International Journal of Engineering Applied Science and Management
ISSN (Online): 2582-6948
Vol. 4 Issue 10, October 2023

Machine Learning and Deep Learning Techniques for the Classification and Detection of Plant Diseases: A Survey

Poorvi Vishwakarma[1], Chetan Agrawal[2], Pramila Lovanshi[3]
CSE Department, Radharaman Institute of Technology and Science, Bhopal, India1,2,3
poorvativish@gmail.com[1], chetan.agrawal12@gmail.com[2], rits.pramila@gmail.com[3]

Abstract: *Today agriculture is very important because of the country's growing population and rising food needs. Therefore, it is necessary to increase crop productivity. One of the major factors contributing to reduced crop yields is the presence of bacterial, fungal, and viral diseases. Plant disease detection techniques can be used to handle and prevent this. Since machine learning primarily uses information from the data itself and provides excellent methods for detecting plant diseases, it will be used in the process of identifying diseases in plants. Machine learning methods can be used to identify diseases because they primarily focus on data superiority outcomes for a certain goal. This method uses deep learning and machine learning based on artificial intelligence (AI) to provide a thorough review of all the methods used in plant disease identification. In the realm of computer vision, deep learning has also grown significantly in importance as a means of providing improved performance results for plant disease detection. The machine learning and computer vision fields have achieved considerable success with the application of deep learning developments across various sectors. In order to demonstrate the superiority of the deep learning model over the machine learning model, a comparative analysis of the performance and application of machine and deep learning approaches is made in a number of research publications. The deep learning technology can be used to identify leaf diseases from collected photos, thereby preventing significant crop losses.*

Keywords: *Plant disease prediction, Crop productivity, Visual symptoms, precision agriculture, deep learning, machine learning.*

1. Introduction

Any nation's economic development depends significantly on agriculture. Meeting the population's current food needs has become a difficult problem because of the growing population, frequent changes in weather, and limited resources. On top of the aforementioned challenges, crop diseases have been increasing in severity and scale. Crop diseases cause production losses that can be mitigated with continuous monitoring. Researchers from the Food and Agriculture Organization of the United Nations predicted that plant diseases alone cost the global economy about US$220 billion annually [1]. Developing new methods to detect diseases on plants or leaves at an early stage can significantly increase yield potential.

Precision agriculture is a rapidly developing field aimed at addressing current concerns about agricultural sustainability. Machine learning (ML) is the cutting edge technology underpinning precision agriculture, allowing the machine to learn without having to be programmed directly, and in conjunction with Internet of Things (IoT) enabled farm equipment, is the future of agriculture. There are several studies that have employed or proposed ML methods to detect or classify plant diseases. The majorities of these works take as input a plant/leaf image and detect whether or not there is a disease. These works treat the problem as a classification one, either binary classification (healthy or diseased plant/leaf) or multi-class classification (where various diseases are targeted). Classical ML methods, like Random Forest (RF) and Deep Learning (DL) ones have been utilized for this purpose. On the other hand, there are fewer works that aim to detect both the type of the disease and the diseased regions.

In this paper, we aim to review ML and DL methods that have been applied to either classify or detect plant diseases. The issue of the proficient plant diseases protection is closely linked to viable change in climate and agriculture [2]. Studies show that climate change may vary pathogenic stages and rates; host resistance may also be altered, leading to physiological variations in host-pathogen co-operations [3]. The actuality that now-a-days, diseases more freely transferred around the globe than ever before complicates the situation. New diseases may occur where they have not been identified previously and, inherently, where local expertise to combat them is not available [4] (see Table 1).

The unreliable use of pesticides may cause long-term pathogens to develop resistance and seriously decrease the ability to combat it. One of the pillars of precision farming [5] is the prompt and exact interpretation of diseases in plants. It is crucial that the financial and other re- sources are not unnecessarily wasted and thus that the production is healthier by addressing the problem of developing long lasting pathogenic resistance and alleviating the adverse effects of climate change.

Table 1: A review on plant disease segmentation techniques

Plant disease Segmentation techniques	Merits	Demerits
Threshold Method [9]	1. Any prior knowledge of the picture is not needed. 2. Fast, basic, and low-cost computationally. 3. It's easy to use and acceptable for real-life scenarios.	1. The resulting image cannot guarantee that the segmented regions are contiguous since spatial information can be overlooked. 2. The choice of a threshold is important. 3. Extremely sensitive to noise.
Clustering method [10]	1. It is easy to obtain homogeneous regions. 2. Faster in terms of computation. 3. The smaller the value of K, the better K-means operates.	1. Worst-case scenario conduct is bad. 2. It necessitates clusters of similar size.
Edge detection method [11]	1. It works well with pictures that have a higher contrast among regions.	1. Worked badly for a picture with a lot of edges. 2. It's difficult to find the right object edge.
Regional method [12]	1. It helps you to choose between interactive and automated image segmentation techniques. 2. The movement from the inner point to the outer region establishes more distinct entity boundaries. 3. In comparison to other approaches, it provides more reliable performance.	1. More computation time and memory was required, and the process was sequential. 2. User seed selection that is noisy results in faulty segmentation. 3. Splitting segments appear square due to the region's splitting scheme

