International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Machine Learning-based Detection and Extraction of Crop Diseases: A Comprehensive Study on Disease Patterns for Precision Agriculture

Prasad Chaudhari[1], Ritesh V. Patil[2], Parikshit N. Mahalle[3]

Submitted: 24/11/2023 **Revised**: 29/12/2023 **Accepted**: 10/01/2024

Abstract: This paper presents a comprehensive study on the application of machine learning for the detection and extraction of crop diseases, with a focus on understanding disease patterns in the context of precision agriculture. The research explores the integration of advanced machine learning techniques, particularly Convolutional Neural Networks (CNNs), for accurate and efficient identification of crop diseases. The study encompasses an extensive literature review, surveying the evolution of machine learning applications in agriculture, and critically examines the effectiveness of these methods in addressing the challenges associated with traditional disease detection methods. The proposed research investigates diverse crop disease patterns, leveraging state-of-the-art machine learning architectures to enhance the precision of disease identification. Through an in-depth comparative analysis, we assess the performance of machine learning models against traditional methods, shedding light on the advancements and limitations in the field. Furthermore, the study explores the potential of transfer learning, data augmentation, and interpretable machine learning techniques to improve the robustness and interpretability of disease detection models. This research contributes to the growing body of knowledge in precision agriculture, offering insights that can inform future research directions and practical implementations in the quest for sustainable and optimized crop management.

Keywords: *Precision Agriculture, Crop Disease Detection, Machine Learning, Convolutional Neural Networks, Disease Patterns, Transfer Learning, Sustainable Agriculture.*

1. Introduction

In the realm of precision agriculture, the accurate identification and timely extraction of crop diseases play a pivotal role in ensuring optimal yield and resource utilization. Traditional methods of disease detection often prove inadequate, necessitating a shift towards innovative technologies. This survey explores the application of machine learning techniques in the detection and extraction of crop diseases, offering a comprehensive examination of disease patterns[1]. By synthesizing advancements in image processing, sensor-based approaches, and data-driven models, this study aims to provide valuable insights for the enhancement of precision agriculture practices.

1.1 Background

Ensuring crop health and optimising production provide major problems for the agriculture sector. A significant hazard, crop diseases can cause significant financial losses as well as negative effects on the environment[2]. Conventional illness detection techniques frequently fail to deliver precise and timely information. Using machine learning presents a viable path for accurate crop disease extraction and early diagnosis, especially when combined with cutting-edge image and sensor technology[3]. The purpose of this project is to investigate and assess how well machine learning approaches work to address these issues and further precision agriculture.

[1]*Research Scholar at SKNCOE Research Center, Pune*
[2]*SKNCOE Research Center, Pune*
[1]*Vishwakarma Institute of Information Technology, Pune, Maharashtra, India*
[2]*PDEA's College of Engineering, Pune, Maharashtra, India*
[3]*Vishwakarma Institute of Information Technology, Pune, Maharashtra, India*
[1]*prasad.chaudhari29@gmail.com*
[2]*rvpatil3475@yahoo.com*
[3]*parikshit.mahalle@viit.ac.in*

1.2 Importance of Crop Disease Detection in Precision Agriculture

Crop disease detection has a diverse role in precision agriculture, encompassing a range of factors that enhance the overall productivity, sustainability, and efficiency of agricultural methods[4]. The following main ideas emphasise how important crop disease detection is to precision agriculture:

- **Early Detection and Intervention:**

Early detection of crop illnesses enables timely action, stopping the disease's spread across the crop population. By taking early action, farmers can minimise potential produce losses by implementing targeted and timely interventions like removing damaged plants, modifying irrigation, or spraying pesticides.

- **Optimized Resource Management:**

The accurate and effective use of resources, such as water, fertilisers, and pesticides, is essential to precision agriculture. Farmers may apply resources more efficiently by accurately detecting diseases, ensuring that medicines are only provided where and when they are required. This minimises wasteful resource use, cutting expenses and negative environmental effects.

