International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(10s) 2026 pp. 1027-1045
https://doi.org/10.70917/ijcisim-2026-3694
Article

WheatDisease-HRY: A Real-Field Wheat Disease Dataset with Baseline Deep Learning Benchmarks for Automated Disease Detection

Sheetal Choudhary[1,2], Sunila Godara[1], Sanjeev Kumar[1]

[1]Department of Computer Science & Engineering, Guru Jambheshwar University of Science & Technology, Hisar 125001, Haryana, India
[2]Computer Section, Chaudhary Charan Singh Haryana Agricultural University, Hisar 125004, Haryana, India
Corresponding author: Sheetal Choudhary
Email: sheetal.phougat@hau.ac.in

Abstract: In India, to increase the wheat yield production and support sustainable agricultural practices, timely and correct identification of wheat crop diseases such as yellow rust, brown rust, etc., is very much essential. Further, the majority of deep learning algorithms show high efficiency in plant disease detection and recognition, but these studies rely on handcrafted datasets captured under controlled laboratory conditions and accordingly will not perform well in the real world domain. Furthermore, there is a lack of real field wheat crop disease datasets collected in India, particularly from Haryana state, being the largest producer of wheat crops. Therefore, to address the issue of lack of region specific crop diseases dataset, which are region-specific, the wheat crop disease data set is curated, consisting of 2672 images of healthy and diseased leaves of wheat crops collected from the real field of CCS Haryana Agricultural University, Hisar. The dataset includes typically three classes of leaves of wheat crop, i.e., Yellow Rust (910), Brown Rust (922) and Healthy leaves (840) collected using both a DSLR camera and different smartphone cameras under natural lighting and field conditions. All images were validated and labeled by taking the expertise of wheat pathologists to ensure reliability on the dataset. The study also provides a comprehensive, systematic workflow for transforming raw data into a high-quality benchmark dataset for training using image preprocessing and augmentation techniques. Besides this, a comparative benchmarking analysis is performed under identical experimental conditions using three widely adopted deep learning architectures, which includes ResNet50, MobileNetV2 and EfficientNet-B0. The experimental results show that ResNet50 and EfficientNet-B0 achieve similar performance i.e. approximately 91% classification accuracy on real-field data. However, MobileNetV2 offers a lightweight alternative suitable for mobile and edge deployment. Furthermore, Grad-CAM based explainability analysis was performed to validate model predictions and highlight disease specific regions in wheat leaves. Therefore, this study contributes a practical region specific WheatDisease-HRY dataset and baseline benchmarking framework for developing robust AI based wheat disease diagnosis tools for real world agricultural applications in India.
Keywords: Deep learning; Wheat disease classification; Real-field dataset; Benchmarking; Explainable AI; Grad-CAM; Precision agriculture;

1. Introduction

Agriculture remains a cornerstone of the Indian economy and is the principal source of livelihood for a large proportion of the population. The agriculture and allied sectors contribute approximately one-fifth of India's gross domestic product (GDP), underscoring their importance in food security and rural development [1, 2]. Haryana is one of the foremost wheat-producing states of India due to the favorable agro-climate conditions and improved irrigation facilities of the state due to which the state fulfills significantly the requirement of the central gain procurement system as well as supports local livelihood and national food reserves.

Wheat crops are exposed to multiple stressors, like changing climate conditions, pest infestations, soil degradation, fungal and bacterial infections, etc. Among these, crop diseases represent one of the major threats to yield loss. Crop disease affects both the quality and quantity of the yield, which is a serious challenge for the small farmers whose livelihood completely depends upon the agricultural output. In extreme cases, this loss is the main cause for the economic setbacks, threatening household food security and disturbing overall income stability.

Major wheat crop foliar diseases are yellow rust and brown rust, which typically cause very harsh effects on the crop [3–5]. These diseases are actually responsible for continuous yield loss of crops, particularly in the area belonging to South Asia, if not detected at early stages [6, 7]. The disease symptom and intensity are also strongly influenced by environmental conditions and crop management practices used in that region [8, 9].

Traditionally, disease identification has been carried out by the expert farmers, extension workers and the persons from the related organizations or other institutions that have specialties in this sector. However, this approach is labor-intensive, time consuming and dependable on experts. Sometimes, specifically at an early stage or when multiple diseases exhibit similar visual characteristics, the diseases symptoms may be misdiagnosis or overlooked [10, 11]. Therefore, the delay and inaccuracy of disease identification force the farmers to use excessive or inappropriate amounts of pesticides, thus increase production costs and environmental degradation [7, 12]. Moreover, sometimes the manual inspection becomes wrong under heterogeneous field conditions and different environmental conditions. These constraints motivate us to provide some automated, reliable and scalable disease detection systems that can help farmers to get timely solutions for their crop disease. This will also decrease the dependencies on an agriculture expert.

Recent advancements in the field of computer vision and artificial intelligence will provide us with a deep learning based automated plant disease recognition system [13, 14]. Further, the Deep convolutional neural networks (CNNs) have shown very impressive performance on image data in comparison of traditional machine learning approaches that depends upon the handcrafted features. Thus these technologies then support precision agriculture by enabling disease detection, continuous crop monitoring and informed decision making at field level.

However, the major problem faced by the researchers that these techniques are heavily reliant on the datasets which were captured under controlled laboratory conditions. These datasets, particularly PlantVillage, have significantly supported plant disease recognition research to get an idea but fail to handle the real-field variability challenges [15, 16]. Accordingly, the models trained on such datasets often give poor generalization when deployed in real field applications. The data collected from the real field area are mainly affected by lighting effect, background confusion and occlusions [18–20]. Furthermore, distinguishing visually similar diseases such as yellow rust and brown rust still become a critical challenge which demands robust and interpretable models. These challenges significantly affect model reliability in real field deployment where misclassification between similar diseases can lead to incorrect treatment decisions. Additionally, explainable AI techniques are increasingly required to improve trustworthiness and transparency in agricultural decision support systems.

Previous studies have also emphasized that real field collected datasets are essential for increasing the practical deployment of plant disease recognition systems and supporting precision agriculture applications [18–20]. Nevertheless, region-specific real-field wheat disease datasets that specifically belong to Haryana state are very limited. The wheat disease is influenced by regional climate conditions, local environmental factors and crop management practice used by the farmers. Therefore, dataset collected from one geographical region may not have adequate disease symptoms as observed in other region. Despite Haryana being one of the India top wheat producing states, there is a lack of region specific real field wheat disease datasets representing local disease involvement under natural conditions. Therefore, non-availability of such a dataset limits standardized benchmarking and affects the disease detection system for the deployment in local farm areas [23, 24].

To the best of our knowledge, limited studies have simultaneously addressed real field dataset development, systematic benchmarking and explainability analysis for wheat disease classification under Indian agricultural conditions. Many studies focus on a single architecture or employ different experimental setting due to which comparison will become difficult [21 – 27]. Therefore, establishing baseline benchmark results on real field datasets is important for future methodological developments and researches.

