International Journal of Intelligent Communication and Computer Science
Vol. 2, No. 3, 2024, Pages 62-80.

Plant Disease Detection Using Machine Learning: A Comprehensive Review

Shivani Mishra[1] and Rohitashwa Pandey[2]

[1]Research Scholar, Department of Computer Science and Engineering, Bansal Institute of Engineering and Technology, Lucknow, Affiliated to AKTU, Lucknow
[2]Department of Computer Science and Engineering Bansal Institute of Engineering and Technology, Lucknow,, Affiliated to AKTU, Lucknow

Review Paper

Email: shivani1mishra1996@gmail.com
Received: 27Aug. 2024, Revised: 13 Nov. 2024 Accepted: 11 Dec. 2024

Abstract:

Early and accurate plant diseases detection is quite critical in order to have good yield and healthy crops. Conventional techniques of disease detection, such as visual inspections by experts, take much too time, subjective, and require extensive labour. At present, machine learning (ML) acts as a powerful tool for automating the plant disease detection process. Here, an in-depth review of the various machine learning approaches used for plant disease detection, including supervised, unsupervised, and deep learning techniques. We examine the key ML algorithms applied to plant disease recognition, discuss the challenges associated with their deployment, and present promising future directions for further research in this field. Moreover, we explore the role of sensor technologies, computer vision, and data fusion in improving the results of these diagnostic systems.

Keywords: Plant Disease, Machine Learning and Artificial Intelligence

1. Introduction

In agriculture, plant diseases are a major concern that can significantly reduce crop yield and quality, with far-reaching consequences for both food security and the economy. Crop diseases can lead to the degradation of plant health, weakening the plants and making them more susceptible to other environmental stresses. This ultimately results in reduced productivity, lower market value, and, in some cases, crop failure [1]. Now, the requirement of food continues to rise as the population grows, making it increasingly important to develop effective strategies for disease management. Timely and accurate disease detection is a key factor in ensuring healthy crops and optimizing agricultural practices. Early intervention allows farmers to take preventive or corrective measures that can reduce the spread of diseases, minimize losses, and prevent unnecessary use of harmful chemicals like pesticides.
Many farming operations have traditionally relied on human visual examination by agricultural specialists as the usual way of detecting plant diseases. These methods involve inspecting plants individually or in small groups to identify visible symptoms of disease. While effective in small-scale operations, these methods are quite costly and time taking and accuracy is also a major concern with these methods. Furthermore, as farm sizes increase and the demand for faster, more efficient practices grows, traditional methods become unsustainable. Large agricultural fields or greenhouses often make it difficult to visually inspect every plant or to detect diseases early enough, leading to delays in diagnosis and potentially severe economic losses. The complexity of identifying diseases early, especially when symptoms are subtle or at initial phase, adds to the

challenge. As a result, there is a growing need for automated, scalable solutions that can make timely plant disease detection.

ML [2], a subset of artificial intelligence (AI) [3], is a promising solution for these issues. ML techniques enable the automation of disease detection by training models to recognize patterns in data, including images, sensor readings, and environmental factors. Unlike traditional methods, ML models can be very quick and accurate, identifying disease symptoms even before they are visibly noticeable to humans. Large datasets of images and sensor data collected from agricultural fields are used for training these models, enabling them to learn complex relationships between plant health and disease indicators. By using features such as color changes, texture variations, and abnormal patterns on leaves or stems, ML models can detect specific diseases with high accuracy and speed. The ability to process data in real-time allows for continuous monitoring of plant health, empowering farmers to respond promptly to emerging issues and reduce the need for widespread pesticide use, which can be harmful to the environment and human health.

ML-based plant disease detection can scale the intensity of disease spread in the crops. With the increasing availability of low-cost sensors, drones, and high-resolution cameras, vast amounts of image data can be captured and analysed across large fields in real time. This capability enables farmers to keep an eye on their crops regularly and detect disease outbreaks early in the growing season, preventing the spread of disease to unaffected areas. Moreover, by identifying disease patterns with high accuracy, ML models help farmers take more targeted actions, such as applying pesticides only where needed or choosing appropriate treatment options for specific diseases. This not only helps preserve crop quality but also provides a better way to monitor the crops continuously.

Here, a detailed overview of ML techniques which can be used for detecting plant disease, with a focus on the most commonly used approaches, including supervised learning, unsupervised learning, and deep learning. Supervised learning techniques, such as decision trees, support vector machines (SVM), and random forests, rely on labelled datasets to learn the characteristics of plant diseases and predict the disease status of new, unseen images. Unsupervised learning techniques, on the other hand, do not require labelled data and instead look for patterns or clusters in the data, which can be useful for detecting unknown diseases or novel symptoms. Deep learning, particularly convolutional neural networks (CNNs) is very helpful in this domain. This is due to its feature of automatically learning and extracting complex features from raw image data, significantly improving detection accuracy and efficiency.

The review also highlights emerging trends in plant disease detection, such as transfer learning, generative adversarial networks (GANs), and multimodal data fusion. Transfer learning allows models to leverage knowledge gained from one task or dataset and apply it to a different but related task, making it easier to develop disease detection systems in regions with limited training data. GANs are increasingly being used to generate synthetic data for training models, helping to overcome the issue of insufficient labeled data, especially for rare plant diseases. Multimodal data fusion combines information from multiple sources, such as images, environmental sensors, and weather data, to provide a more comprehensive understanding of plant health and disease dynamics.

However, despite the significant progress made in applying machine learning to plant disease detection, there are still challenges that need to be addressed. One of the primary challenges is the availability and quality of data. High-quality annotated datasets are essential for training ML models, but in many cases, such data is scarce, especially for rare diseases. Furthermore, many of the existing models face difficulties in generalizing across different plant species, growing conditions, or geographical locations, which can limit their applicability in diverse agricultural settings. Real-time implementation in large-scale agricultural environments is another challenge, as ML models often require substantial computational resources, which may not be readily available in all farming contexts. There are also concerns about model interpretability, as farmers and agricultural experts need to understand why a particular disease diagnosis was made to trust and act on the recommendations generated by the models.

2. Plant Disease Detection Methodology

It is quite important to detect plant disease detection in agricultural management, aiming to monitor plant health, minimize crop losses, and upgrade yield quality. This procedure of detecting plant diseases using images typically includes numerous stages: Image Acquisition, Image Pre-processing, Image Segmentation, Feature Extraction, and Classification as shown in Figure 1. The detailed explanation of each step is detailed below:

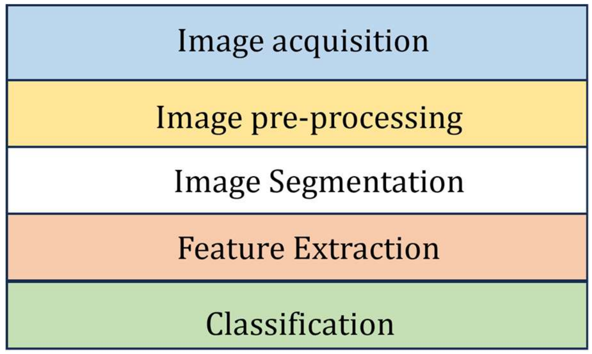

Figure 1: Plant disease detection process

2.1 Image Acquisition

This is the first step in the plant disease detection process. It involves capturing images of plants, typically using digital cameras, smartphones, or specialized equipment like drones or sensors. These images can be taken in various environments and under different lighting conditions. For accurate disease detection, high-quality images are essential, as they need to capture enough detail to identify symptoms such as spots, lesions, or discoloration caused by the disease. In modern agricultural practices, multispectral or hyperspectral imaging is sometimes used, which provides additional information beyond the visible spectrum, such as infrared or ultraviolet data, to better assess plant health.

Figure 2: Sample images for plant diseases

Figure 2 presents a collection of sample images depicting various plant diseases. These images showcase different types of plant diseases, each with distinct symptoms such as leaf discoloration, spots, lesions, or wilting. The purpose of these sample images is to illustrate the visual diversity of plant diseases, which can vary based on the pathogen (fungal, bacterial, viral) or environmental factors. By analysing these images, researchers and practitioners can better understand how different diseases manifest on plants, helping in the development of more effective detection and treatment methods. The images may serve as a reference for distinguishing between healthy and diseased plants, an important task in agriculture and crop management.