Adequate and timely identification of diseases, which includes early impediment, has never been more significant in this changing environment. Plant pathologies can be detected in a number of ways. Some diseases have no apparent syndromes or the response is too late to act and a refined examination is required in those situations. However, most illnesses create a manifestation of some sort in the spectrum visible, so that a trained professional examination is the primary technique for the detection of plants. A plant pathologist should have better observational proficiency to identify characteristic symptoms in order to obtain exact plant disease diagnostics [6]. Alterations in symptoms determined by sick plants may result in inaccurate diagnosis because it could be more difficult for amateur and hobbyists to determine than for a professional pathologist. The beginners in gardening and the experienced specialists can greatly benefit from an automated system devised to detect plant conditions with the appearance and visual symptoms of the plant as a verifier of disease diagnoses.

Improvements in computer vision offer opportunities to increase and strengthen the practice of accurate plant protection and to broaden the market for precise agriculture computer vision applications. In order to detect and classify plant diseases, the utilization of common technology for digital image processing like color detection and threshold [7] was employed.

Different approaches to deep learning are recently being used for plant diseases detection and the most popular of these are CNN. Deep learning is a new trend in machine learning, with state-of-art results in many areas of research, including computer vision, pharmacy and bio- informatics. Deep learning benefits from the capacity to use raw data directly without the use of handcrafts32]. The use of deep learning, for two main reasons, has recently produced good results both academically and industrially [8]. First, every day is generated large amounts of data. These data may therefore be used to develop a profound model. Second, the computational power of the Graphics Processing Unit allows deep models to be trained and leveraged in computing parallelism.

Rest of the paper is structured as: in section II various actors of plant disease are discussed, in section III reviews research works that proposed/utilized ML and DL algorithms for predicting/recognizing plant diseases. Section IV provides an overview of the datasets that are commonly used in the literature for plant disease detection and classification using both objects detection and classification techniques. Section V includes the open challenges that need to be addressed, while Section VI includes the future directions for addressing these open challenges. Finally, Section VII summarizes the work of this paper.

2. Factors responsible for plant diseases

A wide range of agricultural diseases can arise at various stages of plant development and harm the plant's growth, which can have a negative impact on overall crop production [13]. Plant diseases are caused by a variety of conditions at various phases of plant development [14]. As summarized in [15], crop disease-causing variables are categorized into two: biotic factors and abiotic factors. Biotic factors such as viruses, fungi, bacteria, mites, and slugs emerge as a result of microbial infection in plants, whereas abiotic variables such as water, temperature, irradiation, and nutritional deprivation damage plant growth [16]. Accordingly, some sample plant leaf images with different diseases from the PlantVillage dataset and different images from other datasets showing healthy and diseased plant leaves have been included in the study [17] and different images from other datasets showing healthy and diseased plant leaves have been summarized in the work of [18] accordingly. Additionally, the detail computer vision-based techniques and processes including field crops, image acquisition, leaf image datasets, image preprocessing (test set, training set, and validation sets), data splitting, and performance assessment methods) for plant disease detection and classification have been clearly indicated in the work of [17]. The details of the factors respon- sible for plant diseases have been depicted in Fig. 1. Additionally, some sample plant leaf images with different diseases from the PlantVillage dataset and different images from other datasets showing healthy and diseased plant leaves have been depicted in Figs. 2 & 3 respectively.

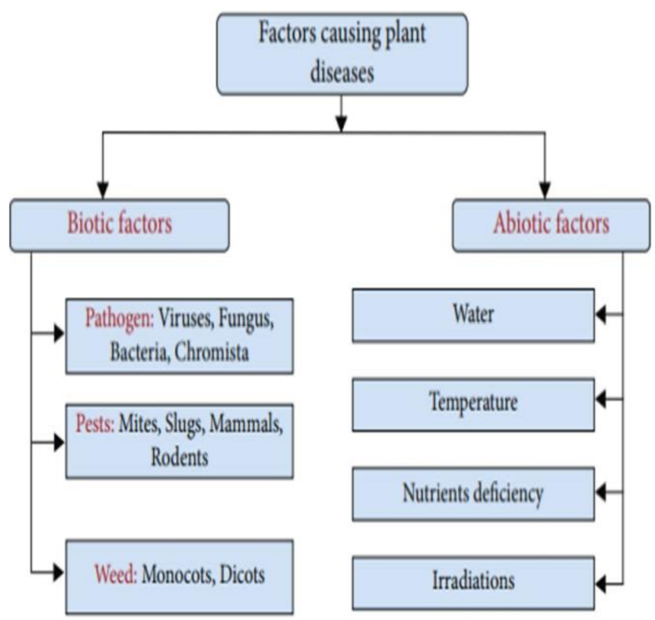

Fig. 1 Factors responsible for plant diseases

Some sample plant leaf images with different diseases from the PlantVillage dataset and different images from other datasets showing healthy and diseased plant leaves have been depicted in Figs. 2, 3 respectively.