- **Increased Crop Yield and Quality:**

Crop quality is generally improved and high crop yields are maintained when crop diseases are promptly identified and managed. Plants can grow, develop, and fully produce high-quality harvests possible if illnesses are prevented or their effects are lessened.

- **Economic Sustainability:**

Crop diseases can have a significant negative financial impact on farmers' livelihoods and the agricultural sector. Precision agriculture helps farmers safeguard their financial investments, sustain profitability, and contribute to the stability of the agricultural industry by reducing production losses through early disease diagnosis.

- **Data-Driven Decision-Making:**

Farmers can obtain useful information and insights from disease detection systems based on machine learning. Making decisions based on data enables farmers to choose crop management plans with knowledge, resulting in more productive and sustainable farming operations.

- **Reduced Environmental Impact:**

Minimising the impact of farming operations on the environment is the goal of precision agriculture. The potential harm to ecosystems and biodiversity is lessened when targeted disease identification and treatment eliminate the need for agrochemicals to be applied widely.

- **Enhanced Sustainability and Food Security:**

Precision agriculture supports sustainable agricultural practices by reducing the effects of crop diseases. To satisfy the growing demand for food production to support expanding populations, it is imperative to ensure the health and productivity of crops to contribute to global food security.

1.3 Objectives of the Survey

This survey papers major goals are to thoroughly investigate and assess the most recent machine learning algorithms for crop disease extraction and detection, with a focus on applications in the field of precision agriculture. The purpose of the survey is to give a comprehensive understanding of the different kinds of crop diseases that impact agricultural production and, in turn, crop quality and yield. Through an analysis of conventional techniques for illness detection, such as visual inspection, laboratory testing, and remote sensing, the survey will lay the groundwork for a fundamental comprehension of current methods and their shortcomings.

In addition, the study will explore the developments and uses of machine learning in this field, with particular attention on data-driven models, computer vision, and image processing. The objective is to assess how successfully deep learning strategies, along with supervised and unsupervised learning techniques, identify and categorise agricultural diseases. The survey attempts to shed light on the advantages and disadvantages of the machine learning techniques that are currently in use by examining pertinent datasets, assessment measures, and case studies. The survey will also cover future trends and emerging technologies that have the potential to improve crop disease detection accuracy and efficiency in the context of precision agriculture, as well as challenges encountered in the annotation and collection of datasets for machine learning model training. In the end, the study looks for areas where research is lacking and where it can be strengthened, which will help to continue developing reliable and useful solutions for crop disease detection.

1.4 Scope and Limitations

This research project aims to investigate the application of machine learning methods in crop disease diagnosis and extraction, with a focus on precision agriculture. The study aims to examine several crop illnesses that are commonly observed in agriculture and analyse the potential of machine learning methods, namely those that rely on image processing, computer vision, and data-driven models, to enhance the precision and promptness of disease detection. In the context of precision agriculture techniques, the paper seeks to present a comprehensive examination of several machine learning techniques, their suitability for handling a range of datasets, and their potential influence on enhancing crop output and quality.

It is imperative to recognise specific limitations that fall within the purview of this study. First, the quality and accessibility of datasets may have an impact on how effective machine learning models are. The ability of the models to be applied to various crop types or geographical areas may be impacted by incomplete or biassed datasets. Second, while the study will concentrate on prevalent and important diseases affecting major crops, it might not cover every potential crop disease. Thirdly, because precision agriculture procedures require a certain amount of technology infrastructure to be implemented, the research's conclusions could not be directly applicable to areas or farmers that have limited access to cutting-edge technologies. Notwithstanding these drawbacks, the study intends to highlight opportunities for further investigation and development as well as offer insightful information about the use of machine learning for crop disease diagnosis in precision agriculture.

2. Types of Crop Diseases

Crop diseases have an impact on the quantity and quality of agricultural products, which is a danger to global food security. Numerous pathogens, including as nematodes, bacteria, viruses, and fungus, can cause these illnesses[5]. Russets, blights, and mildews are examples of common fungal diseases that can affect a variety of crops. Large crop losses can also result from bacterial diseases like leaf spot and bacterial wilt. It is known that viruses can infect crops including maize, tomatoes, and potatoes. These viruses are spread by insects or other plants. Microscopic worms called nematodes that live in the soil can harm plants by causing root infections. Detecting and managing crop diseases can be difficult for farmers due to their distinct features, symptoms, and means of transmission.