2. Image Pre-processing

Once, the input images are available, these input images are passed through pre-processing stage. This is done in order to improve the quality and make them suitable for further analysis. This stage involves various techniques to improve the contrast, remove noise, and standardize the image. Common pre-processing steps include:

1. **Noise reduction**: Removing random noise that might distort the image and interfere with disease detection.
2. **Contrast enhancement**: Improving the visibility of important features like disease symptoms by adjusting brightness or contrast.
3. **Resizing**: Standardizing the image size to make processing more efficient and consistent.
4. **Colour normalization**: Adjusting colour variations across different images or lighting conditions to ensure consistency in feature analysis.

5. **Filtering**: Applying techniques like Gaussian blur to reduce unnecessary detail or smooth out small, irrelevant details.

These pre-processing steps ensure that the image is clear, standardized, and suitable for effective analysis.

3. Image Segmentation

In the process of image processing, this process is a method for dividing a picture into discrete, significant areas or segments. To identify objects or regions of interest within the image, this technique entails assembling pixels that have comparable properties, such as color, intensity, or texture. The goal is to isolate the areas of the image that are of interest. Segmentation helps reduce the complexity of the image and focuses on relevant portions like leaves, stems, or fruit. Common segmentation techniques include:

1. **Thresholding**: Dividing the image into regions based on pixel intensity values.
2. **Edge Detection**: Identifying the boundaries of objects within the image to separate diseased from healthy areas.
3. **Clustering**: Grouping pixels with similar characteristics (such as colour or texture) to form clusters that represent distinct regions of the image.
4. **Region Growing**: Starting with a seed point and expanding the region based on pixel similarity criteria.

Effective segmentation is critical in isolating diseased areas from the healthy parts of the plant, ensuring that only relevant regions are analysed further.

4. Feature Extraction

Feature extraction involves identifying and quantifying distinctive characteristics of the segmented regions that may indicate the presence of disease. These features could be based on colour, texture, shape, or patterns found in the diseased areas. Commonly extracted features include:

1. **Colour features**: These can include the colour intensity or variations in colour patterns (e.g., yellow or brown spots on leaves) that signify disease.
2. **Texture features**: Patterns such as spots, lesions, or irregularities on the plant surface can be analysed with the help of methods such as the Gray-Level Co-occurrence Matrix (GLCM), which measures the texture by analysing pixel intensity relations.
3. **Shape features**: The size, shape, and boundaries of the diseased regions can provide important clues. For example, the shape of lesions or the outline of affected leaves can be distinctive for certain diseases.
4. **Statistical features**: These may include measures of pixel distribution, contrast, and homogeneity within the diseased region.

By extracting these features, the algorithm can develop a set of numerical values which are used to describe the important aspects of the diseased regions, which are later used for classification.

5. Classification

Classification is the final stage in the plant disease detection process. In this stage, the image is categorized into predefined classes with the help of the extracted features. Various ML and DL techniques are applied to classify the plant images based on the extracted features. Common classification methods include:

1. **Traditional Machine Learning**: Techniques like Support Vector Machines (SVM), Decision Trees, k-Nearest Neighbours (k-NN), and Random Forests are often used, where a model is trained on labelled data to recognize patterns associated with different disease types.
2. **Deep Learning**: CNNs are particularly effective for image classification, as they are capable of automatically learning the features from raw image data and are highly efficient in detecting complex patterns in plant diseases.
3. **Neural Networks**: These models can analyse large sets of extracted features and classify the image based on learned patterns.

The classification step outputs the final result, which can be classified into diseased or healthy or may be classified into a number of diseases. The model's accuracy and precision relies on the quality of the features extracted, the training data, and the chosen classification method.

2. Machine Learning Approaches in Plant Disease Detection

2.1 Supervised Learning

This is widely adopted methods in plant disease detection. In this method, a model is trained using labelled data. Here, a machine learning algorithm learns to map input features (such as plant image characteristics) to predefined categories (such as healthy or diseased). Once model is trained, it can easily estimate the disease status of new, unseen images.

2.1.1 Decision Trees

Decision trees are a basic yet effective supervised learning algorithm. It is very effective in classification tasks in plant disease detection. In this algorithm, data is divided into various classes as per the decision rules derived from the input features. These rules are constructed by recursively dividing the dataset into smaller subsets, leading to a binary decision that categorizes the image into either a healthy or diseased class. Despite being simple, decision trees are interpretable and useful in low-data scenarios (Figure 3). However, they can suffer from overfitting and may not generalize well to complex datasets.

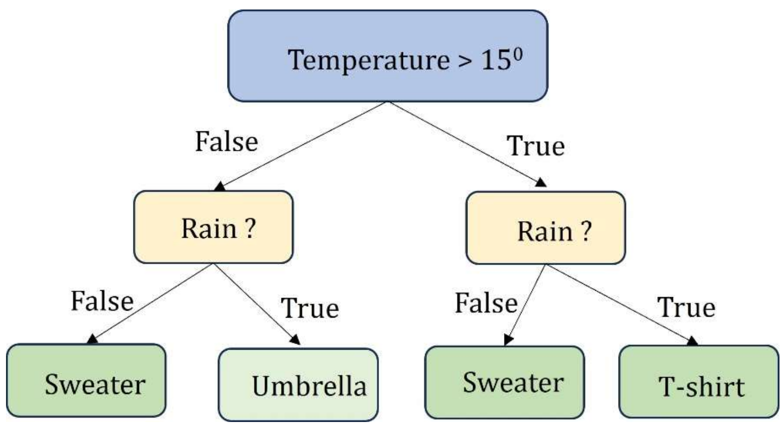

Figure 3: Schematic of Decision Tree process

The detection and classification of plant diseases have become a vital area of research due to their impact on agricultural productivity. Various machine learning techniques, particularly decision tree classifiers, have been employed for accurate identification of diseases affecting crops. The following studies explore the use of these methods in plant disease detection.

Rajesh et al. (2020) [4] presented a comprehensive approach for detecting and classifying leaf diseases by utilizing decision tree algorithms. The authors employed image processing techniques to extract features from leaf images and then applied decision tree classifiers to categorize the diseases. Their findings highlighted the effectiveness of decision trees in distinguishing between different types of leaf diseases with high accuracy. This study demonstrated how decision tree models could be used for real-time plant health monitoring in agricultural practices.

Chopda et al. (2018) [5] applied a decision tree classifier to detect diseases in cotton crops. By collecting data from cotton plant images, they used the decision tree algorithm to identify various disease symptoms. The study emphasized the adaptability of decision trees in agricultural environments, where the simplicity and interpretability of the model offer a significant advantage in identifying diseases with minimal computational resources.

Ramesh et al. (2018) [6] explored the use of ML techniques, including decision trees, to detect a variety of plant diseases. They combined image processing methods with ML algorithms for classifying the plant diseases. The research illustrated that decision trees, when combined with other techniques like support vector machines, could be very effective in detecting plant disease.

Mengistu et al. (2018) [7] focused on coffee plant disease identification using a hybrid approach combining image processing with decision tree classifiers. The study integrated both colour and texture-based features extracted from coffee plant images to identify diseases. Their research demonstrated that the hybrid approach improved the detection accuracy compared to using decision trees alone, showing promise for automated systems in agriculture.

Ahmed et al. (2019) [8] examined rice leaf disease detection using various ML algorithms, including decision trees. By processing images of rice leaves, they developed a model capable of identifying several diseases, such as bacterial blight and leaf streak. The authors concluded that decision tree classifiers could be integrated into mobile applications for on-the-go disease detection in rice crops.

2.1.2 Support Vector Machines (SVM)

SVM are powerful classifiers that work by finding the optimal hyperplane that separates data into different classes (Figure 4). SVM is particularly effective in high-dimensional feature spaces, such as those generated by images, and is well-suited for plant disease classification tasks where the dataset may be small or moderately large. SVM is particularly useful when the boundary between healthy and diseased plants is not linear and can handle a number of kernel functions to improve performance. However, SVM models can be computationally intensive and require careful parameter tuning.