To address these challenges, this paper introduces the WheatDisease-HRY dataset consisting of 2672 real field wheat leaf disease images collected under natural agricultural conditions from the experimental field areas of CCS Haryana Agricultural University, Hisar, Haryana, India with a near balanced distribution across disease classes. The datasets consist of RGB images belongings to three agriculturally important classes' Yellow rust, Brown rust and Healthy wheat leaves. Images were acquired using both smartphones camera and DSLR camera under natural field conditions, preserving realistic variability in lighting, background complexity, leaves overlapping, orientation and disease severity level. The dataset was designed to represent real field agricultural environment and supports the development of AI based disease diagnosis systems.

A comparative analysis is also conducted using three deep learning architectures MobileNetV2, ResNet50 and EfficientNet-B0 to balance computational efficiency and feature extraction capability [25–27]. The dataset and benchmarking results are used as a reference for future study on wheat disease detection under real field conditions. In addition to this, model interpretability is investigated using Grad-CAM to visualize the regions which contributes to classification decisions. This enhances trust and transparency on AI based agricultural decision making systems.

Main Contributions

The main contributions of this study are summarized as follows:

1. A real field wheat leaf disease image dataset named Wheat Disease-HRY is developed using 2672 annotated images which were collected under natural agricultural conditions from Haryana, India.
2. Comprehensive dataset characterization including image acquisition protocols, disease classes, labeling procedure, environmental variability and dataset compositions.
3. Establishment of a standardized benchmarking framework including dataset preparation, augmentation, evaluation protocols and reproducible baseline benchmark results.
4. Further a systematic and fair benchmarking of three deep learning architectures (MobileNetV2, ResNet50 and EfficientNet-B0) was conducted under identical training conditions to evaluate performance tradeoffs between classification accuracy and computational efficiency for real field wheat disease classification.
5. Using Grad-CAM visualization, model interpretability was examined to highlight discriminative leaf regions contributing to classification decisions.

2. Related Work

2.1 Wheat Disease Datasets and Real-Field Imaging

The performance of AI based plant disease detection systems depends on the quality and variety of the dataset used for model training and evaluation. Recently, various publicly available plant disease datasets have used for research in various agricultural applications like automated disease detection, precision farming, robotics and automation etc. Among these, the PlantVillage dataset and PlantDoc datasets have become widely used resources for plant disease classification and benchmarking [15-18].

In spite of its importance, PlantVillage and other similar datasets were primarily collected under controlled conditions using detached leaves, uniform and white backgrounds and having controlled illumination environments. Consequently, the models trained on these datasets generally failed to handle the variability found in real field including background confusion, leaf overlapping, lighting effects, occlusions and differing disease severity levels [18–20].

To address these limitations, recent studies have focus on the importance of agricultural datasets collected from the real field/farm areas that preserve natural environmental variability. Real-field datasets provide more realistic benchmarks for evaluating model robustness and generalization capability during practical deployment [18–20]. Mahmood et al. demonstrated that real field acquired wheat disease dataset collected under natural environmental conditions significantly improves the robustness of deep learning based disease recognition systems [32]. Similarly, Nigam et al. highlighted the importance of real field collected dataset having realistic disease severity variations for reliable wheat rust disease assessment and monitoring applications [33].

Nevertheless, as compared with publicly available datasets for several horticultural crops, the real field wheat disease datasets are remains relatively limited. Furthermore, the datasets consisting of images according to Indian agro-climatic conditions particularly from major wheat-producing regions such as Haryana are also very limited which restricts the development and evaluation of disease recognition systems intended for deployment in local agricultural environments.

2.2 Deep Learning-Based Wheat Disease Detection

In the previous studies, the traditional image processing techniques are combined with a classical machine learning classifier. These approaches use handcrafted features like color, texture, shape etc. using handcrafted feature extraction methods. These features was then combined with classical machine learning classifiers for classification like support vector machine, decision tree etc.[16, 28]. However, these approaches were effective under controlled conditions only and their performance degrades significantly when deployed in complex field environments due to noise and lighting variability in data [29].

Nowadays, the deep learning techniques used for the image based plant disease recognition system which can automatically extract the important features from the raw image data [13, 14]. Moreover, the Convolutional neural network (CNN) based approaches showed superior performance as compared to handcrafted methods. Studies by Ferentinos, Too et al. and subsequent researchers have shown that architectures such as ResNet, MobileNet, EfficientNet and DenseNet can achieve high classification performance when trained on sufficiently large datasets [13, 25–27].

Recently various studies show the use of deep learning approaches for the development of automated wheat disease detection systems [21, 22, 31]. However, these studies reported motivating classification results but many of these studies relied on controlled datasets or evaluate only single model architecture makes comparison difficult. Moreover, various factors like differences in dataset composition, experimental protocols and evaluation methodologies bound reproducibility and cross study comparison. Therefore, launching standardized benchmark results on real field collected datasets provides scope for advancing wheat disease recognition research.

2.3 Explainable AI for Agricultural Disease Recognition

Furthermore, the deep learning models have achieved high classification accuracy but it is difficult to interpret the decision making process of these models. This issue decreases the trust and reliability on the models intended to deploy in real world agricultural scenarios. The farmers, agronomists and agricultural stakeholders also showing less interest in adoption of these agricultural applications where trust and reliability is the major concern.

Therefore, explainable artificial intelligence (XAI) techniques have increasingly been combined with automated plant disease recognition systems to address this issue. Methods like Gradient-weighted Class Activation Mapping (Grad-CAM) enable visualization of image regions which contribute most strongly to model predictions. Recent studies have clearly demonstrated that Grad-CAM can help to verify that whether classification decisions are based on disease symptoms or some irrelevant background information. This improves the model transparency and trustworthiness [31, 34, 36].

In real field automated agricultural applications where images data were used, the integration of explainability mechanisms is particularly very important for disease diagnosis systems as the environmental complexity influences the model behavior and prediction reliability due to background noise.

2.4 Research Gap and Motivation

The literature review highlights several important limitations that continue to hamper the development of robust automated wheat disease recognition systems.

First, there is a limited availability of region specific wheat disease datasets collected under natural agricultural conditions especially from major wheat-producing regions of India. Second, many existing studies rely on controlled datasets that lacks real field variability and then restricting model deployment in real agricultural applications. Third, there is lack of standardized benchmark resources which enable reproducible comparison of deep learning architectures on realistic wheat disease datasets. Finally, very few studies simultaneously combine real-field dataset development, baseline benchmarking and explainability analysis within a integrated framework.

To address these limitations, the present study introduces WheatDisease-HRY, a real-field wheat disease dataset collected under natural agricultural conditions in Haryana, India. In addition to comprehensive dataset development and characterization, the study establishes baseline benchmark results using representative deep learning architectures and employs Grad-CAM based explainability analysis to support transparent and reproducible evaluation of wheat disease recognition systems.

