REVIEW ARTICLE Indian Journal of Agricultural Research, Volume 59 Issue 5: 681-690 (May 2025)

Leveraging Artificial Intelligence for Disease Diagnosis in Agricultural Crops: A Review

Anita Rani Mehta[1], Pardeep Kumar[1], Guru Prem[2], Shalini Aggarwal[3], Rakesh Kumar[4] 10.18805/IJARe.A-6363

ABSTRACT

The agricultural sector contributes significantly to the gross domestic product of emerging nations. One of the primary factors that reduce agricultural productivity is plant diseases. Detecting plant diseases is a crucial step requiring attention for a productive economy and agriculture. Traditional techniques of crop disease detection are laborious and time-consuming, requiring a great deal of skill and effort. Academicians, researchers and practitioners have recently shown a great interest in the new area of automatic plant disease identification. As soon as they arise on plant leaves, early detection of plant diseases can be aided by machine learning (ML) and deep learning (DL). A comprehensive analysis has been conducted in this review paper to assess the viability to identify plant diseases applying machine learning models. This paper also considers the illnesses and viruses that affect different crops: black gram, chickpea, mungbean and potato. A detailed procedure has been done to recognise and classify plant diseases. The analysis of various ML and DL algorithms adopted by researchers to classify diseases has been explained. Finally, challenges are highlighted to recognise and classify plant diseases utilising ML and DL algorithms, particularly in legume crops. Developing and integrating real-time datasets on plant diseases using advanced models is the future for monitoring, detecting and managing plant ailments.

Key words: Agricultural crops, Deep learning, Machine learning, Plant disease detection.

Food security in India has been a primary national concern since the mid-1960s. The major cereal crops, such as rice and wheat, had substantial expansion from 1960 to 1970 and led to the green revolution mainly due to the emergence of high-productivity crop cultivars, availability of fertilizers and development of irrigation facilities (Roy *et al.,* 2016). However, food grain production has increased significantly from 80 to 250 million ton (Mt) from 1965 to 2015, respectively, in India (Bhattacharjya *et al.,* 2017). Nonetheless, we are self-dependent on food consumption and production, but this expansion occurred at the expense of a reduction in the area of leguminous crops. The pulses are essential in enhancing soil fertility through nitrogen fixation, reducing soil pathogens and environment-friendly crops as they require less fertilizers, chemicals and irrigation water (Nimbrayan and Tripathi, 2019). Meeting the dietary protein requirement through meat requires higher energy, land and water consumption and causes higher greenhouse gas emissions than plant-based alternatives. Therefore, to conserve natural resources, maintain nutritional security in humans and environmentally sustainable agricultural regions, we need to increase the productivity of Pulses. India is producing around 19 Mt pulses, which is insufficient to meet the domestic requirement of about 21 Mt. We need to enhance the production annually by a 2.2% growth rate to achieve a self-reliance target of 39 Mt by 2050 (Pandey *et al.,* 2019). The major pulses grown are chickpea (*Cicer arietinum*), pigeon pea (*Cajanus cajan*) and urd bean (*Vigna mungo*) cultivated on 40%, 20% and 12% acreage out of total area cultivated in India (Pal *et al.,* 2018).

[1]Department of Computer Science and Applications, Kurukshetra University, Kurukshetra-136 119, Haryana, India.
[2]School of Core Engineering, Shoolini University, Solan-173 229, Himachal Pradesh, India.
[3]SUS Government College, Indri-132 041, Karnal, Haryana, India.
[4]Natural Resource Management Division, Indian Council of Agricultural Research, New Delhi-110 012, India.

Corresponding Author: Anita Rani Mehta, Department of Computer Science and Applications, Kurukshetra University, Kurukshetra-136 119, Haryana, India.
Email: anita_mehta.dcsa@kuk.ac.in
ORCID: 0009-0005-7940-5653

How to cite this article: Mehta, A.R., Kumar, P., Prem, G., Aggarwal, S. and Kumar, R. (2025). Leveraging Artificial Intelligence for Disease Diagnosis in Agricultural Crops: A Review. Indian Journal of Agricultural Research. 59(5): 681-690. doi: 10.18805/IJARe.A-6363.

Submitted: 09-01-2025 **Accepted:** 31-03-2025 **Online:** 26-04-2025

Potato (*Solanum tuberosum* L.), is the third most prominant agricultural crop globally (Raymundo *et al.,* 2017) and can potentially mitigate the challenge of nutritional security. The majority of potatoes in India are grown in plains and the three states that produce the most are Uttar Pradesh, West Bengal and Bihar. Together, they account for over 75% of the country's total production, with their respective contributions of 32.38%, 26.94% and 14.56% (Pradel *et al.,* 2019). Abiotic and biotic stresses have significantly reduced legume yields in South Asia and Southeast Asia. Biotic stresses, caused by pathogens like

bacteria, viruses and fungi, are highly transferable and dangerous, whereas physiological factors like nutrient deficiency, sunburn and problematic soils cause abiotic diseases. Fungal infections among biotic stresses may reduce pulse crop yield by 40% to 60% (Kaur *et al.,* 2011). Among various biotic stresses, fungal infections such as Fusarium wilt (*Fusarium oxysporum* f. sp. ciceri) (Nene *et al.,* 2012) and Ascochyta blight caused by necrotrophic fungus *Ascochyta rabiei* (Singh *et al*., 2022) and dry root rot (*Cicer arietinum*) (Sinha *et al.,* 2021) are the major diseases, which may reduce 100% crop yield of chickpea (Gupta and Sharma, 2015). While powdery mildew, yellow mosaic and cercospora leaf spot are considered the major serious diseases in urdbean (Pandey *et al.,* 2019). Early blight of potato is a fungal infection caused by a pathogen named *Alternaria solani* (Murmu *et al.,* 2017) and late blight caused by oomycete, Phytophthora infestans (Chowdappa *et al.,* 2015) have been considered critical bio-stress diseases may cause more than 50% reduction in the tuber yield.

