Research Article

Multiagent and Grid Systems
2024, Vol. 20(3-4) 265–282

DOI: 10.1177/15741702241304087
journals.sagepub.com/home/mgs
Sage | IOS Press

Plant disease detection and classification techniques: A review

Smita R Sankhe[1] and Asha Ambhaikar[1]

Abstract

Crop production can be greatly affected by different diseases, which seriously threaten food security. Consequently, detecting plant diseases earlier and preventing the spread of plant diseases is necessary to avoid the economic imbalance. Nevertheless, the manual detection of plant diseases is a time-consuming and error-prone process. Numerous existing techniques are adopted that are unreliable in terms of accuracy and fail to identify the infected region due to non-uniform complex backgrounds resulting in mispredictions. Hence, this research presents a review that focuses on enhancing plant disease detection accuracy and early intervention. The comparison of various machine learning and deep learning techniques, data acquisition, segmentation methods, and feature extraction techniques are presented, particularly in disease prediction. The basic prediction approach encloses the classification model which is trained and the publicly available dataset or the collection of real-time data from the farms is tested. Those datasets include images of normal as well as plants with disease spots which serve as benchmarking datasets for the research. The paper recognizes challenges like limited data, scalability, and accuracy but seeks to leverage previous technologies to advance agricultural practices. Ultimately, the aim is to improve upon existing methodologies for more effective plant disease prediction, contributing to a more robust agricultural sector.

Keywords

plant disease detection, segmentation, feature extraction, machine learning, deep learning

Received: 23 April 2024; accepted: 17 November 2024

1 Introduction

The ultimate objective of wealth is Agriculture which has a great contribution to the lives of the earth. plant disease prediction plays a major concern in the agricultural field as it decides the yield as well as the economy of agriculture.[1] In particular, the disease found in the food crops such as rice, wheat, Tomato, and so on, can lead to food insecurity and economic fall around the world. Formerly, the crops were inspected by experienced farmers to spot the disorders of the plants that sometimes created the misconception of the disease.[2] Due to that misconception, the diseases are not treated with proper medications, in turn; they spread rapidly.[3] Thanks to the severity of the disease spot in the plants and the inconsistent existing approach, plant disease prediction with the recent technologies emerged. The basic idea of automatic plant disease prediction is based on Artificial Intelligence (AI), where machines act with human intelligence.[4] Various research in AI and their modifications are carried out in order to establish the effective performance in plant disease prediction at their early stage.[5]

The AI mechanism includes various machine learning approaches that are not fully automated initially.[6] Later with several algorithms and classifiers, the machine learning can address critical conditions of the plant disease[7] with methods such as K-nearest neighbor (KNN), Support vector machine (SVM),[8] Random Forest[9] and Logistic Regression.[10] These methods inhibit benefits as well as drawbacks. The SVM[8] finds complexities when working with large datasets and is

[1]Computer Science & Engineering, Kalinga University, Naya Raipur, Chhattisgarh, India

Corresponding author:
Smita R Sankhe, Computer Science & Engineering, Kalinga University, Kotni, Near Mantralaya, Naya Raipur - 492101, Chhattisgarh, India.
Email: smitars357@gmail.com

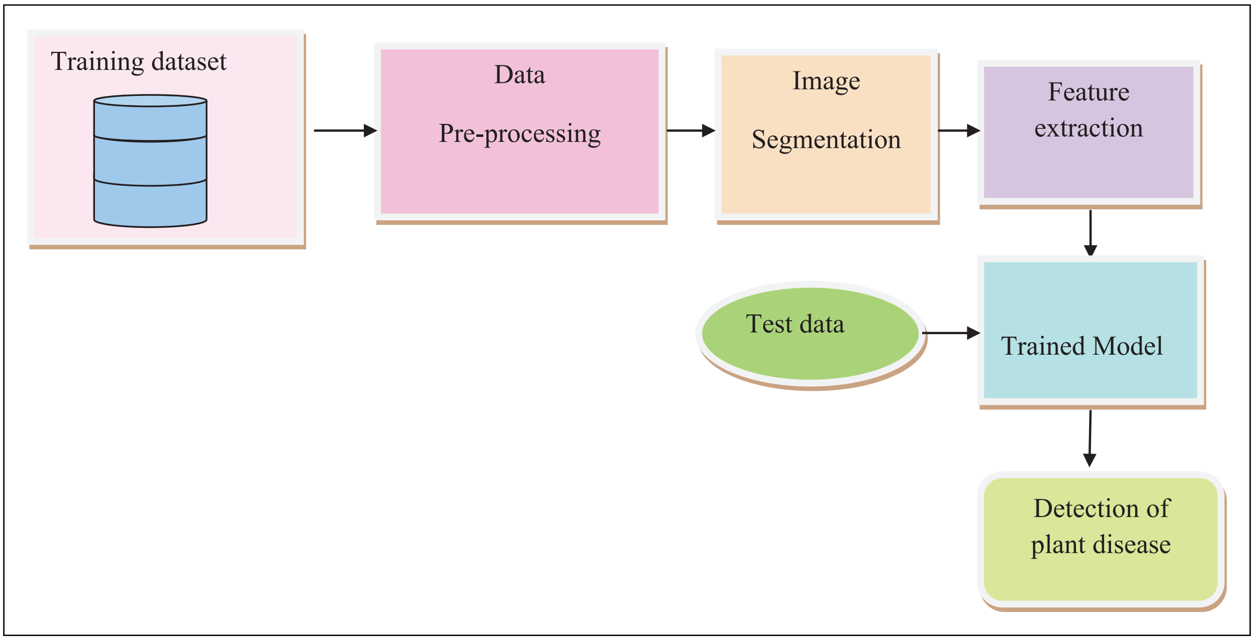

Figure 1. Block diagram of plant disease detection.

highly sensitive to noise, whereas the Random forest[9] requires more computational power and creates the overfitting problem. The logistic regression[10] only aids with the linear functions and not with the non-linear regressions as it creates the linear boundaries that cause overfitting. deep learning achieves better in comparison with other existing approaches of machine learning. A few methods of deep learning are Convolutional Neural Networks (CNN), Artificial Neural Networks (ANN), Recurrent Neural Networks (RNN), Gradient Descent, and so on. deep learning inhibits various advantages such as the ability to make use of the raw data without the skilled features, and high computational power that leverages the parallel processing.[6] Thus the deep learning methods become the greater companion to the farmers through which the increased productivity can be achieved. To date, several kinds of research are under process with advanced techniques to aid plant disease detection.

The paper elaborates on techniques or algorithms utilized by the deep learning models to detect the disease spots of the plants prior to overcoming the effects of them. The overview of the paper depends on the comparison of certain things such as the feature extraction methods, segmentation methods, classifiers, algorithms of optimization, and majorly on the datasets utilized in the few existing researches.

The rest of the paper contains sections 2 – section 8. Section 2 explores the literature review of the research papers, and section 3 elaborates on the Analysis and Discussion, which includes the comparison of 35 papers in terms of the dataset, feature extraction, segmentation, deep learning, and machine learning methods. Section 4 illustrates the standard benchmarks and the information of the disease related to the dataset whereas Section 5 explains the role of deep learning models in disease prediction. Section 6 presents the challenges that are identified for the discussed models. Sections 7 and 8 end the survey with the conclusion and the directions for future works respectively.

2 Classification models for plant disease classification

The block diagram of the plant disease detection is depicted in Figure 1, in which the input data for plant disease prediction is collected and fed to the data preprocessing to enhance the quality of the images. Further image segmentation is carried out to separate the infected region from the plant image. Following the image segmentation, feature extraction is carried out to identify the disease pattern effectively. Further, the test data is given to the trained model based on machine learning or deep learning to detect the plant disease.

Former traditional technologies were utilized to identify pest or plant diseases, and later machine learning and deep learning methods emerged which shows a drastic growth in the past few decades. These researches are based on image processing with several unique algorithms, techniques, and models incorporated. This section elaborates on the existing methods with their benefits and drawbacks. The taxonomy of the techniques considered for a comprehensive survey is given in the following section.