3. Dataset Description

3.1 Dataset Overview

This study introduces WheatDisease-HRY, a region specific real field wheat leaf disease image dataset collected from the experimental farm area of CCS Haryana Agricultural University (CCS HAU), Hisar, Haryana, India. The dataset is designed to address the lack of publicly available real field wheat disease datasets from major wheat producing regions of India, particularly Haryana. The primary objective of collecting the dataset from this geographical region is to capture disease occurrence and symptom variability influenced by local agro climatic conditions.

WheatDisease-HRY was developed to address the limited availability of region specific real field wheat disease datasets representing Indian agro climatic conditions. Unlike datasets collected under controlled laboratory environments, WheatDisease-HRY preserves natural field variability encountered during practical agricultural operations, thereby providing a realistic resource for developing and evaluating disease recognition systems.

Dataset / Study	Year	Crop Focus	Images	Classes	Acquisition Environment	Region	Benchmark Baselines
PlantVillage (Mohanty et al.)	2016	Multiple crops	54,306	38	Controlled laboratory conditions	USA	Limited benchmarking
PlantDoc (Singh et al.)	2019	Multiple crops	2,598	Up to 17	Real-world images	Multiple regions	Yes
Wheat Leaf Dataset (Getachew et al.)	2021	Wheat	407	3	Real wheat farm, uncontrolled environment	Ethiopia	CNN benchmarks provided
Wheat Disease Field Dataset (PPA, UK & Ireland)	2022	Wheat	19,160	5	Real-field conditions	UK & Ireland	Multiple CNN benchmarks
WheatRust21 Dataset	2024	Wheat	Multi-class rust dataset	4+	Real agricultural imagery	Multi-region	Yes
WheatDisease-HRY (Proposed)	This study	Wheat	2,672	3	Real-field conditions	Haryana, India	MobileNetV2, ResNet50, EfficientNet-B0

Table 1: Comparison of WheatDisease-HRY with representative plant and wheat disease datasets reported in the literature

This table compares WheatDisease-HRY with representative plant and wheat disease datasets reported in the literature. While widely used resources such as PlantVillage have significantly advanced plant disease recognition research, many existing datasets were collected under controlled conditions or lack representation of Indian wheat-growing environments. In contrast, WheatDisease-HRY was developed using images acquired directly from agricultural fields under natural conditions and provides region-specific benchmark data for wheat disease recognition research[15-20].

Yellow Rust (a)	Brown Rust (b)	Healthy Wheat Leaf (c)
		

Figure 1. Sample images of (a) Yellow Rust, (b) Brown Rust and (c) Healthy wheat leaves from the WheatDisease-HRY dataset collected under natural illumination and background variability

The dataset includes 2672 RGB images of wheat leaves belonging to three agriculturally relevant classes: yellow rust, brown rust, and healthy leaves. The dataset includes substantial intra class variability due to differing disease severity levels and field acquisition condition.

The dataset shows a near balanced distribution by preserving naturally occurring disease incidence patterns shown during field collection. This provides fair benchmarking and comparative evaluation of disease recognition algorithms by keeping agricultural realism.

Images were captured under natural real field conditions. These collected images having natural variations in lighting, background variation, leaf orientation, leaf overlapping and disease severity. This variability is important for developing and evaluating disease detection systems used for deployment in real life agricultural applications as highlighted in previous studies [11–13]. The dataset is designed to handle real field agricultural variability faced during practical deployment.

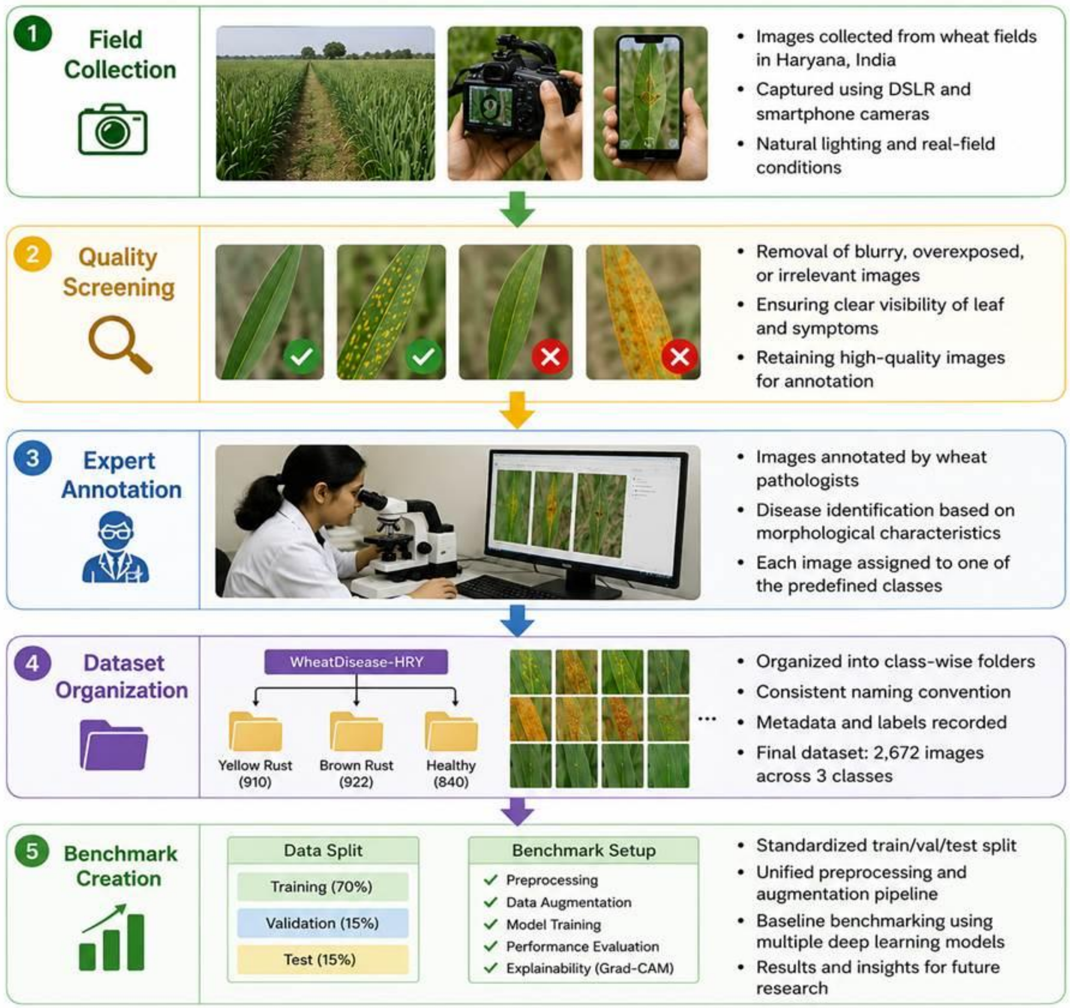

Figure 1. Workflow of WheatDisease-HRY dataset development from field collection to benchmark creation.

Figure 2: Workflow of WheatDisease-HRY dataset development from field collection to benchmark creation.