Detecting and diagnosing plant diseases is crucial for global quantitative and qualitative food security standards. Traditional techniques like fluorescence *in situ* hybridization, polymer chain reaction, immunofluorescence and flow cytometry are traditional techniques used for disease detection (Vishnoi *et al.,* 2023). Artificial intelligence (AI) technologies in agriculture offer fast, accurate results through computerized detections and image processing, reducing labour costs and time inefficiencies and improving crop yield and quality. They also facilitate disease control plans by utilizing early crop health data and disease location. These AI advancements have also brought new opportunities to automatic plant disease detection. AI techniques, especially Deep Learning (DL) and Machine Learning (ML), can learn from large datasets to detect diseases in real-time, offering scalable solutions for modern agriculture. ML methodologies use training data to perform tasks using examples described by attributes like nominal, binary, ordinal, or numeric variables. The trained model can classify, predict, or cluster new examples utilizing the training experience obtained after learning. The widely used ML and DL-based approaches include Random Forest (RF), Support Vector Machines (SVM), K-Nearest Neighbor (KNN) and Artificial Neural Networks (ANN) for identifying and detecting plant diseases.

Contribution of the study

These are the following contributions of this study:

- Plant diseases that affect legume and potato crops.
- Procedures to recognise and classify plant diseases.
- Major ML and DL techniques involved in plant illness identification.
- Performance of ML and DL approaches regarding accuracy for disease detection.
- Challenges and issues in using AI-powered disease detection systems.

Organization of the study

This review focuses AI techniques, especially ML and DL approaches used for ailment identification in vegetables and legume crops. It compares various ML and DL techniques and discusses data collection and model deployment challenges. Future research directions are also explored to improve ML and DL-based agricultural crop disease identification systems. The step-wise procedure to identify and classify leaf infections in agricultural crops is given in section 2. Section 3 represents the AI techniques, consisting SVM, RF, K NN and CNNs. Performance assessment of ML and DL techniques to detect disease in legume and vegetable crops are discussed in section 4. Section 5 discusses the challenges that arise in the adoption of AI techniques.

Procedure to identify and classify agricultural crop diseases

The process of detecting plant ailments by applying image processing techniques and soft computing involves acquisition of pictures, pre-processing of pictures, extraction of features and classification or identification (Fig 1). The first step in disease classification using AI is the acquisition of images used as input for models (Abuhayi and Bezabh, 2023). High-resolution digital cameras and smartphones can capture images in formats like jpg, png, jpeg and tiff. Some open-access datasets are also available, *i.e.* Fig 2 shows sample pictures of unhealthy and healthy leaves of black gram from the BPLD dataset. This is an essential step in digital image processing, converting visible pictures into binary records for further processing on a computer. Image preprocessing is the subsequent step, utilizing mathematical/statistical models for transforming images into common formats. It also enhances their geometric characteristics and appearance by using features like noise removal, image smoothing, distortion removal, color conversion and cropping to minimize computational costs and convert the image to required benchmark resolutions. The next step is segmentation, in which images are distributed into various segments according to their characteristics. Image pixel partitioning, grouping, labelling and further classification based on specified labels are all possible with segmentation techniques. Researchers extensively use k-means clustering, Thresholding and Edge-based segmentation techniques in various studies.

Objects can be differentiated from one another by their features, which are relevant and distinctive characteristics. The feature extraction stage, which attempts to extract the suitable characteristics that define each class, is crucial to building the classification/recognition model. Principal component analysis (PCA), histogram of orientated gradients (HOG), gray-level cooccurrence matrix (GLCM), discrete wavelet transformations (DWT) and local binary patterns approaches are typically used as feature extractors

in a variety of investigations (Singh *et al.,* 2022; Harakannanavar *et al.,* 2022). The classifiers use the feature vectors that combine the characteristics derived from DWT, GLCM and PCA as input samples to identify and categorize the images. The classification stage is also important for diagnosing plant diseases with computer vision and image processing methods. In this step, the disease is initially detected and then categorized into known classes, popularly known as classification. The categorization is subdivided into healthy and diseased objects of the studied crop. Nowadays, various ML and DL classifiers are used and in the next section, SVM, RF, K NN and CNNs are described.

AI techniques for disease detection in legume and vegetable crops

Due to the notable developments in AI techniques, the automation of agricultural crop disease identification has recently become feasible and offers novel opportunities for reducing manual labour reliance, enhancing the accuracy of diagnostics and improving cost-effectiveness. AI helps to learn and predict diseases from large datasets using instantaneous detection, ML and DL algorithms, thus providing a feasible solution for modern agriculture scaling. This section focuses on the most commonly utilized AI techniques and their applications for disease diagnosis in legume and vegetable crops.

Support vector machines (SVM)

SVM is an advanced supervised method primarily applied for classification purposes. SVM can detect plant leaf diseases by categorizing leaf images according to attributes like texture, color and shape into healthy and unhealthy categories. In SVM, the main objective is to identify a hyperplane that maximally separates the margin into two classes (healthy and unhealthy). The margin is the distance, expressed in terms of support vectors, between the nearest data points from each class and the hyperplane. If we have a given linearly separable input training dataset, (Xi, Yi).

Where,

$Xi \in \mathbb{R}^n$ = Input feature vector (color and texture features of leaf images).

$Yi \in \{-1, +1\}$ = Class label (diseased or healthy).

The equation of hyperplane can be expressed as:

$$W^T x + b = 0 \quad ....(1)$$

Where,

W = Denotes the weight vector (normal to hyperplane).

b = Represents the bias term.