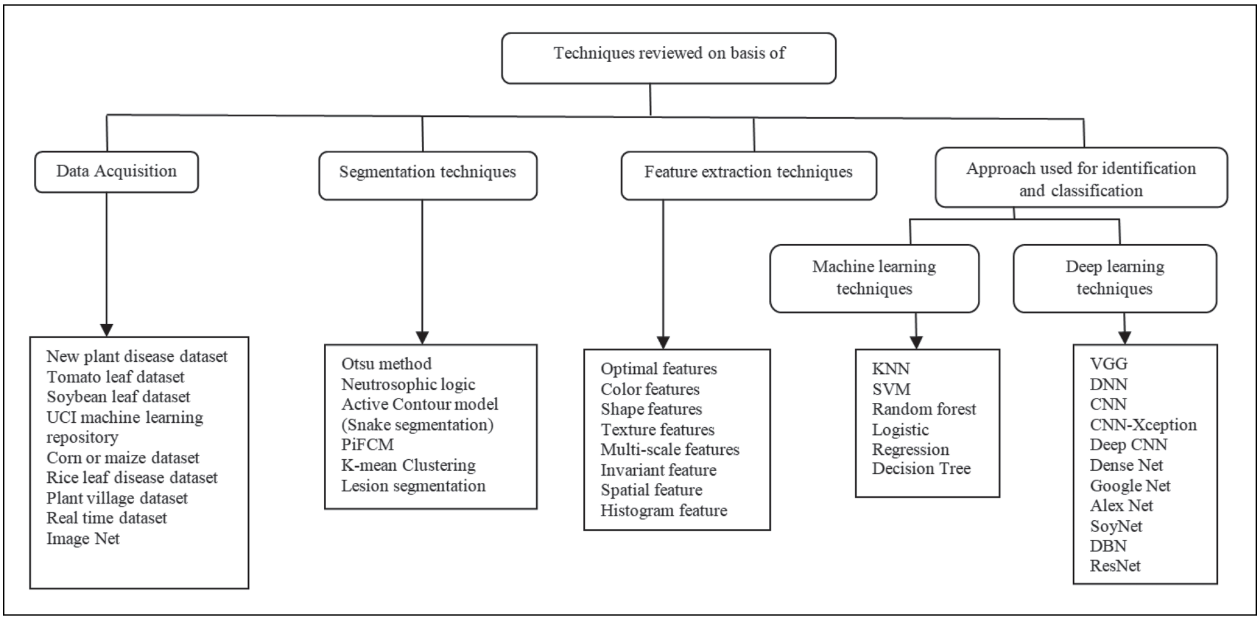

Figure 2. Techniques considered for a comprehensive survey and future prospects.

2.1 Taxonomy

Deep learning, machine learning, and computer vision techniques are critical to the agriculture industries for the detection and classification of crop illnesses.[1] To automatically detect and diagnose agricultural plant diseases, Figure 2 represents the techniques considered for a comprehensive survey and future prospects.

2.1.1 Data acquisition. The data is obtained from the new plant disease,[11] tomato leaf disease detection,[12] soybean leaf dataset,[13] UCI machine learning repository,[14] corn or maize dataset,[15] rice leaf disease dataset,[16] plant village,[17] and ImageNet dataset[6] for the identification and classification of the plant disease detection.

2.1.2 Segmentation techniques. Segmentation is a process that divides an image into regions to identify the infected areas of a plant leaf. The segmentation technique includes the Otsu method, Neutrosophic logic, Active contour model (Snake segmentation), Probabilistic Intuitionistic Fuzzy c-Means Algorithm (PiFCM), and K-means clustering and is demonstrated as follows,

Otsu method: The segmentation by the Otsu method provides the particular region of interest in the image. The Otsu method specifies only certain diseases that affect the rice crops.

Neutrosophic Logic: The neutrosophic segmentation divides the image into three regions such as True, False, and Intermediate segments. The segmented image also provides a factor 'I' for the degree of uncertainty.

Active Contour model: The segmentation is carried out by the external constraints and fits the object contour. Moreover, the snake segmentation provides less robustness.

Probabilistic Intuitionistic Fuzzy c-Means Algorithm (PiFCM): The piFCM is crucial for unsupervised image segmentation. However, it does not have local information on the image content which leads to misconception of the pixels and provides inefficient performance.

K-means clustering: K-means clustering provides higher dimensions that make it applicable to multiple features. In place of K-means, alternative segmentation techniques can be employed for more precise extraction of the lesion.

2.1.3 Feature extraction techniques. Feature extraction is the process of evaluating features or the properties of the particular segmented region to minimize the data of the image through which the algorithm enables effective analysis. The features should highlight all the dominant characteristics of the image with reduced size, and the features depend on the color, texture, geometry, histogram, pixel, and space factors of the image.

Optimal features: The optimal features are refined and elected by the whale optimization algorithm (WOA).[18] Additionally, the optimal feature for plant disease detection includes color, texture, and shape. These features are extracted from images to perform classification tasks.

Color features: The color features are the predominant characters of the image to classify the plant disease through the different types of color spaces and the parameters.[19] Color spaces such as RGB, HSV, HIS, LAB, and LUV are utilized in the process of extracting the color appearance of the infected region. These color spaces provide distinct information whereas RGB is a prime color space that enhances the Red, Green, and Blue color variants accurately.

Texture features: The texture features are extracted to understand the image intensities combined in the segmented image. These features usually measure the shape and the structure of the disease spots, the distance between the infected areas, the range, and the number of pixels.[9] Homogeneity, cluster prominence, shade, energy, and so on, are some of the features that enable to evaluation of the texture of the image. Gray-level co-occurrence matrix (GLCM), Local Binary Pattern (LBP), Histogram of Oriented Gradients (HOG), and entropy are some of the familiar techniques employed to extract the texture features.

Geometric features: The features that are extracted in terms of area, aspect ratio, major and minor axis, perimeter, orientation, and solidity[19] in diseased plants are referred to as geometric features. The geometric features elaborate the global and local information regarding the shape, size, and position of the affected area in the plant recommended for disease prediction. The above-mentioned features in fusion with other features provide better accuracy.

Multi-scale features: The features used in plant disease detection are used to identify the plant disease signatures more precisely. A multi-scale feature[3] is used for simultaneous segmentation and identification of tomato leaf diseases.

Histogram features: Histogram Information Content (HIC)[9] is the parameter for characterization of the global and local color distributions and this method is easy to compute and requires minimal time. HIC is evaluated as the probability of occurrence of the relative information about each plane in the plant diseased image.

Spatial features: Spatial features are obtained due to the change observed in the movement of the image. In particular, spatial features are the amount of spaces captured by the segmented diseased plant image.[9]

2.1.4 Classification models. The classification model of the plant disease prediction is categorized as machine learning and deep learning.

i) Machine learning

Asraa Safaa Ahmed et al.[7] applied KNN for plant disease prediction. This method effectively detects the plant disease of the soybean plant. The KNN method handles the multiclass cases and can deal with the non-linear data, but they are the slow predictors and they require high storage memory so the model is computationally high in cost. The authors of[8,19–21] proposed the various forms of the SVM. SVM predicts the plant disease in a better way compared to other methods but provides fewer efficacies due to the unavailability of the labeled data in the dataset and is sensible to noise and enabled overlapping. Gittaly Dhingra *et al.*[9] utilized Random forest as the method to reduce the overfitting issue concentrate on the dimensionality and reduce the feature space. The drawbacks found are the approach is highly complex due to the trees count which increases the time to predict the output. In,[10] Divyansh Tiwari *et al*. presented the Logistic regression method which is an easily interpretable, implementable model that can be trained fast. The method can classify the unknown records but the methods provide the irregular responses with the non-linearity. When the observations count is less than the features count then the method leads to overfitting.

ii) Deep learning

The research discussed above falls into the machine learning category and the enhancements are featured in the deep learning methods. The authors of[3,4,22–26] utilized CNN from the deep learning methods. CNN plays a major part in deep learning methods having advantages of automatic feature extraction, classification, and recognition. Through hierarchical learning, the model minimized the computation and handled the large datasets easily. Due to the high computational requirements and the interpretability challenges, the CNN faces challenges such as, the model was slow and provided limited effectiveness. Junde Chen *et* al. utilized VGGNet in,[5] the model accepted fewer parameters that work with the 1×1 CNN which can be used to decide the non-linear function without modifying the receptive fields. The model required a large storage memory and a huge network to train which took more time. Thippa Reddy Gade Kallu *et al.*[18] suggested the DNN model with the whale optimization that enhanced the performance by providing the solution to the problem by the adaptive ability and with WOA the model is structured but created the slow convergence speed and low precision that moved the model to local optima easily.

