

Machine Learning Framework for Early Detection of Crop Disease

**Dr. Naveen Kumar H N¹, Aditya A H², Amith N P³, Anish R Jois⁴,
Surya Prakash S P⁵**

¹Associate Professor, Dept. of Electronics and Communication Engineering, Vidyavardhaka College of Engineering, Mysore, Karnataka, India

^{2,3,4,5}Dept. of Electronics and Communication Engineering, Vidyavardhaka College of Engineering, Mysore, Karnataka, India

Abstract

This review paper looks at recent advancements in crop disease detection through deep learning techniques. Crop diseases significantly lowers agricultural productivity, and accurate diagnosis is essential for effective disease management. In order to identify crop illnesses, the study offers a thorough examination of a number of deep learning models, such as hybrid architectures, recurrent neural networks (RNNs), and convolutional neural networks (CNNs). The prospects, challenges, and future directions of incorporating deep learning for precise and quick crop disease identification are also explained in this research. The insights provided offer hope for the development of sustainable agricultural practices through the application of cutting-edge technologies in disease diagnosis and management.

Keywords: Deep Learning, Image processing, Leaf disease, Machine Learning, Plant Disease Prediction.

1. INTRODUCTION

The combination of deep learning and image processing is a powerful tool for detecting plant diseases. As the first stage, image processing entails preprocessing operations including resizing and contrast adjustment to improve the quality of the image for further analysis. Delineating and isolating regions of interest is made easier by features relevant to illness detection that are recovered through contour and texture analysis. Morphological processes such as dilatation and erosion smooth out forms and get rid of noise. Concurrently, deep learning—especially with Convolutional Neural Networks (CNNs)—takes the lead in feature learning. Using convolutional layers to identify spatial hierarchies and pooling layers to minimize dimensionality without sacrificing critical information, CNNs can automatically identify hierarchical features from images. Backpropagation and activation are used in training functions, enhancing the network's capacity to identify complex patterns. With the use of large datasets and pre-trained models that are tailored to certain disease detection tasks, transfer learning improves efficiency. The synergistic technique for precise and efficient plant disease detection is produced by the joint impact of various methodologies, which result in a powerful system where image processing prepares raw data for deep learning's proficiency at understanding complicated patterns. This combination has the power to completely transform agriculture by offering accurate, timely information that is essential for crop protection and long-term food production.

Plant disease identification has advanced significantly in recent years attributable to the combination of deep learning and image processing, which presents a revolutionary method for preserving agricultural production. This review paper takes a deep dive into the environment where cutting-edge technologies come together to tackle the pressing problem of accurately and early plant disease identification. Using deep learning and image processing techniques, the initiative seeks to extract knowledge from numerous research projects between 2015 and 2022. The core of the research is the combination of deep learning models like convolutional neural networks (CNNs) and deep belief networks (DBNs) with image processing techniques like contour feature extraction, adaptive thresholding, and morphological analysis. The introduction provides context for comprehending the development of methods for detecting plant diseases, highlighting the drawbacks of conventional techniques and the promise of image-centric and deep learning-driven approaches. The review attempts to shed light on both the achievements and difficulties in the search for more precise, effective, and scalable plant disease detection systems as we delve into the nuances of diverse datasets, performance evaluation standards, and the nuanced use of deep learning algorithms. In the end, this project not only captures the latest technological developments, but it also emphasizes how crucial they are to changing the precision agriculture scene and guaranteeing world food security.

Significant contributions of our proposed work:

- **Develop an Automated Detection System:** Create an automated system that utilizes cutting-edge technology, such as computer vision, machine learning, and sensor data, to detect crop diseases in their early stages.
- **Enhance Disease Identification Accuracy:** Improve the accuracy of disease identification by developing algorithms that can distinguish between healthy and infected crops, as well as differentiate between various types of diseases.
- **Integrate Multiple Data Sources:** Integrate data from diverse sources, such as images, environmental sensors, and weather data, to provide a comprehensive analysis of crop health and disease risk factors.
- **Enable Real-time Monitoring:** Develop a system capable of real-time or near-real-time monitoring of crops to detect disease outbreaks as soon as they occur, allowing for timely intervention.