To address the issue of non-linearly separable data, kernel function may be employed to transform non-linear input data to high-dimensional feature space. Initially, we need to optimize the problem by maximizing the objective function:

$$Q(a) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad ....(2)$$

based on the terms:

$$\sum_{i=1}^{m} \alpha_i y_i = 0 \text{ and } \alpha_i \geq 0$$

Where,

i = 1,2,3,.. .., m

The next step in training the model, the classification of disease with a new set of samples, can be done by applying:

$$f(x) = \sum_{i=1}^{m} \alpha_i y_i K(x_i, x) + b \quad ....(3)$$

Where,

a = Lagrangian multiplier.

m = Represents training samples.

K-nearest neighbors (K-NN)

The K-NN technique is a supervised ML technique widely applied in pattern recognition and data classification applications (Rajaguru and Sannasi Chakravarthy, 2019). This technique assumes that comparable (homogeneous) objects are located adjacent to one another. Initially, we need to choose k, the number of neighbors and determine the distance between each training data point and the new point by applying distance matrix methods like Euclidean, Manhattan, or Hamming distance. Then, choose K nearest neighbors, count the total data points and allocate the new data point to the class with the maximum K neighbors (Goel and Nagpal, 2023). A labelled set is represented as:

$$T = \{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$$

Where,

X_i = Feature vector for the i^{th} leaf.

y_i = Its corresponding disease type label.

The distance metric $d(x_i, x_j)$ can be calculated using the Euclidean equation.

$$d(x_i, x_j) = \sqrt{\sum_{i=1}^{n} (f_i - f_{ij})^2}$$

Where,

f_{ij} = i^{th} feature of the j^{th} leaf vector.

Random forest (RF)

RF is an ensemble approach, as the decision is made by averaging or voting from several decision trees. By lowering variance, the overfitting issue is also significantly reduced in this approach (Hatuwal *et al.,* 2020). The bagging technique generates a number of decision trees from a random selection of data. The output of each decision tree is combined to create the final decision tree. Predictions for new unlabelled data points can be made using the forest once it has been trained. The prediction of each tree in random forest can be determined by using:

$$f_n^{(N)}(x) = \frac{1}{N} \sum_{i=1}^{N} f_n^{(i)}(x)$$

Where,

$f_n^{(N)}(x)$ = Overall prediction of random forest for input x and n denotes number of samples.

N = Total number of trees in the forest.

$f_n^{(i)}(x)$ = Prediction performed by i^{th} tree for the input x.

Convolutional neural networks (CNNs)

Advancements in DL have significantly improved disease detection accuracy. CNNs have emerged as a promising technique for plant-specific applications because they automatically generate low and high-level features. For identifying plant diseases, pre-trained models like AlexNet, CaffeNet, GoogLeNet and VGG can be fine-tuned by applying transfer learning. These models, trained for general object detection, can be used for disease identification in various plants. Convolution is an elementary operation in CNNs that enables the model to effectively determine spatial hierarchies of features, building them particularly powerful for image classification tasks. The expression generally represents the convolution operation for two-dimensional input image P and filter Q:

$$(P*Q)(i, j) = \sum_m \sum_n (P(i - m, j - n)\, Q(m,n)$$

Where,

(P*Q) (i, j) = Output feature map value at position (i, j).

P (i-m, j-n) = Pixel value of the input image.

Q (m, n) = Kernel value.

Analysis of ML and DL algorithms for disease detection

Legume crops (black gram, chickpea and mung bean)

The protein rich important pulse crop cultivated in India is the black gram (*Vigna mungo* L.), used extensively in Indian cooking. The two main diseases that seriously reduce black gram yields for farmers are Anthracnose and powdery mildew. Random Forest Classifier and Multinomial Logistic Regression were used in this study to help farmers identify plant leaf illnesses in black gram plants early in the growth cycle. Images were gathered from the Thanjavur block in the Thanjavur district of Tamil Nadu for the current study. The study's findings demonstrated that the Random Forest Classifier had a higher accuracy, with 97.00% test accuracy and 99.17% training accuracy (Palanichamy *et al.,* 2024). Powdery mildew and Anthracnose illnesses can cause yield reductions in black gram production up to 40-67%. In this study, 2002 images of diseased leaf specimens were gathered from several agricultural regions in the Tanjore district of Tamil Nadu, India. The RMSprop and Adam optimizer have been utilized to improve the learning rate dynamically. CNN is used to detect Powdery mildew and Anthracnose and attained 92.50% accuracy, 97.14% precision 87.17% recall and 91.89% F1 score (Kalpana *et al.,* 2023).

Four deep learning models, SqueezeNet, DarkNet-19, GoogLeNet and AlexNet, have been used in this study to detect yellow mosaic disease of black gram. 1100 images were collected and classified as healthy, moderate and susceptible. The image dataset was split into two parts: 70% to train and 30% to validate the model. DarkNet19 achieved the highest accuracy, 96.09%, followed by AlexNet with 94.41%, GoogleNet with 93.85% and SqueezeNet with 60.74% (Kumar *et al.,* 2023). In this study, DeepLabv3 and MobileNetV2 were used for feature extraction. Various augmentation methods such as correction, rotation, illumination, mirror symmetry, noise injection and randomshifting were applied to increase the dataset to 15000 images. DCNN (Deep CNN) was applied to detect and classify



Fig 1: Steps involved in plant disease detection.

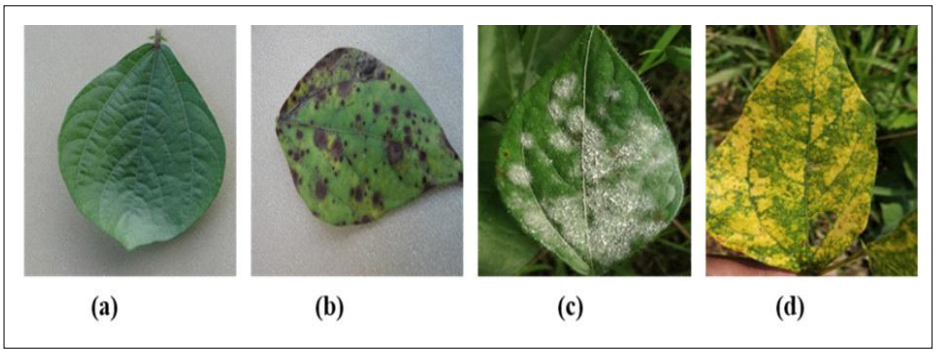


Fig 2: (a) Healthy, (b) Anthracnose, (c) Powdery mildew, (d) Yellow mosaic.

leaf infections in black gram. To determine the model's effectiveness in detecting the infections, the 5-fold cross-validation procedure was applied. The suggested approach attained an accuracy of 99.54%, precision of 98.78%, recall of 98.82% and F1 score of 98.80% (Talasila *et al.,* 2023).