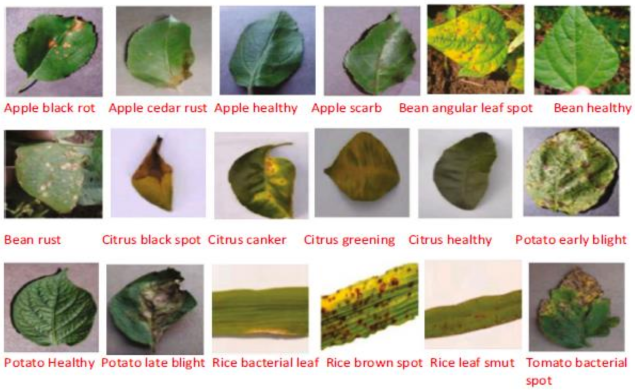

Fig. 2 some sample plant leaf images with different diseases from the Plant Village dataset

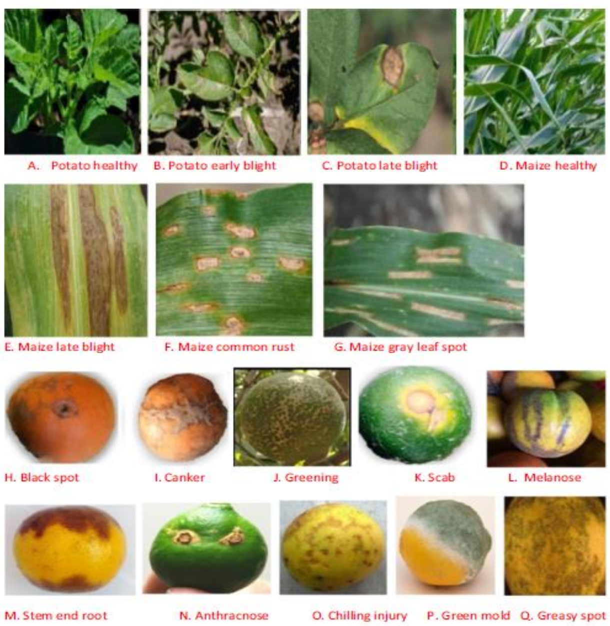

Fig. 3 Different images from other datasets showing healthy and diseased plant leaves

3. Literature Review

This section provides a detailed literature review of disease detection techniques. The author has analyzed the literature on various plant leaf disease detections and classifications, as well as the models/techniques that have been used

According to [19], DL-based solutions for real-time insect detection and identification in the soybean crop have been proposed. The performances of various transfers learning (TL) models were investigated to determine the feasibility and reliability of the proposed approach for determining the insect's identification and detection accuracy. The proposed approach achieved 98.75%, 97%, and 97% accuracy using YoloV5, InceptionV3, and CNN, respectively. Among these, the YoloV5 algorithm performs quite well in the solution and can run at 53 fps, making it suitable for real-time detection. Furthermore, a dataset of crop insects was collected and labeled by mixing images taken with various devices. The proposed study reduced the workload of the producer, was considerably simpler, and produced better results. The authors of [20] have proposed a system that uses DL approaches to classify and detect plant leaf diseases. They collected the images from the PlantVillage dataset website. They used the CNN to classify plant leaf diseases in the suggested method. There were 15 classes, including 12 classes for diseases of various plants that were found, such as bacteria, fungi, and so on, and three classes for healthy leaves. As a result, they achieved high accuracy in both training and testing, with an accuracy of 98.29% in training and 98.029% in testing for all data sets used.

In the study of [16], an effective method for recognizing and identifying rice plant disease based on the size, shape, and color of lesions in a leaf image has been presented. The suggested model uses Otsu's global threshold technique to perform image binarization to remove image background noise. To detect the three rice diseases, the proposed technique based on a fully connected CNN was trained using 4000 image samples of each diseased leaf and 4000 image samples of healthy rice leaves. The results revealed that the proposed fully connected CNN approach was fast and effective, with an accuracy of 99.7% on the dataset. This accuracy far exceeded that of the existing plant disease detection and classification methods. The authors of [21] have presented a model based on CNN to identify and classify tomato leaf disease using a public dataset and complement it with images taken on the country's farms. To avoid overfitting, generative adversarial networks were used to generate samples that were similar to the training data. The results reveal that the proposed model performed well in the detection and classification of diseases in tomato leaves, with accuracy greater than 99% in both the training and test datasets.