Pest and diseases cause over INR 290 billion per annum losses of crops in India. Out of 30,000 plant diseases recorded from different countries, around 5000 occur in India. The fungal infections related decline in crop yield in India is believed to be 5 million tons per year, approximately.

Indian agriculture is considered a global powerhouse. It is the second largest producer of rice, wheat, sugarcane, fruits, vegetables, cotton, and tea. In India, the agri sector employs around 60% of the population and contributes about 17% to the total GDP. One of the major constraints Indian agricultures facing is its low yield, which is 30-5% lower than those of developing countries. The challenges stagnating agricultural productivity in India include outbreaks of pests and diseases, poor soil fertility, unavailability of sufficient water, and climate change.

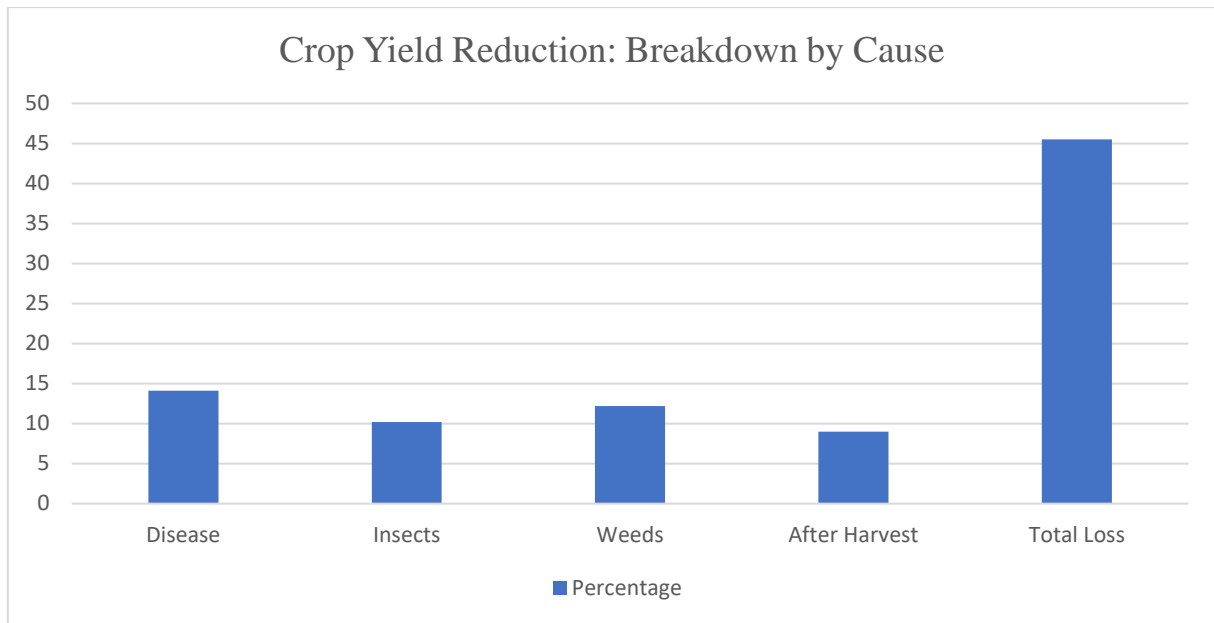


Figure 1: Various Reasons for Plant Loss (Graphical representation)

2. DATA SET USED

The papers highlight the importance of large-scale datasets for creating successful deep learning models by giving an overview of datasets utilized in the domains of weed detection and plant disease identification. It highlights datasets such as Weed, Sugar beets, Plant seedlings, Deep weeds, Grass clover, and others and talks about the difficulties in weed detection, such as changing environmental conditions. Various picture datasets for research have been made accessible, including PlantVillage, PlantClef, Open Plant Disease Dataset, Plant Disease Detection in Cotton Images, and AGRONOMI-Net, which are mentioned in the context of plant disease and pest detection

This paper explores the usage of the PlantVillage dataset, which includes more than 54,000 photos of both healthy and impaired plant leaves from 38 different crop species and disease forms. With a thorough class distribution, it concentrates especially on tomato, pepper bell, and potato species. The dataset is used for preprocessing, feature extraction, and model performance evaluation in order to train and assess machine learning and deep learning models for illness diagnosis. The methodology of the document involves a comprehensive literature review on machine learning for leaf disease classification, utilizing databases like EBSCO host, Scopus, and Google Scholar. The study filters papers based on metrics like citations, publication venue rank, and relevance, focusing on recent academic articles from 2015 to 2022. The increasing interest in plant leaf detection and classification is reflected in the selected papers.

