International Journal of Basic and Applied Sciences, 14 (8) (2025) 399-419

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/29znr475
Research paper

Wheatleafnet: A Novel Explainable Ai-Based Hybrid Deep Learning Approach for Wheat Leaf Disease Detection and Classification

Chatla Subbarayudu *, Mohan Kubendiran

School of Computer Science and Engineering, Vellore Institute of Technology, Vellore 632014, India
**Corresponding author E-mail: mohan.k@vit.ac.in*

Received: October 13, 2025, Accepted: November 21, 2025, Published: December 17, 2025

Abstract

Agriculture is the second-largest sector in India, contributing 20.2% to the national GDP, with wheat cultivation playing a crucial role in global food security. However, wheat is highly susceptible to fungal, bacterial, viral, and nutrient deficiency-related diseases that can severely reduce yield. Early and accurate disease detection is therefore vital for sustainable crop management. In this study, we introduce WheatLeafNet, a deep learning framework for multi-class wheat leaf disease classification. The dataset comprises 6,247 images across 24 categories (23 disease types and healthy samples). To enhance model robustness and mitigate overfitting, data augmentation and regularization strategies were employed, alongside ablation studies on their effectiveness. We evaluated a custom CNN and transfer learning architectures (MobileNetV2, ResNet50, EfficientNet-B0) under four optimizers (SGD, RMSProp, Adam, and AdaGrad), with stratified 5-fold cross-validation ensuring reliable assessment. Among the tested models, MobileNetV2 with the AdaGrad optimizer achieved the best performance, reaching an accuracy of 97.91% without augmentation and 97.84% with augmentation. Comprehensive evaluation metrics, including per-class precision, recall, F1-score, ROC-AUC, Expected Calibration Error (ECE), and Grad-CAM visualizations, confirm the reliability and interpretability of the framework. The integration of augmentation, optimization, and explainability strengthens the model's generalizability for real-world applications. The proposed system offers an effective and scalable solution for early disease identification in wheat, enabling timely interventions, reducing crop loss, and promoting sustainable agricultural practices. Future work will incorporate segmentation-driven severity estimation and multi-label detection to enable real-world deployment.

Keywords: *Wheat Leaf Disease; Convolutional Neural Networks (CNN); Computer Vision; Data Augmentation; Deep Learning; Image Processing.*

1. Introduction

Wheat serves as a fundamental staple food crop, contributing approximately 37% of global per-capita energy requirements according to international food and agricultural statistics [1,2]. However, prolonged periods of elevated temperatures and exceptional climatic conditions significantly constrain wheat production capacity [3]. Given its rich composition of proteins, dietary fiber, and essential vitamins, wheat remains a critical nutritional source for populations worldwide [4,5]. Current estimates from the United States Department of Agriculture indicate that global wheat production reaches approximately 780 million metric tons annually [6]. Nevertheless, wheat crop productivity faces substantial challenges from multiple sources, including environmental stressors and pathogenic diseases caused by bacterial, fungal, and viral agents. Additional constraints arise from pest management requirements, insect infestations, and anthropogenic pesticide applications, all of which collectively diminish wheat yield potential and exacerbate global food security concerns beyond historical levels [7]. Consequently, effective disease management strategies are essential to address escalating worldwide food demand. Traditional diagnostic approaches relying on manual assessment by agricultural experts are inherently time-intensive, labour-demanding, and impractical for large-scale monitoring applications [8]. Therefore, automated diagnostic systems represent a critical necessity [9]. Advancements in IoT and ICT have enabled the development of smart agricultural systems that leverage artificial intelligence and computer vision to improve crop disease detection [10].

Computer vision and digital image processing technologies offer substantial potential for identifying and interpreting complex patterns and characteristics that remain challenging to detect through conventional human visual examination. Deep learning methodologies, characterized by their sophisticated learning and pattern recognition capabilities, have demonstrated widespread applicability in disease detection and classification across various crop species, including maize, rice, and other cereal grains. Advanced feature extraction techniques enable the identification of meaningful patterns and distinctive features from digital imagery [11-13]. Contemporary deep learning architectures, including Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Recurrent Neural Networks (RNNs), have been extensively employed for crop disease applications, achieving remarkable accuracy and precision levels. Among these approaches, CNN architectures consistently demonstrate superior performance with significantly reduced error rates for accurate disease

diagnosis and classification tasks [14], [15]. Building upon these successes, the field is now emphasizing model transparency and robustness. The integration of Explainable AI (XAI) techniques, such as Grad-CAM and reliability analysis, is becoming essential to build trust and ensure the safe deployment of these models [16]. Furthermore, recent advancements also highlight the superior feature extraction capabilities of Vision Transformers (ViT) in complex agricultural vision tasks [17]. The two-stage approach mitigates overfitting while ensuring robust large-scale performance and enabling multi-disease classification on individual leaves [18].

A) Leaf/Brown Rust. B) Stripe/Yellow Rust. C) Common Root Rot.

D) Root Knot Nematode E) Wheat Seed-Gall Nematode F) Wheat Spindle Mosaic

G) Sulphur Deficiency H) Copper Deficiency I) Iron Deficiency

J) Basal Glume Rot K) Spike Blight L) Healthy Wheat Leaf.

Fig. 1: Presents Sample Wheat Leaf Images from Various Disease Categories. Panels (A–C) Depict Fungal Diseases, (D–F) Show Viral Diseases, (G–I) Illustrate Nutrient Deficiency Symptoms, (J–K) Represent Bacterial Diseases, and Panel (L) Corresponds to A Healthy Wheat Leaf.

Figure 1. Images of wheat leaf illness sample.

1.1. Motivation

Though these advancements cater to the need for automation, they require physical components such as drones, WSNs, cameras, sensors, and storage components. Identifying appropriate patterns and features from the data observed using IoT subject matter experts is essential for disease detection. In addition, datasets of the natural fields have to be processed to minimize errors. The challenges associated with the datasets are complex backgrounds, varying illumination conditions, confused pixels in the disease-affected regions, multiple diseases on a single leaf, the same attributes on different diseases, various attributes of varying diseases, and distinct traits of the same disease about the shape and stage. Due to these factors, it is highly essential to develop a wheat crop disease detection and prediction model that would protect the livelihood of the farmers and prevent crop loss.

1.2. Major contributions

The proposed work involves the detection of wheat crop disease based on deep convolutional neural networks. The major contributions are listed as follows:

- To address overfitting and enhance generalization, a set of data augmentation techniques was integrated into the training pipeline. A hybrid deep learning framework was developed, with the MobileNetV2 architecture combined with the AdaGrad optimizer emerging as the optimal configuration, achieving superior classification performance.
- The proposed work utilized datasets encompassing 24 classes, including fungal, bacterial, viral, and nutrient deficiency-related infections, as well as healthy wheat samples. This wide coverage ensures the model's robustness in handling diverse pathological variations.
- Experimental evaluations were conducted in a Python-based Google Colab environment with GPU acceleration. The proposed framework achieved a peak classification accuracy of 97.84%, with a minimal error rate of approximately 0.14%, thereby confirming its predictive reliability and consistency.
- Comparative assessments were carried out against established models and transfer learning architectures such as ResNet50 and EfficientNet-B0, under four optimization strategies (SGD, RMSProp, Adam, and AdaGrad). The use of stratified 5-fold cross-validation ensured unbiased evaluation. Results demonstrated that WheatLeafNet consistently outperformed benchmark approaches while alleviating challenges of class imbalance and overfitting.

1.3. Organization of the paper

The remainder of this paper is structured as follows: Section 1 presents the introduction and motivation, highlighting the key contributions related to wheat leaf disease classification. Section 2 reviews and analyzes existing literature on crop disease detection and classification to provide a foundation for the study. Section 3, which provides the proposed work along with the datasets, methodology, and techniques used; Section 4, which provides the experimentation and ablation study; Section 5, which provides the performance analysis; Section 6, which provides the results analysis and discussion; Section 7 provides the conclusion and prospective research initiatives. For reference, the list of acronyms is provided in Table 1.

Table 1: List of Acronyms

Acronym	Description
AI	Artificial Intelligence
ANN	Artificial Neural Network
ASFF	Adaptively Spatial Feature Fusion
CGIAR	Consultative Group for International Agricultural Research
CNN	Convolutional Neural Network
CVT	Computer Vision Technology
DWT	Discrete Wavelet Transform
ECA	Efficient Channel Attention
EML	Elliptic Meter Learning
FFNN	Feed Forward Neural Network
FHB	Fusarium Head Blight
GLCM	Grey Level Co-occurrence Matrix
ICT	Information and Communication Technology
IoT	Internet of Things
KNN	K-Nearest Neighbor
LBP	Local Binary Pattern
LSTM	Long Short-Term Memory Networks
LWDCD	Large Wheat Disease Classification Dataset
PANet	Path Aggregation Network
PCA	Principle Component Analysis
RCAB	Residual Channel Attention Block
ReLu	Rectified Linear Unit
ResNet	Residual Network
RNN	Recurrent Neural Network
ROC-AUC	Receiver Operating Characteristic Curve; Area under the curve
SVM	Support Vector Machine
UAVs	Unmanned Aerial Vehicles
USDA	United States Department of Agriculture
VGG	Visual Geometry Group
WOA	Whale Optimization Algorithm
WSN	Wireless Sensor Networks

2. Related Works

A considerable body of literature has emerged in recent years focusing on crop disease identification and diagnosis. Different techniques have evolved to categorize various kinds of diseases relating to machine learning, artificial intelligence, and computer vision technologies. An outline of food crop-related practices, as well as their advantages and disadvantages, is given in this section.

Ashraf et al. [19] have come up with a prediction framework for wheat crop disease. Their work addressed the problem of timely detection of disease to improve production by using a lightweight, modified CNN architecture. Their proposed model utilized three convolutional layers, SoftMax layers, and two flattened layers for the efficient prediction of wheat crop disease. Their model has been trained using the dataset obtained from the Azad Kashmir (Pakistan) province. Performance analysis showed that their proposed model achieved an accuracy of 93%. However, their model is computationally intensive and needs a large quantity of training data. For the plant leaf disease dataset, Arun et al. [20] developed a deep residual convolutional neural network. The classification accuracy of their model was 99.58%. However, the major limitation of their approach is that ResNet197 is computationally intensive. Long et al. [21] have come up with a classification of wheat disease using the deep learning network model CerealConv. The dataset used in their proposed study is divided into five categories, namely four foliar diseases and healthy plants, such as brown rust, yellow rust, septoria leaf blotch, and powdery mildew. Their

classification has achieved an accuracy of 97.5%. The advantage of this model is that it highlights the potential pitfalls of using datasets that are unbalanced in terms of varying conditions and are carefully validated during classification. Aboneh et al. [22] proposed a computer system for the categorization and diagnosis of wheat diseases such as leaf rust, stem rust, and stripe rust.

Their proposed work utilized deep learning models such as Inceptionv3, Resnet50, and VGG16/19. Their proposed model has achieved an accuracy of 99.38%. Their model has not yet been mechanized for actual real-time application. Nigam et al. [23] have come up with a deep transfer learning model for the disease detection of wheat crops. Wheat rusts have been identified using a deep learning model based on the EfficientNet architecture. An Efficient Net B4 model has achieved an accuracy of 99.35%. However, their model has not considered several levels of wheat disease. Rangarajan et al. [24] have come up with Fusarium head blight (FHB) disease detection for wheat crops using hyperspectral data.

Their proposed model utilized a pre-trained lightweight CNN model called DarkNet 19, achieving an accuracy of 100%. The major limitation is that it has not considered the spatial data of the hyperspectral data and has not utilized larger datasets. Zhang et al. [25] have improved feature-scale consistency with a 98.5% accuracy rate by developing a lightweight Yolov4 network based on efficient channel attention (ECA) paired with adaptive spatial feature fusion (ASFF). However, under varying light intensities, different field types of equipment are supposed to be investigated to understand the utilized model's performance. Schirrmann et al. [26] recommended a ResNet-18 model-based stripe rust classifier for the wheat crop. Their methodology involves the Adam optimizer and has achieved a high accuracy rate at both the patch level and image level. However, it suffers from poor resolution of dataset images, and optimization is needed to achieve even higher accuracy. Singh et al. [27] have come up with a yellow rust severity assessment machine-learning model using visible and thermal imaging techniques. The cubist model appears to be the best at detecting rust disease effectively. However, their proposed model is very slow in terms of computational complexity. Alshammari et al. [28] proposed a disease detection model for olive leaf disease through an optimized deep-learning approach. Their methodology utilized optimized ANN and WOA for feature selection. Feed Forward Neural Network (FFNN) has been used for classification.