Aditya Khampria *et al*.[27] suggested the deep encoder network that reduces the complexity by extracting the accurate features from the large and high-dimensional dataset. The drawback was that the model eliminated the vital information due to the misclassification that occurred on the input data. Jihen Amara *et al*.[28] demonstrated the LeNet to forecast the plant disease. The advantages of the LeNet are majorly used in the identification of handwritings which utilizes a simple architecture as well as working with the dataset containing millions of datasets. This advantage sometimes leads to overfitting which becomes the disadvantage of the model. Jun Liu and Xuewei Wang[29] delivered the YOLOV3 model to predict plant disease. The model was fast and efficient and provided high accuracy with generalization but failed to work with smaller object detection and needed more time due to its heavy storage memory. Rubini PE and Dr.Kavitha P[30] suggested the Dense Net to progress the efficacy of plant disease prediction by providing gradient flow and access of features to each layer, which enhances the performance. The drawbacks of the model were that spliced feature maps can concatenate with every layer creating duplicate entries or multiple entries.

The authors Sanjeev S Sannakki *et al*[31] and PR Rothe *et al*[32] presented the Back propagation neural network which acted as the fastest method. The method contains no tuning parameters other than the inputs given which makes the model more flexible. The model was sensitive to noise and had a matrix-based approach that minimized the accuracy of the model. In,[33] the author Sachin B Jadhav *et al*, described the AlexNet to predict plant disease. Since the model was faster to train, the model possesses some limitations such as gradient vanishing, and normal distribution.

iii) Role of optimizations in training deep learning and machine learning models

Optimization techniques are a strong tool to enhance the efficacy of the research by managing the input resources which maximizes the performance of the model. Though training the model in deep learning is a critical task, the optimizers aid in better training with minimal loss of training error and generalization error.

Rider-CSA optimizer: The Rider-CSA optimizer[1] is the combination of the Rider Optimization Algorithm (ROA) and the Cuckoo Search Algorithm (CSA). ROA is inspired by riders who race to the target location and CSA is inspired by Cuckoos which possess unique breed behavior. The Rider CSA works with fictional computing, through which the convergence is improved in terms of local optima, and in addition, solves the issues with unknown search spaces of the real-world applications.

K-means clustering: The clustering algorithm[31] is a popular unsupervised machine learning method that segments similar data into clusters. This method produces distinct, non-overlapping cluster work, which makes it easy to interpret and visualize. Choosing the ideal cluster counts (K) is a difficult undertaking that may call for domain expertise or trial and error.

Backpropagation Neural Network: The backpropagation algorithm[32] is proposed to evaluate the gradient descent in accordance with the weights of the several inputs. The algorithm is highly adaptable and does not require prior knowledge and it does not lay on any tunable parameters except the inputs of the algorithm.

PCA-WOA optimizer: The Whale Optimization Algorithm (WOA) merges with the Principal Component Analysis (PCA) to form the PCA-WOA[18] optimizer. The WOA has the unique hunting pattern known as the humpback whale hunting pattern which is utilized to achieve better efficacy whereas the PCA aids in real-time implementations as they are versatile and simple to proceed the further processing. In addition, the combination of both reduces overfitting and increases the performance.

3 Analysis and discussion

This section provides the analysis of the plant disease detection and classification techniques in terms of different analysis including the dataset analysis, analysis of segmentation techniques, deep learning based techniques described as follows,

3.1 Dataset analysis

Table 1 shows the dataset analysis including the details about the attributes information in the dataset, class information, and techniques utilized for the enhancement.

In this survey, the datasets named "new plant disease dataset" and "Soybean.leaf.dataset" are the datasets commonly used for Tomato and Soybean respectively. The new plant disease dataset is the majorly utilized dataset that contains various plant species and has many updated versions almost enclosing all the diseases. This dataset contains 38 classes and 87k RGB images of training, testing, and validation directories. The training and validation are in the ratio of 80:20, where the plants included are Apple, Potato, Tomato, Soybean, Raspberry, Pepper bell, Orange, Peach, Grape, corn (Maize), and Strawberry. All plants included have their own test, train, and validation images that depict the diseases arising through bacterial or mite infection. The soybean.leaf.dataset is the dataset collected across various regions of the world and contains 6410 images of Soybean leaf with dimension 500×500. These images are collected together as 3 classes namely Caterpillar, Diabrotica Speciosa, and healthy containing 3309, 2205, and 896 images respectively. The other datasets found during

Table 1. Dataset analysis.

Dataset	Data attribute information	Class information	Techniques for data enhancement
New plant disease dataset[11] – Tomato plant	Tomato mosaic virus, target Spot, Bacterial spot Tomato Yellow Leaf Curl Virus, Late blight, Leaf Mold, Early blight, Spider mites, Two-spotted spider mite, Septoria leaf spot, Tomato healthy.	38 classes-87k images	Preprocessing
Tomato leaf disease detection[12] – Tomato plant	Tomato mosaic virus, Target Spot, Bacterial spot Tomato Yellow Leaf Curl Virus, Late blight, Leaf Mold, Early blight, Spider mites, Two-spotted spider mite, Septoria leaf spot, Tomato healthy.	10 classes-11000 images	Preprocessing
Soybean. Leaf. Dataset[13] – Soybean plant	Caterpillar, Diabrotica Speciosa, and healthy	3 classes - 6410 images	Preprocessing
UCI machine learning repository[14]- Soybean plant	diaporthe-stem-canker, charcoal-rot, rhizoctonia-root-rot, phytophthora-rot, brown-stem-rot, powdery-mildew, downy-mildew, brown-spot, bacterial-blight,bacterial-pustule, purple-seed-stain, anthracnose, phyllosticta-leaf-spot, alternarialeaf-spot,frog-eye-leaf-spot, diaporthe-pod-&-stem-blight,cyst-nematode, 2–4-d-injury, herbicide-injury.	19 classes	Preprocessing
Corn or Maize dataset[15] – Maize plant	Common Rust, Gray Leaf Spot, Blight, Healthy	4 classes – 4188 images	Preprocessing
Grape disease dataset[34] – Grape plant	Black Rot, ESCA, Leaf Blight, Healthy.	4 classes – 9027 images	CNN
Mango leaf disease dataset[35] – Mango plant	Anthracnose, Bacterial Canker, Cutting Weevil, Die Back, Gall Midge, Powdery Mildew, Sooty Mould, Healthy	8 classes - 4000 images	Preprocessing
Banana Leaf disease[36] - Banana leaf plant	Xanthomonas, Sigatoka, Healthy.	3 classes – 29 images	CNN
Rice Leaf diseases dataset[16] – Rice plant	Leaf smut, Bacterial leaf blight, brown spot, Healthy.	4 classes – 120 images	CNN
Plant Village[17] - Potato, Strawberry, and Apple plants	Potato -Early Blight, Late Blight, Healthy.	3 classes – 7128 images	Preprocessing
	Strawberry - Leaf scorch, Healthy.	2 classes – 4498 images	
	Apple - Apple scab, Black rot, Cedar apple rust, Healthy.	4 classes – 7771 images	

the survey are UCI soybean, Banana leaf disease, Rice leaf disease, Mango leaf disease, Corn (Maize) dataset, Grape leaf dataset, and plant village.

3.2 Comparison of plant disease detection techniques (2013–2016)

Table 2 demonstrates the analysis of plant disease by segmentation models such as K-means clustering and Active contour model (snake segmentation), Genetic algorithm and Otsu method respectively. The clustering algorithms are the supervised algorithms to identify different clusters in the image data. The active contour is a segmentation process that separates the pixel of identification from the image for further analysis. The classification tree is the deep learning model used for plant disease analysis, but cannot be interpreted for large datasets and is represented in Table 3.