The authors of [18] have used the dataset "PlantVillage" to depict four bacterial infections, two viral diseases, two mold diseases, and one mite-related ailment. Images of unaffected leaves were also shown for a total of 12 crop species. For the development of prediction models, ML approaches such as SVMs, grey-level co-occurrence matrices (GLCMs), and CNNs were used. AI for classification has evolved alongside the development of the backpropagation of ANNs. Based on the real-time leaf images gathered, a KMC operation was also performed to detect diseases. Finally, the proposed approach achieved an overall accuracy of 99% and 98% for rice trees and apples, respectively, and 96%, 94%, 95%, and 97% for tomato trees. Multi-class classification problems, such as the one in this study, were evaluated using precision, recall, and fmeasure metrics for a set containing only one symptom pool for each class. The authors of [22] have proposed the use of an enhanced CNN technique to detect rice disease. DNNs have had a lot of success with image classification tasks. In this study, they have demonstrated how DNNs can be used for plant disease detection in the context of image classification. Finally, this research compares existing techniques in terms of accuracy of 80%, 85%, 90%, and 95% for TL, CNN + TL, ANN, and ECNN + GA techniques, respectively. The work in [23] has addressed numerous ML and DL techniques. SVM, KNN, RF, LR, and CNN were the ML approaches used in the study effort for disease prediction in plants. Then, a comparison of ML and DL approaches was carried out. Among the ML techniques, the RF has the best accuracy of 97.12%; however, when compared to the DL model presented in the study, the CNN technique has the highest accuracy of 98.43%.

The capacity to identify rice leaf disease was limited by the image backgrounds and the conditions under which the images were acquired [24]. DL models for automated identification of rice leaf diseases suffer significantly when evaluated on independent rice leaf disease data. The results of well-known and frequently used TL models for detecting rice leaf disease were examined in this study. There were two methods for accomplishing this: frozen layers and fine-tuning. The DenseNet169 findings produced an excellent testing accuracy of 99.66%, and when the results of the fine-tuned TL models were analyzed, Xception performed well and achieved 99.99% testing accuracy. The authors of [25] have presented Ant Colony Optimization with Convolution Neural Network (ACO-CNN), a novel DL technique for disease detection and classification. ACO was used to assess the effectiveness of disease diagnostics in plant leaves. The CNN classifier was used to sub- tract color, texture, and plant leaf arrangement geometries from the given images. Some of the effectiveness metrics used for analysis and providing a proposed method demonstrate that the proposed approach outperforms previous techniques with an accuracy rate. Aoncert measurements were utilized for the execution of these approaches. Finally, the ACO-CNN model outperformed the C-GAN, CNN, and SGD models in terms of accuracy, precision, recall, and f1-score. The accuracy rates of C-GAN, CNN, and SGD were 99.6%, 99.97%, and 85%, respectively. The accuracy rate in the ACO-CNN model was 99.98%; therefore, precision, recall, and F1-score have higher rates in the ACO-CNN technique compared to other models, and the F1-score has the highest rate compared to other models. The authors of [26] have presented a DL model (PPLCNet) that includes dilated convolution, a multi-level attention mechanism, and GAP layers. The model used novel weather data augmentation to expand the sample size to enhance the generalization and robustness of feature extraction. The feature extraction network uses saw-tooth dilated convolution with a configurable expansion rate to extend the perceptual field of the convolutional domain, effectively addressing the problems of insufficient data information extraction. The lightweight CBAM attention mechanism was located in the feature extraction network's middle layer. It was used to improve the model's information representation. By reducing the number and complexity of parameters computed by the network, the GAP layer prevents overfitting of the model. The validation of the retained test dataset reveals that the PPLC-Net model's recognition accuracy and F1-score were 99.702% and 98.442%, respectively, and that the number of parameters and FLOPs were 15.486 M and 5.338G, respectively, which can meet the requirements of accurate and fast recognition. Furthermore, the proposed integrated CAM visualization approach fully validates the efficiency of the proposed model. According to the study [27], an effective CNN model was proposed to categorize tomato leaf diseases and detect the name of the disease affecting tomato leaves. An approach to a 2-dimensional Convolutional Neural Network (2DCNN) model with 2-Max Assembling covers and completely related layers has been proposed. The experimental results show that the model was successful enough to detect the disease with an accuracy of 96% when compared to other classification models such as SVM, VGG16, Inception V3, and Mobile Net CNN model.

To extract different features [28] have used model engineering (ME). To improve feature discrimination and processing speed, several SVM models were used. In the training process, the kernel parameters of the radial basis function (RBF) were computed depending on the selected model. Six leaf image sets encompassing healthy and sick leaves of apple, corn, cotton, grape, pepper, and rice were analyzed using PlantVillage and UCI databases. Accordingly, the categorization procedure yielded almost 90,000 images. The findings of the experimental implementation phase reveal the potential of a powerful model in classification activities, which would be useful for a variety of future leaf disease diagnostic applications in the agricultural business. In terms of stability, the dilated learning model outperforms the typical ResNet-18 design. On the test set, the model had an average accuracy of 98.5% for leaf disease recognition models. In recognizing grape or cotton leaf diseases, a test set accuracy of 97.93% is less than the proposed structure's accuracy of 97.93%. The authors of [29] have presented an image segmentation algorithm for the automatic detection and classification of plant leaf diseases. It also includes an overview of various disease classification techniques that can be used to detect plant leaf disease. The genetic algorithm was used for image segmentation, which was vital for disease detection in plant leaf disease.

The ensemble classifiers (EC) in [30] were developed by using various approaches to preparation, feature extraction, and classification. The performance of these multiple ensemble techniques was then compared to select the best ensemble classifiers. The suggested technique's precision and reliability were tested in both controlled laboratory settings and real-world conditions using two databases, namely PlantVillage and Taiwan tomato leaves. The top EC, which achieved 96% accuracy, was determined by the consideration of shadow, brightness fluctuations, disease similarities, background clutter, multiple leaves, and diverse textures. Here, the proposed ensemble models were presented and linked to several DL techniques. Furthermore, the proposed solution out- performed the most recent state-of-the-art DL technique.