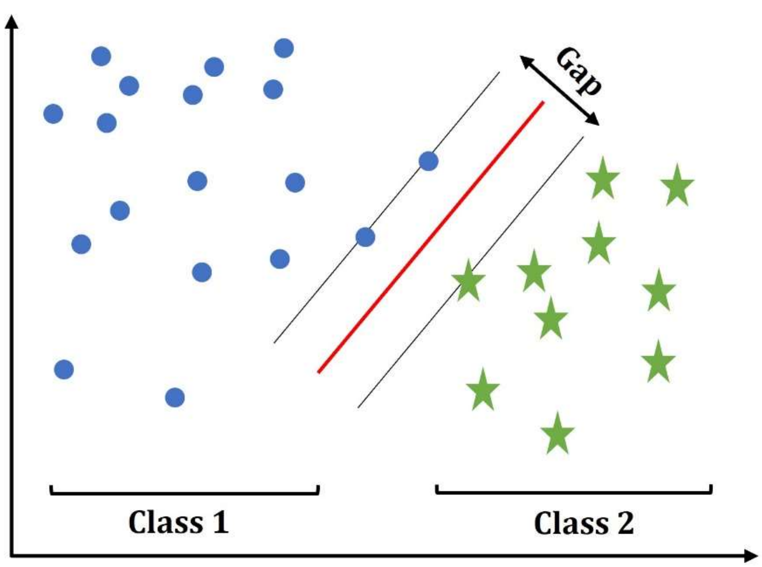

Figure 4: Schematic of SVM process

One of the most used strategies for detecting plant diseases is ML, particularly SVM. SVM is a prominent option in many research devoted to the classification and identification of plant diseases because of its well-known robustness and efficacy in managing high-dimensional data. Important research that looks at the use of SVM and other machine learning techniques in this field are highlighted below.

Das et al. (2020) [9] use SVM for classifying plant diseases. The authors applied image processing techniques to extract relevant features from leaf images and then used SVM to classify the diseases. The study demonstrated that SVM, when paired with feature extraction methods, could efficiently classify different leaf diseases.

Shruthi et al. (2019) [10] made a review of ML classification methods for plant disease detection, with a focus on SVM. The paper evaluated various methods and discussed how SVM, due to its high accuracy and precision, is often the algorithm of choice for disease classification tasks. This review also highlighted the advantages of SVM in handling complex and noisy datasets, which is common in agricultural data, which makes it very useful in the case of large-scale agricultural applications.

Sahu and Pandey (2023) [11] proposed an optimal hybrid multiclass SVM model for plant leaf disease detection. They combined SVM with the spatial fuzzy C-means clustering algorithm to improve the detection accuracy. This hybrid model was particularly effective in handling multiple classes of diseases in plant leaves, enhancing the system's ability to accurately categorize various plant diseases.

Thaiyalnayaki and Joseph (2021) [12] explored the combination of SVM and DL models for plant disease classification. The study found that combining SVM with deep learning techniques improved the overall result and accuracy. SVM acted as a strong classifier for disease categories, while deep learning models helped in automatic feature extraction from raw data. This hybrid approach has shown to outperform traditional methods, especially in cases with complex and large datasets.

Mokhtar et al. (2015) [13] detected tomato leaf diseases using SVM. The research utilized SVM to classify different disease types affecting tomato plants based on their leaf images. The results showed that SVM could successfully detect various tomato leaf diseases with high classification accuracy, thus proving its effectiveness in agricultural applications. This study highlighted the potential of SVM as a reliable tool for monitoring plant health in tomato cultivation.

2.1.3 Random Forests

Random forests are a robust ensemble learning technique that combines the predictions of multiple decision trees to improve classification or regression performance. During the training phase, the algorithm constructs numerous decision trees, each trained on a random subset of the data and features. Random forests are robust against overfitting and perform well even with noisy data. They can handle both continuous and categorical features, which makes them versatile for different plant disease detection tasks. Random forests are also known for their ability to assess feature importance, helping researchers identify which features (such as leaf texture or colour) are most critical for disease classification.

In the context of agricultural technology, plant disease detection plays a crucial role in improving crop yield and ensuring sustainable farming practices. Various machine learning algorithms, including Random Forest (RF), K-Nearest Neighbours (KNN), SVM, and CNN, have been employed for this purpose. The following studies highlight the use of these algorithms in detecting and classifying plant diseases, focusing on Random Forest and its comparison with other machine learning methods.

Govardhan and Veena (2019) [14] proposed a method for diagnosing diseases which mostly affect the leaves of tomato plant. They used Random Forest algorithm for detecting the diseases. By analysing images of tomato leaves, the authors were able to classify different diseases based on extracted features. The Random Forest classifier was found to be effective in handling complex data, with high accuracy in identifying diseases such as bacterial spot and early blight. The study emphasized the robustness of Random Forest in managing large datasets and its ability to handle multiple disease classes.

Hatuwal et al. (2020) [15] compared multiple machine learning algorithms for plant leaf disease recognition, including Random Forest, KNN, SVM, and CNN. The study concluded that while CNNs offered superior performance in terms of accuracy, Random Forest was highly competitive, especially in terms of interpretability and computational efficiency. RF's ability to aggregate results from multiple decision trees made it a strong contender for large-scale disease recognition tasks. The study also highlighted that the choice of algorithm depended on the computational resources available and the complexity of the dataset.

Sujatha et al. (2021) [16] compared DL models (particularly CNNs) with traditional ML algorithms, including Random Forest and SVM, for disease detection. While DL models, such as CNNs, outperformed MLalgorithms in terms of detection accuracy, Random Forest was shown to provide competitive results with significantly lower computational requirements. This paper concluded that for real-time detection with limited resources, Random Forest remains a practical and efficient alternative, particularly in settings where deep learning models are not feasible.

Patra, Chakraborty, and Gupta (2023) [17] examined the use of the Random Forest algorithm for plant disease prediction. This study focused on the application of RF to predict plant diseases in various crops by analysing leaf images. The authors found that RF is quite fruitful in prediction accuracy, with the ability to handle a large variety of disease types across different crops. This study also highlighted the flexibility of RF, which can be adapted to different agricultural systems and disease categories, making it a reliable tool for plant health monitoring.

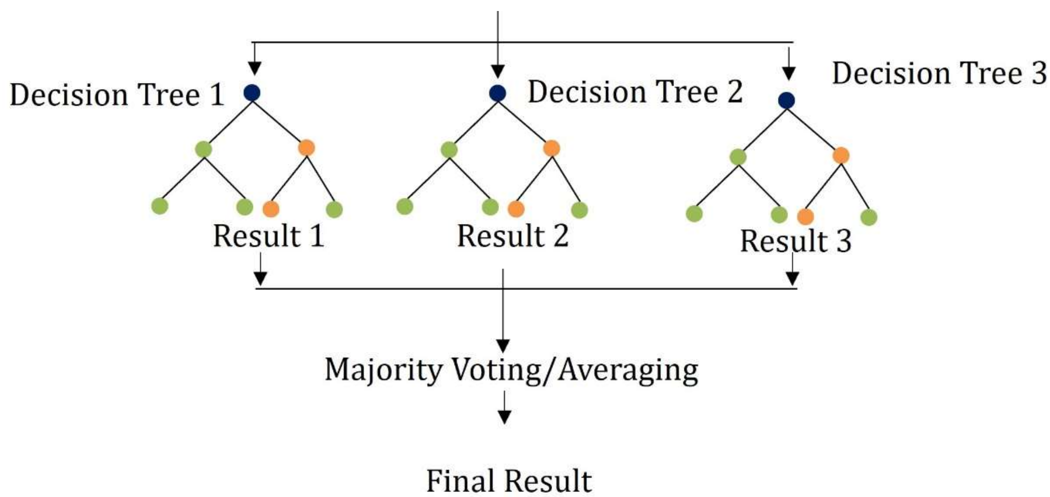

Figure 5: Schematic of Random Forest process

2.2 Unsupervised Learning

A subfield of ML known as "unsupervised learning" uses data without labelled outputs—that is, input data without any predetermined categories or goal values—to train the model. Instead, the model analyzes the data to uncover hidden patterns, structures, or relationships without explicit guidance. One of the basic applications of this technique is in clustering, where data points are grouped based on similarity, and dimensionality reduction, which simplifies data while preserving essential information. Unsupervised learning is particularly useful for exploring datasets, discovering insights, and pre-processing data for other machine learning tasks. This type of learning is useful when labelled datasets are scarce or expensive to obtain.

2.2.1 K-Means Clustering

K-Means clustering is a widely used unsupervised learning algorithm for segmentation tasks. It partitions data points into K clusters based on similarity, where K is a user-defined number (Figure 6). When we discuss about plant disease detection, K-Means can be used to segment regions of an image that correspond to diseased and healthy parts of the plant. This helps in identifying affected areas of the plant, although K-Means can struggle with complex or overlapping clusters in the data.