Table 2. Comparison of segmentation techniques (2013–2016).

Segmentation	Benefits	Limitations	Future works
K-means clustering (Sannakki et al., 2013)[31]	K-means is useful for datasets with several features since it scales well to larger dimensions.	Adversarial attacks, in which well-crafted inputs cause inaccurate predictions, can affect neural networks.	For a more accurate lesion extraction, different segmentation approaches can be used instead of K-means clustering.
Active Contour Model (Snake segmentation) (Rothe and Kshirsagar, 2015)[32]	The segmentation is carried out by external constraints and fits the object contour.	The model is less robust due to the snake segmentation as well as the fewer features extracted.	The method can be extended to other crops such as Orange, Wheat, Corn, and so on.
Genetic algorithm (Singh and Misra, 2017)[8]	The algorithm provides the number of optimum solutions and not a single solution.	SVM finds it difficult to work with large datasets	The classification process can be improved with the various classifiers.
Otsu (Islam et al., 2017)[21]	The segmentation provides the particular region of interest in the image.	The dataset only specifies certain diseases which can be extended to obtain better accuracy.	Future work can include estimation of the severity of the disease automatically.

Table 3. Comparison of deep learning techniques (2013–2016).

Deep Learning	Total number of layers	Limitations	Future works
Classification Tree (Sabrol and Satish, 2016)[37]	13	The classification tree is prone to overfitting and the large datasets are hard to interpret.	The availability of the classification models in the current scenario is large and they can be utilized for plant disease detection.
DeepCNN(Sladojevic et al., 2020)[38]	16	However, the fine-tuning had no significant change in the accuracy which limited the performance.	Future work will be collecting images for enhancing the database and improving the accuracy of the model.
AlexNet (Mohanty et al., 2016)[39]	9	The drawback was that the model classified only single leaves on a uniform background.	Requires collecting more data to enhance the accuracy.

3.3 *Comparison of plant disease detection techniques (2017–2019)*

Table 4 displays the feature extraction techniques used for plant disease detection are color and texture features and statistical texture features. The disease in the plant is classified by the color of the affected region and the texture feature examines the shape and size of the disease spot. The deep learning and the machine learning algorithm used for plant disease detection are represented in Table 5 and Table 6.

3.4 *Comparison of plant disease detection techniques (2020–2022)*

Table 7 depicts the feature extraction techniques like VGG19, Darknet-53, GoogleNet, and PCA-WOA that are utilized for plant disease detection. VGG19 is the most widely used feature extraction approach with less error rate. Further, the DarkNet-53 is utilized for low-level and high-level feature extraction. Furthermore, the GoogleNet is utilized for multi-scale feature extraction. Table 8 displays the segmentation techniques such as PCM and Segmentation user interface for plant disease detection. CNN is utilized as an excellent method in the context of deep learning. Table 9 shows the comparison of deep learning techniques. KNN and logistic regression are the machine learning techniques utilized in Table 10 for plant disease detection.

Table 4. Comparison of feature extraction techniques (2017–2019).

Features	Feature extraction	Benefits	Limitations	Future works
Color and texture features (Brahimi et al., 2017)[6]	GLCM, Google Web Toolkit (GWT)	The features extracted are hand-crafted features that provide seamless features.	Computational time and cost are high which makes the model less effective in certain regions.	The visualization of features can be utilized for the future gradation in the model.
Color and texture features (Singh and Misra, 2017)[8]	Color co-occurrence method	Over traditional gray scale representation, the color feature provides a better light spectrum.	SVM finds it difficult to work with large datasets	The classification process can be improved with the various classifiers.
Statistical texture features (Islam et al., 2017)[21]	GLCM	The histograms of the color planes that were extracted provide a path to better accuracy.	The dataset only specifies certain diseases which can be extended to obtain better accuracy.	Future work can include estimation of the severity of the disease.

Table 5. Comparison of deep learning techniques (2017–2019).

Deep Learning	Total number of layers	Limitations	Future works
Google Net (Brahimi et al., 2017)[6]	22	Computational time and cost are high which makes the model less effective in certain regions.	The visualization of features can be utilized for the future gradation in the model.
LeNet (Amara et al., 2017)[28]	7	Lack of severity estimation of the plant disease.	The future works can be extended to several varieties of plants.
CNN (Liu et al., 2017)[40]	15	The model works with the particular disease since the dataset is collected under a specific season.	Other Deep Neural networks can be utilized instead of CNN, such as YOLO, Faster RCNN, and so on.

3.5 Overall Analysis of the plant disease detection techniques

The achievements of the dataset utilized in plant disease detection are categorized as machine learning and deep learning.

1. Machine learning: Table 11 represents the overall analysis of the machine learning techniques including the dataset information, sample plant utilized for the classification, algorithm utilized, advantages and disadvantages.
2. Deep learning: Table 12 represents the overall analysis of the deep learning techniques including the dataset information, sample plant utilized for the classification, algorithm utilized, advantages, and drawbacks.

4 Standard benchmarks and information on disease

4.1 Classes of new plant disease dataset

The new plant disease dataset is the majorly utilized dataset in disease prediction and this dataset contains various plant species among them tomato is taken into consideration. Figure 3 depicts the classes of the tomato species and Table 13 shows the classes of tomato from the new plant disease dataset.

Table 6. Comparison of machine learning techniques (2017–2019).

Machine learning	Total number of layers	Limitations	Future works
SVM (Singh and Misra, 2017)[8]	2	SVM finds it difficult to work with large datasets	The classification process can be improved with the various classifiers.
M-SVM (Islam et al., 2017)[21]	2	The dataset only specifies certain diseases which can be extended to obtain better accuracy.	Future work can include estimation of the severity of the disease automatically.

Table 7. Comparison of feature extraction techniques (2020–2022).

Features	Feature extraction	Benefits	Limitations	Future works
Pixel deviation (Tiwari et al., 2020)[10]	VGG19	The technique provides accurate feature extraction with less error.	The model provides less robustness and accuracy with certain datasets.	The model can be embedded with the IoT to connect them with smartphones which enhances real-world agricultural production.
Low-level and high-level features (Liu and Wang, 2020)[29]	Darknet-53	The technology depicts the effect of each stage of feature extraction.	The algorithm cannot maintain a high detection rate.	Various diseases should be collected with high resolution for the database to predict the various diseases accurately.
Multi-scale features (Ji et al., 2020)[41]	Google Net	Google Net is the fastest method of feature extraction with higher accuracy.	To train the dataset the computational cost is high	Effective pruning mechanisms should be utilized in the future to reduce the computational resources.
Optimal features (Gadekallu et al., 2021)[18]	PCA-WOA	Helps in removing the correlated features.	No comparative analysis to demonstrate better outcomes.	Use various datasets to enable proactive predictions.

Table 8. Comparison of segmentation techniques (2020–2022).

Segmentation	Benefits	Limitations	Future works
PCM (Cristin et al., 2020)[1]	The particular segmentation model poses the global convergence.	The algorithm has a slow convergence speed.	Various algorithms can be employed to progress the model's performance.
Segmentation user interface (Karthik et al., 2020)[23]	The segmentation model provides the accurate brightness, contrast, stretch, and rotational characteristics of the image.	The S-CNN model fails to predict the disease if the given sample contains more than one disease and is sensitive to segmentation.	The model can further proceed with the Digital object identifiers which link the material data facility.

Table 9. Comparison of deep learning techniques (2020–2022).