3.2 Image Acquisition Protocol

Image acquisition was carried out during the wheat growing season to capture the various disease symptoms. Data were collected in two ways which was directly from standing crops in the farm area and some by detaching leaves instantly to preserve natural visual characteristics.

Images were captured using:

- DSLR camera for high resolution image acquisition
- Smartphone cameras to reflect practical field usage

All images were captured under natural daylight conditions without artificial lighting and background control. The variation in sunlight intensity, shadow patterns, soil background and surrounding vegetation was retained to enhance dataset realism and generalization capability [12,13].

Parameter	Description
Location	CCS Haryana Agricultural University, Hisar, India
Environment	Real-field agricultural conditions

Image Type	RGB color images
Capture Devices	DSLR camera and smartphone cameras
Lighting Condition	Natural daylight
Background	Natural field background
Image Orientation	Varying leaf angles and orientations
Acquisition Period	Active wheat growing season

Table 2. Image Acquisition Details

Images were acquired using both DSLR and smartphone cameras under natural lighting to maximize dataset diversity and realism. Consequently, the dataset also captures substantial variability in viewing angles, image resolution, leaf orientation and environmental conditions that are commonly encountered during practically deployment.

Images acquisition was conducted in experimental wheat fields of CCS Haryana Agricultural University, Hisar, Haryana, India. Images were collected directly from standing crops during the growing season under naturally occurring environmental conditions without artificial lighting, background manipulation or controlled image acquisition settings. The dataset reflects real field characteristics having varying sunlight intensity, partial shadows, overlapping of leaves and heterogeneous backgrounds. The images were collected at the various stages of the disease symptoms growth level to generalize the model applicability for practically deployment.

3.3 Disease Categories and Dataset Composition

The WheatDisease-HRY dataset includes three major classes of wheat crop leaves:

- Yellow rust: having yellow to orange pustules arranged in stripes along leaf veins.
- Brown rust: shows reddish brown pustules scattered irregularly on leaf surfaces.
- Healthy leaves: leaves showing no visible disease symptoms.

These diseases are important wheat diseases and responsible for significant yield losses if not detected timely and managed [3–5].

Class	**Number of Images**	**Percentage (%)**
Yellow Rust	910	34.06
Brown Rust	922	34.51
Healthy	840	31.43
Total	2672	100

Table 3. Class-wise Distribution of WheatDisease-HRY Dataset

The dataset contains almost same number of diseases for each disease category collected according to their natural occurrence in real field as the real field disease occurrence very rare follows uniform patterns [11].

The selected disease categories shown major wheat crop diseases and important for developing and evaluating automated disease diagnosis systems for practical agricultural applications. In addition to this, the healthy leaf samples also included to trained the model difference between diseased and non diseased plants.

3.4 Annotation and Validation

A structured annotation workflow with image screening, disease identification, expert verification and quality assessment was implemented during dataset development to ensure dataset reliability and benchmarking.

The WheatDisease-HRY dataset images were manually validated by wheat plant pathologists from CCS Haryana Agricultural University, Hisar, Haryana, India. The disease labels were assigned according to disease symptoms and severity. Low quality images were also reviewed for labeling consistency and accuracy.

Accurate annotation for dataset is important as labelling noise can negatively affect the performance and reliability of image based automated disease recognition models [6, 11]. They also provide a reliable base for benchmarking and comparison evaluation.

During quality control the unclear or poor quality samples were excluded. The final annotation process assured that each image was assigned to a single disease category based on visually shown symptoms and expert domain knowledge.

3.5 Dataset Challenges and Real-Field Variability

WheatDisease-HRY dataset has intentionally consists the complexity of practical real field agricultural environments. Images in datasets have contain required natural variability arise due to natural illumination, leaf orientation, background confusion, disease severity, camera viewpoints and environmental conditions in real field. In addition to this, overlapping leaves, soil backgrounds, partial occlusion and neighboring vegetation are also present in the dataset. These characteristics increase classification difficulty but provide a more real benchmark for evaluating the robustness and generalization capability of disease recognition systems [30].

3.6 Data Preparation Pipeline

A consistent preprocessing procedure was applied to all images before model development for the standardized and reproducible performance evaluation framework. After that, the model specific preprocessing functions corresponding to pre trained architecture were applied during training. The preprocessing steps include:

- Image resizing to model specific input dimensions
- Removal of corrupted or low quality samples
- Model specific normalization and preprocessing functions corresponding to pre-trained architectures

These steps ensures disease related visual features and reduce inter device variability.

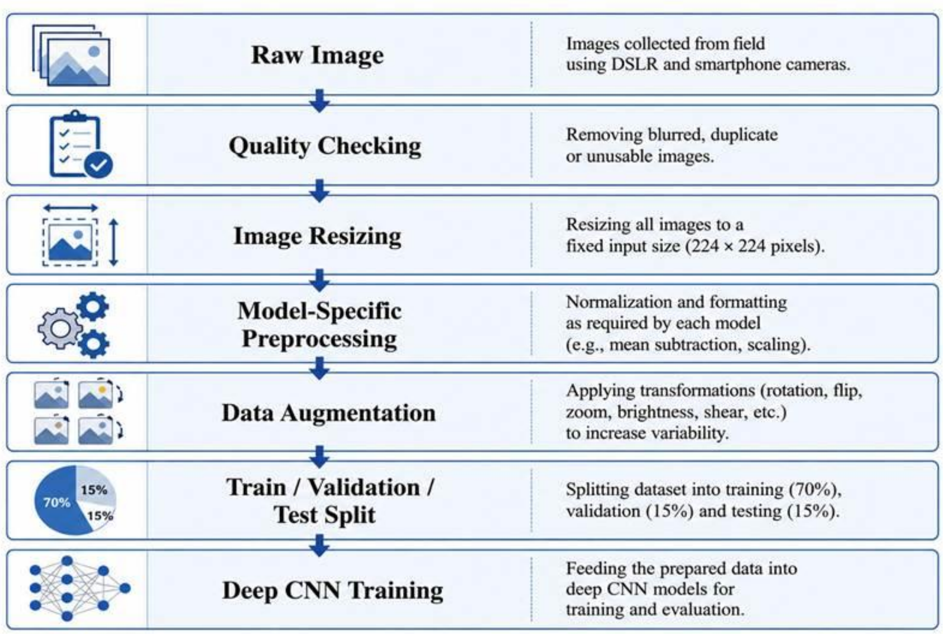

Figure 3. Data preparation and preprocessing workflow used for WheatDisease-HRY dataset standardization and model training

3.7 Data Augmentation Strategy

Various data augmentation techniques are applied during training of the model to improve generalization and reduce overfitting. This ensures real world variations in the dataset and improves model robustness under unseen field conditions [18–20]. The following transformations were carried out to increase the volume of dataset:

- Random rotations
- Horizontal and vertical flipping
- Zooming and scaling
- Brightness and contrast adjustment

The above augmentation techniques were applied only to the training subset. The augmentation operations were applied only on dataset during benchmark model development. The original dataset kept as unchanged and serves as the primary benchmark resource introduced in this study.