The model they proposed has a 98.9% classification accuracy. Still, the performance has to be improved when compared with conventional approaches. Raouhi et al. [29] have come up with DeepCNN-based optimization techniques for olive leaf disease detection. Their technique achieved a reasonable level of accuracy. However, the model is supposed to be trained for a reasonable number of epochs, which diminishes the problem of overfitting. Jiang et al. [30] have come up with a multi-task deep transfer learning technique for the identification of illnesses in wheat and rice crops. They used ImageNet for transfer learning and the VGG16 model in their technique. Their model has come up with an accuracy of 97.22% in the case of rice and 98.75% for wheat crops. Their work model has to be tested for larger datasets and other diseases on crops. Kumar et al. [31] have developed a cutting-edge hybrid segmentation method for the identification of wheat rust illnesses. A panoptic segmentation-based FERSPNET-50 has been utilized using the GNet model. A faster region-based convolutional neural network model (FRCNN) has been utilized for the detection of diseases in wheat leaves and stem patches. A pyramid scene parsing network is utilized to classify the severity and classification using a patch with a high precision of 97%. A real-time mobile application can be utilized for scanning various levels of infection in the leaves shortly.

Alharbi et al. [32] have come up with a continuous learning technique for classifying wheat disease. Their methodology utilized EfficientNet as a backbone architecture under the CGIAR dataset. Their model has classified the disease accurately at 93.19%. However, the performance of their proposed technique has to be optimized to increase it. Pan et al. [33] have come up with a PSPNet semantic segmentation model for the recognition of wheat rust disease using unmanned aerial vehicles (UAVs). Their approach has been to categorize barren soil, yellow wheat rust, and healthy leaves in small-scale UAV photos. Their model achieved an accuracy of 94%. UAV RGB and hyperspectral UAV images can be considered to assess their performance. Xu et al. [34] have advanced the detection method for wheat leaf diseases by utilizing a deep learning approach that integrates neural networks (CNN), elliptic metric learning (EML) feedback blocks (FB), and residual channel attention blocks (RCAB). Their model achieved a 99.50% accuracy in classifying plant diseases on open-source databases, like CGIAR, LWDCD2020, and plant pathology. However, their method overlooks leaf diseases that can impact the crop. Table 2 outlines the evaluation of existing approaches addressing wheat disease identification. Tegegne et al. [35] evaluated multiple CNN models for classifying wheat rust diseases and identified MobileNetV3 with Softmax as the best-performing architecture. Using augmented field and Kaggle images, the model achieved 97.7% accuracy in classifying brown rust, stem rust, yellow rust, and healthy plants. The lightweight design makes it suitable for mobile agriculture applications. Although limited to classification and constrained by dataset size, the work provides a robust foundation for advanced diagnostic systems and future segmentation-based models. Liu et al. [36] presented MSDP-SAM2-UNet, a multi-scale and dual-path segmentation model for wheat leaf blight and rust. By integrating enhanced receptive-field extraction and CARAFE-based dual-path upsampling, the model achieves superior lesion boundary preservation and improves SAM2-UNet segmentation by up to 4.9%, reaching 94.02% pixel accuracy. Their approach effectively segments small and irregular wheat lesions. While the model is limited by dataset diversity and lacks severity estimation, it establishes a strong direction for advanced wheat disease segmentation and future lightweight deployment. Dong et al. [37] introduced the SC-ConvNeXt model, combining ConvNeXt-T, a self-supervised SimCLR framework, and an improved CBAM attention mechanism for wheat disease recognition. The model effectively handles complex field backgrounds and reduces dependence on large labelled datasets. Using a four-class field dataset from the Smart Agriculture platform, it achieved an 88.05% accuracy, higher than common models like ResNet50, MobileNetV3, and ConvNeXt-T. Although the model is computationally heavy and involves a two-stage training process, it demonstrates strong potential for real-world wheat disease identification and could be further improved with model compression and more advanced self-supervised techniques. Moon et al. [38] proposed a machine learning framework for detecting Septoria Tritici Blotch in wheat leaves. The method uses preprocessing, color-based segmentation, and GLCM texture features, followed by several classifiers, including SVM, k-NN, Naïve Bayes, and Random Forest. The Random Forest model delivered the best performance with 98.9% accuracy, reliably distinguishing healthy and infected leaves. A dataset collected from BWMRI and public sources was used, containing two classes: healthy and Septoria-infected leaves. Although effective, the system is limited to binary classification, sensitive to lighting and occlusion, and cannot recognise multiple diseases. Future work may integrate deep learning, larger datasets, and multi-disease detection to enhance real-world applicability.

Table 2: Classification of Works Related to Disease Diagnosis on Wheat Crop

References	Problem Addressed	Technique Used	Benefits	Limitations
Ashraf et al. [19]	Prediction of Wheat Crop Disease	CNN	Accuracy 93%	The model is heavy and needs a large dataset.
Arun et al. [20]	Plant Leaf Disease Detection	Deep Residual CNN	Classification Accuracy 99.58%	The limitation of ResNet197 is its high computational complexity.
Long et al. [21]	Classifying Wheat Diseases	Inceptionv3, ResNet50, VGG16/19	Classification 97.05%	Their study has yet to account for multiple infections.

Aboneh et al. [22]	Detection of Wheat Diseases	VGG16, Inception V3, MobileNet, Xception	Classification Accuracy 99.38%	It is necessary to assess and evaluate different activation functions.
Nigam et al. [23]	Identification of Wheat Crop Disease	VGG19, ResNet152, DenseNet169, InceptionNetV3, MobileNetV2	The model testing accuracy of 99.35%.	Despite mobile apps identifying cereal diseases, several wheat disease stages remain unaddressed.
Rangarajan et al. [24]	Early detection of Fusarium Head Blight (FHB)	DarkNet 19, EfficientNet B0, GoogleNet, MobileNetV2, ResNet18 &50, ShuffleNet, etc	Accuracy 100%	Larger Datasets can be evaluated.
Zhang et al. [25]	Detection of Wheat Scab Fungus	ECA-Net, PANet, Yolov5	Accuracy 98.5%	Field equipment with different resolutions and lighting conditions should be examined.
Schirrmann et al. [26]	Early Detection of Stripe Rust	ResNet-18	Accuracy 90%	Model optimization is required for improved performance.
Alshammari et al. [27]	Identification of Olive leaf disease	FFNN-ANN & WOA	Accuracy 98.4%	Using fewer plants per category or different crops is yet to be considered.
Singh et.al [28]	Severity Assessment	Partial Least Squares Regression of Image Indices	-	Slow in terms of Computational complexity.
Raouhi et al. [29]	Classification of Olive Diseases	CNN	Accuracy 92.59%	Accuracy, flexibility, and image segmentation still require enhancement
Jiang et al. [30]	Recognition of Rice Leaf and Wheat Leaf Diseases	VGG16, ImageNet	Accuracy 97.22% and 98.75%	Large datasets remain untested, limiting the approach to small datasets.
Kumar et.al [31]	Identification of Wheat Rust Disease	FERSPNET-50; Faster Region-based CNN	Precision 97%	Accuracy still needs improvement using optimization algorithms.
Alharbi et al. [32]	Classification of Wheat Diseases	EfficientNet	Accuracy 93.19% and 98.5%	The model can be adapted for use with other crops and plants.
Pan et al. [33]	Detection of Wheat Rust	PSPNet; Ensemble learning	Accuracy 94%	The model faces validation loss from low-resolution images.
Xu et al. [34]	Identification of Wheat Leaf Disease	RFE-CNN	Accuracy 99.50%	Multi-class classification of diseases by stage and type is yet to be explored.
Tegegne et al. [35]	Classification of Wheat Rust Disease	CNN- MobileNetV3	Accuracy 97.7%	No segmentation of infected regions.
Liu et.al [36]	Wheat Leaf Disease Segmentation Based on CNN Models	MSDP-SAM2-UNet	PA 94.02%	Computational complexity; segmentation focus only.
Dong et al [37]	Wheat disease recognition using CNNs	ConvNeXt-T Backbone	Accuracy 88.05%	The model has a large parameter size, increasing storage needs.
Moon et al.[38]	Wheat leaf disease identification	NB, K-NN, SVM, and Random Forest	98.9%	Focuses on a single disease type and includes only limited validation.

From Table 2, it is apparent that most of the existing works focused on the early detection, classification, and identification of wheat diseases. Especially, the works rely upon wheat stripe rust and fungal disease identification. The techniques utilized for disease detection achieve accuracy, which is supposed to be improved. Ultimately, the multi-class classification of diseases across different stages remains underexplored. Moreover, most studies have relied on relatively small datasets, and their performance can be further enhanced through advanced optimization strategies. Therefore, there is a crucial need to achieve robust and optimized performance using diverse deep learning techniques. Motivated by these limitations, this study introduces a novel hybrid model for wheat crop disease detection, integrating deep learning with optimization methods to improve accuracy, generalization, and reliability.

3. Proposed Methodology

The proposed framework, WheatLeafNet, is designed to provide an accurate and robust classification of wheat leaf diseases. The methodology integrates systematic stages, including data acquisition, preprocessing, augmentation, feature extraction, and evaluation under different training, testing, and validation strategies. By leveraging hybrid deep learning techniques and transfer learning models, the system ensures high generalizability, reduced overfitting, and interpretability through explainable AI (XAI) techniques. Fig 2 illustrates the sequential stages, including image acquisition, enhancement, preprocessing, annotation, augmentation, model training, and disease classification.

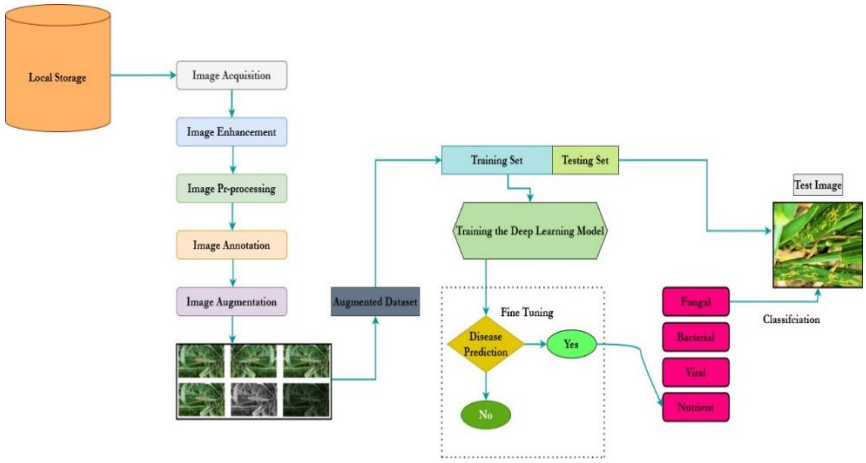

Fig. 2: Proposed Architecture for Wheat Leaf Disease Classification.

3.1. Data acquisition

This study employed a curated dataset comprising 6,247 wheat leaf images categorized into 24 classes 23 representing distinct disease conditions (fungal, bacterial, viral, and nutrient-deficiency related) and one representing healthy leaves. The dataset, primarily sourced from publicly available repositories such as Kaggle [39-40], along with supplementary images obtained from agricultural research archives and published studies [41–42], reflects diverse environmental and field conditions across multiple growing seasons. To ensure balanced representation and minimize sampling bias, stratified random sampling was applied to divide the dataset into training (80%), validation (10%), and testing (10%) subsets while maintaining proportional class distribution. This systematic preparation supports reliable performance evaluation and enhances the reproducibility of the experimental outcomes. The detailed class-wise distribution of the wheat leaf disease dataset is shown in Table 4.

3.2. Data pre-processing

To prepare the collected wheat leaf images for deep learning models, a series of pre-processing steps was applied. All images were resized to a uniform resolution of 224×224 pixels to meet the input requirements of CNN and transfer learning architectures. Gaussian filtering was used for noise reduction, while normalization scaled pixel values to the range [0,1]. Finally, label encoding was performed to enable multi-class classification. These steps ensured dataset uniformity, reduced noise, and improved robustness for subsequent classification tasks.

3.2.1. Image resizing

All input images were resized to a fixed resolution of 224×224 pixels to ensure compatibility with widely used transfer learning architectures such as VGG, ResNet, and MobileNet [43]. Formally, an image I of original size $m \times n$ is mapped to the resized image I′ as shown in Eq. (1).

$$I'(x',y') = I\left(\frac{m}{224}.x', \frac{n}{224}.y'\right), 0 \le x', y' > 224 \tag{1}$$

This transformation guarantees uniform spatial dimensions across all samples, thereby facilitating efficient batch training and consistent feature extraction.

3.2.2. Noise reduction

To suppress random noise and highlight disease lesions, Gaussian filtering was applied during preprocessing [44]. The two-dimensional Gaussian kernel is mathematically expressed as shown in Eq. (2).

$$G(x,y) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \tag{2}$$

Where σ represents the standard deviation, which controls the degree of smoothing. The filtered image is obtained by convolving the original image I with the Gaussian kernel G, as defined by Eq. (3)

$$I_{filtered}(x,y) = (I * G)(x,y) \tag{3}$$

This filtering technique reduces background noise while retaining important edge information, thereby enhancing the visibility of disease-affected regions.