In summary, the document refers to Standard Area Diagrams (SADs) developed for various crops and diseases, aiding accurate visual estimation of disease severity. Another document explores the historical development, best practices, and opportunities for improving visual estimates of plant disease severity, covering aspects such as accuracy, estimation processes, scale types, and the impact of factors like experience and training on accuracy.

Table 1: Plant diseases and no of image samples from benchmark dataset

Name of the Dataset	Crops	Classes (No. of image samples available)	Image Resolution	URL
Plant Village []	Tomato	Mosaic virus (373), Yellow Leaf Curl Virus (3209), Target Spot (1404), Spider mites Two spotted spide (1676), Septoria leaf spot (1771), Leaf Mold (952), Late blight (1909), Early blight (1000), Bacterial spot (2127), Healthy (1591)	256*256	https://www.kaggle.com/datasets/emmarex/plantdisease
	Potato	Early blight (1000), Late blight (1000), Healthy (152)		
	Pepper	Bacterial spot (997), Healthy (1478)		
Plant Doc []	Apple	Apple leaf (9), Apple scub leaf (10), Apple rust leaf (10)		https://www.kaggle.com/datasets/abdulhasibuddin/plant-doc-dataset

	Bell Pepper	Bell pepper leaf (8), Bell pepper leaf spot (9)		
	Blueberry	Blueberry leaf (11)		
	Cherry	Cherry leaf (10)		
	Corn	Corn Gray leaf spot (4), Corn leaf blight (12), Corn rust leaf (10)		
	Potato	Potato leaf early blight (8), Potato leaf late blight (8)		
	Tomato	Tomato Early blight leaf (9), Tomato Septoria leaf spot (11), Tomato leaf bacterial spot (9), Tomato leaf late blight (10), Tomato leaf mosaic virus (10), Tomato leaf yellow virus (6), Tomato mold leaf (6),		
	Grape	grape leaf black rot (8)		
New Plant Disease []	Apple	Apple scab (2016), Apple Black rot (1987),		

		Cedar apple rust (1760), apple healthy (2008)		
	Cherry	Cherry Powdery mildew (1683), Cherry healthy (1826)		
	Corn	Corn Cercospora leaf spot gray leafspot (1642), Corn Common rust (1907), Corn Northern Leaf Blight (1908), Corn healthy (1859)		
	Grape	Grape Black rot (1888), Grape Esca (1920), Grape Leaf blight (1722), Grape healthy (1692)		
	Orange	Orange Huanglongbing (2010)		
	Peach	Peach Bacterial spot (1838), Peach healthy (1728)		

	Pepper	Pepper bell Bacterial spot (1913), Pepper bell healthy (1988)		
	Potato	Potato Early blight (1939), Potato Late blight (1939), Potato healthy (1824)		
	Strawberry	Strawberry Leaf scorch (1774), Strawberry healthy (1824)		
	Tomato	Tomato Bacterial spot (1702), Tomato Early blight (1920), Tomato Late blight (1851), Tomato Leaf Mold (1882), Tomato Septoria leaf spot (1745), Tomato Spider mites Two-spotted spider mite (1741), Tomato Target Spot (1827), Tomato Yellow Leaf Curl Virus (1961),		

		Tomato mosaic virus (1790), Tomato healthy (1926),		
Rice diseases image dataset	Rice	Brown Spot (523), Hispa (565), Leaf Blast (779), Healthy (1488)	1338*1338	https://www.kaggle.com/datasets/minhhuy2810/rice-diseases-image-dataset
OSF	Maize	Northern Leaf Blight (18222)		https://osf.io/p67rz/