Augmentation Type	Purpose
Rotation	Simulate leaf orientation variability
Horizontal/Vertical Flip	Improve invariance to viewpoint
Zoom & Scaling	Address distance variation
Brightness Adjustment	Handle illumination changes
Contrast Adjustment	Enhance robustness to lighting conditions

Table 4. Data Augmentation Techniques Used

3.8 Benchmarking Readiness and Research Utility

The WheatDisease-HRY dataset has been designed as a reusable research resource for agricultural computer vision and precision agriculture applications. The dataset can support future usage in disease classification, transfer learning, explainable artificial intelligence, lightweight model development and mobile assisted disease diagnosis systems. Recent studies have shown that real field datasets having natural lighting variation, background effect and disease severity variety are important for improving the real world application deployment capability of AI based agricultural systems [32].

The dataset and benchmarking results will be used by the researchers working on AI based disease detection systems. By providing a region specific real field dataset along with standardized benchmark results, the WheatDisease-HRY dataset becomes a reference platform for future comparative evaluation of wheat disease recognition systems under realistic agricultural conditions.

4. Benchmarking Framework

4.1 Experimental Objective

The main purpose of the study is not to advise a novel classification architecture but to establish reproducible baseline benchmarks for the WheatDisease-HRY dataset. These benchmarks provide reference performance levels to facilitate future comparative studies. This can also support the evaluation of disease recognition approaches under realistic field conditions.

All benchmark architectures were trained and tested using identical dataset partitions, preprocessing procedures, augmentation settings, optimization strategies and evaluation metrics to ensure a fair and reproducible evaluation. The resulting benchmark performance serves as a reference baseline for future studies using the WheatDisease-HRY dataset.

4.2 Dataset Splitting Strategy

The dataset is divided into three groups training (70%), testing(15%) and validation(15%) using a stratified sampling technique to ensure unbiased evaluation and reproducibility. Stratification technique preserves the original class distribution across all subsets which is particularly important for preserving the original class distribution and ensuring representative evaluation across disease categories. [18, 20]. The said splitting is carried out using a fixed random seed to maintain class distribution and to ensure reproducibility of experimental results. All the benchmarked deep learning models were evaluated under identical experimental conditions.

Class	**Training (70%)**	**Validation (15%)**	**Testing (15%)**	Total
Yellow Rust	637	136	137	910
Brown Rust	645	138	139	922
Healthy	589	127	124	840
Total	1871	401	400	2672

Table 5. Train–Validation–Test Split of WheatDisease-HRY Dataset

A standardized dataset partitioning strategy was adopted to ensure fair and reproducible comparison across all benchmark architectures. The validation subset was also used for hyperparameter monitoring and early stopping and the test subset remains unseen during the training and validation phases.

4.3 Deep Learning Models for Benchmarking

Three representative deep learning architectures were selected for benchmarking to provide a comparison across different model complexities:

- ResNet50
- MobileNetV2
- EfficientNet-B0

These architectures were used in plant disease detection systems and serve as standard baselines for comparative evaluation [9, 13, 18].

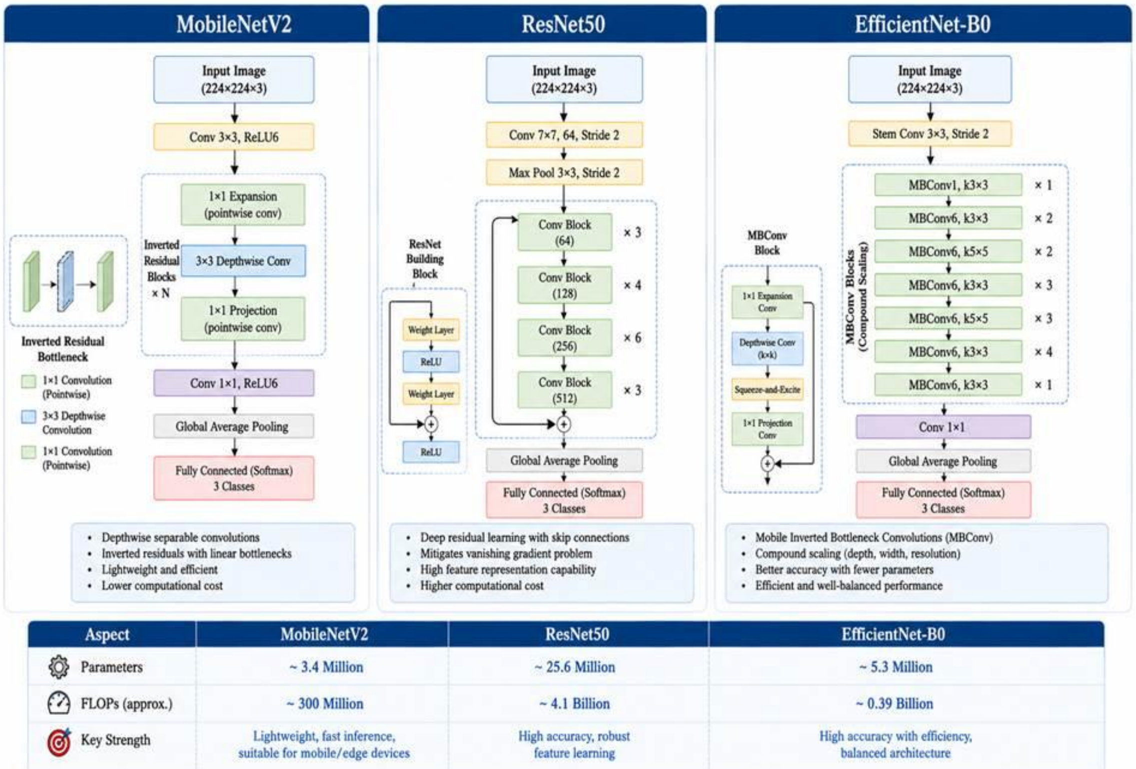

Figure 4. Overview of benchmarked deep CNN architectures used in this study i.e. MobileNetV2, ResNet50 and EfficientNet-B0.

All three architectures were initialized using ImageNet pretrained weights and finetuned under similar experimental conditions for fair benchmarking. The selected architectures signify diverse design ideas and deployment scenarios. MobileNetV2 was chosen as a lightweight architecture suitable for mobile and edge based agricultural applications. ResNet50 signifies a widely adopted residual learning framework with strong feature extraction capability. EfficientNet-B0 provides a balanced trade off between model complexity and predictive performance. Therefore, these models establish representative baseline benchmarks for evaluating the WheatDisease-HRY dataset.

4.4 Hardware/Software details

The experiments were conducted using the Python programming language along with some standard scientific libraries eg NumPy, OpenCV and Matplotlib. All experiments were conducted using the Google Colab cloud computing platform. The deep learning models were implemented using the TensorFlow framework with the Keras backend. NVIDIA Tesla T4 GPU acceleration, TensorFlow and Python were used for model training and evaluation. The Colab runtime environment provided sufficient system memory for handling image preprocessing, batch loading and model optimization during training. All experiments were executed using a fixed random seed to ensure reproducibility of results.