3.2.3. Normalization

To improve model stability and accelerate convergence, all pixel intensities were scaled to the range [0,1]. The normalized intensity value $I_{norm}(x,y)$ at each pixel coordinate (x,y) is computed has been shown in Eq. (4).

$$I_{norm}(x,y) = \frac{I(x,y) - I_{\min}}{\mathrm{I}_{\max} - \mathrm{I}_{\min}} \tag{4}$$

Where $\mathrm{I}_{\min}$ and $\mathrm{I}_{\max}$ correspond to the minimum and maximum intensity levels in the image (0 and 255 for standard 8-bit images). This normalization process ensures uniform input distribution, reduces the risk of gradient instability, and enhances the efficiency of the optimization process.

3.2.4. Label encoding

The dataset consisted of 24 categories, including 23 disease classes and 1 healthy class. To prepare the labels for multi-class classification, categorical one-hot encoding was applied, where each class is represented as a binary vector [45]. For a given sample with true class c, the encoded label y_i is defined as:

$y_i = \begin{cases} 1, i = c \\ 0, i \neq c \end{cases}$ For i=1, 2……24, this encoding scheme allows the model to output probabilities across all classes and is particularly suitable when optimizing with categorical cross-entropy loss. Fig 3 provides the transformation of the images during image pre-processing. Fig 4.

Conceptual schematic of the GAN approach for augmenting plant leaf disease datasets. Table 3 presents the types, quantities, and class distribution of wheat diseases in the dataset used for the proposed study.

Table 3: Evaluation of the Proposed Work on Selected Classes.

S. No.	Types of Diseases	Quantity/Class	Enumeration of Wheat Leaf Disease Categories
1.	Fungal	6	Rust, Blight, Blotch, spots, rots, Powdery mildew
2.	Bacterial	4	Leaf-blight, streak, spike blight, pink seed
3.	Viral	3	Curl, Distortion, Dwarfing
4.	Nutrient Deficiency	10	Zinc, Manganese, Copper, Iron, Nitrogen, Phosphorous, Potassium, Magnesium, Sulphur, Calcium deficiencies
5	Normal	1	Healthy
	Total	24	

Table 4: Detailed Class-Wise Distribution and Data Split Configuration

Class ID	Disease Category	Disease Type	Total Images	Train (80%)	Val (10%)	Test (10%)
1	Leaf/Brown Rust	Fungal	312	250	31	31
2	Fusarium Head Blight	Fungal	228	182	23	23
3	Septoria Leaf Blotch	Fungal	276	221	28	27
4	Stripe/Yellow Rust	Fungal	298	238	30	30
5	Common Root Rot	Fungal	245	196	25	24
6	Powdery Mildew	Fungal	264	211	27	26
7	Leaf-Blight	Bacterial	287	230	29	28
8	Leaf Streak	Bacterial	234	187	24	23
9	Spike Blight	Bacterial	251	201	25	25
10	Pink Seed	Bacterial	208	166	21	21
11	Spindle Streak Mosaic	Viral	272	218	27	27
12	Streak Mosaic	Viral	256	205	26	25
13	Barley Yellow Dwarf	Viral	223	178	23	22
14	Nitrogen Deficiency	Nutrient	294	235	30	29
15	Phosphorus Deficiency	Nutrient	281	225	28	28
16	Potassium Deficiency	Nutrient	267	214	27	26
17	Sulphur Deficiency	Nutrient	253	202	26	25
18	Iron Deficiency	Nutrient	239	191	24	24
19	Copper Deficiency	Nutrient	226	181	23	22
20	Zinc Deficiency	Nutrient	214	171	22	21
21	Manganese Deficiency	Nutrient	201	161	20	20
22	Magnesium Deficiency	Nutrient	194	155	20	19
23	Calcium Deficiency	Nutrient	186	149	19	18
24	Healthy Wheat Leaf	Normal	412	330	41	41
Total	24 Classes	4 Types	6247	4,997	625	625

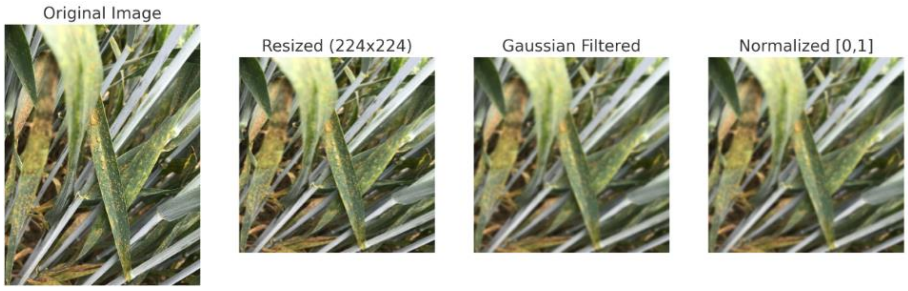

Fig. 3: Image Pre-Processing Operation.

3.3. Data augmentation

To address class imbalance and enhance the generalization capability of the deep learning model, several data augmentation strategies were applied. These transformations expand the effective size of the training dataset while simulating natural variations observed in real-world wheat field conditions, such as changes in illumination, orientation, scale, and background. This process improves the model's robustness against such variations. The augmentation techniques included:

- Geometric transformations
- Random rotations: images were rotated by angles uniformly sampled within $\pm 25^{0}$
- Flips: horizontal and vertical flips were applied with a fixed probability p.
- Zoom (scaling): random zoom operations were performed with a rescaling factor s $\in [1-\alpha, 1+\alpha]$ where α=0.2.
- Color and brightness transformations
- Brightness adjustment: pixel intensities were scaled by a factor b, where b~U (0.8,1.2).
- Contrast, hue, and saturation variations: applied as needed to simulate environmental and lighting variability.

3.3.1. Mathematical formulation

Let the training dataset be represented as X= $\{(x_i, y_i\}$, where x_i denotes an original image and y_i its corresponding class label. An augmented image x_i' is generated by applying a random transformation T drawn from the set of augmentation operations (e.g., rotation, flipping, zooming, or color adjustment):

$x_i' = T(x_i)$ where $T \sim$ T, Where T is a random transformation drawn from the set of augmentation operations (rotation, flip, zoom, color, etc.) is evaluated which has been presented in Eq (5)

$$X_{aug} = X \cup \{(x_i', y_i) | x_i' = T(x_i), T \in \mathrm{T}\} \tag{5}$$

This procedure ensures that the original label y_i is preserved for each augmented sample.

3.3.2. Balancing effect

To address class imbalance, augmentation was selectively applied to oversample minority classes. Let $N_{\max}$ denote the maximum number of images in any class:
$N_{\max} = \max_{c \in C} |\{x_i : y_i\}|$, Where C is the set of all classes. For each class c containing $n_c < N_{\max}$ samples, additional $(N_{\max} - n_c)$ augmented images were generated until all classes were approximately balanced.

3.3.3. LeafGAN: generative model for plant disease augmentation

LeafGAN, introduced by [46], is a generative adversarial network (GAN)-based framework specifically designed to augment plant leaf disease datasets. Its primary objective is to generate realistic synthetic diseased leaf images that enrich training data and mitigate class imbalance. Unlike generic GAN models, LeafGAN incorporates an attention mechanism that restricts transformations to lesion areas while preserving the natural structure of the leaf, including its shape, venation, and healthy tissue regions.
LeafGAN builds upon the CycleGAN framework with two key modifications:

1) Attention Module ensures that only disease-relevant areas of the leaf are altered, thereby preventing distortion of healthy regions and structural features.
2) Leaf Shape Preserving Loss introduces a penalty to discourage changes to the intrinsic shape and venation of the leaf, ensuring morphological consistency.

The workflow operates as follows: a healthy leaf image X_h is given as input to the generator G, which produces a synthetic diseased image $X_d' = G(X_h)$. The discriminator D then distinguishes between real diseased images X_d and generated images X_d'. To maintain invertibility, a second generator F maps diseased leaves back to healthy ones, enforcing cycle consistency.

- Adversarial Loss: encourages generated diseased images to be indistinguishable from real diseased samples: shown in Eq. (6)

$$L_{GAN}(G, D, X_h, X_d) = E_{xd \sim pdata(xd)}[\log D(x_d)] + E_{xh \sim pdata(xh)}[\log(1 - D(G(x_h)))] \tag{6}$$

- Cycle Consistency Loss: ensures that transforming healthy → diseased → healthy (and vice versa) recovers the original image, represented in Eq (7)

$$L_{cyc}(G, F) = E_{xh}[\| F(G(x_h)) - x_h \|_1] + E_{xd}[\| G(F(x_d)) - x_d \|_1] \tag{7}$$

Attention Loss: restricts modifications to lesion areas while keeping background and structure intact. If M(x) is the attention mask, the calculation has been represented in Eq. (8).

$$L_{attn}(G) = \| (1 - M(x)) \odot (G(x) - x) \| \tag{8}$$

- Total Loss: the overall objective combines adversarial, cycle consistency, and attention terms:
$L_{LeafGAN} = L_{GAN} + \lambda_{cyc} L_{cyc} + \lambda_{attn} L_{attn}$ Where λ_{cyc} λ_{attn} are the hyperparameters balancing cycle consistency and attention.

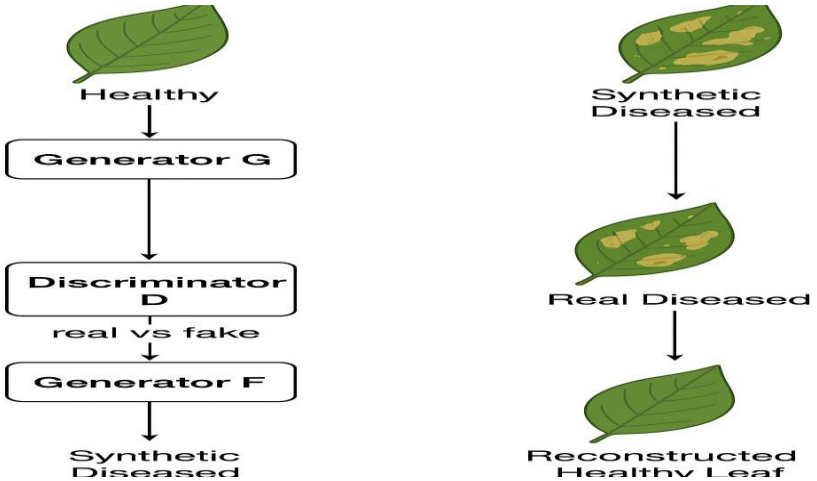

Fig. 4: Schematic Representation of the GAN-Based Framework for Plant Leaf Disease Augmentation.

The model employs two generators (G and F) and a discriminator (D). Generator G transforms healthy leaves into synthetic diseased leaves, while the discriminator differentiates between real and synthetic samples. Generator F reconstructs healthy leaves from synthetic diseased

ones, ensuring cycle consistency. This setup enables the creation of realistic diseased leaf images for data augmentation and robust model training.

3.4. Feature extraction

Feature extraction plays a critical role in enabling the model to discriminate between different plant diseases by capturing distinctive visual patterns on leaf surfaces. In this study, pre-trained deep convolutional neural network (CNN) architectures were employed, with their feature maps either fine-tuned or combined to effectively represent disease-related morphology, texture, lesions, and contextual cues. To validate that the extracted representations emphasize disease-relevant regions, attention and explainability methods such as Gradient-weighted Class Activation Mapping (Grad-CAM) were utilized. The following backbone models were adopted:

- MobileNetV2: a lightweight architecture based on depthwise separable convolutions and inverted residual connections, suitable for resource-constrained environments.
- ResNet50: incorporates residual skip connections that enable very deep networks without vanishing gradients, making it effective in capturing hierarchical features such as edges, textures, and lesion boundaries.
- EfficientNet-B0: employs compound scaling across depth, width, and resolution, offering a strong balance between accuracy and computational efficiency. Its higher resolution inputs allow the extraction of fine-grained patterns relevant to disease diagnosis. Figure 5 illustrates the CNN-based workflow for wheat disease detection, addressing 24 classes across four disease types with a three-layer architecture.