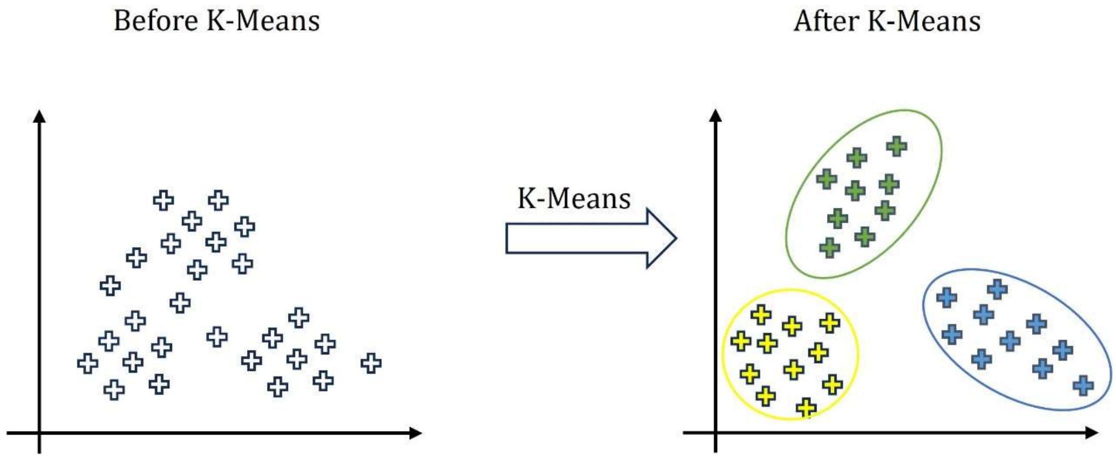

Figure 6: Schematic of K-Means process

One important area of precision agriculture research has been the use of image analysis and ML approaches to identify plant diseases. A combination of K-means clustering for feature extraction and Artificial Neural Networks (ANNs) for classification has been increasingly used for effective leaf disease detection. This literature survey reviews studies that explore the application of K-means clustering and ANN in plant disease detection.

Kumari, Jeevan Prasad, and Mounika (2019) [18] proposed a hybrid model for leaf disease detection that combines K-means clustering for feature extraction and ANN for classification. The study focused on segmenting the plant leaf images using K-means clustering, which helped in isolating the diseased regions. These segmented features were then passed to an ANN model for classification. The hybrid approach is very accurate in detection of various plant diseases. The authors emphasized the robustness of K-means clustering in feature extraction and the effectiveness of ANN in handling the classification task, which can be particularly useful in automated agricultural systems.

Tete and Kamlu (2017) [19] utilized a similar approach by combining thresholding, K-means clustering, and ANN for plant disease detection. The K-means algorithm was first used to segment the plant leaf images, and thresholding techniques were employed to highlight the diseased areas. These features were then fed into an ANN model for classification. The study demonstrated that this approach is effective in detecting diseases with high accuracy, especially when the dataset included variations in leaf shapes and disease types. The combination of thresholding with clustering and ANN allowed for better segmentation of the images, leading to improved classification performance.

Al Bashish, Braik, and Bani-Ahmad (2011) [20] explored the use of K-means clustering for leaf disease detection and classification. In this study, K-means-based segmentation was used to separate the diseased areas of the leaf from the healthy regions. The extracted features were then classified to identify the disease. The authors found that K-means clustering was particularly effective in segmenting leaf images into meaningful regions, making it easier to isolate symptoms of diseases. The approach showed decent accuracy in diseases detection, and the study underscored the importance of effective image segmentation in this process.

2.2.2 Principal Component Analysis (PCA)

PCA is a dimensionality reduction method. It can be used in conjunction with other ML algorithms in order to minimize the complexity of the data without harming vital information. In plant disease detection, PCA is often used for extracting key attributes from images like the shape, texture, and colour of plant leaves. As we decrease the number of features, PCA simplifies the data, helping to improve the results and accuracy of the detection models (Figure 7).

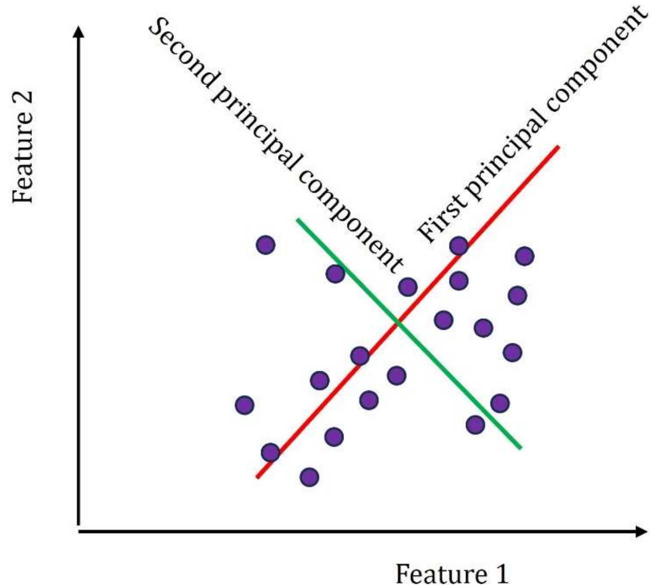

Figure 7: Schematic of PCA process

The detection and classification of plant diseases by using image processing and ML techniques have gained significant attention in precision agriculture. Principal Component Analysis (PCA) is a widely used dimensionality reduction technique, while wavelet transform is utilized for feature extraction. These methods, when combined with ML algorithms like neural networks and SVM, provide effective tools for detecting diseases. This literature survey examines the use of PCA and wavelet-based feature extraction in plant disease recognition.

Wang et al. (2012) [21] explored the use of PCA for dimensionality reduction and neural networks (NNs) for image recognition. The study applied PCA to reduce the complexity of the input data, followed by classification using neural networks. PCA helped in extracting the most relevant features from plant images while maintaining significant information related to disease symptoms. The authors found that combining PCA with neural networks yielded high accuracy in detecting plant diseases, especially for crops with subtle disease symptoms. This method was particularly beneficial in cases where large-scale image datasets were involved, as PCA effectively handled the high-dimensional data, making the neural network's task easier.

Pujari et al. (2013) [22] introduced a process for automatic fungal disease detection. This method used wavelet feature extraction combined with PCA analysis. The wavelet transform was used to extract detailed features of leaf images, such as texture and edges, which are often indicative of fungal infections. These features were then reduced in dimensionality using PCA to remove irrelevant information and improve classification efficiency. The study demonstrated that this hybrid approach led to high accuracy in detecting fungal diseases in commercial crops, showcasing the effectiveness of wavelet and PCA in handling complex disease patterns.

Harini and Bhaskari (2011) [23] utilized wavelet transforms and PCA for identifying leaf diseases in tomato plants. Wavelet transforms helped extract multi-resolution features from the leaf images, which were quite vital in order to identify various diseases, particularly fungal and bacterial infections. PCA was then applied to decrease the feature space and upgrade the classification process. The study showed that this combination improved the accuracy of disease detection in tomato plants and provided a robust approach for handling complex patterns in plant images.

Dhinesh and Jagan (2019) [24] applied PCA for feature extraction and used a linear SVM for the classification of leaf diseases. PCA reduces the dimensionality of the extracted attributes from leaf images, making the classification process more efficient. The reduced features were then passed through a linear SVM for disease classification. The authors found that this combination of PCA and linear SVM achieved high accuracy and efficiency in identifying diseases in plant leaves, especially for distinguishing between different disease types.

2.3 Deep Learning Approaches

DL, particularly CNNs, is quite effective in plant disease detection by enabling automated feature extraction and classification. Unlike traditional machine learning algorithms, DL models can learn hierarchical features directly from raw input data (such as images), and therefore it is very useful in the case of disease detection.

2.3.1 Convolutional Neural Networks (CNNs)

CNNs are designed to automatically learn spatial hierarchies of features from images through convolutional layers. CNNs is broadly used for plant disease detection as it can recognize fine-grained patterns that are characteristic of specific diseases. CNNs have shown excellent performance in tasks like leaf disease classification, where the model can differentiate between different diseases based on visual patterns like lesions, discoloration, and deformation.
The ability of CNNs to learn features automatically has led to breakthroughs in disease detection systems that require minimal manual feature engineering. However, CNNs can be computationally intensive and require large labelled datasets for training. Additionally, ensuring that the model generalizes well to unseen data is a common challenge.