Deep learning	Total number of layers	Limitations	Future works
CNN (Darwish et al., 2020)[4]	6	CNN requires a large set of labeled data, which requires high computational cost and time	The dataset can be collected from various real-time sources in different conditions.
VGG-Net (Chen et al., 2020)[5]	18	VGG-Net creates gradient gradient-exploding problem	The broader range of plant diseases is identified with mobile device applications.
Attention-based residual CNN (Karthik et al., 2020)[23]	27	The results obtained through the attention mechanism vary on every dataset.	The model can utilize various classifiers to enhance the detection performance.
Deep convolutional Encoder network (Khamparia et al., 2020)[27]	22	The algorithm is vulnerable to adversarial attacks.	Some hyper-parameters such as dropout and regularization can be added to the method.
YOLO V3 CNN (Liu and Wang, 2020)[29]	106	The algorithm cannot maintain a high detection rate.	High-quality images of different types of diseases should be collected for the database to predict the various diseases accurately.
S-CNN (Sharma et al., 2020)[42]	6	The S-CNN model fails to predict the disease if the given sample contains more than one disease and is sensitive to segmentation.	The model can further proceed with the Digital object identifiers which link the material data facility.
United model CNN (Ji et al., 2020)[41]	12	To train the dataset the computational cost is high	Effective pruning mechanisms should be utilized in the future to reduce the computational resources.
SoyNet (Karlekar and Seal, 2020)[43]	14	SoyNet is the newly adopted technology that did not include the other plants.	The other soybean dataset can be included in the research to evaluate the performance of the SoyNet.
CNN (Chohan et al., 2020)[44]	06	The method is tested on the single plant disease type.	The system can integrated as the 3-layer model where each layer performs an individual function.
DNN (Gadekallu et al., 2021)[18]	5	The method is not compared with the existing methods to provide a better view of the results.	Use various datasets to enable proactive predictions.
Dense Net (Rubini and Kavitha, 2021)[30]		The VGG in the method has its fully connected nodes which takes almost 533 MB	The dataset with various plant features can be used for training along with different deep-learning algorithms.
AlexNet-CNN (Jadhav et al., 2021)[33]	8	The Alex Net increases the gradient vanishing issue.	The future works can include the performance metrics that enhance the accuracy with varying parameters.
Mobile-Atten (Chen et al., 2021)[45]	33	The method depends on the particular type of rice and not all the types.	Real-world applications can be utilized with this method.

4.2 *Classes of "soybean. leaf. dataset"*

The soybean. leaf. dataset is the dataset that contains the classes of the Soybean disease. The dataset comprises 3 classes including the healthy class and the same classes of images are utilized for both training and validation. Figure 4 illustrates the classes of Soybean from the Soybean. leaf. dataset., and the Table 14 shows the classes of Soybean from the Soybean.leaf. dataset.

Table 10. Comparison of machine learning techniques (2020–2022).

Machine learning	Total number of layers	Limitations	Future works
KNN (Ahmed et al., 2022)[7]	3	Only a single plant dataset is utilized	Various datasets can be included along with different classifiers.
Logistic regression (Tiwari et al., 2020)[10]	12	The model provides less robustness and accuracy with certain datasets.	The model can be embedded with the IoT to connect them with smartphones which enhances real-world agricultural production.

Table 11. Overall analysis of the machine learning techniques.

Dataset	Algorithm	Advantages	Drawbacks
Plant village- Potato leaf (Tiwari et al., 2020)[10]	Logistic Regression	The method aids in detecting the disease early stage which in turn increases the crop yield.	The model provides less robustness and accuracy with certain datasets.
Real-time data- Basil plants (Dhingra et al., 2019)[9]	Random forest	The method provides better accuracy with new segmentation and a new set of features.	The method includes real-time data which is captured in a specific atmosphere providing inefficient performance.
Real-Time Data- Rice (Shrivastava et al., 2019)[20]	SVM	The methods utilized aid in extracting the features and classifying them more accurately.	The performance of the approach is limited due to the unavailability of standard labeled images as the dataset.
Plant Village - Potato leaves (Islam et al., 2017)[21]	MSVM	The model aids in developing an automated and easily accessible system.	The dataset only specifies certain diseases which can be extended to obtain better accuracy.
Plant village and real-time data- Citrus fruits (Sharif et al., 2018)[19]	M-SVM	The method provides the maximum classification rate and the higher accuracy.	The dataset with minimal data provides inefficient performance.
Real-time dataset - Leaves of 10 different plants (Singh and Misra, 2017)[8]	SVM	The plant disease detection is carried out at the earliest stage.	The classification process can be improved with the various classifiers.
Real-time data- Tomato (Sabrol and Satish, 2016)[37]	Classification Tree	The classification tree is prone to overfitting and the large datasets are hard to interpret.	The availability of the classification models in the current scenario is large and they can be utilized for plant disease detection.
UCI machine learning repository - Soybean (Ahmed et al., 2022)[7]	KNN	The computational time is much less and is easy to interpret.	Only a single plant dataset is utilized.
Real-time data- Grape (Sannakki et al., 2013)[31]	K-means clustering	K-means is useful for datasets with several features since it scales well to larger dimensions.	The data points with small changes or initial conditions can result in significantly different cluster assignments.
Real-time data- Cotton leaf (Rothe and Kshirsagar, 2015)[32]	Backpropagation Neural Network.	The algorithm utilized is the fastest algorithm that provides higher accuracy.	The model is less robust due to the snake segmentation which takes a longer time as well as the fewer features extracted.
Real-Time Data - Rice plant (Chen et al., 2021)[45]	Mobile-Attention	The model has a small size which in turn provides a higher accuracy.	The method relies on the particular category of rice and not all the types.

Table 12. Overall analysis of the deep learning techniques.

Dataset	Algorithm	Advantages	Drawbacks
Plant village– Maize (Chen et al., 2020)[5]	VGG-Net	The algorithm provides better consistency and reliability	VGG-Net creates gradient gradient-exploding problem
Plant village dataset - Tomato (Gadekallu et al., 2021)[18]	PCA-WOA with DNN	The time required for implementation is low but the accuracy is high.	No comparative analyses are shown for better outcomes.
Plant village dataset- Potato, Tomato, Maize (Khamparia et al., 2020)[27]	Deep convolutional encoder network	The algorithm provides better consistency and reliability.	Some hyperparameters such as dropout and regularization can be added to the method.
Corn or maize disease dataset- Maize (Darwish et al., 2020)[4]	CNN	CNN requires a large set of labeled data, which requires high computational cost and time	The dataset can be collected from various real-time sources in different conditions.
Plant village dataset- Potato, Tomato, Rice, Corn, Apple, Grape (Kabir et al., 2021)[3]	CNN-Xception	CNN requires a large set of labeled data, which requires high computational cost and time	The multi-label disease classification dataset is currently available, So the dataset should be created with the diverse plant species.
Plant village dataset (Geetharamani and Pandian, 2017)[22]	Deep CNN	The algorithm provides better consistency and reliability.	Predictions can be made including the stem, flower, fruits, and so on.
Real-time dataset- Tomato (Liu and Wang, 2020)[29]	YOLO V3	The method creates strong robustness and has high positioning accuracy.	Various diseases should be collected with high resolution for the database to predict the various diseases accurately.
New plant disease dataset] - Tomato (Rubini and Kavitha, 2021)[30]	DenseNet	The model adapts the most diversified features which in turn increases the accuracy.	The dataset with various plant features can be used for training along with different deep-learning algorithms.
Plant Village.org dataset- Tomato (Brahimi et al., 2017)[6]	GoogleNet	The techniques handled in the research localized the disease regions which makes the model user-friendly.	Computational time and cost are high which makes the model less effective in certain regions.
Real-time data - Tomato (Sharma et al., 2020)[42]	S-CNN	The usage of segmented images instead of the full part in the detection elevates the accuracy of the results.	The S-CNN approach fails to predict the disease if the given sample contains more than one disease and is sensitive to segmentation.
Plant village- Tomato leaf (Karthik et al., 2020)[23]	Attention-based residual CNN	The model accepts any input size due to the extensibility of the design.	The results obtained through the attention mechanism vary on every dataset.
Real-time data- Cucumber leaves (Lin et al., 2019)[25]	CNN	The input is the unbalanced positive and negative images that produce high-accuracy detection.	The collection of data from the field under various conditions as the proposed method utilized the randomly picked images.
Plant Village- Tomato leaf (Prajwala et al., 2018)[24]	CNN	The method works with fewer resource constraints and minimal data.	CNN requires more labeled data to train the model.
CIFAR 10 - Tea (Ahmed et al., 2019)[26]	CNN	This approach lowers loss and accurately forecasts tea leaf disease.	This approach needs training for huge data.
Plant Village- Grape Leaf (Ji et al., 2020)[41]	United model CNN	The model reduces the over-fitting as well as improves the generalization ability with the data augmentation techniques.	To train the dataset the computational cost is high.

