A Survey of Deep Learning-Based Plant Leaf Disease Identification and Classification

Rutik Chaudhari
Department of Computer Engineering
COEP Technological University
Pune, India

Dr. Vahida Z. Attar
Department of Computer Engineering
COEP Technological University
Pune, India

***Abstract*—The agricultural sector plays a vital role in sustaining its growing population and economy. The escalating demand for food due to population growth un-derscores the importance of optimizing crop yield. The main reason behind significant impact on agricultural productivity is due to bacteria, fungi, and viruses that cause plant illnesses. Leveraging deep learning techniques, par-ticularly CNNs, has demonstrated potential as an effective approach to over-come these problems. This survey provides a comprehensive overview of leaf diseases affecting various crops, detailing common pathogens and their early symptoms. Utilization of various datasets which includes real field or controlled environment images of leaf, we highlight essential data prepro-cessing methods including data augmentation, segmentation, normalization and feature extraction. It also examines machine learning and deep learning models such as SVM, CNNs used for accurate disease detection. Perfor-mance metrics like accuracy, precision, recall, and F1-score are used to eval-uate model effectiveness. This paper aims to guide future research in opti-mizing leaf disease detection and improving agricultural productivity through advanced computational techniques.**

***Index Terms*—leaf disease classification, deep learning, machine learning, convolutional neural network (CNN)**

I. Introduction

Plants play a fundamental role in sustaining life on Earth, serving as the primary source of oxygen and a crucial component of ecosystems. However, Plant diseases are an imminent threat to the health of ecosystems and the world's food supply. Appropriate and immediate identification of these illnesses is essential for efficient plant care and sustainable agriculture.

The need to optimize agricultural methods has increased due to the growing demand for food production and environmental issues. Plant diseases can significantly affect crop productivity and quality as they are spread on by pathogens including fungi, bacteria, and viruses. Traditional methods of disease detection often involve manual inspection, which is labour-intensive and time-consuming.

Agriculture is now a significant driver of economic growth. Due to a number of issues, such as population growth and weather variations, the agriculture sectors started searching for novel strategies to increase food production.

Precision agriculture, a transformative paradigm in modern farming, harnesses the synergy of technology, data, and analytics to revolutionize agricultural practices at a localized level. It also empowers farmers with real-time insights, allowing them to make informed decisions that optimize resource usage while mitigating the challenges inherent in traditional methodologies.

In recent years, advancements in computer vision, specifically Convolutional Neural Networks (CNNs), have revolutionized the field of plant pathology by providing robust tools for automated detection and classification of crop leaf diseases.

II. Related work

Farah et al. [1] adopted pre-trained VGG19 deep learning model in a transfer learning based approach to classify healthy and infested soybean leaves. Three parts of the dataset were divided out: 70% as training, 20% as testing, and 10% as evaluation. To follow a binary approach, they developed three cases and achieved accuracy between 93.71% and 94.16%.

Tetila et al. [2] worked on the segmentation of images using the Simple Linear Iterative Clustering (SLIC) superpixels algorithm. Images of soybean leaf diseases were collected using Unmanned Aerial Vehicle (UAV). In this experiment, a comparison of four deep learning architectures was done using various training strategies such as Transfer learning (TL) and Fine Tuning (FT) with different parameters for the detection of plant leaves as a classification step. For Inceptionv3, Resnet-50, VGG19, and Xception, they obtained an accuracy of 99.04%, 99.02%, 99.02%, and 98.56% respectively.

Reddy et al. [3] suggested a machine learning approach for the detection of plant leaf-based diseases using Random Forest (RF) and Support Vector Machine (SVM). The Euclidean distance method was used for disease-affected areas of the leaf. For the cotton leaf dataset, they achieved 75% accuracy.

Algani et al. [4] came up with an approach of using Ant Colony Optimization with Convolution Neural Network (ACO-CNN) for the detection and classification of leaf diseases. The achieved accuracy by this proposed method is 99.98%.