The authors of [31] have proposed a hybrid DL approach for the early detection and classification of tomato plant leaf diseases. A CNN, a convolutional attention module (CBAM), and SVM were combined in the hybrid system. The proposed approach was evaluated using a database of tomato leaf images. The suggested model can initially detect nine distinct tomato diseases; however, it is not limited to this. The obtained findings were highly encouraging, with an accuracy of up to 97.2%, which can be improved by improving learning processes. Here, the proposed approach outperformed better than the state-of-the-art DL approaches. The proposed system was lightweight and efficient, so the farmer may install it on any smart device with a digital camera and processing capabilities. A farmer can detect any disease immediately with a little training, allowing them to take timely preventive measures.

4. Datasets

This section provides an overview of publicly available datasets. These datasets serve different purposes; some of them are used for classification to determine if a plant image is healthy or infected with a disease (discussed in Subsection IV-A), while others are used for object detection to identify diseases on plants (discussed in Subsection IV-B).

4.1 Classification

The PlantVillage dataset [32] comprises 54,303 leaf images, healthy and diseased, categorized into 38 classes based on species and diseases. The dataset includes images of 14 crop species, such as apple, blueberry, cherry, corn, grape, orange, peach, bell pepper, potato, raspberry, soybean, squash, strawberry, and tomato. It covers 17 fungal-related diseases, four bacterial diseases, two mold (oomycete) diseases, two viral diseases, and one mite-related disease. Additionally, it provides images of disease-free healthy leaves from 12 crop species.

The iBean leaf image dataset [33] comprises images of bean leaves captured from field conditions. The National Crops Resources Research Institute (NaCRRI), the national organization in charge of research in agriculture in Uganda, and the Makerere AI lab collaborated to capture these images in several areas of Uganda. Images were captured from the field or garden using a simple smartphone, which were then analyzed by NaCRRI experts who determined which illness was present in each image. The dataset consists of 1, 296 images and three classes. 428 images are for the healthy class, 432 images are for the angular leaf spot, and the remaining 436 for the bean rust.

The PlantLeaves dataset [34] is comprised of 4, 503 images of plant leaves, both healthy and diseased, categorized into 22 categories based on the species and the state of health. The dataset includes 2, 278 healthy leaf images and 2, 225 diseased ones. The images were captured using a basic digital camera.

The PlantaeK dataset [35] is a leaf database of indigenous plants found in Jammu and Kashmir. It comprises 2, 153 images of healthy and diseased plant leaves, categorized into 16 groups by species and health status. The images feature various crop species, including apple, apricot, cherry, cranberry, grapes, peach, pear, and walnut. The dataset comprises 1, 223 healthy leaf images and 934 diseased leaf images.

The Plant Pathology 2020 challenge dataset [36] is a classification dataset for the foliar disease of apples. The creators of the dataset manually captured 3, 651 real-world symptoms of several apple foliar diseases with varied lighting, angles, surfaces, and noise. The dataset includes 865 healthy leaves, 187 cases of complex diseases, 1, 200 cases of apple scab, and 1, 399 cases of cedar apple rot.

The citrus leaf images dataset [37] contains images of healthy and infected citrus plants with diseases such as black spot, canker, scab, greening, and melanosis. The dataset includes 609 images from citrus leaves, of which 58 are healthy images, and 150 images from citrus fruits, of which 22 are healthy images.

The Kaggle dataset [38] contains 9, 436 annotated images and 12, 595 unlabeled images of cassava leave. The dataset contains five classes, one is the class for healthy plants and the other four are for diseases (cmd, cgm, cbsd, and cbb). NaCRRI in collaboration with the AI lab at Makerere University captured and annotated these images.

IJEASM
International Journal of Engineering Applied Science and Management
ISSN (Online): 2582-6948
Vol. 4 Issue 10, October 2023

The dataset of Rice leaf images [39] includes 120 images collected from a village in India. The dataset contains 40 images of each disease, for a total of 120 images. The NLB dataset [130] is comprised of 234 images of leaf spot disease in maize crops.