(continue)

Table 12. Continued.

Dataset	Algorithm	Advantages	Drawbacks
Plant village and real-time data- Mango leaves (Singh et al., 2019)[46]	MCNN	The method is computationally efficient.	The research is only on the mango leaves which can be extended to other plants also.
Real-time data- Soybean (Jadhav et al., 2021)	AlexNet- CNN	AlexNet can handle large-scale datasets which makes the method work faster than others.	The Alex Net increases the gradient vanishing issue.
New plant disease dataset- Soybean (Karlekar and Seal, 2020)[33]	SoyNet	The method increases the diversity of the pooling operation as well as the dropout and the activation layers which aids in achieving good accuracy.	SoyNet is the newly adopted technology that did not include the other plants.
Soybean. Leaf. dataset - Soybean (Nagasubramanian et al., 2018)[43]	DCNN xPlNet	Improves the performance accurately.	Understanding DCNN prediction can be challenging due to the complex nature of the features.
Plant village and Real-Time data- Maize/Corn, Strawberry, Tomato (Chohan et al., 2020)[47]	CNN	The method provides the automatic detection and diagnosis of plant diseases.	The method is tested on the single plant disease type.
Plant village data and new plant disease dataset - Leaves (Cristin et al., 2020)[1]	Rider-CSA based DBN	The model provides higher accuracy.	The algorithm has a slow convergence speed.
Plant Village Data - Banana Leaf (Amara et al., 2017)[28]	Le Net	Provided high accuracy with less computational effort.	Lack of severity estimation of the plant disease
Real-Time Data- Apple leaf (Liu et al., 2017)[40]	CNN	The model has a faster convergence rate.	The model works with the particular disease since the dataset is collected under a specific season.

5 Role of deep learning in disease prediction

machine learning approaches are the ancestors of deep learning but the deep learning approaches overcame the machine learning methods and set a benchmark for various research. They also experimented with several unique methods, algorithms, and techniques to enhance the efficacy of the approach but they exhibit disadvantages too. The KNN is a model that requires no training period so it finds the easy implementation. However, the approach does not employ well in a huge dataset and high dimensionality and is highly sensitive to noise.[7] Though decision trees contain several advantages such as being fast and easily interpretable, capturing the non-linear relationships, scale-invariant, and non-parametric, they are prone to overfitting since it takes more time and the reusability of the model leads to inefficient performance.[37] Even though the Naïve Bayes aids real-time predictions and in solving the multi-class prediction problems, it assumes that the predictions are independent and face the zero-frequency problem.[26] Deep Belief Network (DBN) enables high robustness and the efficient usage of the hidden layers in the architecture. Moreover, DBN hates the time-picking techniques and provides high performance with large data. Though DBN is advantageous, it has a few disadvantages such as being expensive to train, requiring several hardware equipment, and so on.[1] Deep CNN is a well-known method with huge benefits. The key factor of the method is better consistency, requires fewer parameters, and is computationally efficient. The drawbacks of the method are it is limited to generalization, is not efficient with small datasets, and requires high computational resources.[22] The Cuckoo optimization algorithm (COA) is easier to apply and has fewer tuning parameters but has a slower convergence rate as it sooner falls into the local optima.[1]

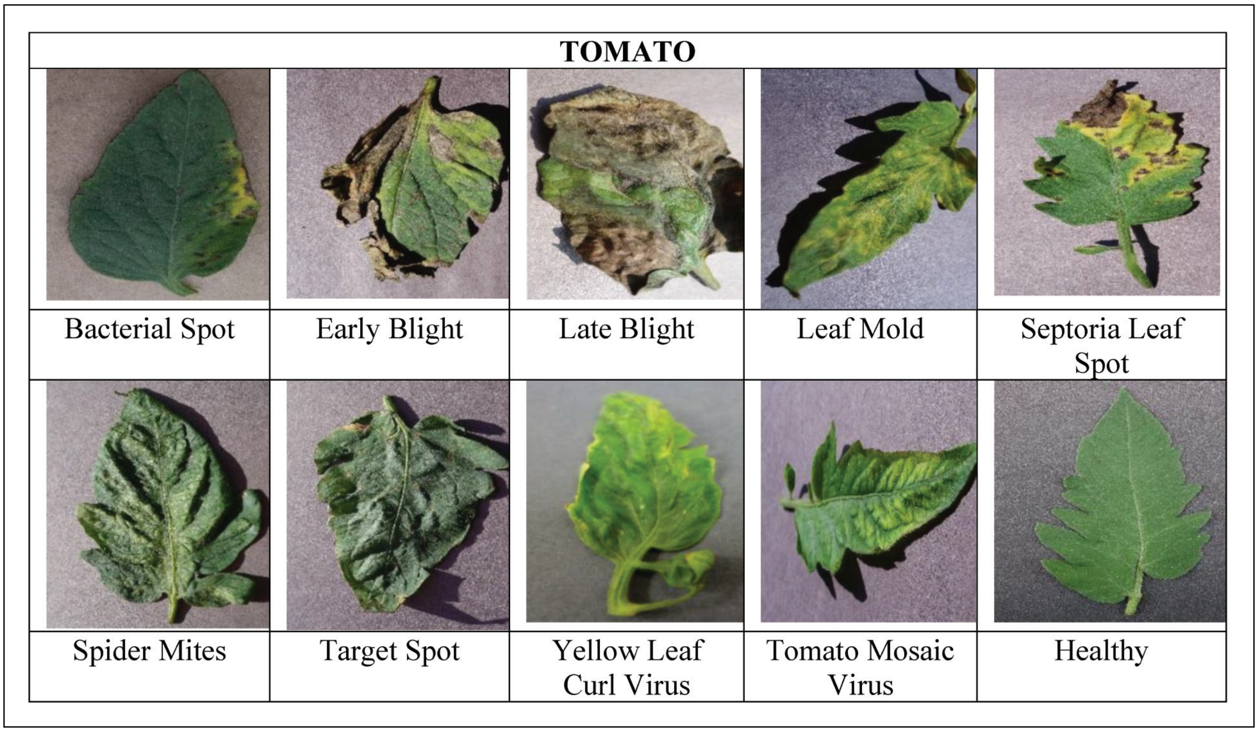

Figure 3. Classes of tomato from the new plant disease dataset.

Table 13. Classes of tomato from the new plant disease dataset.

	Number of images	
Classes	Training	Validation
Bacterial Spot	1702	425
Early Blight	1920	480
Late Blight	1851	463
Leaf Mold	1882	470
Septoria Leaf Spot	1745	436
Spider Mites	1741	435
Target Spot	1827	457
Yellow Leaf Curl Virus	1961	490
Tomato Mosaic Virus	1790	448
Healthy	1926	481

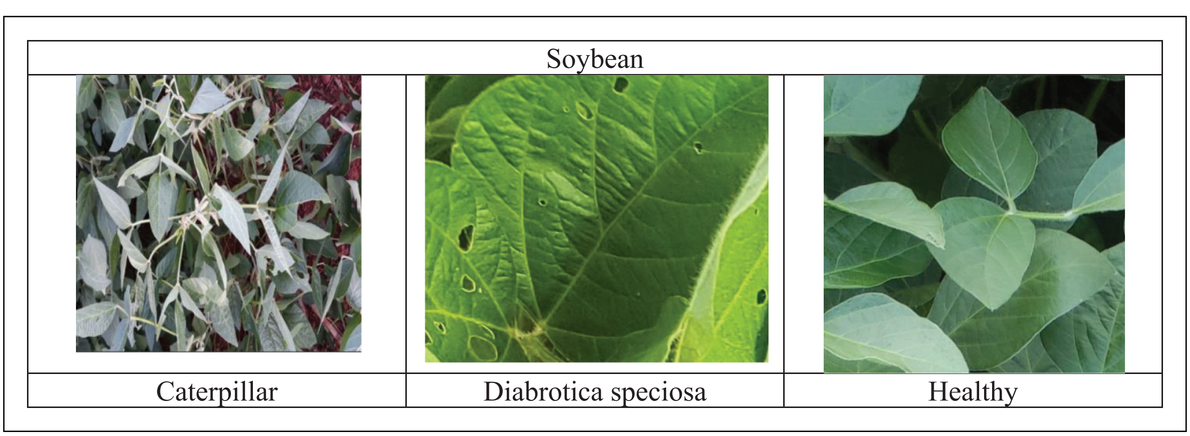

Figure 4. Classes of soybean from the soybean.leaf. dataset.