Due to widespread incidence of fusarium wilt infection in chickpea leaves causes lower crop production, which puts farmers in a difficult financial situation. Chickpea yields can be raised by implementing appropriate safeguards and detecting diseases early. This research uses a convolutional neural network to provide an enhanced method for severity-level-based Fusarium wilt disease prediction using 4339 leaf images from Kaggle by analyzing changes in color and shape of chickpea leaves and achieved 74.79% accuracy (AlZubi *et al.,* 2024). Researchers extensively used bilateral filtering and non-local means filtering for preprocessing. GLCM and color histogram were used for feature extraction. After that, Multi-Layer Perceptrons (MLPs), KNN, SVM and RF were utilized for classification and the proposed model (SVM+GLCM) gained a testing accuracy of 95.49% (Abuhayi and Bezabh, 2023). More recently, (Kanade *et al.,* 2025) developed a real-time image dataset of legume crop containing 4480 total leaf pictures. The researchers used different versions of the CNN based model YOLOv8 and found YOLOv8s have comparatively higher precision, recall and F1-score of 97.2, 78.6 and 86.9% respectively in real-world situations.

The study explores the use of machine learning grounded texture examination technique to find the degree of Fusarium wilt in chickpea. The model employs feature extraction methods such as Gray-Level Run-Length Matrix and Gray-Level Occurrence Matrix. The model was trained on 15,000 images, 70% of which were used to train, 20% to validate and 10% to test it. Finally, the proposed model, *i.e.* GLRLM-HSV with KNN, achieved 94.5% accuracy (Hayit *et al.,* 2024). Mallick *et al.,* 2023 conducted a deep learning-based technique to identify the 6 types of diseases in mung bean. Initially, 234 images were collected from different sources. After applying various data augmentation methods, the number of images were enhanced to 8448. CNN was used to identify and classify multiple diseases, namely Cercospora Leaf Spot, Fusarium Wilt, Powdery Mildew, Halo Blight, Yellow Mosaic and Charcoal Rot. The model achieved 93.65% average accuracy. The accuracy of different ML and DL models in legume crops leaf illness classification had been shown in Fig 3.

Discussion

Table 1 summarizes various ML and DL algorithms to identify diseases in legume crops, highlighting the integration of advanced technologies to improve crop health monitoring. This paper studied different legume crops, including black gram, chickpea and mung bean, for various diseases, including Anthracnose, fusarium wilt, yellow mosaic and powdery mildew. In terms of techniques, both ML and DL methods are widely employed. CNNs were the most commonly used DL models because they extract relevant features from images automatically. The evaluation metrics presented in the table emphasize the overall effectiveness of these techniques, with training and testing accuracy consistently high in many studies, particularly in DL approaches. Studies by (Kalpana *et al.,* 2023) and (Mallick *et al.,* 2023) demonstrate CNN's effectiveness, yielding high accuracy of 92.50% and 93.65%, respectively, in detecting diseases in black gram and mung bean. In contrast, machine learning techniques such as RF (N *et al.,* 2024) and SVM (Abuhayi and Bezabh, 2023) have also shown promising results, with RF achieving a high testing accuracy of 97%.

Vegetable crops (potato)

In recent studies, the authors developed a model to classify diseases, including green tuber and black scarf, using 78 potato tuber images collected from fields in potato tuber. Various ML/DL models such as SVM, KNN, decision tree, logistic regression, VGG19, MobileNetV2 and DenseNet201 were used to classify potato tuber diseases. In machine learning, SVM attained the highest accuracy of 80% and deep learning model DenseNet201 attained 99% accuracy (Moawad *et al.,* 2023). By utilizing images from the dataset by plant village, the authors evaluated CNN to identify early

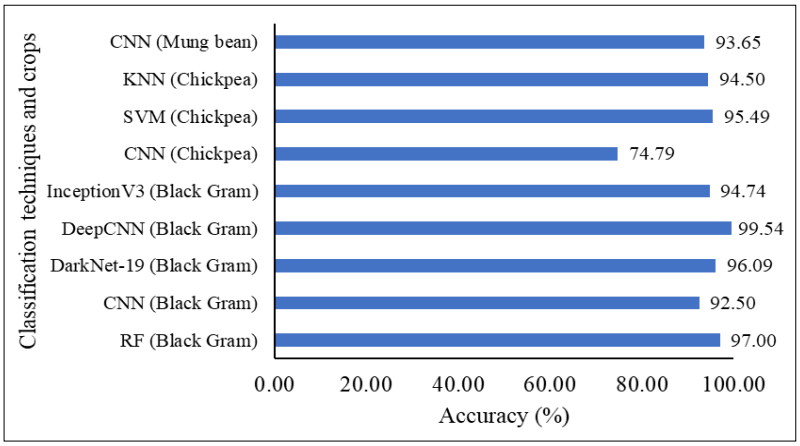


Fig 3: Comparison of classification algorithms (Legume crops).

blight and late blight disease in leaf images of potato. The model was trained using 2108 augmented images and 288 actual images. The Relu activation function was used to minimise linearity and vanishing gradient problems. The proposed model gained 99% overall accuracy (Al-Adhaileh *et al.,* 2023). In this study, the authors employed k-means for image segmentation to separate the affected and healthy parts. For feature extraction, GLCM was used and SVM was utilized to classify early blight, late blight and healthy leaves of potato with an accuracy of 95.99% (Singh and Kaur, 2021).