3. REVIEW METHODOLOGY

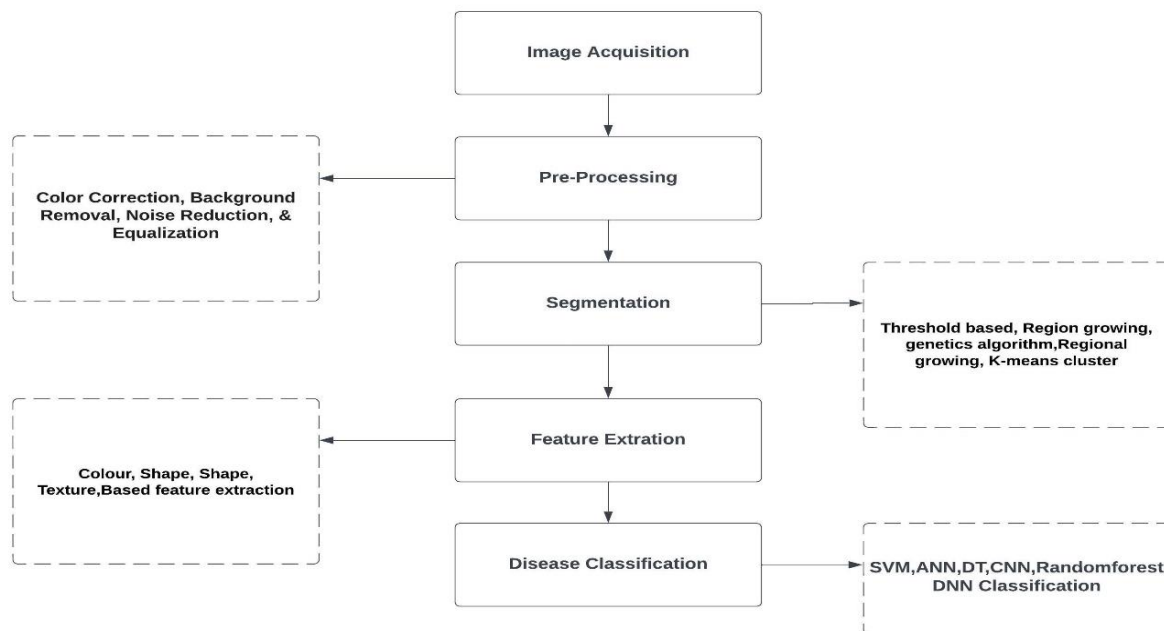


Figure 2: Block diagram of proposed system

The methodology outlined for plant disease classification draws on insights obtained from a comprehensive review of relevant academic papers in the field. The initial steps involve data collection, dataset selection, and preprocessing techniques, following methodologies explains studies such as [1] and [8]. Image enhancement and feature extraction are integral to the process, aligned with findings from [12] and [3], emphasizing the importance of morphological parameters and segmentation for disease identification. Feature selection and dimensional reduction, as detailed in [13], contribute to reducing the dataset for better model performance. The model selection and training steps involves methodologies recommended by [5], which highlights the advantages of specific machine learning and deep learning

techniques for plant disease classification. Model quality and fine-tuning strategies, as suggested by [10], ensure a robust and accurate classification process. The emphasis on results interpretation and deployment aligns with the practical insights derived from [7], which highlights the importance of continuous monitoring and integration into real-world applications. This methodology synthesis draws from a collaborative examination of various papers, identify optimal practices for effective plant disease classification.

4. RELATED WORKS

Plant disease classification follows a structured process with unique processing steps. It initiates with the careful selection and preprocessing of relevant datasets, involving resizing, contrast adjustment, and labelling for supervised learning. Image enhancement techniques, including erosion, dilation, and Gaussian filtering, are applied, followed by segmentation to localize disease regions, and subsequent cropping for targeted analysis. Feature extraction, encompassing colour histograms, texture features, and morphological parameters, contributes to refining the dataset. Feature selection and dimensional reduction are undertaken to identify the most discriminative features and optimize the computational load.