Table 14. Classes of soybean from the soybean.leaf. dataset.

Classes	Number of images
Caterpillar	3309
Diabrotica speciosa	2205
Healthy	896

6 Research gaps

Research gaps identified in the different literature are described in this section.

6.1 Gaps identified for the segmentation models

The Otsu method used for the segmentation assumes equal variances in the foreground and background leading to poor segmentation. In addition, the method is not robust to noise and the variation in the lighting of the surroundings can be reflected in the results of segmentation.[37] Further, the Genetic algorithm in segmentation finds difficulty with the repetitive calculation of the fitness value and does not provide efficient performance in segmentation. Since the algorithm does not require the derivative information, they find it difficult to solve simple problems.[8] Specifically, piFCM is majorly used for unsupervised image segmentation and it does not have local information on the image content which leads to misconception of the pixels and provides inefficient performance.[1] The active contour model has inadequate contour convergence for concave borders and curve flow is commenced at a great distance from the minimum. The model is noise-sensitive as well as highly complex.[32]

6.2 Gaps identified in the machine learning model

The KNN model does not work well with a large dataset and high dimensionality providing fewer efficacies. The model is highly sensitive to noise and requires high memory since the model is computationally expensive.[7] Subsequently, the decision boundary generated by a Support Vector Machine (SVM) can be quite intricate, posing difficulties in the interpretation and comprehension of the model's prediction process in high-dimensional domains.[20] Moreover, logistic regression underperforms on the multiple or non-linear decision boundaries and is not flexible enough to obtain complex relationships. In addition, the approach relies on the feature count and a smaller number leads to overfitting.[10]

6.3 Gaps identified in the feature extraction models

Gray-level co-occurrence matrix requires high computational time and cost and also the method struggles to work with the textures which creates an imbalance of the classes. The method in turn provides the inaccurate or less biased texture feature extraction.[21] Further, Scale-Invariant Feature Transform (SIFT) is based on the histogram of gradient which is computationally heavy and mathematically complicated. Each pixel should be patched and requires a high-power execution since does not work for low-powered devices.[32] Furthermore, VGG19 is very slow to train as well as the model requires a lot of space and bandwidth as the storage. The method contains millions of parameters which leads to the gradient exploding problem, thus providing inefficient performance.[10]

7 Conclusion

The exploration of databases, algorithms, and future prospects in plant disease prediction using plant images highlights the potential for technology to revolutionize agriculture. Significant progress has been made in illness identification and prevention thanks to the availability of a variety of datasets and creative machine-learning algorithms. The complicated interactions between data, algorithms, and domain knowledge that have shaped the landscape of plant health monitoring have been highlighted in this overview. Datasets such as the soybean leaf dataset, comprising thousands of images depicting various soybean diseases, and the new disease dataset have been used to train and assess predictive models. These models, which range from simple deep-learning architectures to complex classical algorithms, have shown remarkable accuracy and efficiency in detecting diseases at an early stage and supporting farmers in reaching justifiable decisions. Future developments in plant disease prediction offer tremendous promise. Collaborations involving agricultural experts, data scientists, and technology developers are expected to improve real-time disease monitoring, improve dataset variety, and improve the accuracy of current models. A more thorough understanding of disease dynamics is expected to result from the use of multispectral and hyperspectral imaging, environmental factors, and genomic data.

8 Future works

Machine learning models may have various disadvantages such as, overfit and underfit. The training data collects noise instead of genuine information; in contrast, if the model is too simplistic to represent the data, it may underfit. Particularly, when the model cannot generalize the testing data without the proper learning in the training data patterns leads to under-fitting. These machine learning drawbacks lead to the use of deep learning to forecast plant diseases, which has several benefits. Due to various factors, there are some restrictions such as contextual imbalance, computational intensity, black box problems, and so on. These restrictions can be overcome with emerging algorithms and optimizations as well as techniques in each processing area. The deep learning models with the new technologies will be excellent tools for predicting even minor indications of disease in the plants.

Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
COA	Cuckoo optimization algorithm
CSA	Cuckoo Search Algorithm
DBN	Deep Belief Network
DL	Deep Learning
DNN	Deep Neural Network
GLCM	Gray-level co-occurrence matrix
GWT	Google Web Toolkit
HIC	Histogram Information Content
HOG	Histogram of Oriented Gradients
KNN	K-nearest neighbor
ML	Machine Learning
PCA	Principal Component Analysis
PCM	Pulse code modulation
RNN	Recurrent Neural Networks
ROA	Rider Optimization Algorithm
SIFT	Scale-Invariant Feature Transform
SVM	Support Vector Machine
VGGNet	Visual Geometry Group Network
WOA	Whale Optimization Algorithm

References

1. Cristin R, Kumar BS, Priya C, et al. Deep neural network-based Rider-Cuckoo Search Algorithm for plant disease detection. *Artificial Intelligence Review* 2020; 53: 4993–5018.
2. Fuentes A, Yoon S, Kim SC, et al. A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition.Sensors. 2017; 17: 2022.
3. Kabir MM, Ohi AQ and Mridha MF. A multi-plant disease diagnosis method using a convolutional neural network. Computer Vision and Machine Learning in Agriculture, pp.99–111. 2021.
4. Darwish A, Ezzat D and Hassanien AE. An optimized model based on convolutional neural networks and orthogonal learning particle swarm optimization algorithm for plant disease diagnosis. *Swarm and Evolutionary Computation* 2020; 52: 100616.
5. Chen J, Chen J, Zhang D, et al. Using deep transfer learning for image-based plant disease identification. *Computers and Electronics in Agriculture* 2020; 173: 105393.
6. Brahimi M, Boukhalfa K and Moussaoui A. deep learning for tomato diseases: classification and symptoms visualization. *Applied Artificial Intelligence* 2017; 31: 299–315.
7. Ahmed AS, Obeas ZK, Alhade BA, et al. Improving prediction of plant disease using k-efficient clustering and classification algorithms. *IAES International Journal of Artificial Intelligence (IJ-AI)* 2022; 11: 939–948.
8. Singh V and Misra AK. Detection of plant leaf diseases using image segmentation and soft computing techniques. *Information Processing in Agriculture* 2017; 4: 41–49.
9. Dhingra G, Kumar V and Joshi HD. A novel computer vision-based neutrosophic approach for leaf disease identification and classification. *Measurement* 2019; 135: 782–794.
10. Tiwari D, Ashish M, Gangwar N, et al. Potato leaf disease detection using deep learning. In 2020 4th International Conference on intelligent computing and Control Systems (ICICCS) (pp. 461–466). IEEE. 2020, May.
11. New plant disease dataset "https://www.kaggle.com/datasets/vipoooool/new-plant-diseases-dataset" (accessed on August 2023).