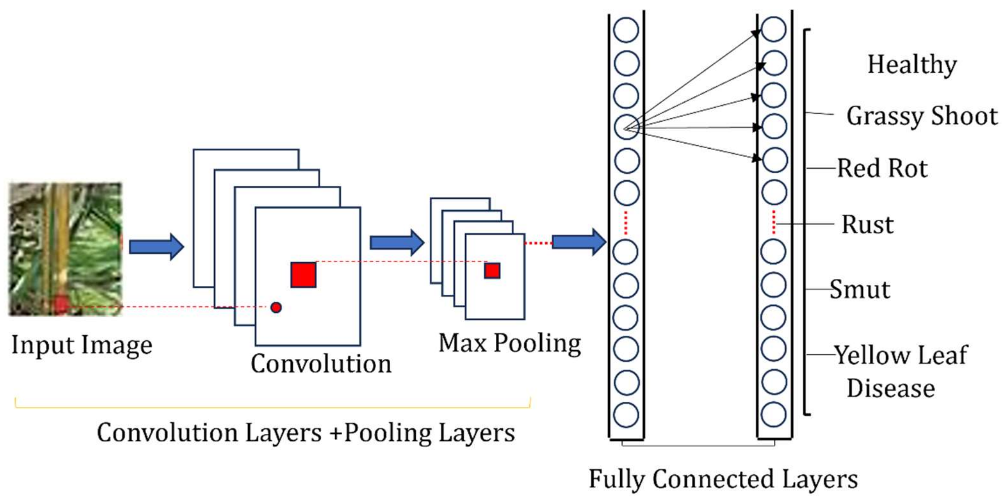

Figure 8: Schematic of CNN process

CNNs is very useful in image recognition and classification tasks that makes it very effecting in detecting plant diseases. These models excel at learning hierarchical patterns in image data, making them highly fruitful for identifying and classifying plant diseases from leaf images (Figure 8). This literature survey explores several studies that leverage CNNs for plant disease detection.

Shrestha, Das, and Dey (2020) [25] investigated the application of CNNs for plant disease detection in their study presented at the IEEE Applied Signal Processing Conference (ASPCON). The authors employed CNNs to detect various plant diseases based on leaf images. As per the study, it can be concluded that CNN is able to accurately identify diseases, offering significant improvements in precision compared to traditional machine learning methods. The research also highlighted the potential for CNNs to handle large and diverse datasets, making them highly scalable for real-world applications in agriculture.

Islam (2020) [26] explored the use of a CNN model integrated with image processing techniques. The model was trained on images of plant leaves, where preprocessing steps like image resizing and normalization were used to enhance the input data. The study found that CNNs were highly effective in distinguishing between healthy and diseased plant leaves, achieving high classification accuracy. The study underscored the importance of combining CNNs with advanced image processing to improve detection performance, especially in the presence of noise or variable lighting conditions.

Shelar et al. (2022) [27] examined the use of CNNs for plant disease detection in their work presented at ITM Web of Conferences. The study focused on the automatic classification of plant diseases using CNNs, demonstrating that the model could effectively classify plant diseases. The authors reported high accuracy in their experiments, highlighting CNNs' ability to learn complex patterns and detect subtle disease symptoms, making them suitable for large-scale agricultural applications where manual inspection is not feasible.

Deepalakshmi, Lavanya, and Srinivasu (2021) [28] used CNN algorithms for detecting diseases in plant leaves. The study employed a custom CNN architecture designed to process leaf images and classify them based on disease symptoms. The results showed that CNNs could provide reliable and accurate disease detection, even in challenging conditions where the plant images varied in terms of quality and lighting. This work reinforced the power of CNNs in handling real-world agricultural problems where data variability is common.

Sharma, Berwal, and Ghai (2020) [29] focused on evaluating various deep learning CNN models for plant disease detection, incorporating image segmentation techniques. Image segmentation was used to isolate the diseased areas of the plant leaves, and CNN models were applied to classify the segmented regions. The study compared the performance of different CNN architectures, highlighting their strengths and weaknesses in disease detection. The analysis indicated that CNNs, when combined with segmentation, significantly enhanced detection accuracy and speed, proving to be highly effective for automated plant disease diagnosis in agricultural fields.

2.3.2 Transfer Learning

This learning involves leveraging pre-trained deep learning models (such as VGG16, ResNet, or Inception) that have been trained on large, general-purpose datasets (e.g., ImageNet). These models can be fine-tuned on smaller plant disease datasets, significantly reducing the need for large annotated data and computational resources. Transfer learning has been particularly successful in plant disease detection, where labelled data is often scarce (Figure 9). By utilizing pre-trained models, researchers can achieve high accuracy in classifying diseases with fewer data and faster training times.
It is a deep learning method that uses a pre-trained model on a new but similar task. This approach is particularly beneficial for plant disease detection with small labelled datasets. By leveraging models trained on massive datasets, transfer learning allows for better generalization and faster convergence with less labelled data. This literature survey explores studies that have applied transfer learning for plant disease detection.

Chen et al. (2020 [30]) applied deep transfer learning to image-based plant disease identification. The authors used pre-trained CNNs to detect diseases in plants, taking advantage of the knowledge gained from large image datasets. Transfer learning allowed for better performance in identifying diseases even with limited labelled plant disease images. The results indicated that deep transfer learning could significantly improve the accuracy of disease identification and reduce the time and resources needed to train a model from scratch, making it a powerful tool in plant disease diagnostics.

Abbas et al. (2021) [31] introduced a novel approach that combines transfer learning with conditional Generative Adversarial Networks (C-GANs) to generate synthetic images for tomato plant disease detection. The C-GAN generated synthetic disease images, augmenting the training data for the transfer learning model. By using this combination, the study was able to overcome the challenge of limited labelled data and achieved high accuracy in detecting diseases in tomato plants. This approach showed how synthetic data generation through C-GANs could complement transfer learning in handling real-world agricultural problems, where datasets are often small and imbalanced.

Vallabhajosyula et al. (2022) [32] proposed an ensemble deep learning model based on transfer learning for plant leaf disease detection. By combining multiple pre-trained CNN models through an ensemble approach, the study aimed to enhance the detection performance for a variety of plant diseases. Transfer learning allowed the models to leverage pre-trained features from large image datasets, and the ensemble approach combined the strengths of different models to improve accuracy. The results showed that this ensemble approach could effectively identify plant diseases, achieving high classification accuracy and robustness across various types of plant diseases.

Mukti and Biswas (2019) [33] used the ResNet50 architecture for transfer learning in plant disease detection. ResNet50, a deep CNN model known for its residual connections, was fine-tuned on a smaller dataset of plant disease images. The authors demonstrated that using a pre-trained model like ResNet50 allowed for effective disease detection in plants, even with a limited dataset. The study highlighted that the use of deep transfer learning with a powerful pre-trained model like ResNet50 could lead to accurate disease classification while reducing the computational resources needed for training.

Hassan et al. (2021) [34] explored the use of CNNs in combination with transfer learning for the identification of plant leaf diseases. The study used a pre-trained CNN model and fine-tuned it with a small dataset of plant leaf images. The authors found that the transfer learning approach improved classification performance and generalization when compared to training a CNN from scratch. This approach also enabled faster convergence, which is critical for real-time disease identification in agricultural practices.

