11, pp. 719–726, 2022, doi: 10.14569/IJACSA.2022.0131183.

[9] M. Ahmad, F. Ullah, A. R. A. Hamza, M. Usman, M. Imran, I. Batyrshin, A. Gelbukh, and G. Sidorov, "Cotton leaf disease detection using vision transformers: a deep learning approach," *African Journal of Biomedical Research*, pp. 5760–5769, 2024, doi: 10.53555/ajbr.v27i3s.3421.

[10] S. Cheng *et al.*, "A high performance wheat disease detection based on position information," *Plants*, vol. 12, no. 5, 2023, doi: 10.3390/plants12051191.

[11] S. Nigam, R. Jain, V. K. Singh, S. Marwaha, A. Arora, and S. Jain, "EfficientNet architecture and attention mechanism-based wheat disease identification model," *Procedia Computer Science*, vol. 235, pp. 383–393, 2024, doi: 10.1016/j.procs.2024.04.038.

[12] A. Alharbi, M. U. G. Khan, and B. Tayyaba, "Wheat disease classification using continual learning," *IEEE Access*, vol. 11, pp. 90016–90026, 2023, doi: 10.1109/ACCESS.2023.3304358.

[13] D. S. Joseph, P. M. Pawar, and K. Chakradeo, "Real-time plant disease dataset development and detection of plant disease using deep learning," *IEEE Access*, vol. 12, pp. 16310–16333, 2024, doi: 10.1109/ACCESS.2024.3358333.

[14] E. A. Nigus, G. B. Taye, D. W. Girmaw, and A. O. Salau, "Development of a model for detection and grading of stem rust in wheat using deep learning," *Multimedia Tools and Applications*, vol. 83, no. 16, pp. 47649–47676, 2024, doi: 10.1007/s11042-023-17434-y.

[15] Z. Zhang *et al.*, "Enhancing wheat disease diagnosis in a greenhouse using image deep features and parallel feature fusion," *Frontiers in Plant Science*, vol. 13, 2022, doi: 10.3389/fpls.2022.834447.

[16] X. Fang, T. Zhen, and Z. Li, "Lightweight multiscale CNN model for wheat disease detection," *Applied Sciences*, vol. 13, no. 9, 2023, doi: 10.3390/app13095801.

[17] X. Wen *et al.*, "MnasNet-SimAM: an improved deep learning model for the identification of common wheat diseases in complex real-field environments," *Plants*, vol. 13, no. 16, 2024, doi: 10.3390/plants13162334.

[18] J. Nagpal, L. Goel, and P. S. Shekhawat, "A hybrid deep learning neural network for early plant disease diagnosis using a real-world wheat–barley vision dataset: challenges and solutions," *International Journal of Data Science and Analytics*, vol. 20, no. 3, pp. 2141–2162, 2025, doi: 10.1007/s41060-024-00578-x.

[19] S. Yu, L. Xie, and L. Dai, "ST-CFI: Swin Transformer with convolutional feature interactions for identifying plant diseases," *Scientific Reports*, vol. 15, 2025, doi: 10.1038/s41598-025-08673-0.

[20] A. Khamparia, G. Saini, D. Gupta, A. Khanna, S. Tiwari, and V. H. C. de Albuquerque, "Seasonal crops disease prediction and classification using deep convolutional encoder network," *Circuits, Systems, and Signal Processing*, vol. 39, no. 2, pp. 818–836, 2020, doi: 10.1007/s00034-019-01041-0.

[21] P. Ashwini, N. Suguna, and N. Vadivelan, "Improved bald eagle search optimization with entropy-based deep feature fusion model for breast cancer diagnosis on digital mammograms," *Multimedia Tools and Applications*, vol. 83, no. 14, pp. 41785–41803, 2024, doi: 10.1007/s11042-023-17144-5.