4.2 Object Detection

The PlantDoc dataset [40] includes 2, 345 images. These images contain 13 plant species and 18 classes of diseases. This dataset is publicly available for download and it can also be used as an open dataset for benchmarks. The classes of the PlantDoc dataset are the following: Cherry leaf, Peach leaf, Cherry leaf, Peach leaf, Corn leaf blight, Apple rust leaf, Potato leaf late blight, Strawberry leaf, Corn rust leaf, Tomato leaf late blight, Tomato mold leaf, Potato leaf early blight, Apple leaf, Tomato leaf yellow virus, Blueberry leaf, Tomato leaf mosaic virus, Raspberry leaf, Tomato leaf bacterial spot, Squash Powdery mildew leaf, Grape leaf, Corn Gray leaf spot, Tomato Early blight leaf, Apple Scab Leaf, Tomato Septoria leaf spot, Tomato leaf, Soyabean leaf, Bell pepper leaf spot, Bell pepper leaf, grape leaf black rot, Potato leaf, and Tomato two-spotted spider mites leaf. PlantDoc contains 8, 851 annotations with an average of 3.4 annotations per image. The average image size is 0.53 mp and the distribution of the image sizes starts from 0.01 mp to 24.00 mp. The median image ratio is 800 675. The classes Blueberry leaf, Tomato leaf yellow virus, and Peach leaf are overrepresented with more than 600 images for each class and the classes Tomato leaf late blight, Tomato Early blight leaf, Apple rust leaf, Apple Scab Leaf, grape leaf black rot, Corn rust leaf, Corn Gray leaf spot, Soybean leaf, Potato leaf, and Tomato two-spotted spider mites leaf are under-represented with less than 220 images for each class.
The CropDeep dataset [41] is composed of 31, 147 images containing over 49, 000 annotated instances from 31 different classes. The images were captured in greenhouses under various conditions using different cameras. Additionally, the IP102 dataset [42] is a comprehensive benchmark dataset for recognizing insect pests. It comprises over 10, 000 images divided into 102 categories, with insect pests that mainly target one agricultural product grouped together in the same top-level category. The IP102 dataset has a hierarchical taxonomy.

Table 2 presents the statistics for each dataset including: (i) the name of the dataset, (ii) the type of the crop, (iii) the number of images, (iv) the number of classes, and (v) whether classification (C) or object detection is targeted (C).

Table 2: Datasets Statistics

Dataset	Crop	Images	Classes	Method
PlantVillage	Various crops	54, 303	38	C
iBean	Bean	1, 296	5	C
PlantLeaves	Various crops	4, 502	22	C
PlantaeK	Various crops	2, 153	16	C
Plant Pathology	Apples	3, 651	4	C
Citrus leaf images	Citrus fruits	759	6	C
Kaggle	Cassava	22, 031	5	C
Rice leaf images	Rice	120	3	C
NLB	Maze	234	1	C
PlantDoc dataset	Various crops	2, 345	18	O
CropDeep	Various crops	31, 147	31	O
IP102	Corn	10, 000	102	O

5. Challenges in Plant Disease Detection

After a detailed review of ML and DL algorithms for plant disease detection and classification and the detailed computational study on five state-of-the-art object detection algorithms for plant disease detection and eighteen state-of-the-art classification algorithms for plant disease classification on a widely-used dataset, we have identified several challenges in practical applications of plant disease detection:

1) There is a lack of models that handle non-image data. Most existing classification and object detection algorithms focus solely on image data, neglecting other relevant information such as temperature and humidity. Developing techniques to incorporate non-image data is essential for more accurate predictions.

2) There are only a few completely annotated open datasets. Many studies rely on the PlantVillage dataset, which was obtained in a controlled laboratory setting. Generating larger datasets under real-world conditions is crucial. Collaborative efforts are needed to create representative datasets.

3) Most works treat the disease detection problem as a classification problem, either binary classification or multi-class classification. While many works treat disease detection as a classification problem, more emphasis should be placed on object detection to identify both the disease type and affected regions in the image.

4) Most papers use a single dataset used to train and test the model. Models trained on a single dataset often perform poorly on different datasets. It is essential to consider diverse datasets to improve model robustness.

5) *Overreliance on CNN architectures:* While CNNs yield good results, exploring other neural network architectures like recurrent neural networks can enhance disease detection methods.

6) *Small leaf and early-stage disease recognition*: Current datasets mainly consist of images with large leaves. Annotating datasets for early-stage disease detection and small leaf recognition is necessary.

7) *Challenges with illumination and occlusion:* Existing algorithms struggle with images under varying lighting conditions and occlusion. More robust methods are needed to address these issues.

8) *Computational efficiency*: Many models are computationally intensive, hindering real-time applications. Researchers should focus on improving the computational efficiency of their models.

6. Future Directions in Plant Disease Detection

In addition to the challenges mentioned above, there are several promising directions for future research in plant disease detection:

1) *Integration of non-image data:* Develop models that can effectively integrate non-image data, such as environmental factors, into disease detection algorithms to improve prediction accuracy.

2) *Creation of diverse and real-world datasets*: Collaborate with experts to generate large, representative datasets under real-world agricultural conditions to enhance the generalizability of models.

3) *Emphasis on object detection:* Explore the potential of object detection methods for predicting plant diseases, which can provide more detailed information about disease localization.

4) *Robustness across datasets:* Develop models that perform consistently well across various datasets to ensure their practical utility.

5) *Exploration of alternative neural network architectures:* Experiment with different neural network architectures beyond CNNs, such as recurrent neural networks, to uncover their potential in disease detection.

6) *Early-stage and small leaf recognition:* Annotate datasets specifically for early-stage disease recognition and the identification of diseases on plants or leaves with small sizes.

7) *Addressing illumination and occlusion challenges:* Implement techniques to enhance the robustness of algorithms in the presence of variable lighting conditions and occluded images.