Considering the early blight and late blight diseases of potato, Hou *et al*. (2021) used a graph cut algorithm to accurately and effectively segment the leaf from the images. The texture characteristics were retrieved using a local binary pattern and the colour features of the pictures were recovered using the individual channels of the L*a*b* colour space on the region of interest. Four machine learning techniques (KNN, SVM, RF and Artificial Neural Networks) were employed to assess the performance for potato illness recognition; the SVM technique had the most significant overall accuracy of 97.4% for the disease classification. In another study, an ensemble CNN is proposed by combining MobileNetV2, VGG16 and ResNet50 to classify fungal diseases in potato. 6644 potato leaf images were collected from three distinct datasets of different sites. The proposed model attained 98.49% accuracy (Singh *et al.,* 2024). InceptionV3, the DL model was employed to classify early blight, late blight and healthy leaves of potato. The dataset having 2152 pictures was splitted into two portions *i.e*. training set and validation set in 80:20 ratio and 98.60% accuracy was achieved. This study also supports the use of AI-powered diagnostic tools in farming management, offering notable improvements in crop protection and economic sustainability over the long term (Dutta *et al.,* 2024). Another recent study (Pasalkar *et al.,* 2024) employed VGG16 model to classify 600 images of early blight, late blight and leaf curl. The dataset was splitted into three sections: training (80%), validation (10%) and testing (10%). The model gained 97.4% accuracy after 50 epochs of training.

Table 1: Summary of ML and DL techniques involved in disease detection of legume crops.

Reference	Crop	Disease studied	Techniques	Dataset	Evaluation
N *et al.,* 2024	Black gram	Anthracnose, powdery mildew	Random forest	Images collected from Thanjavur (Tamil Nadu, India)	Training accuracy- 99.17%, Testing accuracy-97.00%
Kalpana *et al.,* 2023	Black gram	Anthracnose, powdery mildew	CNN	Images collected from Tanjore (Tamil Nadu, India)	Accuracy- 92.50%, precision-97.14%, recall-87.17%, F1score-91.89%
Kumar *et al.,* 2023	Black gram	Yellow mosaic	Dark Net-19, Squeeze Net, Alex Net, Google Net	1100 images	96.09%, 60.74%, 94.41%, 93.85%
Talasila *et al.,* 2023	Black gram	Leaf diseases	DeepCNN	Black gram plant leaf disease (BPLD) dataset	Accuracy-99.54%, Precision-98.78%, Recall-98.82% F1 score-98.80%
Mhala *et al.,* 2024	Black gram	Yellow mosaic, Anthracnose, powdery mildew, leaf crinckle	InceptionV3	BPLD dataset (1007 images)	Training accuracy-98.80%Testing accuracy-94.74%
AlZubi *et al.,* 2024	Chickpea	Fusarium wilt	CNN	4339 images	74.79% accuracy
Abuhayi and Bezabh, 2023	Chickpea	Not mentioned	SVM		Testing accuracy-95.49%
Hayit *et al.,* 2024	Chickpea	Fusarium wilt	KNN	15000 images	94.50%
Mallick *et al.* 2023	Mung bean	Cercospora leaf spot, fusarium wilt, powdery mildew, halo blight, yellow mosaic and charcoal rot	CNN	8448 images	Accuracy-93.65%, precision-92.60%, recall-94.40% F1 score-92.80%.

The accuracy of different ML and DL models in vegetable crops leaf illness classification had been shown in Fig 4.

Discussion

Table 2 summarizes various ML and DL algorithms employed to identify diseases in potato, highlighting their effectiveness across different datasets and evaluation metrics. ML techniques like SVM have been widely used, as seen in studies by Moawad *et al*. (2023); Singh and Kaur, (2021) and Hou *et al.* (2021). SVM, combined with DenseNet201, achieved high accuracy (80% and 99%) in detecting Green Tuber and Black Scarf diseases using a small self-captured dataset of 78 images. Similarly, Singh and Kaur, (2021) utilized SVM for early and late blight detection, obtaining an impressive accuracy of 95.99% with high precision, recall and F1-score values, suggesting robust performance on the Plant Village dataset. Hou *et al.* (2021) extended the use of SVM alongside Random Forest, ANN and KNN, achieving accuracies ranging from 89.90% to 92.1% on the AI Challenger Global AI Contest dataset. In our study, the DL models, particularly Convolutional Neural Networks (CNNs), show the highest performance. Al-Adhaileh *et al.* (2023) achieved 99% accuracy for early and late blight detection using the Plant Village dataset. Meanwhile, Ensembled CNNs by (Singh *et al.* (2024) demonstrated a superior performance of 98.49% on a large dataset of 6644 images. Similarly, (Dutta *et al.* (2024) employed Inception V3, achieving 98.60% accuracy, while Pasalkar *et al.* (2024) used VGG16 for detecting multiple diseases and shows with 97.4% accuracy on a self-captured dataset.

Challenges in the adoption of AI techniques

The study highlights the application of AI in agriculture, particularly in disease detection in legume and vegetable crops. In the literature, it is evident that AI has the potential

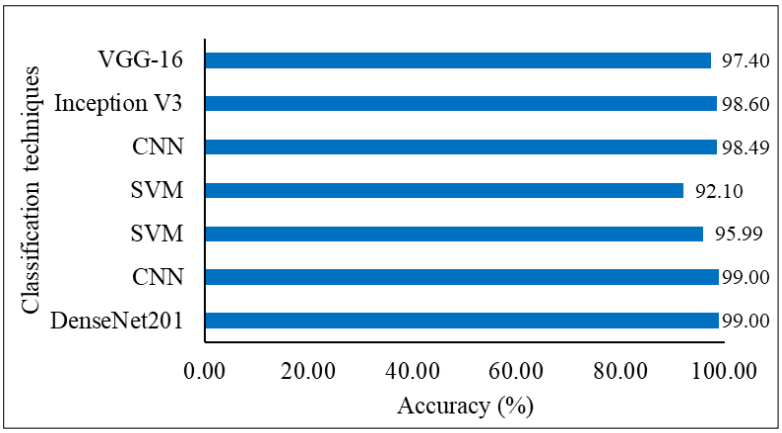


Fig 4: Comparison of classification algorithms (Potato).