Bedi et al. [5] proposed an approach of a hybrid model based on Convolutional Autoencoder (CAE) and a Convolutional Neural Network (CNN) for plant leaf disease detection. The model's obtained testing accuracy is 98.38% using 9,914 parameters for training.

Sachdeva et al. [6] worked on potato, tomato, and pepper bell leaves of plant village dataset using Convolutional Neural Network (CNN) and a Bayesian algorithm and they achieved accuracy of 98.9%.

Atila et al. [10] utilized the EfficientNet deep learning model to classify images of leaves taken from the Plant Village collection, and they also conducted a comparison of the performances of other CNN models.They achieved 99.91% accuracy using EfficientNet B5 on original dataset and 99.97% accuracy using EfficientNet B4 on augmented dataset.

Harakannanavar et al. [8] conducted feature extraction using computer vision and compared classification using machine learning approaches. The proposed method is Discrete Wavelet Transform (DWT), Principal Component Analysis (PCA), Grey Level Co-occurrence Matrix (GLCM) and Convolutional Neural Network (CNN). Tomato samples with six disorders were taken from the plant village dataset. In this experiment, the proposed model achieved an accuracy of 99.09%.

Sardogan et al. [11] concentrated on the identification and categorization of diseases on tomato leaves using a Convolutional Neural Network (CNN) with the Learning Vector Quantization (LVQ) algorithm. A total of 500 tomato leaf images were sourced from the plant village dataset. Their method obtained average accuracy of 86%.

Baser et al. [7] introduced the TomConv model, an enhanced Convolutional Neural Network (CNN) applied to 16,012 images of leaves taken from the Plant Village data collection. Their classification accuracy for tomato plant leaves was 98.19%.

Paymode et al. [9] used transfer learning using VGG19 model for disease detection and classification. For tomatoes and grapes, the accuracy values are 95.71% and 98.40%. The overview of research studies is shown in table I and figure 1 provides the taxonomy for detection and classification of plant leaf diseases.

III. Background

This section delves into the various types of diseases that afflict plant leaves, each with unique symptoms and implications. From fungal and bacterial infections to viral outbreaks, leaf diseases pose a serious threat to agricultural productivity. This overview highlights the critical need for early diagnosis and integrated management practices to mitigate the effects of these pervasive plant pathogens.

A. *Pest infested diseases*

It provides the information related to diseases on soybean given in [1].

1. Diabrotica speciosa Infestation
   Diabrotica speciosa is also known as the cucumber beetle which inflicts damage by feeding on leaves, roots, and stems. Early signs of infestation include small, irregular holes in the leaves and damage to the roots and stems. This can lead to stunted growth and wilting of the plant. It primarily affects crops like cucumbers, beans, and corn.
2. Caterpillar Infestation
   Caterpillar infestations on plants are caused by various species of Lepidoptera which lead to significant leaf damage. Early signs include small holes and irregular chewing patterns on leaves, which can progress to extensive damage, including skeletonization where only leaf veins remain visible. Caterpillar frass (droppings) may also be present on and around affected plants, indicating their presence and feeding activity.

In the below sections we provide the common plant diseases categorized into fungal, bacterial and viral diseases with their early signs or symptoms.

B. *Fungal diseases*

1. Powdery mildew: White powdery spots on leaves, usually starting on older leaves.
2. Downy mildew: Yellow angular spots on upper leaf surfaces with a downy white to purplish gray growth on the undersides.
3. Anthracnose: Small, dark, water-soaked spots that enlarge and develop concentric rings.
4. Rust: Yellow to orange pustules, usually on the undersides of leaves, which later turn brown and release spores.

C. *Bacterial diseases*

1. Bacterial blight: Water-soaked lesions that turn brown and may have a yellow halo.
2. Bacterial spot: Small, water-soaked spots that become angular and necrotic with yellow halos.
3. Crown gall: Swollen, tumor-like growths on roots and stems.

D. *Viral diseases*

1. Tobacco mosaic virus: Mosaic patterns of light and dark green on leaves, stunting, and leaf distortion.
2. Cucumber mosaic virus: Mottling, yellowing, and distortion of leaves, often accompanied by stunted growth.
3. Tomato yellow leaf curl virus: Yellowing and upward curling of leaf margins, leaf narrowing, and reduced fruit size.