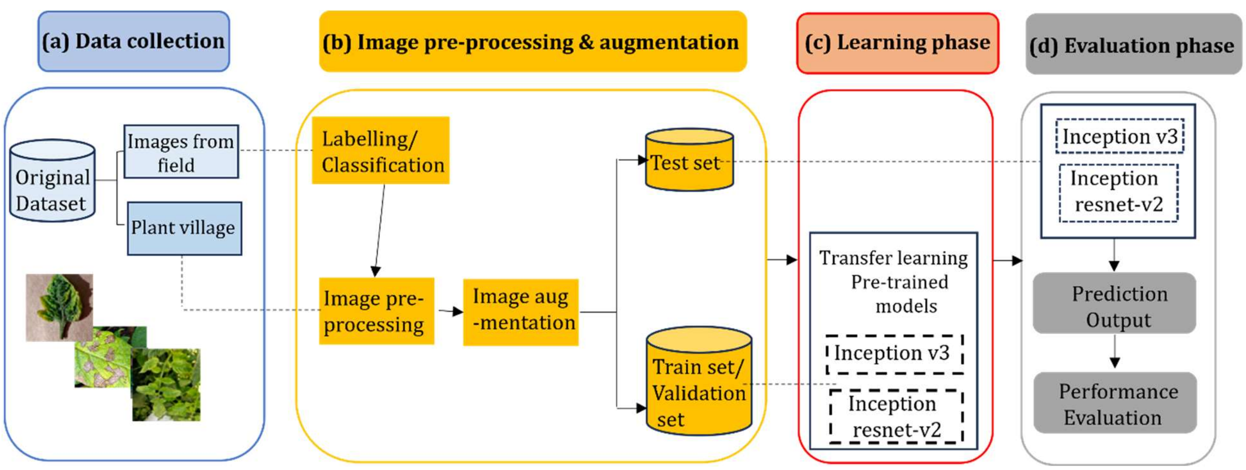

Figure 9: Schematic of Transfer learning process

2.3.3 Generative Adversarial Networks (GANs)

GANs have been explored as a means to generate synthetic plant disease images. GANs consist of two networks: a generator, which creates fake images, and a discriminator, which tries to distinguish between real and fake images. By training both networks in opposition, GANs can generate realistic images of diseased plants, augmenting small datasets and improving the performance of disease detection models (Figure 10). GANs are particularly useful when labelled data is limited or imbalanced.

GANs have gained considerable attention in various fields, including plant disease detection. GANs are particularly useful for data augmentation by generating synthetic images to address the challenges of limited labelled data, which is a common issue in agricultural datasets. By combining GANs with other machine learning techniques, such as CNNs, significant improvements in disease identification and classification accuracy have been achieved. Below is a review of key studies in this area.

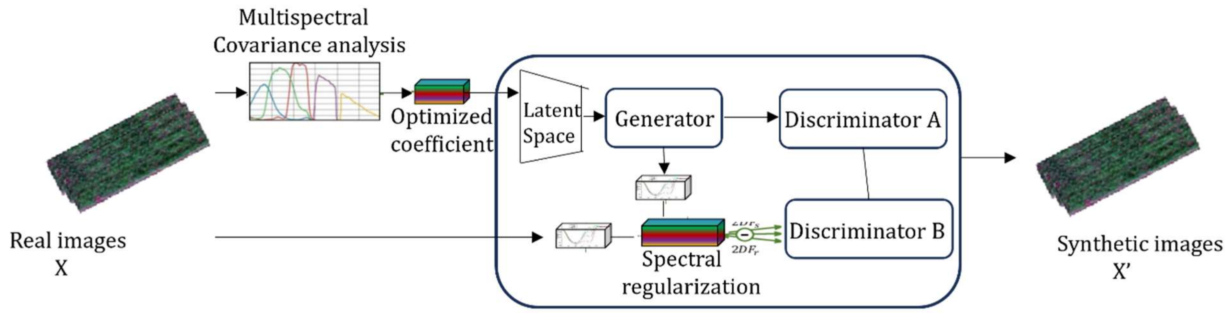

Figure 10: Schematic of GAN process

Liu et al. (2020) [35] proposed a method that uses GANs for data augmentation to improve grape leaf disease identification. The study demonstrated that GAN-generated synthetic images could expand the training dataset, providing more varied examples of grape leaf diseases. This helped to address the issue of limited data and improved the performance of CNN-based classifiers. The results showed that GAN-based augmentation enhanced the model's ability, even with a small number of labelled real-world images.

Chen and Wu (2023) [36] tackled the problem of sparse data in grape leaf disease identification by combining GANs and CNNs. The authors used GANs to generate synthetic images, increasing the diversity of the training set. These synthetic images were then used to train a CNN, which enabled the model to achieve high accuracy in detecting grape leaf diseases. The study highlighted the power of combining GANs and CNNs, particularly when dealing with small datasets, to improve model generalization and performance.

Deshpande and Patidar (2022) [37] used a combination of GANs and Deep CNN (DCNNs) for tomato disease detection. GANs were employed to generate synthetic images of tomato leaves with diseases, which were then used to train the DCNN. The study found that the generated images helped the model learn to recognize tomato leaf diseases with improved accuracy. By augmenting the dataset with GAN-generated images, the model was able to achieve better results compared to using real-world images alone.

Nerkar and Talbar (2021) [38] explored cross-dataset learning with reinforced GANs to upgrade the performance of leaf disease detection. The authors trained GANs on multiple datasets to generate synthetic images that could be applied across different plant species, improving model robustness. This approach enabled the transfer of knowledge from one dataset to another, making the model more adaptable to various plant diseases. The results showed that using reinforced GANs for cross-dataset learning led to significant performance improvements, especially in the detection of leaf diseases in scenarios where datasets were limited or imbalanced.

Ramadan et al. (2024) [39] combined CNNs with GANs for image-based rice leaf disease detection. The authors used GANs to generate synthetic rice leaf images, which were then used to augment the training data for the CNN. The study demonstrated that GANs helped improve the CNN's accuracy in identifying rice leaf diseases, even with a relatively small number of real images. By using synthetic images, the model learned to recognize rice leaf diseases more effectively, showcasing the potential of GAN-based data augmentation in agricultural applications.

3. Datasets for Plant Disease Detection

A number of publicly available datasets have facilitated the progress of machine learning in plant disease detection. These datasets typically contain images of plant leaves and other plant parts, annotated with disease labels. Popular datasets include:

1. **PlantVillage Dataset [40]**: This is one of the most well-known datasets, containing images of 38 plant species and 14 crop diseases. It has over 50,000 images, making it one of the largest publicly available datasets for plant disease classification.

2. **Kaggle Plant Disease Dataset [41]**: This dataset includes images of various crops like apple, tomato, and grape, annotated with different disease types. It is widely used for training and benchmarking disease detection models.
3. **DeepPlant Dataset [42]**: This dataset is designed for training deep learning models and contains plant images affected by several diseases. The dataset is focused on providing a rich source of labeled images for plant disease detection tasks.

These datasets provide the necessary resources to train and evaluate machine learning models, driving advancements in the field.

4. Challenges and Limitations

While ML has proven effective for plant disease detection, several challenges remain:

1. **Data Quality and Quantity:** High-quality, labelled datasets are the foundation of building effective machine learning models, especially for complex tasks like plant disease detection. To train a model capable of accurately identifying diseases, it is crucial to have a large, diverse, and representative dataset that includes a variety of plant species, disease types, and environmental conditions. However, creating such datasets is often both time-consuming and costly. It requires not only extensive data collection but also manual annotation of images, which is labour-intensive and prone to human error. Furthermore, even with a large dataset, issues like noisy data—where images may be blurry, poorly lit, or contain irrelevant background information—can compromise the quality of training. Additionally, many datasets are imbalanced, where certain plant diseases may be underrepresented, further complicating the task of training an accurate model. This imbalance can cause the model to perform poorly on the underrepresented classes, making it less robust in real-world applications where such diseases are less common but still important to detect.

2. **Class Imbalance:** One of the most common issues in agricultural disease detection datasets is class imbalance, where there is a disproportionate representation of healthy plant images compared to diseased images. This imbalance arises because healthy plants are far more common in agricultural fields, whereas diseases are often rare or localized. Consequently, machine learning models trained on such imbalanced data tend to develop a bias towards the majority class (healthy plants), making them less sensitive to the minority class (diseased plants). This bias can lead to a significant reduction in model performance, particularly in detecting rare or emerging plant diseases. In practical terms, this means that the system might miss subtle symptoms of a disease or fail to identify new pathogens that are not well-represented in the training data. Addressing class imbalance is critical, and strategies such as data augmentation, oversampling the minority class, or using loss functions that penalize misclassifications of the minority class can help mitigate this issue.

3. **Environmental Variability:** Agricultural settings are dynamic, and the variability in environmental conditions can significantly impact the quality and consistency of the data used for disease detection. Factors such as lighting conditions, camera quality, weather, and the background in images can introduce substantial variability. For instance, poor lighting can cause shadows, glares, or unclear images, making it difficult for the model to correctly identify plant features. Similarly, variations in camera quality, whether due to resolution, lens distortion, or other technical limitations, can degrade the visual data, resulting in inaccurate or inconsistent information. Furthermore, plants grow in stages, and these stages are often marked by different visual features that may or may not be indicative of disease. Models trained on a narrow range of growth stages may struggle to generalize when faced with plants at different stages. These factors can make it challenging for a model to perform consistently across a wide range of environmental conditions, which is essential for real-world deployment in diverse agricultural settings.