Table 2: Summary of ML and DL techniques involved in disease detection of potato.

Reference	Crop	Disease studied	Techniques	Dataset	Evaluation
Moawad *et al.,* 2023	Potato	Green tuber and black scarf	SVM, DenseNet 201	78 Images (self-captured)	80.00%, 99.00%
Al-Adhaileh *et al.,* 2023	Potato	Early blight, late blight	CNN	Plant village dataset	99.00%
Singh and Kaur, 2021	Potato	Early blight, late blight	SVM	Plant village dataset	Accuracy-95.99%, Precision-96.12%, Recall- 96.25%, F1 score-96.16%
Hou *et al.,* 2021	Potato	Early blight, late blight	SVM, Random Forest, ANN and KNN	AI Challenger Global AI Contest	92.10%,91.00%, 90.80% and 89.90%
Singh *et al.,* 2024	Potato	Fungal diseases	Ensembled CNN	6644 images from three datasets	98.49%
Dutta *et al.,* 2024	Potato	Early blight, late blight	Inception V3	Plant village dataset	98.60%
Pasalkar *et al.,* 2024	Potato	Early blight, late blight, leaf curl	VGG16	600 images (self-captured)	97.40%

through ML and DL models proposed by researchers for the timely identification of disease in agricultural crops. Machine learning and deep learning models have significantly progressed in plant disease detection, but several challenges still impede their widespread and effective implementation. Scarcity of datasets, noisy and imbalanced data, problems in image acquisition, isolation of disease characteristics and complexity and overfitting issues are the main challenges in agricultural crops disease detection. In this concern Kiran and Chandrappa, (2023) highlighted the issues of reduced contrast and noise in the images captured in real-time open environment. The efficiency of proposed models significantly relies upon the availability of adequate and suitable qualitative data to correctly understand the underlying relationship between predictive variables and the predictors. The non-availability of qualitative data can pose a barrier to correctly identifying diseases because of wide variations in plant varieties and climatic conditions over time and space. In disease investigation studies, the noise problems, such as backgrounds in distinct colors and additional factors like grass, roots, soil and neighboring plant leaves during data collection and preprocessing, can introduce noise to machine learning models, leading to unsatisfactory performance.

Current datasets have been developed under controlled environments, but obtaining comparable images may be challenging due to several reasons such as degree of moisture, light intensity and ecological variables. Ngugi *et al.* (2024) also emphasized the problem of deep learning models to generalize through a variety of ecological settings and promise reliable accuracy in real-world situations. Similarly, Raza *et al.* (2025) emphasized the precision farming enables the accurate crop leaf illness diagnosis owing to technological advancements in AI. Researchers also highlighted that extensive progress has been made toward DL techniques but real-time and open field image datasets tend to present limitations including small lesion characteristics along with restricted generalization capabilities and blurring images from backgrounds and imbalanced data distributions that produce overfitting results. The use of big model variables presents obstacles for applications in agriculture due to their restricted resources. Because of overfitting, the model produces excellent results on its data, but the results could be inconsistent when applied to other data. Even while an overfitted model might perform exceptionally well when classifying training data, it might not perform well when introducing fresh data.

Joseph *et al.* (2024) observed that plant disease diagnosis faces major difficulties due to shortage in availability of large and non-laboratory datasets. In laboratory processed image datasets, the images of leaves are reproduced, which omits different disease indicators and diverse background settings with different illumination in the images. It has also been observed that some diseases have unique symptoms that can be captured through specific feature extraction techniques. Extracting features related to these symptoms (*e.g.,* the shape and size of lesions, the texture of infected areas) helps diagnose the specific type of disease affecting the plant. High-quality features can also reduce noise and irrelevant information, allowing the model to distinguish between diseased and healthy plants. The images captured under controlled weather conditions with small size datasets and traditional image pre-processing techniques may substantially reduce the disease classification accuracy and other important parameters of performance evaluation (Joshi *et al.,* 2025). The performance of AI models becomes negatively affected when they encounter environmental noise and differing backgrounds and changing plant conditions. Quality data availability stands essential for disease recognition because natural conditions affecting brightness and moisture can produce false outcomes. The inadequate inclusion of ecological variations in controlled environment datasets causes models to have limited ability for generalization. Enhancing diagnostic accuracy requires researchers to develop both diverse non-laboratory datasets and improved feature extraction techniques which help models address challenges like overfitting while operating in natural agricultural environments.

CONCLUSION AND FUTURE SCOPE

The studies demonstrate a wide range of applied algorithms, datasets and evaluation methods, showing advancements in plant disease detection. Moreover, the dataset size and geographical diversity are crucial in model performance. However, the datasets from specific regions like Tamil Nadu (India) show the importance of localized data for accurate disease detection. However, the non-availability of the datasets on major legume crops is a serious concern for experts who want to integrate new technologies for accurate and timely disease detection in agriculture. Since more variables can affect actual results, future research should investigate how plant diseases are affected under various location-specific climate conditions. This review paper also suggests that there is a need for scientific and technical experts to come forward and develop the datasets on legume crops in real time, which will help the farmers in the timely identification and management of diseases. Ultimately, these studies demonstrated the transformative potential of ML and DL techniques in enhancing crop disease management, promising improved yield and sustainable agricultural practices. Overall, this study highlighted the effectiveness of DL techniques, particularly CNN-based models, in achieving high accuracy. Further, the proposed automatic system should aim to identify plant diseases in agricultural research by removing noise, background and other elements, isolating only the desired segment within captured images and facilitating real-time identification.