Understanding the complexity and diversity of plant leaf diseases is essential for creating effective detection and management methods. Transitioning from theoretical knowledge to practical applications requires extensive use of datasets. These datasets are crucial for training and testing machine learning and deep learning models, as well as for assessing their real-world effectiveness.

TABLE I
OVERVIEW OF RESEARCH STUDIES

Reference	Year	Methodology	Accuracy	Analysis
Farah et al. [1]	2023	Transfer learning using VGG19 model	Between 93.71% and 94.16%	The proposed method uses a pre-trained VGG19 model through transfer learning. Transfer learning significantly reduces the training time and resources required.
Algani et al. [4]	2023	ACO-CNN	99.98%	ACO algorithm is employed for feature extraction and CNN for classification. It outperformed the existing methods regarding accuracy, precision, recall, and F1-score.
Baser et al. [7]	2023	Enhanced CNN	98.19%	Performance metrics are not explicitly detailed beyond accuracy, which could limit the understanding of the model's effectiveness.
Harakannanavar et al. [8]	2022	DWT, PCA, GLCM and CNN	99.09%	The proposed method achieves higher accuracy than existing models.
Paymode et al. [9]	2022	Transfer learning using VGG19 model	98.40%	Their approach used transfer learning and achieved higher accuracy on multiple crop.
Reddy et al. [3]	2021	Machine learning approach using SVM and RF	75%	K-means clustering and euclidean distance methods are used in image segmetation step. SVM is used for the classification task.
Bedi et al. [5]	2021	Hybrid model using CAE and CNN	98.38%	Due to dimensionality reduction using CAE, the number of features is reduced. The suggested approach has less number of training parameters as compared with other approaches.
Sachdeva et al. [6]	2021	CNN and Bayesian algorithm	98.90%	Bayesian learning is used as it enhances pixel dependency.
Atila et al. [10]	2021	EfficientNet B4 model	99.97%	This study focuses on efficientnet deep learning models ranging from B0 to B7.
Tetila et al. [2]	2020	InceptionV3 Resnet-50 VGG19 Xception	99.04% 99.02% 99.03% 98.56%	Their approach focuses on comparing four deep learning models with different strategies. They achieved higher accuracy with fine tuning 100% and 75% strategies.
Sardogan et al. [11]	2018	CNN with LVQ	86% average	CNN is responsible for feature extraction, while LVQ is used for classification.

IV. Datasets

Datasets are essential for deep learning, acting as the base for training and testing models. The quality, size, and variety of a dataset directly affect how well a model can learn and produce precise predictions on new data. Some of the researchers also worked on particular plant species taken from single dataset in the reviewed papers. This section presents the information of datasets utilized by researchers in this survey. The overview of datasets is shown in II.

Dataset consists images of healthy and infested soybean leaves used in [1] which are accessible via the mendeley website. The images were collected and annotated by Mignoni et al. [12] experts. The images were taken using three different devices: two smartphones and an unmanned aerial vehicle (UAV), each equipped with a 48MP AI triple camera. The size of images were standardize to 500×500 pixels in JPEG format. In January 2021, the photos were captured in natural field and weather conditions such as windy, sunny and cloudy in the State of Mato Grosso, Brazil. Dataset split was done on total 6410 images into 896 images as healthy, 3309 images as caterpillar and 2205 images as diabrotica speciosa.

As mentioned in this [13] paper, the soybean plants images were taken in an actual field environment located in the city of DouradosMS, Brazil using DJI Phantom 3 Professional UAV is equipped with a resolution of 2.3 megapixels and 1/2.3-inch Sony EXMOR sensor. Total 300 aerial images were taken over various days and under different climatic conditions. Digital negative (DNG) images were acquired of the targets of interest, positioned 2 meters above the plantation, with the camera oriented at a 90-degree angle relative to the ground.