Method applied	Dataset used
Histogram of Oriented gradient (HOG), and local binary patterns (LBP)	Cassava disease dataset, UCI Rice diseases dataset, Citrus dataset
Weakly Supervised and Unsupervised learning, Transfer Learning	Data feature Weed dataset, weeds and cassava seedling DCD dataset
YOLO, SSD, and Faster R-CNN	PlantDoc, Agrivision, PlantClef
EfficientNetB7, VGG19 and MobileNetV2	Plant Village dataset
Grey level co-occurrence matrix (GLCM), CNN	Plant Village dataset, Saitama
NasNetLarge, Xception, ResNet152V2, CNN	Plant Village dataset
Phytopathometry, Standard Area Diagrams (SAD)	Plant Village dataset
Computer vision fluorescence imaging	Alexnet and GoogleNet
The DSSApple application	-
Deep learning techniques	ImageNet and OUFDD
K-Nearest Neighbour (KNN) and CNN	Mosaic virus (MV) and leaf mold (LM)
K-Nearest Neighbour (KNN) and CNN	Yellow curl (YC), Spotted Spider Mite (SPM) and Tobacco Spot (TS)
CNN, AlexNet, LeNet	Plant Village dataset
KNN, J48, Naïve bayes and Logistic Regression	UCI Machine Learning Repository
CNN	APS image Database
YOLOv3-tiny network	Eleusine indica testing datasets

Author	Yea	Disease
Jianping yao	2023	Bacterial leaf blight, Brown spot, and leaf
Wen Zhang	2023	Weed
Muhammad	2023	Tomato diseases, and
Yogesh	2023	Bacterial spot, early
P.K.	2023	Early and late blight,
C.K.Sunil	2023	Early blight and late
Clive	2023	Early blight and late
Sharada P	2016	Early blight and late
Gabriele	2023	Decayed fruit disease,
Biyun Yang	2023	The fungal diseased peaches in the slight,
Sunil S	2023	early blight, brown spot ,
C.Jackulin	2023	late blight, Bacterial leaf
Kakarla	2023	Apple scab and tomato
C.Jackulin	2023	Brown spots, Bacterial
Javaid Ahmad Wani	2023	Narrow leaf spot, sheath blight, Anthracnose,
Shaun	2023	Weed

5. LITERATURE SURVEY

The approaches in this paper [1] includes comprehensive evaluation from 2023 focuses on the identification of plant diseases in a number of areas, such as leaf mold, bacterial spot, early blight, and late blight. The PlantVillage dataset, a reputable source in the field of plant pathology, was selected for this investigation. Kumar explores the effectiveness of various deep learning-based methods for disease detection, including sophisticated models like MobileNetV2, VGG19, and EfficientNetB7.

This paper [2] has reviewed sophisticated deep learning models for plant disease detection, with a particular emphasis on peach leaf bacteria and tomato illnesses. PlantDoc, AgriVision, and PlantClef are just a few of the datasets included in the review, demonstrating the variety of sources available for model evaluation and training. The fact that 2023 is mentioned as the review year emphasizes how current the study is. Shoaib investigates state-of-the-art techniques for plant disease identification, using Faster R-CNN (Region-based Convolutional Neural Network), SSD (Single Shot Multibox Detector), and Yolo (You Only Look Once).

The recent paper [4] study from 2023 focuses on the identification and categorization of several plant diseases, including brown rust, bacterial leaf blight (BLB), leaf smut, early and late blight, and brown spot. The well-known PlantVillage and Saitama Research Center datasets, which offer a varied selection of samples for training and assessment, are among the datasets used in this study. Midhunraj uses a dual strategy, combining more sophisticated approaches like convolutional neural networks (CNNs) with more conventional machine learning techniques like the grey level co-occurrence matrix (GLCM)

Clive H. Bock's evaluation [5] from 2021 focuses on early and late blight in particular and emphasizes the visual assessment of plant disease severity. The PlantVillage dataset was selected for this analysis, underscoring its value as a standard for plant pathology research. Bock explores a century of study, explaining how methods for visually judging the severity of plant diseases have evolved historically, what works, and how to improve them going forward. The paper introduces Standard Area Diagrams (SADs), a tool frequently used as a reference for visually determining disease severity, and phytopathometry, a method probably utilized for quantitative assessment.

The paper [6] looks to increase productivity and sustainability in agriculture, modern robotic technologies for precision weed management combine autonomous systems with cutting-edge sensors and machine learning algorithms. These cameras and LiDAR-equipped systems can autonomously identify and distinguish between weeds and crops. Additionally, they have precision spraying mechanisms that let robots apply herbicides only to specific weed spots, cutting down on chemical use and environmental impact. The "Data Feature Weed dataset," which consists of 800 x 600-pixel RGB photos of weeds and corn seedlings, is crucial for training machine learning algorithms to recognize weeds correctly. By streamlining weed control, reducing labor-intensive tasks, and optimizing herbicide application, these innovations hope to promote more environmentally friendly farming methods.