[22] J. Feng, W. E. Ong, W. C. Teh, and R. Zhang, "Enhanced crop disease detection with EfficientNet convolutional group-wise transformer," *IEEE Access*, vol. 12, pp. 44147–44162, 2024, doi: 10.1109/ACCESS.2024.3379303.

[23] P. Ashwini, D. Mythili, P. Nagarani, and P. Jyothi, "Analysis of SARS-COV-2 CT-scan images using deep convolutional neural network models," in *2025 3rd International Conference on Intelligent Cyber Physical Systems and Internet of Things*, 2025, pp. 981–986. doi: 10.1109/icoici65217.2025.11253242.

[24] G. Chetan, K. Lavanya, S. B. S. S. S. V. Krishna, and D. V. Lakshmi, "Plant disease identification using MobileNetV2 and InceptionV3," in *Proceedings of the 2024 3rd Edition of IEEE Delhi Section Flagship Conference, DELCON 2024*, 2024. doi: 10.1109/DELCON64804.2024.10867205.

[25] S. Depuru, K. Amala, P. Supriya, A. Basi Reddy, and R. S. Gireesh, "VGG-16 technique to reduce the global food crises and enhance the crop yields: deep learning approaches," in *Proceedings of the 3rd International Conference on Applied Artificial Intelligence and Computing, ICAAIC 2024*, 2024, pp. 596–599. doi: 10.1109/ICAAIC60222.2024.10575562.

BIOGRAPHIES OF AUTHORS

P. Ashwini is working as an assistant professor in the Department of Computer Science and Engineering, Sreenidhi University, India. She received her B.Tech. and M.Tech. from JNTUH, Hyderabad. She has 15+ years of teaching experience in an engineering college. Her research interests are image processing and deep learning. She can be contacted at email: ashwini_cse@mvsrec.edu.in or ashwinireddy90@gmail.com.

D. Mythili is working as assistant professor in the Department of Computer Science and Engineering, Vasavi College of Engineering, India. She received her B.Tech. and M.Tech. from JNTUH, Hyderabad. She has 5+ years of teaching experience in an engineering college. Her research interests are IoT and machine learning. She can be contacted at email: prabhupatil84@gmail.com.

P. Jyothi is working as an assistant professor in the Department of Computer Science and Engineering, Teegala Krishna Reddy Engineering College, India. She received her Ph.D. from VIT University, B.Tech. and M.Tech. from JNTUH, Hyderabad. She has 9+ years of teaching experience in an engineering college. Her research interests are image processing and deep learning. She can be contacted at email: jyothipeta@tkrec.ac.in.

K. Swathi is working as an assistant professor in the Department of Computer Science and Engineering, CVR College of Engineering, India. She received her B.Tech. and M.Tech. from JNTUH, Hyderabad. She has 9+ years of teaching experience in an engineering college. Her research interests are cyber security deep learning. She can be contacted at email: kathiswathi@cvr.ac.in.

G. Rama Krishna is working as an assistant professor in the Department of Computer Science and Engineering, YJR DMS College of Engineering, India. He received his B.Tech. (CSE) degree from Acharya Nagarjuna University, Guntur, and M.Tech. (CSE) from JNTUK, Kakinada. He has 9+ years of teaching experience in an engineering college. He can be contacted at email: rkguduru.cse@yjrdmsce.ac.in.

M. Sowmya, is working as an assistant professor in the Department of Computer Science and Engineering, Stanley College of Engineering and Technology for Women, India. She completed her M.Tech, (SE) from JNTU-HYD and completed her Ph.D. in CSE at Annamalai University. She has 17 years of teaching experience. She is good at programming. She is certified by MTA (Microsoft certification) on cloud computing and certified by ICT Academy and DELL on data science and big data analytics. She organized many workshops and FDPs. Published papers in international journals and conferences. Guided more than 10 M.Tech. projects. Her research areas are image processing and deep learning. She can be contacted at email: msoumya@stanley.edu.in.