A citrus leaves dataset was created by rauf et al. [14] and used in this [4] paper. Images were collected using DSLR (Canon EOS 1300D) and annotated manually by domain experts. After processing, images were resized to 256×256

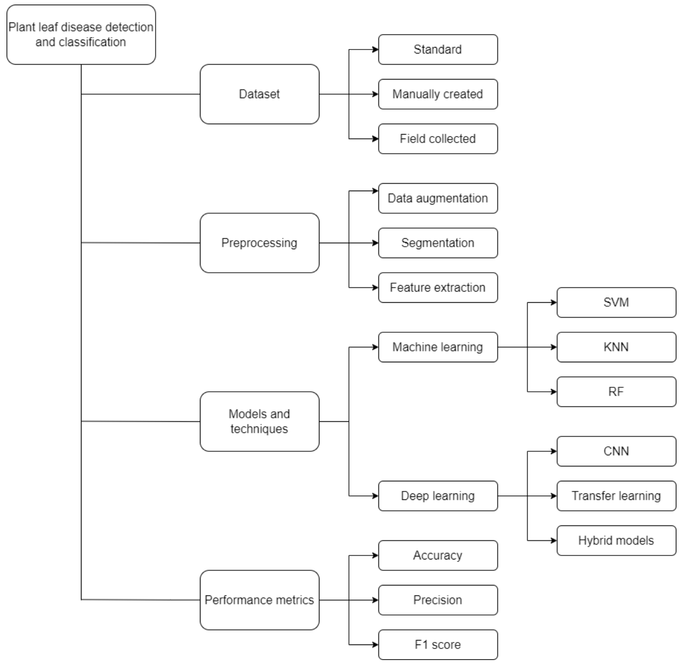

Fig. 1. Taxonomy for detection and classification of plant leaf diseases

TABLE II
OVERVIEW OF DATASETS

Reference	Dataset	Environmental conditions	Crop	Data source website	No. of images
Farah et al. [1]	Soybean leaves	Real field	Soybean	Mendeley	6,410
Tetila et al. [2]	—	Real field	Soybean	—	300
Algani et al. [4]	Citrus fruits and leaves	Real field	Citrus	Mendeley	609
Bedi et al. [5]	Plant village	Laboratory	Peach	Kaggle	4,457
Sachdeva et al. [6]	Plant village	Laboratory	Tomato, potato and pepper bell	Kaggle	20,639
Atila et al. [10]	Plant village	Laboratory	All (augmented)	Kaggle	61,486
Harakannanavar et al. [8]	Plant village	Laboratory	Tomato	Kaggle	600
Sardogan et al. [11]	Plant village	Laboratory	Tomato	Kaggle	500
Baser et al. [7]	Plant village	Laboratory	Tomato	Kaggle	16,012

with a resolution of 72 dpi. There are 609 photos overall, showing both healthy and diseased leaves with melanose, canker, black spot, and greening.

The Plant Village dataset included the images of the leaves of peach plants [15] which was used in this [5] paper to identify bacterial spot disease. The 4457 leaf images of peach plants in the collection are categorized as healthy and unhealthy.

Dataset used in this [6] research work comprises 20,639 leaf images of potato, pepper bell and tomato from plant village having 15 various classes of healthy and diseased leaves.

In [10], they used plant village dataset which comprises 14 distinct plant species and 38 classes in which 12 are healthy and 26 are diseased. Total images in original dataset taken from plant village and augmented dataset which was created by Geetharamani et al. [16] contains 55,448 and 61,486 respectively. To carry out the experimentation work they [8] used tomato leaf samples with six disorders from plant village dataset.

In [11], 500 images of tomato leaves in Plant Village were utilized. The images in the dataset were resized to 512×512. In this [7] research work also they used plant village dataset having 16,012 tomato leaf images with 10 classes.

V. Preprocessing methods

A. Data augmentation

Data augmentation methods are used to increase the size of a dataset with the help of existing data by creating multiple images of same image. This helps improve the efficiency of deep learning and machine learning models by enhancing their ability to generalize. Its benefits include reduced overfitting, improved robustness, and better model accuracy. Common preprocessing techniques include rotation, flipping, scaling, cropping, and color adjustments. Drawbacks can include increased computational cost and potential introduction of noise.