The paper [7] briefs about visual data, machine learning for leaf disease classification creates models that can recognize and classify plant illnesses. Tagged photos showing different plant species and the associated disease statuses are usually included in the datasets utilized in these applications. Because Convolutional Neural Networks (CNNs) automatically learn hierarchical characteristics from photos, they are a popular solution for this kind of assignment.

Deep learning-based methods for the identification and categorization of plant diseases—with a particular focus on early blight, late blight, and brown spot—are the subject of C. K. Sunil's 2023 comprehensive study [8]. The PlantVillage dataset, a popular resource in plant pathology research, was selected for this investigation. Many advanced deep learning models, such as NasNetLarge, Xception, ResNet152V2, EfficientNetB5, EfficientNetB7, VGG19, and MobileNetV2, are used by Sunil.

DSSApple, a hybrid expert system created by Gabriele Sottocornola[12] in 2023, is intended to diagnose post-harvest diseases that impact apples, with a focus on decayed fruit disease and post-harvest apple disease. The DSSApple application, which offers a sophisticated tool for the prompt and accurate identification of diseases affecting harvested apples, marks a significant leap in agricultural technology. The expert system's hybrid design implies the use of both rule-based and possibly machine learning techniques, enabling a thorough and sophisticated assessment of the signs and patterns connected to post-harvest illnesses.

The study [13] helps to create models that can recognize and classify plant illnesses, machine learning for leaf disease classification uses visual data. Typically, the datasets utilized in these applications are collections of labelled photos that show different plant species and the associated disease states. Because they can automatically extract hierarchical information from photos, Convolutional Neural Networks (CNNs) are a popular method for this kind of task.

Javaid Wani [14] carried out a thorough investigation of machine learning and deep learning-based computational algorithms for the detection of automatic agricultural illnesses in 2021. The study focused on a number of diseases, such as early blight, block mold, sheath blight, anthracnose, and narrow leaf spot." The study's dataset, which features a carefully chosen selection of photos created especially for plant pathology research, came from the APS Image Database. Wani's strategy was based mostly on the use of Convolutional Neural Networks (CNN), a potent deep learning technique well-known for its effectiveness in picture categorization exercises.

Shaun M. Sharpe's [16] concentrated on using a convolutional neural network (CNN) to detect goosegrass in tomato and strawberry crops in 2020. *Eleusine indica* was the target weed for identification in this study, and testing datasets were used to evaluate the generated model's performance. The You Only Look Once (YOLO) algorithm's YOLOv3-tiny network variant, intended for real-time object recognition with less computing cost, was used in the methodology. The study sought to improve the efficacy and precision of

goosegrass detection in agricultural environments, namely in tomato and strawberry fields, by utilizing this CNN-based methodology. The findings probably provide light on how well the YOLOv3-tiny network works for real-time weed detection and presented a promising avenue for the development of automated systems to manage and mitigate the impact of weeds on crop yields.

6. CONCLUSION

Upon analysing numerous articles on crop disease detection, it is apparent that significant improvements have been achieved in utilising cutting-edge technology, like machine learning, image processing, and sensor networks, to promptly and precisely identify crop illnesses. Promising outcomes have been observed in automating the detection process, enabling early diagnosis, and providing rapid intervention to decrease crop losses through the combination of various technologies.

Using deep learning methods, especially convolutional neural networks (CNNs), for image-based disease identification is one prevalent trend that has been noted. These models have proven to be highly accurate and reliable in identifying tiny patterns connected to a range of crop illnesses. The utilisation of drone technology, hyper-spectral data, and spectral imaging has additionally improved the accuracy and efficacy of disease identification over extensive agricultural regions. Despite the progress made, certain challenges persist, such as the need for extensive labelled datasets, model generalisation across diverse environments and crop types, and the development of real-time monitoring systems. Future research should focus on addressing these challenges and exploring innovative solutions to enhance the scalability and applicability of crop disease detection technologies.