Crop diseases have an effect that goes beyond lower yields and monetary losses. An increase in the usage of fungicides and pesticides during disease outbreaks can have negative effects on the environment and human health. Furthermore, the transmission of these illnesses may have catastrophic effects on regional and worldwide food resources, thereby making food poverty worse. With the use of machine learning algorithms, precision agriculture presents a viable way to deal with these issues by reducing the need for chemical inputs, enabling timely intervention, and enabling early and accurate identification of crop diseases[6]. Comprehending the wide array of crop diseases is imperative to construct efficacious machine learning models that may facilitate resilient and sustainable farming methodologies.

2.1 Overview of Common Crop Diseases

Common crop diseases have a major impact on crop productivity, quality, and global food security, endangering agriculture worldwide. These illnesses, which can have a significant negative impact on the environment and economy, are brought on by a variety of pathogens, including bacteria, viruses, fungus, and nematodes. Powdery mildew, rusts, and blights are examples of fungal infections, which are a common type of crop disease. These illnesses can affect a variety of crops, such as cereals, fruits, and vegetables, and they frequently flourish in warm, humid environments. Bacterial infections can cause wilting, necrosis, and general plant loss in crops. Examples of these diseases are bacterial wilt and fire blight. Viral infections are another frequent hazard to crops like maize, tomatoes, and potatoes. They can cause symptoms like discoloured leaves, mosaic patterns, and reduced growth. Globalisation, climate change, and monoculture practices are some of the elements that contribute to the development of these illnesses, thus it's critical to create efficient methods for early identification and treatment. To apply precision agriculture techniques that maximise resource efficiency and reduce the impact of pesticides, it is essential to recognise and comprehend the patterns of these diseases. Machine learning has become a promising technique for the early identification and diagnosis of agricultural diseases, having the potential to completely change how farmers monitor and protect their crops. It does this by analysing enormous information and recognising complex patterns.

2.2 Impact on Crop Yield and Quality

In the field of precision agriculture, the application of machine learning-based methods for crop disease diagnosis and extraction has a significant effect on crop output and quality. Through the quick and precise detection of diseases in their early stages, these technologies allow farmers to take immediate action to minimise the risk of crop damage[7]. Early diagnosis contributes to the best possible use of resources, such as the accurate dosing of insecticides, fertilisers, and water, which improves crop health and eventually boosts production.

Moreover, the capacity to precisely identify illnesses directly contributes to the increase in crop quality. By allowing for the differentiation of different illness kinds, machine learning models minimise the need for broad-spectrum medicines and enable targeted treatments. This focused strategy helps produce high-quality crops while simultaneously lessening the negative effects of farming methods on the environment. By implementing these technologies, farmers can make data-driven decisions, improve their farming methods, and guarantee a productive and sustainable agricultural system that optimises yield and quality. All things considered, one of the main factors propelling the broad adoption of machine learning in contemporary agricultural operations is its beneficial effects on crop productivity and quality.

2.3 Economic and Environmental Consequences

Crop diseases have significant negative effects on the environment and economy, which emphasises how important it is for precision agriculture to have efficient ways for detecting and extracting information. Crop diseases have a direct financial influence on farmers' incomes and the viability of the agricultural industry as a whole by causing significant losses in agricultural productivity and quality[8]. Lower yields not only mean less crops available for consumption, but they also cause havoc on regional and international markets, causing price swings and possible food shortages. Farmers' financial burden is further increased by expenses related to disease management, such as the cost of insecticides and fungicides.