B. Segmentation

Segmentation techniques are used to partition an image into meaningful regions, facilitating detailed analysis and understanding of the image content. Benefits include improved accuracy in object detection, precise localization, and better feature extraction. Techniques include thresholding, clustering, edge detection, and deep learning-based methods. High computational cost, sensitivity to noise and variations, and the need for large annotated datasets for training complex models are some of the drawbacks of segmentation.

C. Feature extraction

Feature extraction is a crucial step in the data preprocessing and it involves transforming raw data into a set of attributes or features that are more informative and easier for models to understand. This process helps in simplifying the data, reducing its dimensionality, and highlighting the most relevant information for the learning algorithm. Various approaches which are based upon color, shape, texture and deep learning are available for feature extraction. Texture-based features use methods such as GLCM, LBP, and Gabor filters to extract the texture of the image. Shape-based features use CSS, Fourier descriptors, and the Hough transform to extract the shapes and sizes of the objects in the image. Color moments, color correlograms, and color histograms are just a few of the methods that color-based features employ to extract color information from an image. In CNN layers, the features are extracted automatically.

VI. Classification or detection techniques

Plant leaf diseases detection or classification methods are categorized into deep learning and machine learning based approaches.

A. Machine learning

Some researchers prefer machine learning models for plant disease recognition due to their lower computational requirements and simplicity. These models work well with smaller datasets and they also require less training time and achieve high accuracy with well-selected features. As in [3], they used SVM and RF for classification purpose. Additionally, machine learning algorithms are less prone to overfitting and can be more robust with limited labeled data, making them practical for many agricultural applications. Additionally, these models can be combined with other techniques to enhance performance, creating hybrid models that leverage the strengths of different approaches. This adaptability makes machine learning algorithms a reliable choice in various scenarios.

B. Deep learning

1. Convolutional Neural Networks (CNNs)
   CNNs are a type of deep learning model specifically designed for processing and analyzing visual data. They are widely used due to their ability to automatically learn and extract hierarchical features from images. CNNs consist of multiple layers, including convolutional layers that detect local patterns, pooling layers that reduce dimensionality, and fully connected layers that integrate these features for classification. Activation functions introduce non-linearity which enables the model to learn complex patterns. CNNs can be further enhanced with advanced architectures like ResNet, VGG, EfficientNet and Inception as used in [2], [7], [10], which improve accuracy and efficiency in various image processing tasks.

2. Transfer learning
   Transfer learning using pretrained models such as VGG [1], [9], which has learned features from a large dataset like ImageNet. This technique is particularly useful when dealing with limited data, as the pretrained model's knowledge can be fine-tuned for specific tasks. Transfer learning allows the pretrained model to retain its learned features while focusing on the new task, making it an efficient and effective approach for various applications.

3. Hybrid models
   Hybrid models combining CNNs with other techniques are used to enhance performance and accuracy in tasks. In a hybrid approach, CNNs can be employed for feature extraction from images, leveraging their strength in identifying complex patterns. Hybrid models used by researchers such as ACO-CNN [4], DWT-PCA-GLCM-CNN [8], CAE-CNN [5], CNN-bayesian algorithm [6] and CNN-LVQ [11]. Hybrid models offer improved performance, flexibility, and often better generalization on smaller datasets compared to using CNNs alone.

Conclusion

The application of CNN and deep learning approaches has demonstrated significant promise in improving detection capabilities. This paper highlights various research studies where researchers employed different CNN models and techniques such as transfer learning, fine-tuning pre-trained models, and hybrid models. The choice of model architecture and optimization techniques plays a significant role in determining the effectiveness of disease detection systems. The deployment of CNN and deep learning technologies in crop leaf disease detection represents a promising solution to one of agriculture's most pressing challenges. By continuing to improve data collection methods and refining model architectures, we can achieve even greater precision and reliability in disease detection systems, ultimately contributing to more sustainable agricultural practices and better global food security. Future research efforts should focus on refining these model architectures, optimizing hyperparameters, and expanding training datasets to enhance the robustness and scalability of detection systems.