12. Tomato leaf disease detection “https://www.kaggle.com/datasets/kaustubhb999/tomatoleaf” (accessed on August 2023).
13. Soybean.Leaf.dataset “https://www.kaggle.com/datasets/maeloisamignoni/soybeanleafdataset” (accessed on August 2023).
14. UCI machine learning repository “https://archive.ics.uci.edu/” (accessed on August 2023).
15. Corn or Maize dataset “https://www.kaggle.com/datasets/smaranjitghose/corn-or-maize-leaf-disease-dataset” (accessed on August 2023).
16. Rice Leaf diseases dataset “https://www.kaggle.com/datasets/vbookshelf/rice-leaf-diseases” (accessed on August 2023).
17. plant Village “https://www.kaggle.com/datasets/tushar5harma/plant-village-dataset-updated” (accessed on August 2023).
18. Gadekallu TR, Rajput DS, Reddy MPK, et al. A novel PCA–whale optimization-based deep neural network model for classification of tomato plant diseases using GPU. *Journal of Real-Time Image Processing* 2021; 18: 1383–1396.
19. Sharif M, Khan MA, Iqbal Z, et al. Detection and classification of citrus diseases in agriculture based on optimized weighted segmentation and feature selection. *Computers and Electronics in Agriculture* 2018; 150: 220–234.
20. Shrivastava VK, Pradhan MK, Minz S, et al. Rice plant disease classification using transfer learning of deep convolution neural network. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 2019; 42: 631–635.
21. Islam M, Dinh A, Wahid K, et al. Detection of potato diseases using image segmentation and multiclass support vector machine. In 2017 IEEE 30th canadian conference on electrical and computer engineering (CCECE) (pp. 1–4). IEEE. 2017, April.
22. Geetharamani G and Pandian A. Identification of plant leaf diseases using a nine-layer deep convolutional neural network. *Computers & Electrical Engineering* 2019; 76: 323–338.
23. Karthik R, Hariharan M, Anand S, et al. Attention embedded residual CNN for disease detection in tomato leaves. *Applied Soft Computing* 2020; 86: 105933.
24. Prajwala T, Pranathi A, SaiAshritha K, et al. Tomato leaf disease detection using convolutional neural networks. In 2018 eleventh international conference on contemporary computing (IC3) (pp. 1–5). IEEE. 2018, August.
25. Lin K, Gong L, Huang Y, et al. deep learning-based segmentation and quantification of cucumber powdery mildew using convolutional neural network. *Frontiers in Plant Science* 2019; 10: 155.
26. Ahmed K, Shahidi TR, Alam SMI, et al. Rice leaf disease detection using machine learning techniques. In 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI) (pp. 1–5). IEEE. 2019, December.
27. Khamparia A, Saini G, Gupta D, et al. Seasonal crops disease prediction and classification using deep convolutional encoder network. *Circuits, Systems, and Signal Processing* 2020; 39: 818–836.
28. Amara J, Bouaziz B and Algergawy A. A deep learning-based approach for banana leaf diseases classification. Datenbanksyste-mefür Business, Technologie und Web (BTW 2017)-Workshopband. 2017.
29. Liu J and Wang X. Tomato diseases and pests detection based on improved Yolo V3 convolutional neural network. *Frontiers in Plant Science* 2020; 11: 898.
30. Rubini PE and Kavitha P. deep learning model for early prediction of plant disease. In 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV) (pp. 1104–1107). IEEE. 2021, February.
31. Sannakki SS, Rajpurohit VS, Nargund VB, et al. Diagnosis and classification of grape leaf diseases using neural networks. In 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT) (pp. 1–5). IEEE. 2013, July.
32. Rothe PR and Kshirsagar RV. Cotton leaf disease identification using pattern recognition techniques. In 2015 International conference on pervasive computing (ICPC) (pp. 1–6). IEEE. 2015, January.
33. Jadhav SB, Udupi VR and Patil SB. Identification of plant diseases using convolutional neural networks.International. *Journal of Information Technology* 2021; 13: 2461–2470.
34. Grape disease dataset “https://www.kaggle.com/datasets/rm1000/grape-disease-dataset-original” (accessed on August 2023).
35. Mango leaf disease dataset “https://www.kaggle.com/datasets/aryashah2k/mango-leaf-disease-dataset” (accessed on August 2023).
36. Banana Leaf disease “https://www.kaggle.com/datasets/harshitbana/banana-leaf-disease” (accessed on August 2023).
37. Sabrol H and Satish K. Tomato plant disease classification in digital images using classification tree. In 2016 international conference on communication and signal processing (ICCSP) (pp. 1242–1246). IEEE. 2016, April.
38. Sladojevic S, Arsenovic M, Anderla A, et al. Deep neural networks based recognition of plant diseases by leaf image classification. *Computational Intelligence and Neuroscience* 2016; 2016: 3289801.
39. Mohanty SP, Hughes DP and Salathé M. Using deep learning for image-based plant disease detection. *Frontiers in Plant Science* 2016; 7: 1419.
40. Liu B, Zhang Y, He D, et al. Identification of apple leaf diseases based on deep convolutional neural networks. *Symmetry* 2017; 10: 11.
41. Ji M, Zhang L and Wu Q. Automatic grape leaf diseases identification via UnitedModel based on multiple convolutional neural networks. *Information Processing in Agriculture* 2020; 7: 418–426.

42. Sharma P, Berwal YPS and Ghai W. Performance analysis of deep learning CNN models for disease detection in plants using image segmentation. *Information Processing in Agriculture* 2020; 7: 566–574.
43. Karlekar A and Seal A. SoyNet: Soybean leaf diseases classification. *Computers and Electronics in Agriculture* 2020; 172: 105342.
44. Chohan M, Khan A, Chohan R, et al. plant disease detection using deep learning. *International Journal of Recent Technology and Engineering* 2020; 9: 909–914.
45. Chen J, Zhang D, Zeb A, et al. Identification of rice plant diseases using lightweight attention networks. *Expert Systems with Applications* 2021; 169: 114514.
46. Singh UP, Chouhan SS, Jain S, et al. Multilayer convolution neural network for the classification of mango leaves infected by anthracnose disease. *IEEE access* 2019; 7: 43721–43729.
47. Nagasubramanian K, Jones S, Sarkar S, et al. Hyperspectral band selection using genetic algorithm and support vector machines for early identification of charcoal rot disease in soybean stems. *plant Methods* 2018; 14: 1–13.

Author Biographies

Smita R Sankhe is a dedicated Assistant Professor in the Computer Engineering Department at K. J. Somaiya College of Engineering. With over 16 years of teaching experience and a brief stint in the industry, she has a strong passion for her field. Her expertise lies in Machine Learning, Deep Learning, Natural Language Processing, and Artificial Intelligence. Smita holds a B. Tech in Computer Science from SNDT University, an M.E. in Computer Engineering from Mumbai University and she is currently pursuing her Ph.D. at Kalinga University. Over the years, Smita has guided students through a wide range of courses, from Object-Oriented Programming to Full Stack Development. She's a prolific researcher with over 20 publications, including 12 in prestigious peer-reviewed journals, and has presented her work at renowned IEEE conferences. Some of her standout projects include innovative IoT systems for water management and PH monitoring, supported by grants from the University of Mumbai. Beyond teaching, Smita is an active mentor and evaluator, having guided students in hackathons like Smart India Hackathon and reviewed papers for international journals and conferences. She also contributes to her department as NAAC Criteria 3 In-charge and a member of the Research and Development cell. Smita's enthusiasm for technology and education shines through in her workshops, expert lectures, and commitment to nurturing the next generation of engineers.

Asha Ambhaikar is B.E., M.Tech. & Ph.D. in Computer Science and Engineering. At present she is working as a Professor (CSE), Kalinga University Naya Raipur. India. She has more than 30 years of Academic experience. She has published more than 185 research papers in reputed National and International Journals (Scopus Indexed & SCI), 277 Google citations, 21 patents and 3 patent Grant. 15 PhD scholars completed their PhD under her guidance. She is a member of the Editorial Board and Reviewer of 31 various international journals and conferences. She is also the member of 20 professional societies like life member of IAENG (International Association of Engineers, Hong Kong, IEEE, Indian Society of Technical Education (ISTE), Computer Society of India (CSI), IET, ASDF, Computer Science Teachers Association (CSTA), and Association for Computing Machinery (ACM), New York, USA, IACSIT (International Association of Computer Science and Information Technology, Singapore. Member of SDIWC (The Society of Digital Information and Wireless Communication, USA. She has also chaired various National and International Conferences around various countries as a keynote speaker and session chair. She has also published 9 books by Lambert Publication Germany and 14 book chapter in Springer and Elsevier. She has attended and conducted more than 60 Quality Improvement Programs/FDPs/Refresher courses. She has received 17 National & International Awards.