The use of chemical treatments to control agricultural diseases presents considerable concerns from an environmental standpoint. Pesticide use on a large scale has the potential to contaminate soil and water, which impacts not just the surrounding agricultural area but also ecosystems located outside of farm boundaries. Furthermore, the effects on the environment also include the emergence of disease strains resistant to pesticides, which starts a vicious cycle that calls for the use of stronger and perhaps dangerous chemicals[9]. Precision agriculture can alleviate these economic and environmental effects by enabling focused interventions, lowering the requirement for widespread chemical use, and promoting sustainable farming methods. Precision agriculture is made possible by machine learning-based disease identification and extraction. By providing insights into the economic and environmental effects of crop diseases and suggesting creative remedies through machine learning applications in precision agriculture, this extensive study seeks to solve these intricate dynamics.

3. Traditional Methods for Crop Disease Detection

Traditional Methods for Crop Disease Detection:

3.1 Visual Inspection:

For a very long time, a key technique for identifying crop diseases has been visual inspection. This conventional method entails farmers and agricultural experts making direct observations of the physical characteristics of plants. The following is a thorough explanation of the visual inspection process used to identify crop diseases:

3.1.1 Measurement parameters:

- **Discoloration:** Variations in the hue of leaves, stems, or other plant components may be a sign of several different illnesses. For instance, browning or yellowing of the leaves may indicate fungus infections or dietary deficits.
- **Lesions:** The existence of lesions, spots, or odd markings on plant surfaces could indicate a bacterial, fungal, or viral infection.
- **Wilting:** A number of conditions, including vascular illnesses, bacterial wilt, and water stress, can cause plants to droop or wilt.
- **Aberrant Growth Patterns:** Certain disorders that impact plant development may be associated with abnormal growth patterns, abnormalities, or stunted growth.

3.1.2 Sampling Techniques:

- Samples of plants exhibiting symptoms are frequently collected by farmers and experts for additional analysis. Taking leaves, stems, or other afflicted plant parts may be necessary for this.
- Sampling is essential for gaining a thorough grasp of the features of the disease and facilitates precise identification.

3.1.3 Field Surveys:

- Systematic field surveys are usually used to do visual inspections. As they move through fields, specialists note their observations while examining crops at different stages of growth.
- Periodic surveys may be conducted, particularly during critical growth stages or when environmental factors are conducive to the development of disease.

3.1.4 Training and Experience:

- Farmers or specialists with the necessary expertise and experience are essential for a successful visual evaluation. They must be knowledgeable about the signs and symptoms of various illnesses.
- It takes ongoing training to keep people informed about new illnesses, symptom variations, and crop hazards.

3.1.5 Challenges of Subjectivity:

- The subjective nature of visual assessment is its main disadvantage. Individual differences in symptom interpretation may result in inaccurate information.
- Disease misidentification can happen, particularly when symptoms are caused by several sources or overlap.

3.2 Laboratory Testing:

Laboratory testing is a common component of traditional crop disease diagnosis approaches; this is a thorough but resource-intensive approach. Plant parts, like leaves or stems, have to be physically cut and sent to a specialised lab in order to collect samples. The time-consuming nature of this approach could cause delays before useful information is acquired. Enzyme-Linked Immunosorbent Assay (ELISA) and Polymerase Chain Reaction (PCR) are two frequently used laboratory procedures for illness identification. A molecular biology method called PCR amplifies DNA sequences, making it possible to identify certain diseases in plant samples. In contrast, ELISA finds antigens or antibodies and provides information on the existence of specific diseases. These techniques are accurate, but they take a lot of time, specialised staff, and equipment, so they are not as good for quick on-site detection.

Furthermore, the inability to receive laboratory test results quickly could make it more difficult to take prompt action to stop the spread of illness. Current knowledge is essential for agricultural operations to make well-informed decisions, and crop management may not require laboratory testing's longer turnaround time. To enable timely intervention and reduce the impact of illnesses on crop yield and quality, there is an increasing demand for more effective and real-time crop disease detection techniques.