REFERENCES

1. Kumar, Y. (2023). A Systematic Review of Different Categories of Plant Disease Detection Using Deep Learning-Based Approaches. *International Center for Numerical Methods in Engineering (CIMNE)*, 30:4757–4779.
2. Shoaib M, Shah B, EI-Sappagh S, Ali A, Ullah A, Alenezi F, Gechev T, Hussain T and Ali F (2023) An advanced deep learning models-based plant disease detection: A review of recent research. *Front. Plant Sci.* 14:1158933. doi: 10.3389/fpls.2023.1158933
3. Singh, B.K., Delgado-Baquerizo, M., Egidi, E. et al. Climate change impacts on plant pathogens, food security and paths forward. *Nat Rev Microbiol* 21, 640–656 (2023). <https://doi.org/10.1038/s41579-023-00900>
4. Midhunraj, P.K., Thivya, K.S. & Anand, M. An Analysis of Plant Diseases on Detection and Classification: From Machine Learning to Deep Learning Techniques. *Multimed Tools Appl* (2023). <https://doi.org/10.1007/s11042-023-17600-2>
5. Bock, C.H., Chiang, K.S. & Del Ponte, E.M. Plant disease severity estimated visually: a century of research, best practices, and opportunities for improving methods and practices to maximize accuracy. *Trop. plant pathol.* 47, 25–42 (2022). <https://doi.org/10.1007/s40858-021-00439-z>
6. Zhang, W., Miao, Z., Li, N. et al. Review of Current Robotic Approaches for Precision Weed Management. *Curr Robot Rep* 3, 139–151 (2022). <https://doi.org/10.1007/s43154-022-00086-5>
7. Yao, J., Tran, S.N., Sawyer, S. et al. Machine learning for leaf disease classification: data, techniques and applications. *Artif Intell Rev* 56 (Suppl3),3571–3616(2023). <https://doi.org/10.1007/s10462-023-10610-4>

8. Sunil, C.K., Jaidhar, C.D. & Patil, N. Systematic study on deep learning-based plant disease detection or classification. *Artif Intell Rev* 56, 14955–15052 (2023). <https://doi.org/10.1007/s10462-023-10517-0>
9. Action using computational intelligence and image processing. *J Plant Dis Prot* 128, 19–53 (2021). <https://doi.org/10.1007/s41348-020-00368-0>
10. Sottocornola, G., & Baric, S. (2023). DSSApple: A hybrid expert system for the diagnosis of post-harvest diseases of apple. *Elsevier*, 100070((2023)100070), 22. <https://doi.org/10.3389/fpls.2023.1158933>
11. Tyagi, K., & Karmarkar, A. (2020). Crop Health Monitoring System. *International Conference for Emerging Technology (INCET)*, 978-1-7281-6221–8.
12. Sunil S. Harakannanavar, Jayashri M. Rudagi, Veena I Puranikmath, Ayesha Siddiqua, R Pramodhini, Plant leaf disease detection using computer vision and machine learning algorithms, *Global Transitions Proceedings*, Volume 3, Issue 1, 2022, Pages 305-310, ISSN 2666-285X, <https://doi.org/10.1016/j.gltip.2022.03.016>.
13. Wani, J.A., Sharma, S., Muzamil, M. et al. Machine Learning and Deep Learning Based Computational Techniques in Automatic Agricultural Diseases Detection: Methodologies, Applications, and Challenges. *Arch Computat Methods Eng* 29, 641–677 (2022). <https://doi.org/10.1007/s11831-021-09588-5>
14. C Jackulin, S. Murugavalli, A comprehensive review on detection of plant disease using machine learning, and deep learning approaches, *Measurement: Sensors*, Volume 24, 2022, 100441, ISSN 26659174, <https://doi.org/10.1016/j.measen.2022.100441>.
15. Rajpoot, V., Tiwari, A. & Jalal, A.S. Automatic early detection of rice leaf diseases using hybrid deep learning and machine learning methods. *Multimed Tools Appl* 82, 36091–36117 (2023). <https://doi.org/10.1007/s11042-023-14969-y>
16. Deepti, K. Comparative Analysis of Machine Learning Techniques for Plant Disease Detection-Data Deployment. *J. Inst. Eng. India Ser. B* 104, 837–849 (2023). <https://doi.org/10.1007/s40031-023-00897-w>.