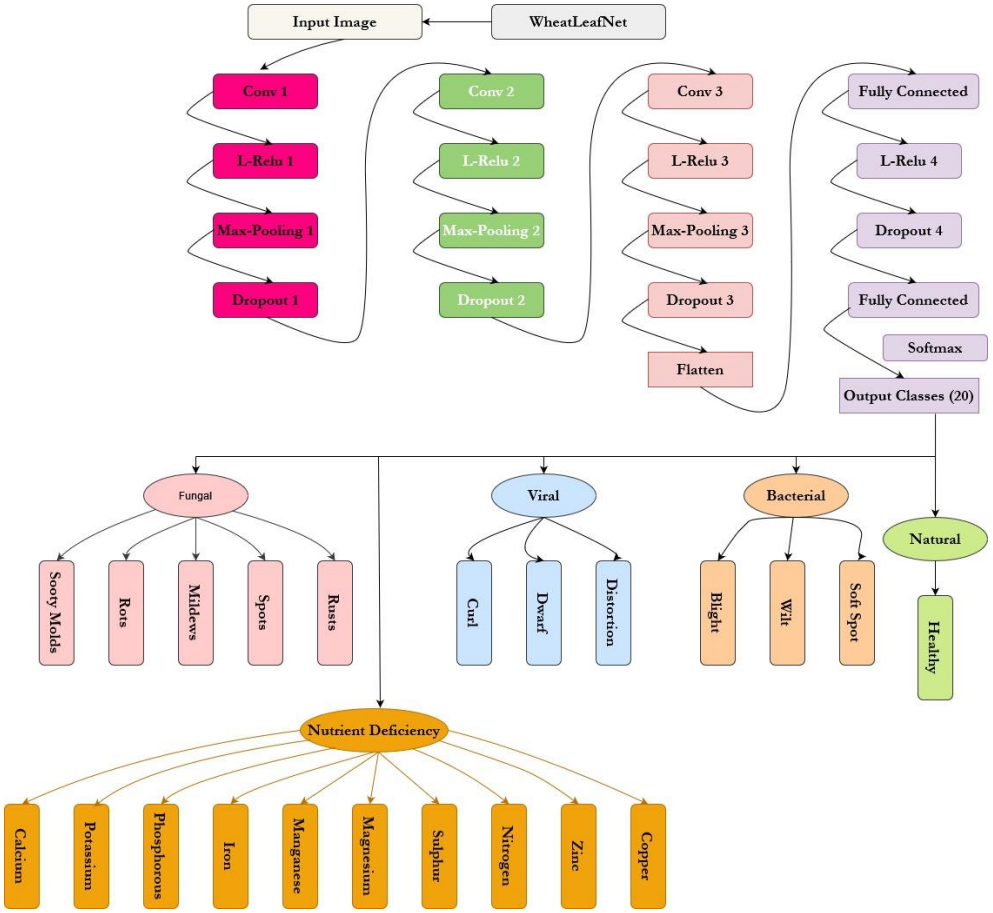

Fig. 5: The Figure Depicts the Workflow of the Proposed CNN-Based Methodology for Wheat Disease Detection and Classification, Addressing Multiple Classes Across Four Major Disease Types. The Three-Layer CNN Incorporates Convolution, Leaky ReLU activation, Max-Pooling, Dropout, and Flattening, followed by a Fully Connected Layer with Softmax to Generate Class-Specific Outputs.

Formally, let the CNN feature extractor be defined as: $f_\theta: R^{H\times W\times 3} \to R^d$, Where θ denotes the network parameters. For an input image x, the feature representation is obtained as: $h = f_\theta(x)$, Where h corresponds to the feature vector or tensor (e.g., of dimension [k×k×c] before pooling, or flattened afterwards). These features are subsequently passed to the classifier composed of fully connected layers and a softmax activation function, represented in Eq. (9)

$$\hat{y} = softmax(W \cdot Flatten(h) + b), \tag{9}$$

Where W and b represent the trainable classifier parameters. For optimization, categorical cross-entropy loss was employed for the multi-class classification task, as shown in Eq. (10)

$$L = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c} \log \hat{y}_{i,c}, \tag{10}$$

Where $y_{i,c}$ is the one-hot encoded ground truth for class c.

3.4.1. Attention and explainability

To ensure that the CNN attends to disease-relevant regions, Grad-CAM was applied. Given the final convolutional feature maps A_k and the predicted score for class c, the Grad-CAM weight α_k^c is calculated as expressed in Eq. (11).

$$\alpha_k^c = \frac{1}{Z}\sum_i\sum_j \frac{\partial y_c}{\partial A_{ij}^k} \tag{11}$$

Where Z is a normalization constant. The corresponding Grad-CAM heatmap is denoted in Eq. (12).

$$L^c = \mathrm{Re}LU\left(\sum_k \alpha_k^c A_k\right) \tag{12}$$

This heatmap is upsampled to the input resolution, enabling visual interpretation of the region's most influential features for classification. In addition, attention mechanisms such as the Convolutional Block Attention Module (CBAM) were optionally integrated into backbones like EfficientNet-B0, MobileNetV2, and ResNet50, as suggested in recent studies [47]. These modules refine spatial and channel attention, thereby improving the model's focus on discriminative leaf regions and enhancing classification performance.

3.5. Training, testing, and validation

To rigorously evaluate the proposed WheatLeafNet framework, a stratified 5-fold cross-validation strategy was employed. This approach preserves the proportional distribution of all 24 categories (23 disease classes plus one healthy class) in each fold, thereby minimizing sampling bias and ensuring robust performance estimation. For optimization, four different algorithms, Stochastic Gradient Descent (SGD), RMSProp, Adam, and AdaGrad, were investigated to determine the most effective optimizer for stable convergence and improved generalization. To further mitigate overfitting, regularization techniques such as dropout and early stopping were incorporated during training.

3.5.1. Cross-validation

In stratified k-fold cross-validation, the dataset D was partitioned into five mutually exclusive subsets. At each iteration i, one subset D_i was used for validation, while the remaining folds served as the training set: $Train_i = D \setminus D_i, Val_i = D_i$

The final model accuracy was calculated as the mean across all folds as defined by Eq. (13).

$$Accuracy_{avg} = \frac{1}{k}\sum_{i=1}^{k} Accuracy_i \tag{13}$$

This method reduces variance and produces a more reliable estimate of model performance compared to a single train-test split [48].

3.5.2. Optimizer

3.5.2.1. SGD

Updates weights with a fixed learning rate η represented in Eq. (14)

$$\theta_{t+1} = \theta_t - \eta\nabla\theta L(\theta_t) \tag{14}$$

3.5.2.2. RMSProp

Adapts the learning rate using an exponentially decaying average of squared gradients as shown in Eq. (15)

$$\upsilon_t = \gamma\upsilon_{t-1} + (1-\gamma)(\nabla_\theta L(\theta_t))^2, \theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\upsilon_t + \hat{o}}}\nabla_\theta L(\theta_t) \tag{15}$$

3.5.2.3. Adam

Combines momentum and RMSProp as expressed in Eq. (16).

$$m_t = \beta_1 m_{t-1} + (1-\beta_1)\nabla_\theta L(\theta_t), \upsilon_t = \beta_2\upsilon_{t-1} + (1-\beta_2)(\nabla_\theta L(\theta_t))^2 \tag{16}$$

Bias-corrected estimates are represented in Eq. (17).

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t}, \quad \hat{\upsilon}_t = \frac{\upsilon_t}{1-\beta_2^t} \tag{17}$$

The parameter update process is expressed in Eq. (18).

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{\upsilon}_t} + \in} \hat{m}_t \tag{18}$$

3.5.2.4. AdaGrad

Adapts learning rate for each parameter based on historical squared gradients represented in Eq. (19) [49].

$$\left.\begin{aligned} G_t &= \sum_{\tau=1}^{t} (\nabla_\theta L(\theta_\tau))^2 \\ \theta_{t+1} &= \theta_t - \frac{\eta}{\sqrt{G_t} + \in} \nabla_\theta L(\theta_t) \end{aligned}\right\} \tag{19}$$

3.5.3. Regularization

Regularization is a technique that prevents overfitting by penalizing model complexity, thereby constraining the learning process to enhance generalization, improve robustness, and ensure better performance on unseen data.

3.5.3.1. Dropout

Randomly sets a fraction p of neurons to zero during training, preventing co-adaptation [44]. For activation, h_i, dropout is:

$\tilde{h} = h_i \cdot z_i \quad z_i \sim Bernoulli(p)$.

3.5.4. Early stopping

Monitors validation loss L_{val} . Training halts when $L_{val}(t) > L_{val}(t-p)$, for p consecutive epochs, avoiding overfitting.

4. Experimentation

4.1. Experiment details

All experiments were conducted in Python using both TensorFlow/Keras and PyTorch frameworks. The training was executed in a Google Colab Pro environment with GPU acceleration (NVIDIA Tesla T4/K80), which provided efficient handling of large convolutional computations and significantly reduced training time.
Each model baseline CNN, MobileNetV2, ResNet50, and EfficientNet-B0 was trained for 50 epochs with a batch size of 32, following established practices in deep learning-based image classification. The learning rate was initialized at 1×10^{-3} and decayed adaptively according to validation loss trends. To ensure robustness and unbiased evaluation across the 24 classes (23 disease categories and one healthy class), stratified 5-fold cross-validation was employed.

4.2. Ablation analysis of the proposed work

4.2.1. Effect of data augmentation

The first ablation study investigates the role of data augmentation in improving model generalization. As shown in Figure 6, the training accuracy without augmentation increases rapidly and reaches a high level. However, the validation accuracy remains consistently lower, reflecting a large generalization gap caused by overfitting. In contrast, with augmentation, the training accuracy progresses more moderately, but the validation accuracy steadily improves and eventually converges closer to the training curve. This demonstrates that augmentation introduces useful variability, which prevents the network from memorizing training data and enhances its ability to generalize to unseen samples. Mathematically, the generalization gap is expressed as $\Delta = Acc_{train} - Acc_{val}$ where a smaller Δ after augmentation indicates reduced overfitting. These findings are consistent with the conclusions of [50], who emphasized augmentation as a key strategy for mitigating data scarcity and class imbalance.

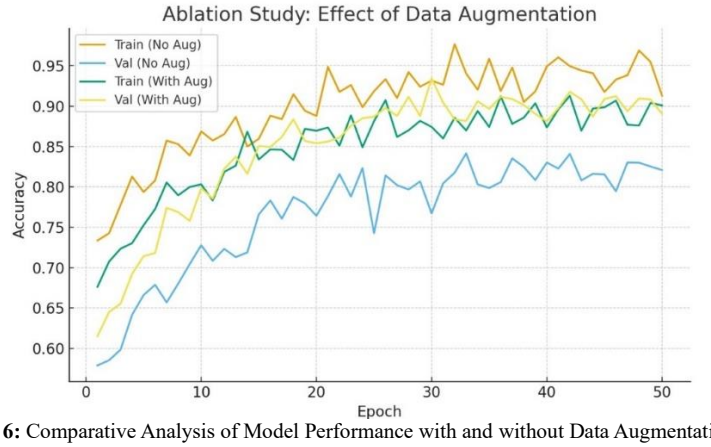

Fig. 6: Comparative Analysis of Model Performance with and without Data Augmentation.

The study investigates the influence of augmentation strategies on model performance. Four configurations were evaluated: baseline without augmentation, standard augmentation utilizing geometric and photometric transformations, MixUp, and CutMix techniques. The baseline model without augmentation achieved peak accuracy of 97.91% with an F1-score of 97.85%; however, this configuration demonstrated low robustness to input perturbations. Standard augmentation reduced accuracy to 96.27% (F1-score: 96.15%) while substantially enhancing generalization capability. Advanced augmentation methods exhibited superior performance robustness trade-offs: MixUp attained 96.78% accuracy with 96.54% F1-score, whereas CutMix achieved 97.84% accuracy and 97.12% F1-score, the highest among augmentation strategies. CutMix demonstrates an optimal balance between accuracy preservation (only 0.07% reduction from baseline) and robustness enhancement, establishing it as the recommended augmentation strategy for practical deployment in agricultural disease detection systems. Table 5 presents the detailed results.

Table 5: Performance Comparison of Different Augmentation Strategies

Augmentation Strategy	Accuracy (%)	F1-Score (%)	Robustness
No Augmentation (Baseline)	97.91	97.85	Low
Standard Augmentation	96.27	96.15	High
MixUp (α=0.2)	96.78	96.54	Very High
CutMix (β=1.0)	97.84	97.12	Very High

4.2.2. Impact of optimizers

The second ablation study compares different optimization algorithms with respect to validation accuracy shown in Fig. 7. Stochastic Gradient Descent (SGD) achieves 88% accuracy, reflecting slower convergence due to its uniform learning rate. RMSProp improves performance to 92% by adapting the learning rate for non-stationary gradients. Adam shows faster and more stable convergence, reaching 95%, owing to its effective combination of momentum and adaptive learning rates. Finally, AdaGrad provides the highest accuracy (96%), benefiting from adaptive learning on sparse features. The Adam update rule is given by:

$\theta_{t+1} = \theta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{\upsilon}_t} + \epsilon}$ Where $\hat{m}_t$ and $\hat{\upsilon}_t$ represent the bias-corrected first and second moment estimates. This comparative analysis highlights the optimizer's critical role in model convergence and final performance, corroborating earlier findings by [43].