3.3 Remote Sensing Techniques:

With the ability to monitor plant health on a broader scale without requiring invasive methods, remote sensing techniques have emerged as significant instruments in the identification of classical crop diseases. The process of gathering data about things or locations without making direct physical contact is known as remote sensing. In the field of agriculture, sensors based on satellites or drones are frequently used to record different electromagnetic radiation wavelengths that crops reflect or emit. Potential disease outbreaks can be identified thanks to these sensors' ability to identify even the smallest changes in plant health. The capacity to deliver information in real-time or almost real-time is a major benefit of remote sensing for crop disease identification. This is quite different from how time-consuming laboratory testing is. Researchers and farmers can identify irregularities linked to illnesses, dietary inadequacies, or other stressors by examining the spectral signatures of crops. Remote sensing is a scalable approach for early detection and intervention because it also makes larger agricultural regions monitorable. Although it is an effective technique, remote sensing is not without its limitations. It takes experience to interpret data, and it can be difficult to discern between various stress variables that affect plant health. Crop disease detection systems based on remote sensing can be made more accurate and dependable by integrating them with other technologies and data sources, such as machine learning algorithms or ground-truthing. This allows for a more thorough approach to precision agriculture.

3.4 Sensor-based Approaches:

Compared with conventional laboratory testing, sensor-based methods provide an on-site, faster way to detect crop diseases. These techniques gather information straight from the crops in the field using a variety of sensors, including spectroscopy, thermal imaging, and hyperspectral imaging. For example, spectroscopy examines how electromagnetic radiation interacts with plant tissue to reveal important details about the crop's health. Comparably, temperature fluctuations that can point to the presence of illnesses are picked up by thermal imaging. Real-time monitoring is a benefit of sensor-based methods, which may be implemented on-site and do away with the requirement for labour-intensive laboratory analysis and sample transfer. Farmers can take pre-emptive steps to lessen the impact of diseases on their crops because of the speedy diagnosis and response to possible outbreaks made possible by the immediate data collecting. Precision farming techniques for increased crop health and output are made possible by these approaches, which also help to manage diseases in agriculture more effectively and promptly.

3.5 Challenges with Traditional Methods:

- **Subjectivity:** Human mistake and variance in interpretation are common in visual inspection.
- **Time-consuming:** Transporting and processing samples for laboratory testing takes time, which delays prompt results.
- **Restricted Coverage:** Sensor-based and remote sensing methods may not provide the fine resolution required for precise disease detection.
- **Resource Intensity:** Lab testing and visual inspection can both be highly resource-intensive, requiring specialised tools and knowledge.

4. Literature Survey

The paper entitled "A systematic analysis of machine learning and deep learning based approaches for identifying and diagnosing plant diseases" by Imtiaz Ahmed and Pramod Kumar[10] Yadav delve into the existing body of research to comprehensively explore the applications, methodologies, and advancements in utilizing ML and DL techniques in the context of plant disease identification. The literature review probably looks at how ML and DL models have been used to improve the precision and effectiveness of disease diagnosis across a broad range of plant diseases in different crops. Several phases of the pipeline for identifying diseases, such as data collection, preprocessing, feature extraction, and model training and assessment, may be included in this analysis. In addition, the writers could point out the benefits and drawbacks of various ML and DL methodologies, providing insight into the difficulties experienced by scholars in this field. The survey should shed light on how various algorithms, structures, and datasets performed in comparison to one another in the research that were found. Furthermore, it can go over new developments, possible research gaps, and future paths for enhancing the scalability and robustness of ML and DL models for plant disease diagnosis. The systematic analysis's overall goal is to add to the expanding corpus of knowledge on precision agriculture by providing insightful information to scholars, practitioners, and legislators who are tackling issues related to plant health.

The paper entitled "A comprehensive review on detection of plant disease using machine

learning and deep learning approaches" by C. Jackulin and S. Murugavalli[11] provides a thorough exploration of the evolving landscape in precision agriculture. The authors explore a broad range of research and approaches used in plant disease detection, emphasising the developments made possible by ML and DL techniques. The literature review covers a range of crops, discussing the difficulties that each present, and emphasises the revolutionary effect that these technologies have in reducing crop losses brought on by illnesses. Reviewing a wide range of ML and DL approaches, the review carefully groups the literature it has studied to provide a comprehensive understanding of the variety of methods available for image classification, feature extraction, and pattern recognition. The authors stress the importance of dataset quality, size, and diversity in impacting the resilience of the illness detection systems as they critically examine the benefits and drawbacks of various models and architectures. The report also emphasises how important it is to use data preprocessing methods and integrate cutting-edge sensors and imaging technology to improve the precision and effectiveness of disease diagnosis. The review adds to the body of knowledge by combining results from many research projects, providing a thorough grasp of the state-of-the-art in plant disease detection using ML & DL applications.