4. **Real-World Deployment**: Deploying machine learning-based plant disease detection systems in real-world environments is inherently more difficult than testing them in controlled, ideal conditions. While models may perform well in a lab or simulation where factors such as lighting, background, and camera settings can be controlled, the unpredictability of outdoor conditions introduces significant challenges.

In outdoor fields, image quality can vary greatly due to inconsistent lighting, camera angles, or the presence of dirt and water droplets on the camera lens. Additionally, environmental factors such as wind, rain, or changes in the plant's exposure to sunlight can affect the appearance of the plants and the clarity of disease symptoms. These environmental challenges can cause a model that was trained in controlled conditions to underperform in the field. Moreover, the system must be able to handle not just variations in image quality but also the diversity of plant species, soil types, and cultivation practices, all of which can impact the disease detection process. For a machine learning model to be effective in real-world agricultural settings, it must be robust enough to generalize across these diverse and changing conditions, which requires continuous refinement and adaptation of the model after deployment.

5. Emerging Trends and Future Directions

The field of plant disease detection is evolving rapidly, and several emerging trends are likely to shape its future:

1. **Edge Computing:** Edge computing is increasingly becoming a critical component in modern agricultural technologies, particularly in plant disease detection. By using edge devices such as drones, robots, and smartphones, data can be collected and analysed in real-time at the source, which significantly reduces the need for centralized cloud-based processing. This not only speeds up the decision-making process but also minimizes latency, enabling more timely and actionable insights. For instance, drones equipped with cameras and sensors can capture images of crops and immediately analyse them for signs of disease, providing instant feedback to farmers. The use of edge computing is particularly valuable in remote agricultural areas where internet connectivity may be limited or unreliable. By processing data locally on the edge device, farmers can receive immediate alerts about potential disease outbreaks and so they can take appropriate actions.

2. **Multimodal Data Fusion:** Multimodal data fusion involves combining information from various sources, such as visual images, hyperspectral images, and environmental sensors, to create a more comprehensive and accurate understanding of plant health. Each data source provides unique insights: for example, visual images capture visible symptoms of disease, while hyperspectral images can detect subtle changes in plant physiology that are invisible to the human eye. Environmental sensors, such as temperature and humidity sensors, can offer additional context, highlighting environmental factors that may influence disease development. By integrating these diverse data types, multimodal data fusion can improve the accuracy, robustness, and reliability of disease detection systems. This holistic approach helps create a more nuanced understanding of plant health, making it possible to detect diseases earlier and with greater precision, which can significantly enhance crop management and yield prediction.

3. **Explainable AI (XAI):** As machine learning models become more sophisticated and are applied in critical areas like agriculture, the need for transparency in decision-making grows. Explainable AI (XAI) techniques aim to make ML models more interpretable, offering insight into how predictions are made. This is particularly important in plant disease detection systems, where farmers and agricultural practitioners must trust the model's decisions to take appropriate actions. For example, an explainable AI system might provide not just a disease diagnosis, but also an explanation of which features in an image led to that diagnosis (such as specific leaf spots or color changes). This transparency fosters trust in the system, allowing users to understand the rationale behind the model's predictions. Furthermore, explainability can assist in identifying potential model biases, ensuring that the system is not only accurate but also fair and reliable across diverse plant species, growth stages, and environmental conditions.

4. **Self-Supervised Learning:** Self-supervised learning is an innovative approach that has the potential to overcome one of the most significant challenges in plant disease detection— the scarcity of labelled data. In traditional machine learning approaches, models require vast amounts of labelled data to learn

from, which is time-consuming and expensive to produce. Self-supervised learning, however, allows models to learn useful representations from unlabelled data by creating tasks that help the model understand the structure of the data itself. For example, a model might predict missing parts of an image or learn to generate future frames of plant growth based on past observations. By leveraging large volumes of unlabelled data, self-supervised learning can reduce the dependency on manually annotated datasets, making it easier to scale plant disease detection systems. This approach can help models improve over time, as they continue to learn from the constantly growing body of data collected in the field, thereby enhancing their accuracy and generalization ability without requiring continuous human intervention for labelling.

6. Conclusion

Machine learning techniques, especially deep learning, have made significant strides in revolutionizing plant disease detection by providing more accurate, efficient, and scalable solutions. These methods can analyse large amounts of data, including images, environmental factors, and sensor data, to detect diseases early in the plant life cycle, which is crucial for effective disease management. Supervised learning, where models are trained on labelled datasets, has been extensively used to detect specific disease symptoms, while deep learning, particularly CNNs, has gained prominence due to its ability to automatically learn hierarchical features from raw image data, improving detection accuracy even in complex or subtle cases. Transfer learning, a technique where a model trained on one dataset is adapted to work on another, is particularly useful in plant disease detection when labelled data is limited, allowing models to generalize across different crops or regions.

However, despite the progress in using machine learning for plant disease detection, several challenges remain. One issue is data scarcity, as high-quality labelled datasets for plant diseases are often difficult to acquire, especially for rare or newly emerging diseases. Furthermore, environmental variability, such as differences in lighting, weather conditions, and plant species, can affect the performance of machine learning models, making it harder to create universally applicable solutions. Real-world deployment of these systems also presents challenges, particularly in terms of computational resources, infrastructure, and model generalization to different agricultural settings.

In order to overcome with these issues and upgrade the scalability, accuracy, and real-world applicability of machine learning-based systems, future advancements in plant disease detection are likely to integrate edge computing, multimodal data, explainable AI, and self-supervised learning techniques. Edge computing will enable the deployment of ML models on local devices, such as drones or sensors, reducing the need for constant data transfer to central servers and allowing for real-time processing in the field. Multimodal data integration, which combines data taken from various domains from various sources like images, environmental sensors, will provide a more comprehensive view of plant health, improving diagnostic accuracy. Explainable AI (XAI) will help increase the trust and interpretability of AI models by offering transparent and understandable reasoning behind disease predictions, which is crucial for farmers and agricultural experts to act confidently on recommendations. Finally, self-supervised learning techniques, which enable models to learn from unlabelled data, hold great promise in addressing data scarcity issues and expanding the capacity of machine learning systems to handle new and unseen diseases without relying heavily on labelled training datasets.

By addressing these challenges, the future of plant disease detection will see even more powerful and adaptable systems capable of supporting farmers in making timely, informed decisions, ultimately leading to better crop management, reduced pesticide use, and more sustainable agricultural practices.

References

1. Fang, Yi, and Ramaraja P. Ramasamy. "Current and prospective methods for plant disease detection." *Biosensors* 5, no. 3 (2015): 537-561.
2. Khirade, Sachin D., and Amit B. Patil. "Plant disease detection using image processing." In *2015 International conference on computing communication control and automation*, pp. 768-771. IEEE, 2015.
3. Ramesh, Shima, Ramachandra Hebbar, M. Niveditha, R. Pooja, N. Shashank, and P. V. Vinod. "Plant disease detection using machine learning." In *2018 International conference on design innovations for 3Cs compute communicate control (ICDI3C)*, pp. 41-45. IEEE, 2018.