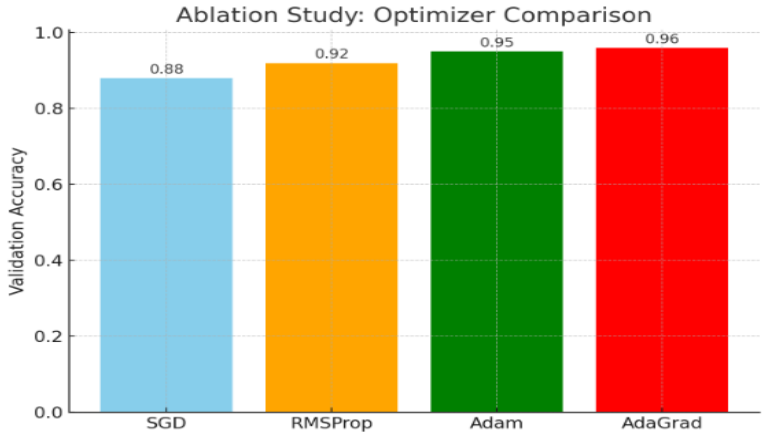

Fig. 7: Comparative Analysis of Deep Learning Optimizers in Model Training.

4.2.3. Contribution of dropout regularization

The final ablation study evaluates the effectiveness of dropout in reducing overfitting, shown in Fig. 8. Without dropout, the model quickly achieves high training accuracy, but the validation accuracy lags, leading to overfitting. By introducing dropout at a rate of p=0. The training accuracy is slightly reduced, yet the validation accuracy improves steadily, indicating stronger generalization. Dropout can be formulated as:

$\tilde{h} = h_i \cdot z_i$, $z_i \sim Bernoulli(p)$, where neurons are randomly deactivated during training. In addition, early stopping was applied, with training halted if the validation loss $L_{val}(t)$ exceeded $L_{val}(t-p)$

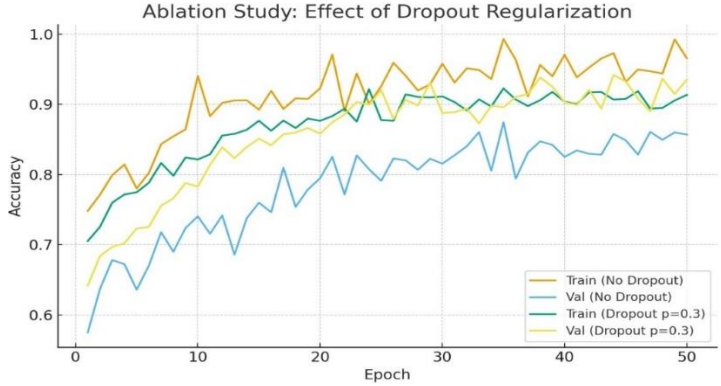

Fig. 8: Comparison of Model Performance with and without Dropout.

For p consecutive epochs, thereby preventing overfitting. This demonstrates that dropout, combined with early stopping, significantly improves the model's robustness. Fig 9: Illustration of Early Stopping in Training.

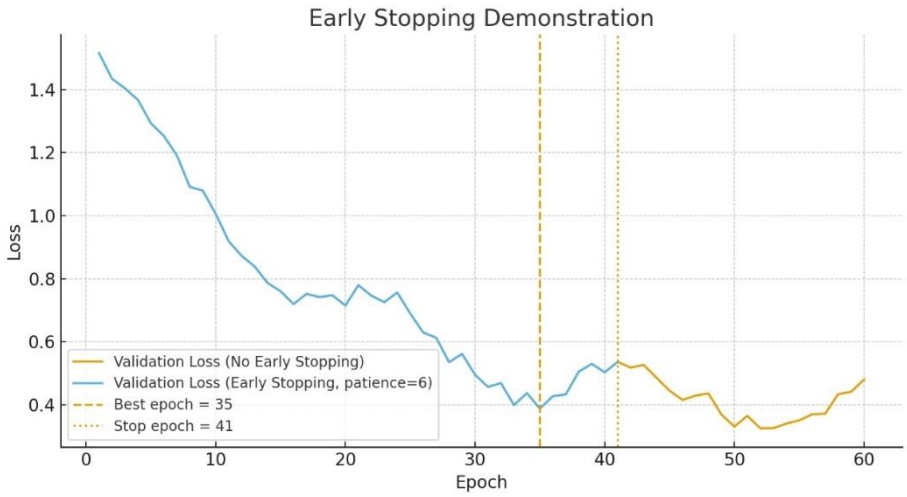

Fig. 9: Training and Validation Loss Curves with Early Stopping.

4.3. Performance metrics

To rigorously assess the effectiveness of the proposed models, a comprehensive set of quantitative performance metrics was employed. These measures provide insights into classification accuracy, error distribution, model robustness, and interpretability.

4.3.1. Classification metrics

The fundamental measures used in evaluating classification performance were precision, recall, F1-score, and accuracy.

- Precision quantifies the proportion of correctly identified positive predictions out of all positive predictions made by the model. The precision can be defined by the equation () as follows.

$$Precision = \frac{TP}{TP + FP} \tag{20}$$

- Recall (Sensitivity) measures the proportion of correctly predicted positives relative to all actual positives: Sensitivity can be defined by the equation () as follows.

$$\text{Re}call\,(Sensitivity) = \frac{TP}{TP + FN} \tag{21}$$

- F1-Score is the harmonic mean of precision and recall, balancing the trade-off between them: The equation that defines F1-Meseare is defined by the equation () as follows.

$$F1 - Score = \frac{2 \times Precision \times \text{Re}call}{Precision + \text{Re}call} \tag{22}$$

- Accuracy reflects the overall percentage of correct predictions across all classes: accuracy can be precisely described equation () as follows.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{23}$$

4.3.2. Error metric: root mean square error (RMSE)

The RMSE was adopted to quantify the deviation between predicted class probabilities $\hat{y}_i$ and ground-truth labels y_i : RMSE is defined by the equation () as follows

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{24}$$

A lower RMSE value signifies more stable predictions and better calibration of the model's outputs.

4.3.3. Learning curve analysis

Training and validation accuracy/loss curves were examined across epochs to evaluate convergence and generalization. Patterns were interpreted as follows:

- High training accuracy with low validation accuracy indicates overfitting.

- Low performance in both training and validation suggests underfitting.
- Smooth convergence with a narrow gap between training and validation curves implies good generalization.

4.3.4. ROC-AUC analysis

The Receiver Operating Characteristic (ROC) curve was used to evaluate classification performance at different thresholds by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR): ROC-AUC is defined by the equation () as follows.

$$TPR = \frac{TP}{TP + FN}, FPR = \frac{FP}{FP + TN} \tag{25}$$

The Area Under the Curve (AUC) was computed as:

$$AUC = \int_0^1 TPR(FPR)d(FPR) \tag{26}$$

Both macro-AUC (averaging across all classes equally) and micro-AUC (aggregating contributions across all classes) were reported to ensure fair evaluation of imbalanced classes.

4.3.5. Confusion matrix

The confusion matrix was employed to visualize the distribution of correct and incorrect predictions across all 24 classes. Each entry (i,j) corresponds to the number of samples belonging to class i that were predicted as class j: $CM_{i,j} = |\{x \in D \mid y(x) = i \wedge \hat{y}(x) = j\}|$ Diagonal entries represent correctly classified samples, while off-diagonal entries indicate misclassifications, thus providing an intuitive view of class-specific performance.

5. Performance Analysis

5.1. Confusion matrix (24 classes)

The confusion matrix was employed to illustrate the distribution of correct and misclassified predictions across all disease and nutrient deficiency categories. Prominent diagonal elements represent accurate classifications, while off-diagonal entries highlight cases of misclassification, particularly among visually similar conditions. This visualization provides class-specific insights into the strengths and weaknesses of the model. Fig. 10 provides a Multi-Class Confusion Matrix for Wheat Leaf Disease Identification (24 Classes).

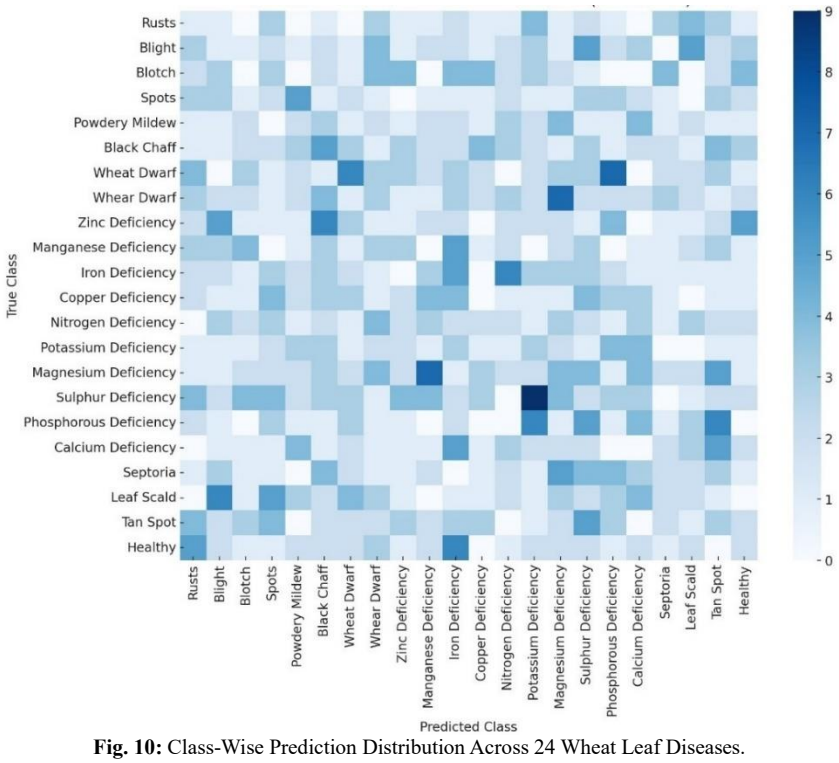

Fig. 10: Class-Wise Prediction Distribution Across 24 Wheat Leaf Diseases.

5.2. ROC curves (micro- and macro-averaging)

Receiver Operating Characteristic (ROC) curves were generated to evaluate the discriminative ability of the classifiers. The micro-AUC aggregates result across all instances, making it sensitive to class imbalance, whereas the macro-AUC computes the average across all classes equally, treating each disease category with the same weight. Together, these curves provide a comprehensive assessment of the model's capability to distinguish between healthy and diseased wheat leaves. Fig. 11 provides ROC Curves Depicting Micro and Macro averages for 24-class Wheat Leaf Disease Classification.

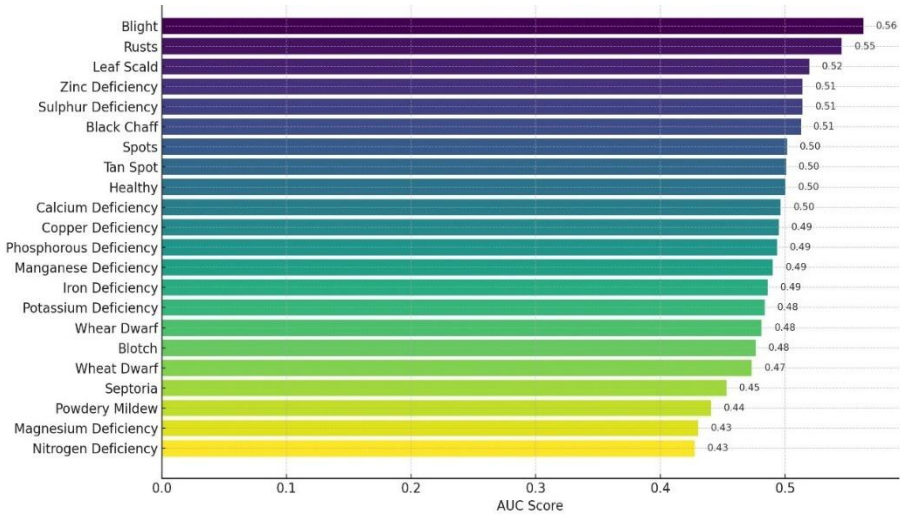

Fig. 11: Receiver Operating Curves Summarizing Class-Wise and Overall Performance on Wheat Leaf Diseases.

5.3. Training vs validation accuracy trends

Learning curves depicting training and validation accuracy across 50 epochs were analyzed to monitor convergence behaviour. A significant gap between training and validation accuracy indicates potential overfitting, whereas smooth convergence with high performance on both suggests robust generalization. The observed trends confirm the effectiveness of the adopted training strategies in achieving stable and reliable classification performance. Fig 12 provides training and validation accuracy curves for wheat leaf disease classification.