The paper entitled “Classification of plant diseases using machine and deep learning” by Monika Lamba et al[12] explores the classification of plant diseases using both traditional machine learning and deep learning techniques. The literature survey likely delves into existing research and methodologies applied in the realm of plant disease classification, presenting an overview of the current state of the field. The authors of this study may go over a variety of methods used in illness diagnosis, from more sophisticated approaches like machine learning and deep learning to more conventional ways like image processing and feature extraction. Conventional techniques could involve identifying visual signs by hand or highlighting disease patterns with image processing algorithms. However, the approaches of machine learning and deep learning—which are probably at the heart of the survey—involve training models on big datasets to automatically identify and categorise diseases based on visual characteristics. The significance of these methods in precision agriculture and their potential for early diagnosis and proactive control of plant diseases would probably be highlighted in the study. To prepare the way for their own contributions and approaches, the writers may also address difficulties and areas of unmet research need in the body of current literature. All things considered, this review of the literature provides a fundamental basis for comprehending the terrain of plant disease classification and lays the groundwork for the authors' contributions and original discoveries in their investigations.

The paper entitled “Machine Learning-Based Tea Leaf Disease Detection: A Comprehensive Review” by Faruk Ahmed et al[13] presents a detailed literature survey on the application of machine learning techniques for detecting diseases in tea leaves. By methodically examining previous research, techniques, and developments in the field, the study sheds insight on the various strategies used by scientists to tackle the difficulties involved in detecting tea leaf disease. Ahmed probably examines the machine learning models used in these investigations, from conventional algorithms to more sophisticated deep learning architectures, and evaluates how well they recognise and categorise different types of tea leaf illnesses. Because the review is so thorough, it looks at all the important aspects that affect how well disease detection models work, including the kinds of diseases that affect tea leaves that are considered, the qualities of the datasets that are used for training and assessment, and the difficulties that are particular to the field of tea plant pathology. To set the path for future research initiatives, Ahmed might also point out any gaps in the literature and explain the drawbacks of current methodologies. The literature review probably seeks to give scholars, practitioners, and stakeholders in precision agriculture and the sustainable management of tea plantations a useful resource by combining information from many sources.

The paper entitled “Machine Learning for Detection and Prediction of Crop Diseases and Pests: A Comprehensive Survey” by Tiago Domingues et al[14] aims to provide a thorough literature survey on the application of machine learning techniques in the agricultural domain for the detection and prediction of crop diseases and pests. The survey thoroughly examines the state of the art in research, emphasising significant developments, approaches, and difficulties found in applying machine learning to precision agriculture. It is possible that authors explore the many machine learning techniques used to detect trends and pinpoint pest or disease outbreaks in crops, including supervised learning, unsupervised learning, and deep learning. The sorts of datasets used for training and testing these models are likely covered in the study, with an emphasis on the significance of representative and diverse datasets for reliable and accurate predictions. The poll also probably covers the difficulties in applying machine learning models in actual agricultural environments, such as problems with data gathering, model interpretability, and incorporating new technologies into current farming methods. The literature review may also discuss the possible drawbacks and advantages of using machine learning to precision agriculture, such as better crop yields, less dependency on chemical treatments, and more effective disease and insect management. The writers most likely point out the gaps in the existing literature, highlighting topics that need more investigation and work. Overall, by bringing together existing knowledge and promoting a clearer understanding of the uses, difficulties, and potential future directions of machine learning in crop disease and pest prediction, this thorough analysis advances the rapidly developing field of agricultural technology.