It should be machine-readable, trustworthy and impartial to ensure data quality and not directly fed to models.

Conflict of interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

REFERENCES

Abuhayi, B.M. and Bezabh, Y.A. (2023). Chickpea disease classification using hybrid method. Smart Agricultural Technology. 6: 100-371. https://doi.org/10.1016/j.atech.2023.100371.

Al-Adhaileh, M.H., Verma, A., Aldhyani, T.H.H. and Koundal, D. (2023). Potato blight detection using fine-tuned CNN architecture. Mathematics. 11(6): 1516. https://doi.org/10.3390/math11061516.

AlZubi, A.A., Raghuramapatruni, R.and Kumari, P. (2024). Classification and severity level assessment of fusarium wilt disease in chickpeas using convolutional Neural Network. Legume Research-An International Journal. 47(11): 1990-1997. https://doi.org/10.18805/LRF-807.

Bhattacharjya, S., Chaudhury, S. and Nanda, N. (2017). Import dependence and food and nutrition security implications: The case of pulses in India. Review of Market Integration. 9(1-2): 83-110. https://doi.org/10.1177/0974929217721763.

Chowdappa, P., Nirmal Kumar, B.J., Madhura, S., Mohan Kumar, S.P., Myers, K.L., Fry, W.E. and Cooke, D.E.L. (2015). Severe outbreaks of late blight on potato and tomato in South India caused by recent changes in the Phytophthora infestans population. Plant Pathology. 64(1): 191-199. https://doi.org/10.1111/ppa.12228.

Dutta, M., Gupta, D., Gulzar, Y., Mir, M.S., Onn, C.W. and Soomro, A.B. (2024). Leveraging inception V3 for precise early and late blight disease classification in potato crops. Traitement Du Signal. 41(2): 705-715. https://doi.org/10.18280/ts.410213.

Goel, L. and Nagpal, J. (2023). A systematic review of recent machine learning techniques for plant disease identification and classification. IETE Technical Review. 40(3): 423-439. https://doi.org/10.1080/02564602.2022.2121772.

Gupta, O.M. and Sharma, M. (2015). Dry root rot of chickpea: An overview. Journal of Food Legumes. 28(4): 267-276.

Harakannanavar, S.S., Rudagi, J.M., Puranikmath, V.I., Siddiqua, A. and Pramodhini, R. (2022). Plant leaf disease detection using computer vision and machine learning algorithms. Global Transitions Proceedings. 3(1): 305-310. https://doi.org/10.1016/j.gltp.2022.03.016.

Hatuwal, B.K., Shakya, A. and Joshi, B. (2020). Plant leaf disease recognition using random forest, KNN, SVM and CNN. Polibits. 62: 13-19. https://doi.org/10.17562/PB-62-2.

Hayit, T., Endes, A. and Hayit, F. (2024). KNN-based approach for the classification of fusarium wilt disease in chickpea based on color and texture features. European Journal of Plant Pathology. 168(4): 665-681. https://doi.org/10.1007/s10658-023-02791-z.

Hou, C., Zhuang, J., Tang, Y., He, Y., Miao, A., Huang, H. and Luo, S. (2021). Recognition of early blight and late blight diseases on potato leaves based on graph cut segmentation. Journal of Agriculture and Food Research. 5: 100154. https://doi.org/10.1016/j.jafr.2021.100154.

Joseph, D.S., Pawar, P.M. and Chakradeo, K. (2024). Real-time plant disease dataset development and detection of plant disease using deep learning. IEEE Access. 12: 16310-16333. https://doi.org/10.1109/ACCESS.2024.3358333.

Joshi, K., Hooda, S., Sharma, A., Sonah, H., Deshmukh, R., Tuteja, N. and Gill, R. (2025). Precision diagnosis of tomato diseases for sustainable agriculture through deep learning approach with hybrid data augmentation. Current Plant Biology. 41: 100437. https://doi.org/10.1016/j.cpb.2025.100437.

Kalpana, M., Karthiba, L., Senguttuvan, K. and Parimalarangan, R. (2023). Diagnosis of major foliar diseases in black gram (*Vigna mungo* L.) using convolution neural network (CNN). Legume Research-An International Journal. 46(7): 940-945. https://doi.org/10.18805/LR-5083.

Kanade, A.K., Potdar, M.P., Balol, G. and Rathod, N. (2025). Artificial intelligence based precise disease detection in soybean using real time object detectors. Legume Research-An International Journal. 1: 6. doi: 10.18805/LR-5431.

Kaur, L., Singh, P. and Sirari, A. (2011). Biplot analysis for locating multiple disease resistance diversity in mungbean germplasm. Plant Disease Research. 26(1): 55-60.

Kiran, S.M. and Chandrappa, D.N. (2023). Plant leaf disease detection using efficient image processing and machine learning algorithms. Journal of Robotics and Control (JRC). 4(6): 840-848.

Kumar, S., Deo, M.M., Kumar, K., Rathore, M., Akram, Mohd. and Pratap, A. (2023). Detection and classification of yellow mosaic disease in *vigna mungo* using convolutional neural network deep learning models. Ecology, Environment and Conservation. 29: S177-S181. https://doi.org/10.53550/EEC.2023.v29i01s.027.

Mallick, M.T., Biswas, S., Das, A.K., Saha, H.N., Chakrabarti, A., and Deb, N. (2023). Deep learning based automated disease detection and pest classification in Indian mung bean. Multimedia Tools and Applications. 82(8): 12017-12041.https://doi.org/10.1007/s11042-022-13673-7.

Mhala, P., Varma, T., Sharma, S. and Singh, B. (2024). Deep transfer learning for enhanced blackgram disease detection: A transfer learning-driven approach (pp: 195-213). https://doi.org/10.1007/978-3-031-64070-4_12.