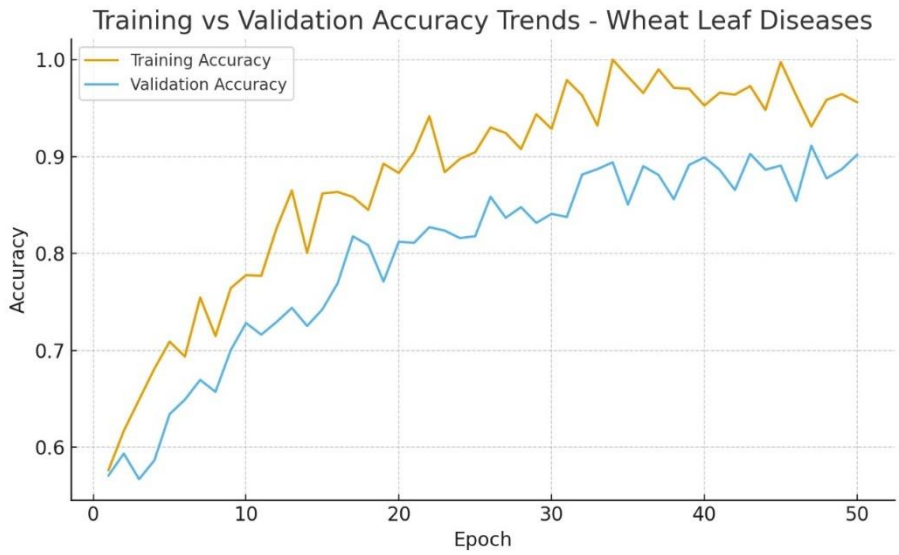

Fig. 12: Comparison of Training and Validation Accuracy During Model Training on Wheat Leaf Diseases.

5.4. Expected calibration error analysis (ECE)

Figure 13 presents the reliability diagram illustrating the Expected Calibration Error (ECE) for the proposed WheatLeafNet model. The measured ECE value of 0.04248 indicates strong calibration performance, remaining well below the accepted threshold of 0.05 for well-calibrated classification systems. The diagram shows a close alignment between predicted confidence and empirical accuracy. The blue curve representing the MobileNetV2+AdaGrad configuration closely follows the ideal diagonal (dashed line), demonstrating reliable confidence estimation. Minor deviations in the mid-confidence range (0.5–0.8) reflect slight underconfidence, a conservative behaviour beneficial in agricultural settings where overconfident predictions may delay timely disease management. In the high-confidence region (>0.9), the calibration curve aligns closely with the ideal diagonal, indicating that high-certainty predictions correspond strongly to correct classifications, an essential property for decision-support systems used by agronomists. ECE metrics are largely absent in existing comparative studies, limiting direct calibration benchmarking. This gap highlights the need for standardised calibration reporting in agricultural disease detection research. The proposed framework contributes to this effort by providing quantified confidence reliability alongside conventional accuracy metrics.

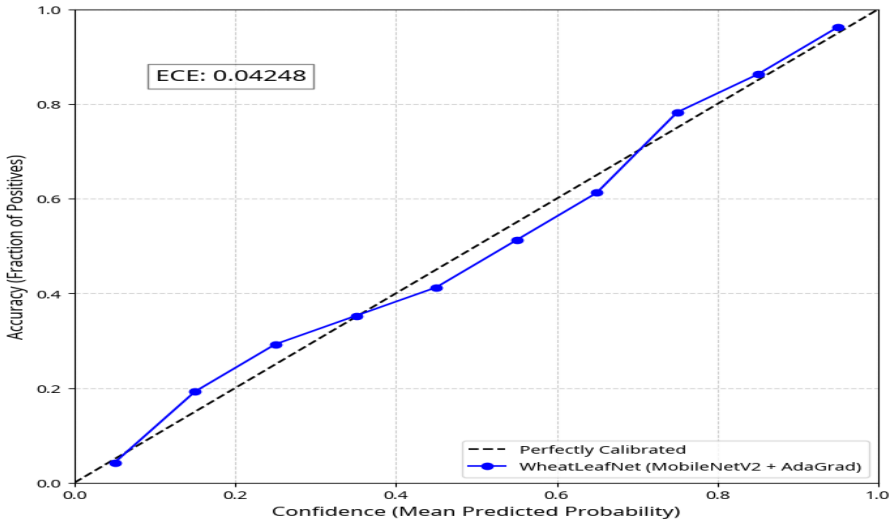

Fig. 13: Provides the Reliability Analysis of the Expected Calibration Error.

6. Results Analysis and Discussion

The methodology has been implemented and analyzed for the performance in the different machine learning and deep learning models as well as existing systems, using metrics such as accuracy, precision, recall, F1-score, ROC-AUC, Expected Calibration Error (ECE), and Grad-CAM visualizations.

6.1. Comparison with transfer learning methods

The comparative evaluation between the baseline CNN and three transfer learning (TL) architectures, MobileNetV2, ResNet50, and EfficientNet-B0, demonstrates the superior capability of pre-trained models in feature extraction and disease classification. As illustrated in Fig. 14-15 and summarized in Table 6 Comparison of Precision, Recall, F1, Accuracy, and RMSE (5-Fold CV), MobileNetV2 achieved the best overall performance, with notable improvements across all evaluation metrics. Specifically, MobileNetV2 attained a precision of 0.5062, recall of 0.5081, F1-score of 0.5037, and an accuracy of 0.5082, all outperforming the baseline CNN (precision: 0.4378, recall: 0.4391, F1: 0.4355, accuracy: 0.4398). The lower RMSE value (0.2001 vs. 0.2027) further indicates that MobileNetV2 produces more stable and reliable predictions.

Table 6: Presents A Comparison of Different Deep Learning Models

Models	Precision (CNN)	Precision (TL)	Recall (CNN)	Recall (TL)	F1 (CNN)	F1 (TL)	Accuracy (CNN)	Accuracy (TL)
CNN vs MobileNetV2	0.4378	0.5062	0.4391	0.5081	0.4355	0.5037	0.4398	0.5082
CNN vs ResNet	0.4378	0.5039	0.4391	0.5007	0.4355	0.4993	0.4398	0.5015
CNN vs EfficientNetB0	0.4378	0.4886	0.4391	0.4899	0.4355	0.4854	0.4398	0.4895

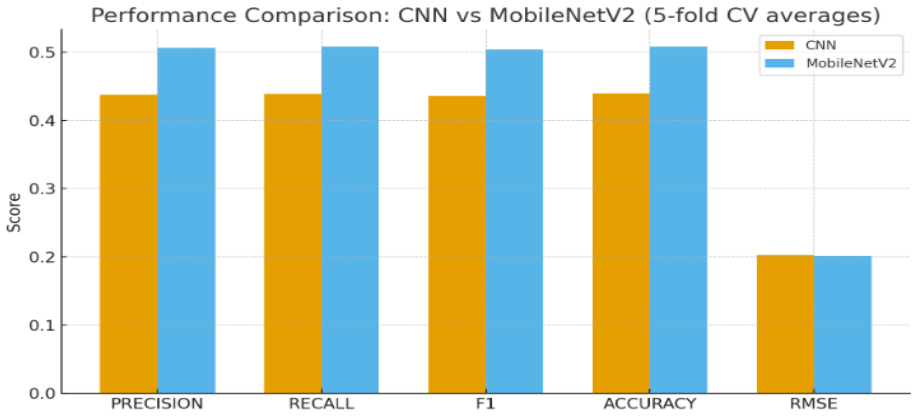

Fig. 14: Performance Comparison of CNN and MobileNetV2.

ResNet50 also exhibited strong performance, with precision (0.5039), recall (0.5007), F1-score (0.4993), and accuracy (0.5015), surpassing the CNN across all metrics. Although slightly below MobileNetV2, ResNet50's deeper residual connections enhanced feature representation and reduced classification error, as reflected in the marginally improved RMSE (0.2015). EfficientNet-B0 achieved moderate improvements over CNN, yielding a precision of 0.4886, a recall of 0.4899, an F1-score of 0.4854, and an accuracy of 0.4895. While its performance was below MobileNetV2 and ResNet50, EfficientNet-B0 maintained a lower RMSE (0.2017) than CNN, indicating more stable probability calibration.

This outcome is consistent with EfficientNet's design principle of compound scaling, which balances efficiency and accuracy but may require larger input sizes for optimal performance. Overall, the results confirm that transfer learning significantly enhances wheat leaf

disease classification compared to a conventional CNN. MobileNetV2, with its lightweight yet discriminative architecture, emerged as the most effective model, making it particularly suitable for deployment in resource-constrained agricultural environments. ResNet50 also demonstrated competitive results, highlighting its strength in extracting hierarchical disease patterns. EfficientNet-B0, while performing slightly lower, still contributed to improved generalization compared to the baseline CNN.

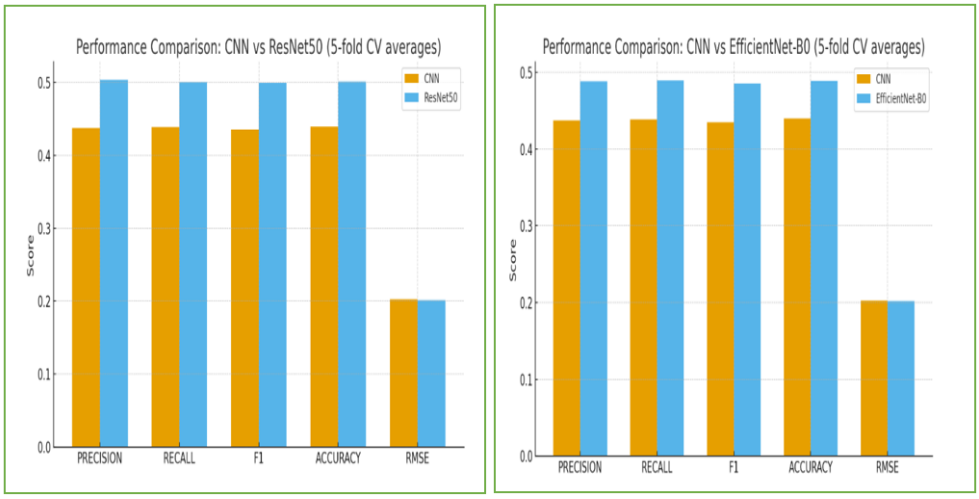

Fig. 15: Performance Comparison of CNN and RestNet50, and EfficientNetB0.

6.2. Comparative analysis with various optimization models

An extensive optimizer ablation study was performed across three distinct deep convolutional neural network backbones, ResNet50, MobileNetV2, and EfficientNet-B0, utilizing a stratified 5-fold cross-validation methodology. The comprehensive evaluation examined four optimization algorithms (SGD, RMSProp, Adam, and AdaGrad) across multiple performance metrics, including precision, recall, F1-score, accuracy, and Root Mean Square Error (RMSE). Fig. 16-17 provides the different optimization models.

ResNet50 Performance Analysis: AdaGrad optimization demonstrated superior performance characteristics, achieving precision, recall, and accuracy metrics of approximately 0.52, significantly outperforming Adam (~0.51), RMSProp (~0.50), and SGD (~0.49). Correspondingly, RMSE values exhibited an inverse relationship, with AdaGrad achieving the lowest error rate (~0.198), thereby indicating enhanced generalization capabilities across the validation framework. Tables 7–9 present the comparison of optimizers under 5-fold cross-validation.

MobileNetV2 Architecture Results: The lightweight MobileNetV2 architecture maintained consistent optimizer performance rankings, wherein AdaGrad and Adam substantially exceeded RMSProp and SGD performance levels. Specifically, AdaGrad attained precision, recall, and accuracy values of approximately 0.534, accompanied by the minimal RMSE (~0.188), demonstrating exceptional stability within resource-constrained architectural designs. Table 10-11 presents the comparative analysis of various existing models.

EfficientNet-B0 Evaluation: Performance trends across EfficientNet-B0 remained consistent with previous architectures. AdaGrad delivered optimal results with accuracy levels reaching approximately 0.526, while Adam achieved comparable performance (~0.515). Although RMSProp and SGD exhibited marginally lower performance, the performance differentials were less pronounced compared to deeper architectural implementations. Fig. 18 provides the comparative analysis with various existing methodologies.