8) *Improved computational efficiency:* Focus on optimizing model architectures and algorithms to make them suitable for real-time applications.

7. Conclusion

The aim of this study is to examine existing research that utilizes ML and DL techniques in precision agriculture, with a particular focus on plant disease detection and classification methods. Furthermore, we present the available datasets for plant disease detection and classification, and provide details about their classes and data, and whether the specific dataset is suitable for classification or object detection. An extensive research study is conceded out on various kinds of machine and deep learning techniques for plant disease recognition and classification. After this, other techniques of classification in machine learning might be employed for may be used for plants disease detection and in the intellect of aiding the farmers an automatic disease detection of all kinds of disease in the crop that were to be detected. This analysis discusses various approaches of DL for the plant diseases detection. Furthermore, several techniques/mappings were summarized for recognizing the disease symptoms. Here the development of deep learning technologies in recent years for the identification of plant leaf diseases. We anticipate that this work will be a useful tool for scientists looking into plant disease detection. Also, a comparative study is also made between machine and deep learning techniques. Though a great deal of noteworthy progress was noticed in recent years, there were still some research gaps that should be addressed and to implement effective techniques for plant disease detection. In future work, we plan to study more algorithms for classification and object detection on more datasets to see whether the results are consistent across different datasets. We also intend to study some image preprocessing and data augmentation techniques to see whether the accuracy of the algorithms can be improved with these techniques.

References

[1] Fao "New standards to curb the global spread of plant pests and diseases" Web page of the Food and Agriculture Organization of the United Nations (2019).

[2] M. Loey, A. ElSawy, M. Afify, Deep learning in plant diseases detection for agricultural crops: a survey, Int. J. Serv. Sci. Manag. Eng. Technol. 11 (2020) 41–58.

[3] U. Shruthi, V. Nagaveni, B. Raghavendra, A review on machine learning classification techniques for plant disease detection, in: 2019 5th International Conference

on Advanced Computing & Communication Systems, ICACCS), 2019, pp. 281–284.
[4] T. Bera, A. Das, J. Sil, A.K. Das, A survey on rice plant disease identification using image processing and data mining techniques, in: Emerging Technologies in Data Mining and Information Security, Springer, 2019, pp. 365–376.
[5] S. Kaur, S. Pandey, S. Goel, Plants disease identification and classification through leaf images: a survey, Arch. Comput. Methods Eng. 26 (2019) 507–530.
[6] J. Shirahatti, R. Patil, P. Akulwar, A survey paper on plant disease identification using machine learning approach, in: 2018 3rd International Conference on Communication and Electronics Systems (ICCES), 2018, pp. 1171–1174.
[7] N. Ganatra, A. Patel, A survey on diseases detection and classification of agriculture products using image processing and machine learning, Int. J. Comput. Appl. 180 (2018) 1–13.
[8] S.S. Kumar, B. Raghavendra, Diseases detection of various plant leaf using image processing techniques: a review, in: 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), 2019, pp. 313–316.
[9] V. Singh, A.K. Misra, Detection of plant leaf diseases using image segmentation and soft computing techniques, Information processing in Agriculture 4 (2017) 41–49.
[10] S. Zhang, Z. You, X. Wu, Plant disease leaf image segmentation based on super pixel clustering and EM algorithm, Neural Comput. Appl. 31 (2019) 1225–1232.
[11] A. Badage, Crop disease detection using machine learning: Indian agriculture, Int. Res. J. Eng. Technol 5 (2018).
[12] J. Singh, H. Kaur, Plant disease detection based on region-based segmentation and KNN classifier, in: International Conference on ISMAC in Computational Vision and Bio-Engineering, 2018, pp. 1667–1675.
[13] Kumar R, Chug A, Singh AP, Singh D. A systematic analysis of machine learning and deep learning based approaches for plant leaf disease classification: a Review. J Sensors. 2022.
[14] Shoaib M, et al. An advanced deep learning models-based plant disease detection: a review of recent research. Front Plant Sci. 2023;14:1–22.
[15] Vishnoi VK, Kumar K, Kumar B. Plant disease detection using computational intelligence and image processing. Berlin Heidelberg: Springer; 2021.
[16] Upadhyay SK, Kumar A. A novel approach for rice plant diseases classification with deep convolutional neural network. Int J Inf Technol. 2022;14(1):185–99
[17] Tiwari V, Joshi RC, Dutta MK. Dense convolutional neural networks based multiclass plant disease detection and classification using leaf images. Ecol Inform. 2021;63: 101289.
[18] Ahmed I, Yadav PK. A systematic analysis of machine learning and deep learning based approaches for identifying and diagnosing plant diseases. Sustain Oper Comput. 2023;4:96–104
[19] Tirkey D, Singh KK, Tripathi S. Performance analysis of AI-based solutions for crop disease identification detection, and classification. Smart Agric Technol. 2023
[20] Jasim MA, Al-Tuwaijari JM. Plant leaf diseases detection and classification using image processing and deep learn- ing techniques. Int Comput Sci Soft Eng Conf. 2020.
[21] Guerrero-Ibañez A, Reyes-Muñoz A. Monitoring tomato leaf disease through convolutional neural networks. Electron. 2023;12(1):1–15
[22] Balaji V, et al. Deep transfer learning technique for multimodal disease classification in plant images. Contrast Media Mol Imaging. 2023;2023:5644727. https://doi.org/10.1155/2023/5644727.
[23] Kirola M, Joshi K, Chaudhary S, Singh N, Anandaram H, Gupta A. Plants diseases prediction framework: a image- based system using deep learning. Proc IEEE World Conf Appl Intell Comput. 2022. https://doi.org/10.1109/AIC55 036.2022.9848899.
[24] S. C. Gopi and H. K. Kondaveeti. Transfer learning for rice leaf disease detection. 2023. https://doi.org/10.1109/ICAIS 56108.2023.10073711.
[25] Algani YMA, Caro OJM, Bravo LMR, Kaur C, Al Ansari MS, Bala BK. Leaf disease identification and classification using optimized deep learning. Meas Sensors. 2023;25:100643. https://doi.org/10.1016/j.measen.2022.100643.
[26] Dai G, Fan J, Tian Z, Wang C. PPLC-Net : neural network-based plant disease identification model supported by weather data augmentation and multi-level attention mechanism. J King Saud Univ Comput Inf Sci. 2023;35(5):101555. https://doi.org/10.1016/j.jksuci.2023.101555.
[27] P. B. R and A. V V. 2022. Tomato leaf disease detection and classification using Cnn. 71 4 2921–2930
[28] M. S. Anari. A hybrid model for leaf diseases classification based on the modified deep transfer learning and ensemble approach for agricultural aiot-based monitoring," vol. 2022.
[29] Singh V, Misra AK. Detection of plant leaf diseases using image segmentation and soft computing techniques Process. Inf Agric. 2016. https://doi.org/10.1016/j.inpa.2016.10.005.
[30] E. Saraswathi and J. FarithaBanu. A novel ensemble classification model for plant disease detection based on leaf images. 2023.
[31] Altalak M, Uddin MA, Alajmi A, Rizg A. A hybrid approach for the detection and classification of tomato leaf dis- eases. Appl Sci. 2022;12:1–14.
[32] Kaggle. (2018). Plantvillage Dataset. Accessed: Nov. 8, 2022. [Online]. Available: https://www.kaggle.com/datasets/emmarex/plantdisease
[33] AIR Lab Makerere University. (2020). iBean Leaf Image Dataset. Accessed: Nov. 8, 2022. [Online]. Available: https://github.com/AI- Lab-Makerere/ibean