REFERENCES

[1] N. Farah, N. Drack, H. Dawel, and R. Buettner, "A deep learning-based approach for the detection of infested soybean leaves," *IEEE Access*, 2023.

[2] E. C. Tetila, B. B. Machado, G. K. Menezes, A. d. S. Oliveira, M. Alvarez, W. P. Amorim, N. A. D. S. Belete, G. G. Da Silva, and H. Pistori, "Automatic recognition of soybean leaf diseases using UAV images and deep convolutional neural networks," *IEEE geoscience and remote sensing letters*, vol. 17, no. 5, pp. 903–907, 2020.

[3] P. C. Reddy, R. M. S. Chandra, P. Vadiraj, M. A. Reddy, T. Mahesh, and G. S. Madhuri, "Detection of plant leaf-based diseases using machine learning approach," in *2021 IEEE International conference on computation system and information technology for sustainable solutions (CSITSS)*. IEEE, 2021, pp. 1–4.

[4] Y. M. Abd Algani, O. J. M. Caro, L. M. R. Bravo, C. Kaur, M. S. Al Ansari, and B. K. Bala, "Leaf disease identification and classification using optimized deep learning," *Measurement: Sensors*, vol. 25, p. 100643, 2023.

[5] P. Bedi and P. Gole, "Plant disease detection using hybrid model based on convolutional autoencoder and convolutional neural network," *Artificial Intelligence in Agriculture*, vol. 5, pp. 90–101, 2021.

[6] G. Sachdeva, P. Singh, and P. Kaur, "Plant leaf disease classification using deep convolutional neural network with bayesian learning," *Materials Today: Proceedings*, vol. 45, pp. 5584–5590, 2021.

[7] P. Baser, J. R. Saini, and K. Kotecha, "TomConv: An improved CNN model for diagnosis of diseases in tomato plant leaves," *Procedia Computer Science*, vol. 218, pp. 1825–1833, 2023.

[8] S. S. Harakannanavar, J. M. Rudagi, V. I. Puranikmath, A. Siddiqua, and R. Pramodhini, "Plant leaf disease detection using computer vision and machine learning algorithms," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 305–310, 2022.

[9] A. S. Paymode and V. B. Malode, "Transfer learning for multi-crop leaf disease image classification using convolutional neural network VGG," *Artificial Intelligence in Agriculture*, vol. 6, pp. 23–33, 2022.

[10] Ü. Atila, M. Uçar, K. Akyol, and E. Uçar, "Plant leaf disease classification using efficientnet deep learning model," *Ecological Informatics*, vol. 61, p. 101182, 2021.

[11] M. Sardogan, A. Tuncer, and Y. Ozen, "Plant leaf disease detection and classification based on CNN with LVQ algorithm," in *2018 3rd international conference on computer science and engineering (UBMK)*. IEEE, 2018, pp. 382–385.

[12] M. E. Mignoni, A. Honorato, R. Kunst, R. Righi, and A. Massuquetti, "Soybean images dataset for caterpillar and diabrotica speciosa pest detection and classification," *Data in Brief*, vol. 40, p. 107756, 2022.

[13] E. C. Tetila, B. B. Machado, N. A. Belete, D. A. Guimarães, and H. Pistori, "Identification of soybean foliar diseases using unmanned aerial vehicle images," *IEEE Geoscience and remote sensing letters*, vol. 14, no. 12, pp. 2190–2194, 2017.

[14] 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 in brief*, vol. 26, p. 104340, 2019.

[15] D. Hughes, M. Salathé *et al.*, "An open access repository of images on plant health to enable the development of mobile disease diagnostics," *arXiv preprint arXiv:1511.08060*, 2015.

[16] G. Geetharamani and A. Pandian, "Identification of plant leaf diseases using a nine-layer deep convolutional neural network," *Computers & Electrical Engineering*, vol. 76, pp. 323–338, 2019.