Moawad, N., Zaki, H., El Moniem, Essa, T. and Said, M. (2023). Detection of potato tuber diseases using machine learning models. International Conference on Artificial Intelligence Science and Applications in Industry and Society (CAISAIS). 1-7. https://doi.org/10.1109/CAISAI S59399.2023.10269994.

Murmu, S., Dey, S. and Chakraborty, A. (2017). Efficacy of different fungicides for management of early blight disease of potato. Journal of Applied and Natural Science. 9(1): 280-285. https://doi.org/10.31018/jans.v9i1.1184.

Nene, Y.L., Reddy, M.V., Haware, M.P., Ghanekar, A.M., Amin, K.S., Pande, S. and Sharma, M. (2012). Field diagnosis of chickpea diseases and their control. information bulletin no. 28 (revised). International Crops Research Institute for the Semi-Arid Tropics.

Ngugi, H.N., Ezugwu, A.E., Akinyelu, A.A. and Abualigah, L. (2024). Revolutionizing crop disease detection with computational deep learning: A comprehensive review. Environmental Monitoring and Assessment. 196(3): 302. https://doi.org/10.1007/s10661-024-12454-z.

Nimbrayan, P.K. and Tripathi, R.K. (2019). Trends and growth rate analysis of pulses in Haryana vis-à-vis India. Journal of Pharmacognosy and Phytochemistry. 8(1): 1724-1729.

Pal, R., Singh, R. and Prasad, C.S. (2018). Pulses in India enhancing the production and productivity. Indian Farming. 67(8): 24-26.

Palanichamy V.N., Kalpana, M., Karthiba, L., (2024). Detection of diseases in blackgram (*Vigna mungo* L.) using machine learning models: A case study. International Journal of Plant and Soil Science. 36(5): 688-695. https://doi.org/10.9734/ijpss/2024/v36i54566.

Pandey, P.R., Gaur, P.M. and Sajja, S.B. (eds.). (2019). Pulses Value Chain Development for Achieving Food and Nutrition Security in South Asia: Current Status and Future Prospects. SAARC Agriculture Centre. p: 181.

Pasalkar, J., Gorde, G., More, C., Memane, S. and Gaikwad, V. (2024). Potato leaf disease detection using machine learning. Current Agriculture Research Journal. 11(3): 949-954. https://doi.org/10.12944/CARJ.11.3.23.

Pradel, W., Gatto, M., Hareau, G., Pandey, S.K. and Bhardway, V. (2019). Adoption of potato varieties and their role for climate change adaptation in India. Climate Risk Management. 23: 114-123. https://doi.org/10.1016/j.crm.2019.01.001.

Rajaguru, H. and Sannasi Chakravarthy, S.R. (2019). Analysis of decision tree and K-nearest neighbor algorithm in the classification of breast cancer. Asian Pacific Journal of Cancer Prevention. 20(12): 3777-3781. https://doi.org/10.31557/APJCP. 2019.20.12.3777.

Raymundo, R., Asseng, S., Prassad, R., Kleinwechter, U., Concha, J., Condori, B. and Porter, C. (2017). Performance of the SUBSTOR-potato model across contrasting growing conditions. Field Crops Research. 20(2): 57-76. https://doi.org/10.1016/j.fcr.2016.04.012.

Raza, A., Pitafi, A.H., Shaikh, M.K. and Ahmed, K. (2025). Optimizing potato leaf disease recognition: Insights DENSE-NET-121 and Gaussian Elimination Filter Fusion. Heliyon. https://doi.org/10.1016/j.heliyon.2025.e42318.

Roy, D., Kishore, A., Joshi, P.K. and Mishra, B. (2016). Pulses for food and nutritional security: The technology perspective. Indian Journal of Genetics and Plant Breeding. 76(04): 375-387. https://doi.org/10.5958/0975-6906.2016.00057.2.

Singh, A.K., Sreenivasu, S.V.N., Mahalaxmi, U.S.B.K., Sharma, H., Patil, D.D. and Asenso, E. (2022). (Retracted) hybrid feature based disease detection in plant leaf using convolutional neural network, Bayesian Optimized SVM and Random Forest Classifier. Journal of Food Quality. 2022(1): 2845320. https://doi.org/10.1155/2022/2845320.

Singh, A. and Kaur, H. (2021). Potato plant leaves disease detection and classification using machine learning methodologies. IOP Conference Series: Materials Science and Engineering. 1022(1): 012121. https://doi.org/10.1088/1757-899X/1022/1/012121.

Singh, G., Kasana, G. and Singh, K. (2024). Improved potato crop disease classification using ensembled convolutional neural network. Potato Research. https://doi.org/10.1007/s11540-024-09787-0.

Singh, R., Kumar, K., Purayannur, S., Chen, W. and Verma, P.K. (2022). Ascochyta rabiei: A threat to global chickpea production. Molecular Plant Pathology. 23(9): 1241-1261. https://doi.org/10.1111/mpp.13235.

Sinha, R., Irulappan, V., Patil, B.S., Reddy, P.C.O., Ramegowda, V., Mohan-Raju, B. and Senthil-Kumar, M. (2021). Low soil moisture predisposes field-grown chickpea plants to dry root rot disease: evidence from simulation modeling and correlation analysis. Scientific Reports. 11(1): 6568. https://doi.org/10.1038/s41598-021-85928-6.

Talasila, S., Rawal, K. and Sethi, G. (2023). Black gram disease classification using a novel deep convolutional neural network. Multimedia Tools and Applications. 82(28): 44309-44333. https://doi.org/10.1007/s11042-023-15220-4.

Vishnoi, V.K., Kumar, K., Kumar, B., Mohan, S. and Khan, A.A. (2023). Detection of apple plant diseases using leaf images through convolutional neural network. IEEE Access. 11: 6594-6609. https://doi.org/10.1109/ACCESS.2022.3232917.