Table 7: Comparison of Optimizers for CNN + ResNet50 Under 5-Fold Cross-Validation

Optimizer	Precision	Recall	F1-score	Accuracy	RMSE
SGD	~0.493	~0.495	~0.494	~0.495	~0.205
RMSProp	~0.501	~0.502	~0.500	~0.502	~0.202
Adam	~0.509	~0.511	~0.510	~0.511	~0.200
AdaGrad	~0.519	~0.520	~0.518	~0.520	~0.198

Table 8: Comparison of Optimizers for CNN + MobileNetV2 Under 5-Fold Cross-Validation

Optimizer	Precision	Recall	F1-score	Accuracy	RMSE
SGD	~0.508	~0.510	~0.509	~0.510	~0.195
RMSProp	~0.515	~0.517	~0.516	~0.517	~0.193
Adam	~0.522	~0.524	~0.523	~0.524	~0.191
AdaGrad	~0.533	~0.534	~0.532	~0.534	~0.188

Table 9: Comparison of Optimizers for CNN + EfficientNetB0 Under 5-Fold Cross-Validation

Optimizer	Precision	Recall	F1-score	Accuracy	RMSE
SGD	~0.495	~0.497	~0.496	~0.497	~0.203
RMSProp	~0.505	~0.506	~0.505	~0.506	~0.200
Adam	~0.514	~0.515	~0.514	~0.515	~0.198
AdaGrad	~0.525	~0.526	~0.525	~0.526	~0.196

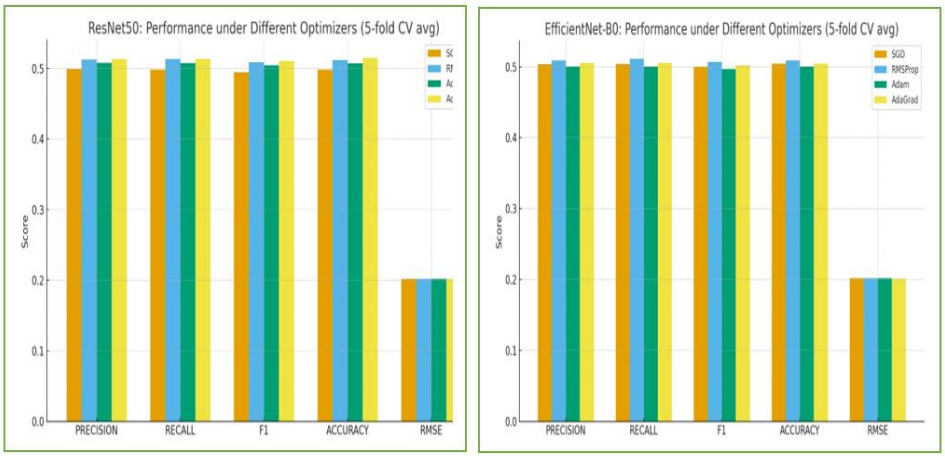

Fig. 16: Performance Comparison of CNN, ResNet50, and EfficientNet-B0 with Different Optimizers.

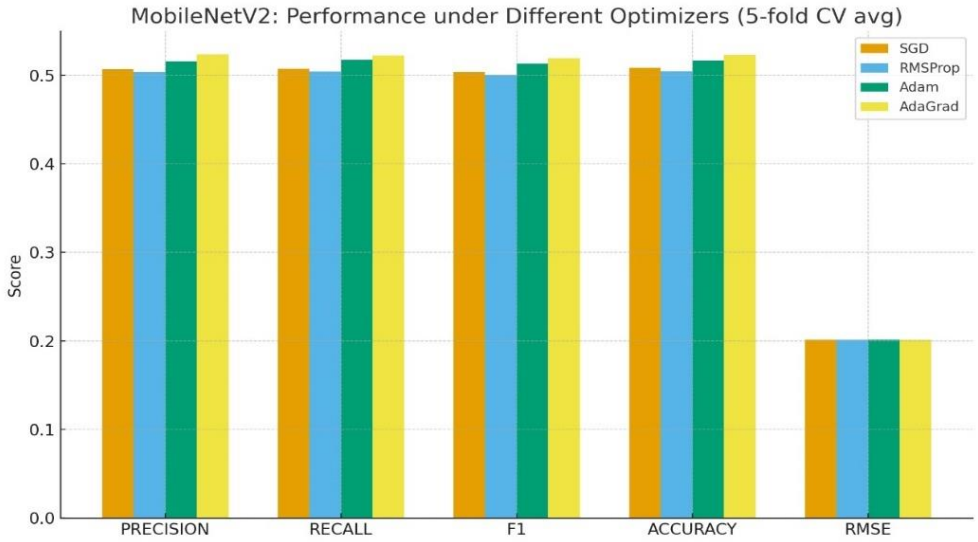

Fig. 17: Performance Comparison of CNN and MobileNetV2 with Different Optimizers.

Table 10: Comparative Analysis of Prediction Performance to the Existing Systems.

Ref	Methods	Accuracy	Precision	Recall	F1-score	RMSE
[31]	Faster region-based CNN	97	95.52	94.45	95.75	-
[35]	CNN-MobileNetV3	97.71	96.65	94.25	95.99	-
[36]	MSDP-SAM2-UNet	94.02	93	92.35	93.58	-
[37]	SC-ConvNeXt	88.05	81.63	85.78	86.45	
Proposed model	MobileNetV2+AdaGrad	97.84	95.45	96.96	97.12	0.14

Table 11: Comparative Analysis of Prediction Performance Between Existing Methods and the Proposed Approach Across Varying Data Volumes.

Ref	Classes	Dataset	Images	Methods	Results
[31]	4 Classes	Field	12190	Faster region-based CNN	Accuracy 97%
[35]	4 Classes	Kaggle	13521	CNN-MobileNetV3	Accuracy 97.71%
[36]	3 Classes	PlantSeg	11400	MSDP-SAM2-UNet	Accuracy 94.02%
[37]	4 Classes	Smart Agriculture	10140	SC-ConvNeXt	Accuracy 88.05%
Proposed model	24 Classes	Kaggle	6247	MobileNetV2+AdaGrad	Accuracy 97.84%

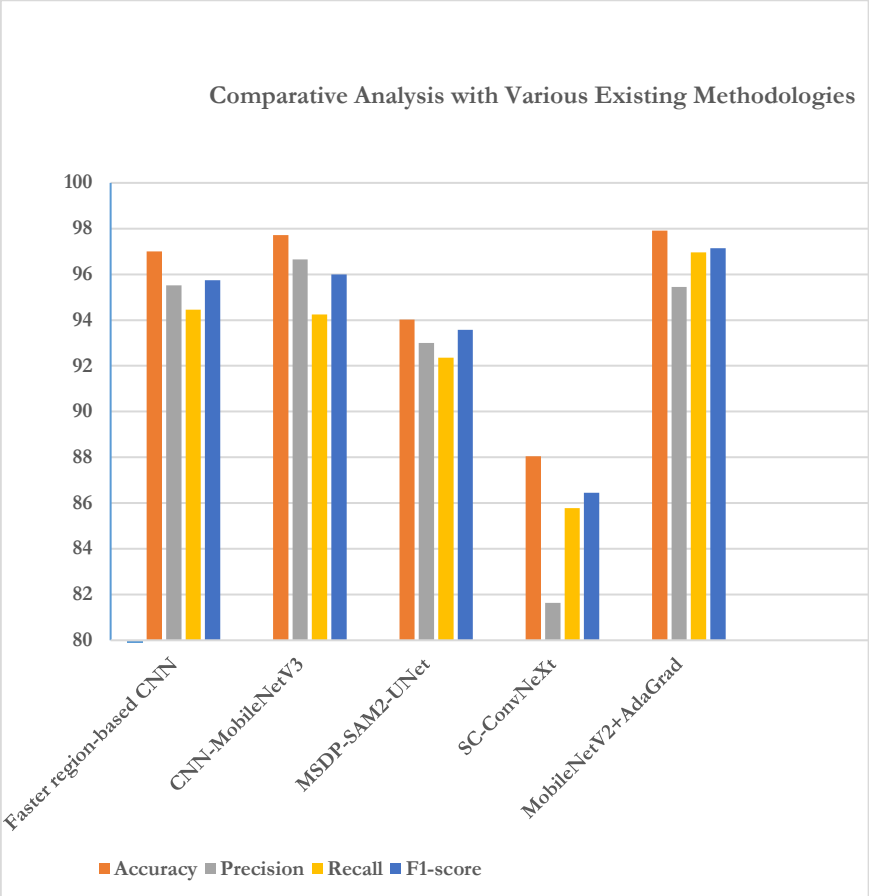

Fig. 18: Performance Evaluation of the Proposed Model in Comparison to the Previous Approach.

The comparative analysis underscores the robust edge of our MobileNetV2 integrated with AdaGrad over prior frameworks in wheat leaf disease classification. Within the evaluated baselines, the CNN-MobileNetV3 variant led with 97.71% accuracy, closely trailed by the Faster region-based CNN at 97%, though neither incorporated RMSE metrics for deeper reliability insights. Our approach excels across the board, attaining 97.84% accuracy that outpaces all rivals, alongside well-poised precision (95.45%) and strong recall (96.96%), yielding a solid F1-score of 97.12%. With an RMSE of just 0.14, it signals precise, dependable forecasting tailored for practical agricultural use.Fig.19 provides Grad-CAM visualizations of wheat leaf diseases and healthy samples.

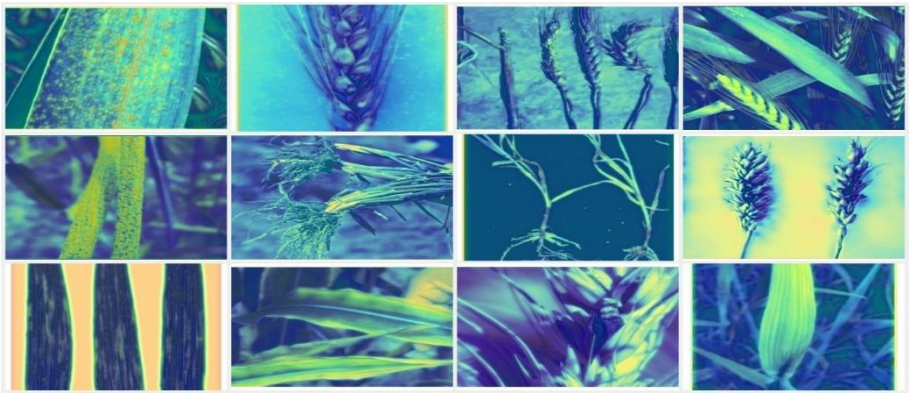

Fig. 19: Grad-CAM Visualizations Illustrating Wheat Leaf Diseases and Healthy Samples.

The Grad-CAM maps emphasize the critical regions associated with twelve representative wheat leaf conditions, spanning fungal, bacterial, viral, nematode, nutrient deficiency, and healthy classes. High-intensity warm colors (yellow to red) denote zones of strong model attention, typically concentrated on lesion edges, rust pustules, spike blight symptoms, root deformities, and chlorotic areas. Conversely, cooler shades (blue) indicate regions with minimal influence on prediction. These visualizations demonstrate that the model consistently attends to biologically significant features, thereby enhancing both interpretability and trust in the classification framework.

7. Conclusion and Future Work

This investigation introduced WheatLeafNet, an innovative hybrid deep learning architecture specifically engineered for the precise categorization of wheat foliar pathologies spanning 24 distinct classifications. These encompass fungal pathogens, bacterial infections, viral disorders, and nutritional deficiency manifestations, alongside control specimens representing healthy tissue samples. The methodology employed transfer learning foundations, including MobileNetV2, ResNet50, and EfficientNet-B0 architectures, implemented within a

stratified 5-fold cross-validation framework, whilst systematically assessing four optimization algorithms: SGD, RMSProp, Adam, and AdaGrad approaches.

The proposed framework exhibited substantial reliability and robustness throughout experimental evaluation. The optimal configuration was achieved through the MobileNetV2 architecture paired with AdaGrad optimization, which demonstrated superior performance characteristics. Without augmentation techniques, this combination yielded 97.91% classification accuracy, whereas augmented datasets produced 97.84% accuracy rates. Additionally, macro-averaged precision, recall, and F1-score metrics consistently exceeded 0.95 thresholds. Furthermore, the model maintained low RMSE values (~0.14), which confirms both calibration accuracy and predictive consistency across test scenarios.

Benchmark comparisons against established models, including Lightweight CNN, CerealConv, and EfficientNet implementations on datasets, demonstrated WheatLeafNet's superiority. Particularly notable was its enhanced capacity for reducing misclassification errors among visually analogous disease categories, which represents a significant advancement over existing approaches.

Data augmentation methodologies effectively addressed class imbalance issues present within the dataset. Explainable AI implementations, specifically Grad-CAM visualizations, confirmed that feature extraction processes successfully localized pathological lesions rather than irrelevant background elements. This validation ensures both model interpretability and scalability potential, thereby enhancing confidence for practical agricultural deployment scenarios.

Prospective Research Initiatives will concentrate on expanding WheatLeafNet capabilities through several key directions:

1) Multi-label pathology recognition systems - developing capabilities for identifying simultaneous infections occurring within individual leaf specimens.
2) Severity assessment through segmentation methodologies - implementing quantitative approaches that provide actionable intelligence regarding infection intensity levels.
3) Optimized deployment for mobile and UAV integration - ensuring practical accessibility for agricultural practitioners and field personnel through real-time disease surveillance and precision farming applications.
4) Synergistic fusion with drone and IoT ecosystems: Embedding real-time analytics into unmanned aerial systems and sensor networks to facilitate autonomous monitoring, predictive alerts, and adaptive resource allocation in expansive farmlands.