4.5 Training Configuration

All models were trained under identical experimental conditions to ensure fair comparison. The following configuration is used:

- Optimizer: Adam
- Loss function: Categorical cross-entropy
- Initial learning rate: 0.0001
- Batch size: 32

- Number of epochs: 20
- Early stopping was applied to prevent overfitting

Data augmentation is applied only to the training set, as described in Section 3.7, to avoid information leakage during evaluation.

Parameter	Value
Optimizer	Adam
Loss Function	Categorical Cross-Entropy
Learning Rate	0.0001
Batch Size	32
Epochs	20
Input Image Size	Model-specific
Augmentation	Training set only

Table 6. Training Parameters Used for Benchmarking

4.6 Evaluation Metrics

The performance of benchmarked models are evaluated using standard classification metrics used for agricultural image analysis [6, 18, 20]:

- Accuracy
- Precision
- Recall
- F1-score

These metrics provide a balanced evaluation of classification under multiclass real field conditions. Confusion matrices were also analyzed to observe class wise prediction behavior. Since the dataset consists of three classes with near balanced representation, accuracy, precision, recall and F1-score jointly provide a comprehensive assessment of classification performance.

4.7 Benchmarking Protocol

To ensure reproducibility and fairness:

- All models were trained using same training, validation and testing partitions.
- No test images were used during training and validation.
- Final performance was reported exclusively on the unseen test subset.
- Evaluation metrics were computed for all models.

These standardized benchmarking provide reliable comparison between different architectures and used as reference for future studies using this dataset. Confusion matrices and Grad-CAM visualizations were also analyzed to interpret class level prediction behavior and model interpretability.

4.8 Reproducibility and Availability

All experiments were conducted under similar environment to ensure consistent and reproducible results. Standardized dataset partitions, training configurations and evaluation metrics were used for every architecture to ensure fair comparison across these architectures. The benchmark results will be used as reference performance baselines for future studies on the WheatDisease-HRY dataset. The dataset will be made available in accordance with institutional data sharing policies and applicable regulations. The WheatDisease-HRY dataset will be made available only as per the institutional data sharing policies.

5. Results and Discussion

5.1 Overall Benchmarking Performance

This section presents the benchmarking results of three deep learning architectures evaluated on the WheatDisease-HRY test dataset. The objective of this analysis is to provide baseline performance levels for real field wheat disease detection under consistent experimental conditions.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
MobileNetV2	88	88	88	88
ResNet50	91	92	91	91
EfficientNet-B0	91	91	91	91

Table 7. Overall classification performance of the benchmarked models in terms of accuracy, precision, recall and F1-score.

The results showed that all three architectures achieve robust performance on the WheatDisease-HRY dataset. It also highlights the effectiveness of deep learning based approaches for real field wheat disease classification. Among all models, ResNet50 and EfficientNet-B0 attains the highest classification accuracy of 91% and superior precision, recall and F1-score. MobileNetV2 shows the competitive results by representing strong generalization capability across disease classes.

These findings will serve as a baseline reference performance level for WheatDisease-HRY when models are evaluated under similar experimental conditions. These benchmarks beneficial for future comparative studies and simplify the unbiased assessment of newly developed disease recognition approaches.

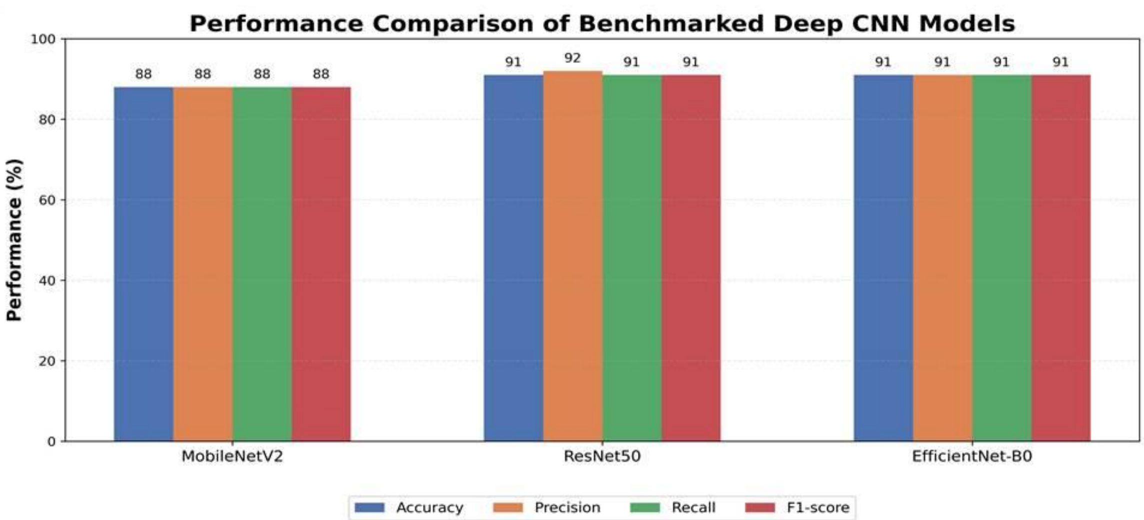

Figure 5. Comparative performance analysis of MobileNetV2, ResNet50 and EfficientNet-B0 using accuracy, precision, recall and F1-score metrics under real-field wheat disease classification conditions.

5.2 Class-Wise Performance Analysis

To gain deeper insight into model behavior, the classwise precision, recall and F1 score are analyzed for the three disease categories viz yellow rust, brown rust and healthy leaves.

Class	Precision (%)	Recall (%)	F1-Score (%)

Yellow Rust	93	85	89
Brown Rust	89	94	91
Healthy	93	94	94

Table 8. Class-wise performance of EfficientNet-B0 on the WheatDisease-HRY test dataset.

The class wise evaluation of EfficientNet-B0 shows strong classification performance across all disease categories. The model achieves the highest precision and recall for the healthy class which clearly shows the difference between diseased and non diseased samples. But slightly lower performance is observed for yellow rust and brown rust. The yellow rust remains comparatively difficult to classify due to visual similarity with brown rust. Moreover, the balanced precision & recall highlights the robustness of EfficientNet-B0 for multiclass wheat disease identification. Furthermore, EfficientNet-B0 is highlighted as a representative model because its performance was comparable to that of the top performing benchmark models.

The observed differences across classes therefore highlight the integral complexity of the WheatDisease-HRY dataset. The presence of disease classes having similar visual patterns symptoms under natural field conditions increases classification complexity and makes the dataset suitable for evaluating model robustness and fine tune the disease classification capability.