and https://www.fao.org/news/story/en/item/1187738/icode/

[34] S. Siddharth, U. Singh, A. Kaul, and S. Jain, ‘‘A database of leaf images: Practice towards plant conservation with plant pathology,’’ Mendeley Data, 2019. [Online]. Available: https://data.mendeley.com/datasets/hb74ynkjcn/1

[35] V. P. Kour and S. Arora, ‘‘Plantaek: A leaf database of native plants of jammu and kashmir,’’ in Recent Innovations in Computing. Singapore: Springer, 2022, pp. 359–368.

[36] R. Thapa, K. Zhang, N. Snavely, S. Belongie, and A. Khan, ‘‘The plant pathology challenge 2020 data set to classify foliar disease of apples,’’ Appl. Plant Sci., vol. 8, no. 9, p. e11390, Sep. 2020.

[37] H. T. Rauf, B. A. Saleem, M. I. U. Lali, M. A. Khan, M. Sharif, and S. A. C. Bukhari, ‘‘A citrus fruits and leaves dataset for detection and classification of citrus diseases through machine learning,’’ Data Brief, vol. 26, Oct. 2019, Art. no. 104340.

[38] Kaggle. (2019). Cassava Disease Classification. Accessed: Nov. 8, 2022. [Online]. Available: https://www.kaggle.com/c/cassava- disease/overview

[39] H. B. Prajapati, J. P. Shah, and V. K. Dabhi, ‘‘Detection and classification of rice plant diseases,’’ Intell. Decis. Technol., vol. 11, no. 3, pp. 357–373, Aug. 2017.

[40] D. Singh, N. Jain, P. Jain, P. Kayal, S. Kumawat, and N. Batra, ‘‘PlantDoc: A dataset for visual plant disease detection,’’ in Proc. 7th ACM IKDD CoDS 25th COMAD, 2020, pp. 249–253.

[41] Y.-Y. Zheng, J.-L. Kong, X.-B. Jin, X.-Y. Wang, and M. Zuo, ‘‘CropDeep: The crop vision dataset for deep-learning-based classification and detection in precision agriculture,’’ Sensors, vol. 19, no. 5, p. 1058, Mar. 2019.

[42] H. Wu, T. Wiesner-Hanks, E. L. Stewart, C. DeChant, N. Kaczmar, M. A. Gore, R. J. Nelson, and H. Lipson, ‘‘Autonomous detection of plant disease symptoms directly from aerial imagery,’’ Plant Phenome J., vol. 2, no. 1, pp. 1–9, Jan. 2019.