4. Rajesh, B., M. Vishnu Sai Vardhan, and L. Sujihelen. "Leaf disease detection and classification by decision tree." In *2020 4th international conference on trends in electronics and informatics (ICOEI)(48184)*, pp. 705-708. IEEE, 2020.
5. Chopda, Jayraj, Hiral Raveshiya, Sagar Nakum, and Vivek Nakrani. "Cotton crop disease detection using decision tree classifier." In *2018 International Conference on Smart City and Emerging Technology (ICSCET)*, pp. 1-5. IEEE, 2018.
6. Ramesh, Shima, Ramachandra Hebbar, M. Niveditha, R. Pooja, N. Shashank, and P. V. Vinod. "Plant disease detection using machine learning." In *2018 International conference on design innovations for 3Cs compute communicate control (ICDI3C)*, pp. 41-45. IEEE, 2018.
7. Mengistu, Abrham Debasu, Seffi Gebeyehu Mengistu, and Dagnachew Melesew. "An automatic coffee plant diseases identification using hybrid approaches of image processing and decision tree." *Indones. J. Electr. Eng. Comput. Sci.* 9, no. 3 (2018): 806-811.
8. Ahmed, Kawcher, Tasmia Rahman Shahidi, Syed Md Irfanul Alam, and Sifat Momen. "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.
9. Das, Debasish, Mahinderpal Singh, Sarthak Swaroop Mohanty, and S. Chakravarty. "Leaf disease detection using support vector machine." In *2020 international conference on communication and signal processing (ICCSP)*, pp. 1036-1040. IEEE, 2020.
10. Shruthi, U., V. Nagaveni, and B. K. Raghavendra. "A review on machine learning classification techniques for plant disease detection." In *2019 5th International conference on advanced computing & communication systems (ICACCS)*, pp. 281-284. IEEE, 2019.
11. Sahu, Santosh Kumar, and Manish Pandey. "An optimal hybrid multiclass SVM for plant leaf disease detection using spatial Fuzzy C-Means model." *Expert systems with applications* 214 (2023): 118989.
12. Thaiyalnayaki, K., and Christeena Joseph. "Classification of plant disease using SVM and deep learning." *Materials Today: Proceedings* 47 (2021): 468-470.
13. Mokhtar, Usama, Nashwa El Bendary, Aboul Ella Hassenian, Eid Emary, Mahmoud A. Mahmoud, Hesham Hefny, and Mohamed F. Tolba. "SVM-based detection of tomato leaves diseases." In *Intelligent Systems' 2014: Proceedings of the 7th IEEE International Conference Intelligent Systems IS'2014, September 24-26, 2014, Warsaw, Poland, Volume 2: Tools, Architectures, Systems, Applications*, pp. 641-652. Springer International Publishing, 2015.
14. Govardhan, Meghana, and M. B. Veena. "Diagnosis of tomato plant diseases using random forest." In *2019 Global Conference for Advancement in Technology (GCAT)*, pp. 1-5. IEEE, 2019.
15. Hatuwal, Bijaya Kumar, Aman Shakya, and Basanta Joshi. "Plant Leaf Disease Recognition Using Random Forest, KNN, SVM and CNN." *Polibits* 62 (2020): 13-19.
16. Sujatha, Radhakrishnan, Jyotir Moy Chatterjee, N. Z. Jhanjhi, and Sarfraz Nawaz Brohi. "Performance of deep learning vs machine learning in plant leaf disease detection." *Microprocessors and Microsystems* 80 (2021): 103615.
17. Patra, Joyjit, Monalisa Chakraborty, and Subir Gupta. "Random Forest Algorithm for Plant Disease Prediction." In *AI to Improve e-Governance and Eminence of Life: Kalyanathon 2020*, pp. 89-101. Singapore: Springer Nature Singapore, 2023.
18. Kumari, Ch Usha, S. Jeevan Prasad, and G. Mounika. "Leaf disease detection: feature extraction with K-means clustering and classification with ANN." In *2019 3rd international conference on computing methodologies and communication (ICCMC)*, pp. 1095-1098. IEEE, 2019.
19. Tete, Trimi Neha, and Sushma Kamlu. "Detection of plant disease using threshold, k-mean cluster and ann algorithm." In *2017 2nd International Conference for Convergence in Technology (I2CT)*, pp. 523-526. IEEE, 2017.
20. Al Bashish, Dheeb, Malik Braik, and Sulieman Bani-Ahmad. "Detection and classification of leaf diseases using K-means-based segmentation and." *Information technology journal* 10, no. 2 (2011): 267-275.
21. Wang, Haiguang, Guanlin Li, Zhanhong Ma, and Xiaolong Li. "Image recognition of plant diseases based on principal component analysis and neural networks." In *2012 8th International Conference on Natural Computation*, pp. 246-251. IEEE, 2012.
22. Pujari, Jagadeesh D., Rajesh Yakkundimath, and Abdulmunaf Syedhusain Byadgi. "Automatic fungal disease detection based on wavelet feature extraction and PCA analysis in commercial crops." *International Journal of Image, Graphics and Signal Processing* 6, no. 1 (2013): 24-31.

23. Harini, D. N. D., and D. Lalitha Bhaskari. "Identification of leaf diseases in tomato plant based on wavelets and PCA." In *World Congress on Information and Communication Technologies*, pp. 978-1. 2011.
24. Dhinesh, E., and A. Jagan. "Detection of leaf disease using principal component analysis and linear support vector machine." In *2019 11th International Conference on Advanced Computing (ICoAC)*, pp. 350-355. IEEE, 2019.
25. Shrestha, Garima, Majolica Das, and Naiwrita Dey. "Plant disease detection using CNN." In *2020 IEEE applied signal processing conference (ASPCON)*, pp. 109-113. IEEE, 2020.
26. Islam, Md Tariqul. "Plant disease detection using CNN model and image processing." *Int. J. Eng. Res. Technol* 9, no. 10 (2020): 291-297.
27. Shelar, Nishant, Suraj Shinde, Shubham Sawant, Shreyash Dhumal, and Kausar Fakir. "Plant disease detection using CNN." In *ITM Web of Conferences*, vol. 44, p. 03049. EDP Sciences, 2022.
28. Deepalakshmi, P., K. Lavanya, and Parvathaneni Naga Srinivasu. "Plant leaf disease detection using CNN algorithm." *International Journal of Information System Modeling and Design (IJISMD)* 12, no. 1 (2021): 1-21.
29. Sharma, Parul, Yash Paul Singh Berwal, and Wiqas Ghai. "Performance analysis of deep learning CNN models for disease detection in plants using image segmentation." *Information Processing in Agriculture* 7, no. 4 (2020): 566-574.
30. Chen, Junde, Jinxiu Chen, Defu Zhang, Yuandong Sun, and Yaser Ahangari Nanehkaran. "Using deep transfer learning for image-based plant disease identification." *Computers and Electronics in Agriculture* 173 (2020): 105393.
31. Abbas, Amreen, Sweta Jain, Mahesh Gour, and Swetha Vankudothu. "Tomato plant disease detection using transfer learning with C-GAN synthetic images." *Computers and Electronics in Agriculture* 187 (2021): 106279.
32. Vallabhajosyula, Sasikala, Venkatramaphanikumar Sistla, and Venkata Krishna Kishore Kolli. "Transfer learning-based deep ensemble neural network for plant leaf disease detection." *Journal of Plant Diseases and Protection* 129, no. 3 (2022): 545-558.
33. Mukti, Ishrat Zahan, and Dipayan Biswas. "Transfer learning based plant diseases detection using ResNet50." In *2019 4th International conference on electrical information and communication technology (EICT)*, pp. 1-6. IEEE, 2019.
34. Hassan, Sk Mahmudul, Arnab Kumar Maji, Michał Jasiński, Zbigniew Leonowicz, and Elżbieta Jasińska. "Identification of plant-leaf diseases using CNN and transfer-learning approach." *Electronics* 10, no. 12 (2021): 1388.
35. Liu, Bin, Cheng Tan, Shuqin Li, Jinrong He, and Hongyan Wang. "A data augmentation method based on generative adversarial networks for grape leaf disease identification." *IEEE Access* 8 (2020): 102188-102198.
36. Chen, Yiping, and Qiufeng Wu. "Grape leaf disease identification with sparse data via generative adversarial networks and convolutional neural networks." *Precision Agriculture* 24, no. 1 (2023): 235-253.
37. Deshpande, Rashmi, and Hemant Patidar. "Tomato plant leaf disease detection using generative adversarial network and deep convolutional neural network." *The Imaging Science Journal* 70, no. 1 (2022): 1-9.
38. Nerkar, Bhavana, and Sanjay Talbar. "Cross-dataset learning for performance improvement of leaf disease detection using reinforced generative adversarial networks." *International Journal of Information Technology* 13, no. 6 (2021): 2305-2312.
39. Ramadan, Syed Taha Yeasin, Md Shafiqul Islam, Tanjim Sakib, Nusrat Sharmin, Md Mokhlesur Rahman, and Md Mahbubur Rahman. "Image-based rice leaf disease detection using CNN and generative adversarial network." *Neural Computing and Applications* (2024): 1-18.
40. Noyan, Mehmet Alican. "Uncovering bias in the PlantVillage dataset." *arXiv preprint arXiv:2206.04374* (2022).
41. Thapa, Ranjita, Kai Zhang, Noah Snavely, Serge Belongie, and Awais Khan. "The Plant Pathology Challenge 2020 data set to classify foliar disease of apples." *Applications in plant sciences* 8, no. 9 (2020): e11390.
42. Lee, Sue Han, Chee Seng Chan, Paul Wilkin, and Paolo Remagnino. "Deep-plant: Plant identification with convolutional neural networks." In *2015 IEEE international conference on image processing (ICIP)*, pp. 452-456. IEEE, 2015.