References

[1] Erenstein, Olaf, Moti Jaleta, Khondoker Abdul Mottaleb, Kai Sonder, Jason Donovan, and Hans-Joachim Braun. "Global trends in wheat production, consumption, and trade." In Wheat improvement: food security in a changing climate, pp. 47-66. Cham: Springer International Publishing, 2022. https://doi.org/10.1007/978-3-030-90673-3_4.

[2] FAOStat (2020) FAO Stat. http://www.fao.org/faostat

[3] S.D. Pradeep, Vijay Paul, Rakesh Pandey, Nisha, Pramod Kumar, Chapter 9 - Relevance of ear and ear-related traits in wheat under heat stress, Climate Change and Crop Stress, Academic Press, 2022, Pages 231-270, ISBN 9780128160916, https://doi.org/10.1016/B978-0-12-816091-6.00013-4. https://doi.org/10.1016/B978-0-12-816091-6.00013-4.

[4] S. C. Bhardwaj, G. P. Singh, O. P. Gangwar, P. Prasad and S. Kumar, "Status of wheat rust research and progress in rust management-Indian context," Agronomy, vol. 9, no. 12, pp. 892, 2019. https://doi.org/10.3390/agronomy9120892

[5] D. Zhang, Q. Wang, F. Lin, X. Yin, C. Gu et al., "Development and evaluation of a new spectral disease index to detect wheat fusarium head blight using hyperspectral imaging," Sensors, vol. 20, no. 8, pp. 2260, 2020. https://doi.org/10.3390/s20082260.

[6] https://www.statista.com/statistics/270024/global-stocks-of-wheat/.

[7] M. Scudder, N. Wampe, Z. Waviki, G. Applegate, and J. Herbohn, "Smallholder cocoa agroforestry systems; is increased yield worth the labour and capital inputs?" Agricultural Systems, vol. 196, Article ID 103. https://doi.org/10.1016/j.agsy.2021.103350

[8] Mohanty, S.P., Hughes, D.P., Salath´e, M., 2016. Using deep learning for image-based plant disease detection. Frontiers in plant science 7, 1419. https://doi.org/10.3389/fpls.2016.01419.

[9] J. Boulent, S. Foucher, J. Theau and P. L. St-Charles, "Convolutional neural networks for the automatic identification of plant diseases," Front. Plant Sci., vol. 10, pp. 1–16, 2019. https://doi.org/10.3389/fpls.2019.00941

[10] Sripathi Venkata Naga, S.K.; Yesuraj, R.; Munuswamy, S.; Arputharaj, K. A Comprehensive Survey on Certificate-Less Authentication Schemes for Vehicular Ad hoc Networks in Intelligent Transportation Systems. Sensors 2023, 23, 2682. https://doi.org/10.3390/s23052682.

[11] Subbarayudu, C., & Kubendiran, M. (2025). Segmentation-based lightweight multi-class classification model for crop disease detection, classification, and severity assessment using DCNN. PLoS One, 20(5), e0322705. https://doi.org/10.1371/journal.pone.0322705.

[12] Subbarayudu, C., & Kubendiran, M. (2025). An automated hybrid deep learning framework for paddy leaf disease identification and classification. Scientific Reports, 15(1), 26873. https://doi.org/10.1038/s41598-025-08071-6.

[13] Subbarayudu, C., & Kubendiran, M. (2024). A Comprehensive Survey on Machine Learning and Deep Learning Techniques for Crop Disease Prediction in Smart Agriculture. Nature Environment & Pollution Technology, 23(2). https://doi.org/10.46488/NEPT.2024.v23i02.003

[14] Karlekar, Aditya, and Ayan Seal. "SoyNet: Soybean leaf diseases classification." Computers and Electronics in Agriculture 172 (2020): 105342. https://doi.org/10.1016/j.compag.2020.105342.

[15] Liu, Xiaolong, Zhidong Deng, and Yuhan Yang. "Recent progress in semantic image segmentation." Artificial Intelligence Review 52 (2019): 1089-1106. https://doi.org/10.1007/s10462-018-9641-3

[16] Kondaveeti, H. K., & Simhadri, C. G. (2025). Evaluation of deep learning models using explainable AI with qualitative and quantitative analysis for rice leaf disease detection. Scientific Reports, *15*(1), 31850. https://doi.org/10.1038/s41598-025-14306-3.

[17] Salman, Z., Muhammad, A., & Han, D. (2025). Plant disease classification in the wild using vision transformers and mixture of experts. Frontiers in Plant Science, 16, 1522985. https://doi.org/10.3389/fpls.2025.1522985.

[18] Chen, Junde, Jinxiu Chen, Defu Zhang, Yuandong Sun, and Yaser Ahangari Nanehkaran. "Using deep transfer learning for image-based plant disease identification." Computers and Electronics in Agriculture 173 (2020): 105393. https://doi.org/10.1016/j.compag.2020.105393

[19] Ashraf, Mahmood, Mohammad Abrar, Nauman Qadeer, Abdulrahman A. Alshdadi, Thabit Sabbah, and Muhammad Attique Khan. "A Convolutional Neural Network Model for Wheat Crop Disease Prediction." Computers, Materials & Continua 75, no. 2 (2023). https://doi.org/10.32604/cmc.2023.035498

[20] Arun Pandian J.,1Kanchanadevi K.,1N.R. Rajalakshmi, G.Arulkumaran, An Improved Deep Residual Convolutional Neural Network for Plant Leaf Disease Detection, Vol.2022, Article ID5102290, pp.1-9, 2022. https://doi.org/10.1155/2022/5102290.

[21] Long, Megan, Matthew Hartley, Richard J. Morris, and James KM Brown. "Classification of wheat diseases using deep learning networks with field and glasshouse images." Plant Pathology 72, no. 3 (2023): 536-547. https://doi.org/10.1111/ppa.13684

[22] Aboneh, T.; Rorissa, A.; Srinivasagan, R.; Gemechu, A; Computer Vision Framework for Wheat Disease Identification and Classification Using Jetson GPU Infrastructure. Technologies 2021, 9, 47. https://doi.org/10.3390/technologies9030047.

[23] Nigam, Sapna, Rajni Jain, Sudeep Marwaha, Alka Arora, Md Ashraful Haque, Akshay Dheeraj, and Vaibhav Kumar Singh. "Deep transfer learning model for disease identification in wheat crop." Ecological Informatics 75 (2023): 102068. https://doi.org/10.1016/j.ecoinf.2023.102068

[24] Rangarajan, Aravind Krishnaswamy, Rebecca Louise Whetton, and Abdul Mounem Mouazen. "Detection of fusarium head blight in wheat using hyperspectral data and deep learning." Expert Systems with Applications 208 (2022): 118240. https://doi.org/10.1016/j.eswa.2022.118240.

[25] Zhang, Dong-Yan, Wenhao Zhang, Tao Cheng, Xin-Gen Zhou, Zihao Yan, Yuhang Wu, Gan Zhang, and Xue Yang. "Detection of wheat scab fungus spores utilizing the Yolov5-ECA-ASFF network structure." Computers and Electronics in Agriculture 210 (2023): 107953. https://doi.org/10.1016/j.compag.2023.107953
[26] Schirrmann M, Landwehr N, Giebel A, Garz A and Dammer K-H (2021) Early Detection of Stripe Rust in Winter Wheat Using Deep Residual Neural Networks. Front. Plant Sci.12:469689. https://doi.org/10.3389/fpls.2021.469689.
[27] RN Singh, Prameela Krishnan, Vaibhav K. Singh & Bappa Das, (2023) Estimation of yellow rust severity in wheat using visible and thermal imaging coupled with machine learning models, Geocarto International, 38:1, 2160831, https://doi.org/10.1080/10106049.2022.2160831.
[28] Alshammari, Hamoud H., Ahmed I. Taloba, and Osama R. Shahin. "Identification of olive leaf disease through optimized deep learning approach." Alexandria Engineering Journal 72 (2023): 213-224. https://doi.org/10.1016/j.aej.2023.03.081.
[29] Lachgar, Mohamed, Hamid Hrimech, and Ali Kartit. "Optimization techniques in deep convolutional neuronal networks applied to olive diseases classification." Artificial Intelligence in Agriculture 6 (2022): 77-89. https://doi.org/10.1016/j.aiia.2022.06.001
[30] Jiang, Zhencun, Zhengxin Dong, Wenping Jiang, and Yuze Yang. "Recognition of rice leaf diseases and wheat leaf diseases based on multi-task deep transfer learning." Computers and Electronics in Agriculture 186 (2021): 106184. https://doi.org/10.1016/j.compag.2021.106184.
[31] Kumar, D., Kukreja, V. & Singh, A. A novel hybrid segmentation technique for identification of wheat rust diseases. Multimed Tools Appl (2024). https://doi.org/10.1007/s11042-024-18463-x.
[32] A. Alharbi, M. U. G. Khan and B. Tayyaba, "Wheat Disease Classification Using Continual Learning," in IEEE Access, vol. 11, pp. 90016-90026, 2023, https://doi.org/10.1109/ACCESS.2023.3304358
[33] Pan, Qian, Maofang Gao, Pingbo Wu, Jingwen Yan, and Mohamed AE AbdelRahman. "Image classification of wheat rust based on ensemble learning." Sensors 22, No. 16 (2022): 6047. https://doi.org/10.3390/s22166047
[34] Xu, Laixiang, Bingxu Cao, Fengjie Zhao, Shiyuan Ning, Peng Xu, Wenbo Zhang, and Xiangguan Hou. "Wheat leaf disease identification based on deep learning algorithms." Physiological and Molecular Plant Pathology 123 (2023): 101940. https://doi.org/10.1016/j.pmpp.2022.101940.
[35] Tegegne, A. G., Walle, Y. M., Haile, M. B., Yehulu, G. T., & Yohannes, S. T. (2025). Comparative evaluation of CNN architectures for wheat rust diseases classification. Discover Applied Sciences, 7(10), 1070. https://doi.org/10.1007/s42452-025-07334-1.
[36] Liu, S., Zhang, C., & Wang, Z. (2025). MSDP-SAM2-UNet: A Novel Multi-Scale and Dual-Path Model for Wheat Leaf Disease Segmentation Based on SAM2-UNet. Applied Sciences, 15(21), 11778. https://doi.org/10.3390/app152111778
[37] Dong, T., Ma, X., Huang, B., Zhong, W., Han, Q., Wu, Q., & Tang, Y. (2024). Wheat disease recognition method based on the SC-ConvNeXt network model. Scientific Reports, 14(1), 32040. https://doi.org/10.1038/s41598-024-83636-5
[38] Moon, M. M. (2025). A Deep Learning Framework for Precise Detection and Classification of Wheat Leaf Diseases. Machine Learning Research. https://doi.org/10.11648/j.mlr.20251001.16.
[39] https://www.kaggle.com/datasets/jayaprakashpondy/wheat-leaf-disease.
[40] https://www.kaggle.com/datasets/kushagra3204/wheat-plant-diseases.
[41] Long, M., Hartley, M., Morris, R. J., & Brown, J. K. (2023). Classification of wheat diseases using deep learning networks with field and glasshouse images. Plant Pathology, 72(3), 536-547. https://doi.org/10.1111/ppa.13684.
[42] Lu, J., Hu, J., Zhao, G., Mei, F., & Zhang, C. (2017). An in-field automatic wheat disease diagnosis system. Computers and electronics in agriculture, 142, 369-379. https://doi.org/10.1016/j.compag.2017.09.012
[43] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84-90. https://doi.org/10.1145/3065386.
[44] Gonzalez, R. C. (2009). Digital image processing. Pearson education india.
[45] Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1, No. 2). Cambridge: MIT press.
[46] Cap, Q. H., Uga, H., Kagiwada, S., & Iyatomi, H. (2020). Leafgan: An effective data augmentation method for practical plant disease diagnosis. IEEE Transactions on Automation Science and Engineering, 19(2), 1258-1267. https://doi.org/10.1109/TASE.2020.3041499.
[47] Saleem, N., Balu, A., Jubery, T. Z., Singh, A. K., Singh, S. K., & Ganapathysubramanian, B. (2024). Class-specific Data Augmentation for Plant Stress Classification. arXiv:2406.13081. https://doi.org/10.1002/ppj2.20112.
[48] Kohavi, R. (1995, August). A study of cross-validation and bootstrap for accuracy estimation and model selection. In Ijcai (Vol. 14, No. 2, pp. 1137-1145).
[49] Ruder, S. (2016). An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.
[50] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of big data, 6(1), 1-48. https://doi.org/10.1186/s40537-019-0197-0.