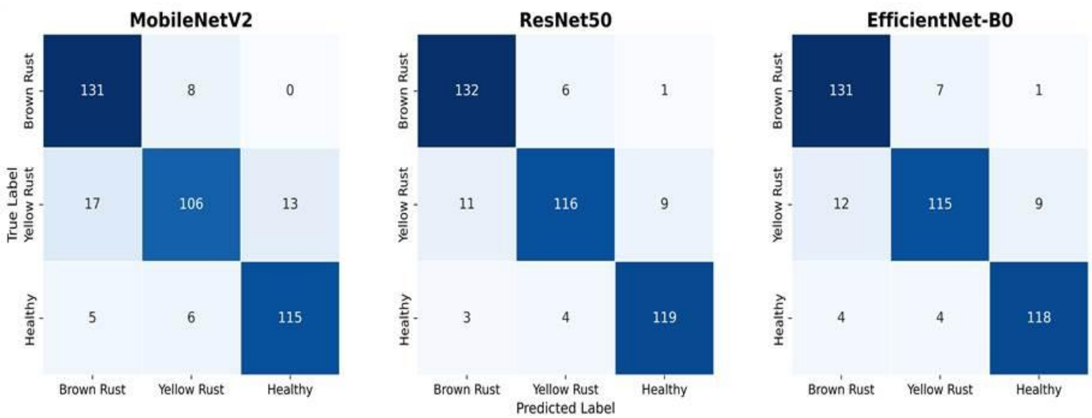

Figure 6. Confusion matrices of MobileNetV2, ResNet50, and EfficientNet-B0 showing class-level prediction behavior on the WheatDisease-HRY test dataset under real-field conditions

5.3 Impact of Real Field Conditions

The WheatDisease-HRY dataset contains agriculture real field conditions like background clutter, leaf orientation, variations in illumination and disease severity. Previous studies also highlighted the importance of real field collected dataset for robust model evaluation [10,11,18,19]. Although, these factors increase the classification complexity but important for the model generalization [11–13].

All three benchmarked models performance highlights the practical relevance of real field collected dataset used in disease detection systems. Models trained and evaluated on real field images demonstrate greater reliability for deployment in real world applications compared with models developed using controlled datasets such as PlantVillage [10–12].

WheatDisease-HRY dataset contains real field environmental variability faced during practical agricultural operations. Therefore, the benchmark results obtained on this dataset provides a more realistic performance and generalization capability of the model to be deploy in real field agricultural applications.

5.4 Comparison Across Model Architectures

The benchmarking results show a clear balance between model complexity and classification accuracy. ResNet50 and EfficientNet-B0 showed comparable performance with approximately 91% classification accuracy under real field conditions. ResNet50 provides more improved sensitivity toward rust affected regions while EfficientNet-B0 provided more balanced class level performance. MobileNetV2 achieved competitive accuracy with lower computational complexity making it suitable for edge based agricultural deployment. This can prove that the deep learning models can successfully classify wheat diseases under real field datasets.

Overall, the three architectures demonstrate that WheatDisease-HRY dataset can support both high capacity and lightweight deep learning models. The availability of reference results provides a useful baseline for future research focused on either maximizing classification accuracy or developing lightweight disease diagnosis systems.

5.5 Discussion and Implications

The main aim of the study is to develop region specific real field wheat disease dataset consisting of practical agricultural conditions in Haryana, India. The dataset also addresses the gap of publicly non availability of wheat disease dataset having real field challenges.

The benchmark results should be interpreted as baseline reference performance levels instead of the maximum achievable accuracy on the WheatDisease-HRY dataset. Future studies using advanced architectures, domain adaption techniques, transformer based approaches or multimodal learning strategies may further improve performance on the WheatDisease-HRY dataset [35-37].

For precision agriculture, the real field acquired datasets are essential for developing deployable crop monitoring systems. WheatDisease-HRY provides a valuable resource for evaluating disease diagnosis models under agricultural real field environmental conditions.

The direct comparison of classification performance among different studies is challenging due to different datasets, data acquisition conditions, disease categories and evaluation frameworks. Therefore, by providing a region specific real field benchmark along with baseline results, WheatDisease-HRY dataset provides a valuable benchmark resource for future comparative research. This also supports the development of robust disease recognition systems for precision agriculture applications.

5.6 Model Interpretability Using Grad-CAM

The study uses Grad-CAM to identify the diseased relevant regions which affects model predictions. It generates class specific heatmaps by analyzing the gradients of the target class in respect to the final convolutional feature maps. This analysis was performed using the same diseased wheat leaf image across all benchmarked architectures to evaluate model decision.

The visualization results also indicate that the three benchmarked models successfully focused on disease relevant regions of wheat leaves. MobileNetV2 exhibited relatively broader activation patterns whereas ResNet50 and EfficientNet-B0 produced more localized attention toward rust affected areas. EfficientNet-B0 particularly indicates stronger focus on lesion dense regions while suppressing irrelevant background information. This indicates the improved discriminative feature learning capability of the model.

These findings indicate that the background complexity did not affect the model predictions. Therefore, the benchmarked models successfully learned meaningful pathological features. This improves the reliability and robustness of the proposed wheat disease classification model under real field agricultural conditions. Similar remarks have also been reported in recent explainable AI studies where Grad-CAM based visualization techniques improve the interpretability and trustworthiness of plant disease classification systems [34, 36].

The Grad-CAM visualizations also support the validity of the WheatDisease-HRY dataset by representing that model predictions are primarily guided by disease-relevant regions rather than background confusions. This observation increases confidence in the reliability of the benchmark results.

Figure 7. Grad-CAM visualization comparison across MobileNetV2, ResNet50 and EfficientNet-B0 using the same diseased wheat leaf image. Highlighted regions indicate disease-specific areas influencing model predictions under real-field conditions.

5.7 Error Analysis

The present study has observed misclassifications between yellow rust and brown rust classes due to overlapping visual texture and color characteristics of the disease in real field conditions. This confirms that the classification errors are due to fine grained disease discrimination not due to background variability.

The observed misclassifications also provide the scope for future research in fine grained disease recognition, attention based architectures, disease severity estimation and multimodal learning approaches etc. These challenges also highlight the practical relevance of WheatDisease-HRY as a benchmark dataset for developing more robust disease recognition systems.

5.8 *Limitations*

The present study has various limitations highlighting scope for future study like dataset expansion, increased diversity in data collection and the interpretation of more deep learning methodologies. These are as follows:

- The WheatDisease-HRY dataset includes only three wheat leaves disease categories.
- The images were collected from the Haryana region and have limited geographic variability.
- Only CNN based architectures are used for benchmarking analysis.

6. Conclusion and Future Work

This study presented WheatDisease-HRY a region specific real field wheat disease image dataset developed for automated disease recognition research under practical agricultural conditions. The dataset consists of 2672 expert annotated images with three classes of wheat crop diseases yellow rust, brown rust and healthy wheat leaves. The data was collected from the farm fields areas of CCS Haryana Agricultural University, Hisar, India. The WheatDisease-HRY dataset has natural field variability like illumination effects, background complexity, leaf overlapping, occlusion and varying disease severity levels. This provides a real field collected dataset resource for agricultural computer vision applications.

Three deep learning architectures MobileNetV2, ResNet50 and EfficientNet-B0 were evaluated using a standardized benchmarking framework to provide reference performance baselines. The findings of the study show that the deep learning models can effectively classify wheat diseases under real field conditions. ResNet50 and EfficientNet-B0 gets the overall highest classification performance. The Grad-CAM based explainability analysis also confirms that the models basically focused on disease relevant regions instead of background confusion.

Both the finding of this study, region specific real field dataset and baseline performance benchmarks will be used as an important resource for further study in wheat disease recognition systems. Futuristic scope for further research will be the expansion of the dataset by including more wheat diseases categories with various disease severity levels, growth stages and geographical locations. Advanced deep learning architectures, multimodal learning approaches, domain adaptation techniques, transformer based models and lightweight deployment may also be explored. The incorporation of UAV imagery, mobile based disease diagnosis systems and real time decision support frameworks also provides directions for enhancing practical field deployment. In nutshell, the continued development and utilization of WheatDisease-HRY can contribute toward robust, scalable and interpretable AI based crop health monitoring systems.

7. Data Availability Statement

The WheatDisease-HRY dataset will be made available upon reasonable request to the corresponding author according to institutional policies and applicable regulations.

8. Ethics Statement

The study did not involve human participants or animals. Plant image data were collected from agricultural research fields with the prior permission of the organization.

References

1. Food and Agriculture Organization (FAO). FAOSTAT Statistical Database. Rome: FAO; 2023.
2. Government of India. Agricultural Statistics at a Glance 2022. New Delhi: Ministry of Agriculture & Farmers Welfare; 2022.
3. Singh RP, Huerta-Espino J, Roelfs AP. The wheat rusts: breeding for resistance and management. Plant Dis. 2002;86(5):465–479.
4. Huerta-Espino J, Singh RP, German S, McCallum BD, Park RF, Chen WQ, et al. Global status of wheat leaf rust caused by *Puccinia triticina*. Annu Rev Phytopathol. 2011;49:457–481.
5. Chen XM. Epidemiology and control of stripe rust (*Puccinia striiformis* f. sp. *tritici*) on wheat. Can J Plant Pathol. 2005;27(3):314–337.
6. Savary S, Willocquet L, Pethybridge SJ, Esker P, McRoberts N, Nelson A. The global burden of pathogens and pests on major food crops. Nat EcolEvol. 2019;3:430–439.
7. Fischer RA, Byerlee D, Edmeades GO. Crop yields and global food security. Philos Trans R Soc B Biol Sci. 2010;365(1554):2831–2840.
8. Singh RP, Hodson DP, Jin Y, Huerta-Espino J, Kinyua MG, Wanyera R, et al. Emergence and spread of new races of wheat stem rust fungus. Annu Rev Phytopathol. 2011;49:465–481.
9. Sharma I, Saharan MS, Sharma P, Kumar J. Climate change and wheat rust diseases in India. J Wheat Res. 2016;8(1):1–10.
10. Mahlein AK. Plant disease detection by imaging sensors – parallels and specific demands for precision agriculture and plant phenotyping. Annu Rev Phytopathol. 2016;54:45–67.
11. Barbedo JGA. Digital image processing techniques for detecting, quantifying and classifying plant diseases. SpringerPlus. 2013;2:660.
12. Pretty J, Bharucha ZP. Sustainable intensification in agricultural systems. Environ Int. 2018;114:157–165.
13. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436–444.
14. Ferentinos KP. Deep learning models for plant disease detection and diagnosis. Comput Electron Agric. 2018;145:311–318.
15. Mohanty SP, Hughes DP, Salathé M. Using deep learning for image-based plant disease detection. Front Plant Sci. 2016;7:1419.
16. Hughes DP, Salathé M. An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv preprint arXiv:1511.08060. 2015.
17. Singh D, Jain N, Jain P, Kayal P, Kumawat S, Batra N. PlantDoc: A dataset for visual plant disease detection. Proceedings of the 7th ACM IKDD CoDS and 25th COMAD. 2020:249–253.
18. Kamilaris A, Prenafeta-Boldú FX. Deep learning in agriculture: a survey. Comput Electron Agric. 2018;147:70–90.
19. Barbedo JGA. Factors influencing the use of deep learning for plant disease recognition. Biosyst Eng. 2018;172:84–91.
20. Lu Y, Yi S, Zeng N, Liu Y, Zhang Y. Identification of rice diseases using deep convolutional neural networks. Comput Electron Agric. 2019;141:278–289.
21. Zhang S, Huang W, Zhang C. Wheat disease recognition using convolutional neural network. IEEE Access. 2019;7:146673–146682.
22. Picon A, Alvarez-Gila A, Seitz M, Ortiz-Barredo A, Echazarra J, Johannes A. Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild. Comput Electron Agric. 2019;161:280–290.
23. Singh A, Ganotra D, Sharma R. Challenges in AI-based crop disease detection under field conditions. Smart Agric Technol. 2022;2:100049.
24. Saleem MH, Potgieter J, Arif KM. Plant disease datasets and benchmarking challenges: a review. Neural Comput Appl. 2022;34:19047–19075.

25. Too EC, Yujian L, Njuki S, Yingchun L. A comparative study of fine-tuning deep learning models for plant disease identification. Comput Electron Agric. 2019;161:272–279.
26. Saleem MH, Potgieter J, Arif KM. Plant disease classification using deep learning architectures. Sensors. 2019;19(23):1–24.
27. Saleem MH, Potgieter J, Arif KM. Benchmarking deep learning models for plant disease recognition. Sensors. 2021;21(9):3025.
28. Patil JK, Kumar R. Feature extraction of diseased leaf images. Int J Comput Appl. 2011;17(1):1–4.
29. Barbedo JGA. A review on the main challenges in automatic plant disease identification based on visible range images. Biosyst Eng. 2016;144:52–60.
30. Severino R, Gomes J, Cunha M. CNN robustness under real-field agricultural conditions. Comput Electron Agric. 2020;175:105589.
31. Lu J, Hu J, Zhao G, Mei F, Zhang C. An in-field automatic wheat disease diagnosis system. Comput Electron Agric. 2021;142:369–379.
32. Mahmood A, Tariq U, Khan S, et al. Deep learning framework using UAV imagery for multi-class wheat disease detection. Sci Rep. 2025;15:10234.
33. Nigam S, Verma P, Kumar R, et al. Automated severity level estimation of wheat rust using deep learning. Front Plant Sci. 2025;16:1489205.
34. Kondaveeti HK, Reddy PS, Basha SHS, et al. Evaluation of deep learning models using explainable AI for plant disease detection. Sci Rep. 2025;15:11342.
35. Elfouly MK, Hassanien AE, Mahdy LN, et al. A deep learning-based framework for large-scale plant disease detection. J Big Data. 2025;12:67.
36. Gopalan K, Nair MS, Kumar S, et al. Enhancing accuracy with ResNet152 and Grad-CAM for plant disease detection. BMC Plant Biol. 2025;25:214.
37. Rashid MRA, Islam MM, Hossain MS, et al. An ensemble learning framework with explainable AI for plant disease detection. Smart Agric Technol. 2025;8:100612.