The paper entitled “Machine Learning Applications for Precision Agriculture: A Comprehensive Review” by Abhinav Sharma et al[15] presents a thorough literature survey on the use of machine learning in the context of precision agriculture. Precision agriculture involves the application of advanced technologies to optimize farming practices, improve crop yields, and minimize environmental impact. The author dive into many facets of machine learning applications in this field, investigating how these technologies support farmers' and stakeholders' data-driven decision-making processes. A wide range of subjects are covered in the literature review, from predictive modelling for disease diagnosis and yield prediction to remote sensing and image analysis for crop monitoring. The writers conduct a thorough analysis of the current state of research, stressing important trends, approaches, and difficulties. The goal of the review is to present a thorough overview of the various ways that machine learning is being applied in precision agriculture, with a focus on how it has the potential to transform farming methods and improve agricultural sustainability. The authors add to the consolidation of knowledge in this quickly developing subject by synthesising data from a wide range of studies. Their insights can guide future research paths and practical applications in precision agriculture.

Understanding the current state of the industry and identifying areas for improvement requires a thorough investigation of machine learning-based crop disease extraction and detection for precision agriculture. Using a gap analysis, one can identify areas in which further research or enhancements to current research and technology are required. To detect agricultural diseases, it is first important to evaluate the variety of machine learning algorithms and methodologies used. Examine these approaches' advantages and disadvantages, considering aspects like precision, scalability, and environmental and crop-specific adaptability. This analysis can point out areas where algorithms perform poorly and direct the creation of stronger models. Examine the datasets that are utilised for machine learning model testing and training, second. Examine how well these datasets represent the variety of crop diseases and environmental influences. Finding holes in the availability of thorough datasets might draw attention to the requirement of conducting data gathering activities in particular areas, for crops, or environmental circumstances.

Additionally, assess how feasible and scalable the current machine learning models are for in-field use. Think about things like processing speed in real time, energy efficiency, and computing needs. More useful solutions for farmers can be developed by identifying gaps in the deployment and integration of machine learning models in precision agriculture systems. Lastly, examine how much the research that have already been done emphasise interdisciplinary cooperation. The convergence of data science, computer science, and agronomy is known as precision agriculture. Evaluate how well researchers, agronomists, and IT developers collaborate. To close the knowledge and application gaps across these domains, future research initiatives aimed at identifying gaps in interdisciplinary collaboration can provide guidance.

In conclusion, algorithmic performance, dataset representativeness, scalability, and interdisciplinary cooperation should all be included in a comprehensive gap analysis for machine learning-based crop disease diagnosis. Closing these gaps can result in more broadly applicable and efficient precision agricultural technologies, which will enhance crop output and health.

5. Machine Learning Applications in Crop Disease Detection

5.1 Image Processing and Computer Vision

Image processing and computer vision are examined in "Machine Learning Applications in Crop Disease Detection: Image Processing and Computer Vision" to see how these techniques can significantly improve the precision and efficacy of crop disease identification. Researchers and professionals have created creative ways to automate the detection and diagnosis procedure by utilising these technologies. When it comes to extracting features from photos of crops afflicted with illnesses, image processing techniques are essential[16]. This entails identifying and measuring visual cues that are symptomatic of diseases, such as colour changes, texture abnormalities, and structural deformities. Convolutional neural networks (CNNs), in particular, are highly effective in discovering and identifying patterns within these extracted data[17]. The combination of machine learning and image processing allows for the development of reliable models that can accurately identify sick from healthy crops. As a branch of machine learning, computer vision improves image interpretability to enable more in-depth analysis and crop health classification. CNNs and other deep learning models have shown impressive results in automating the identification of intricate and nuanced patterns linked to a variety of agricultural diseases[18]. The combination of computer vision and image processing speeds up the detection process and makes it possible to monitor large agricultural areas in real-time and on a scalable scale. Fundamentally, crop disease identification has undergone a paradigm shift with the combination of machine learning and image-based technologies, providing a more dependable and effective method for precision agriculture.

5.2 Sensor-based Approaches

The survey paper's "Sensor-based Approaches" section addresses how crop disease diagnosis in precision agriculture relies on the integration of sensors and probes. These methods involve the strategic placement of a variety of sensors in the field to gather data in real time on plant health and environmental conditions. The existence of agricultural diseases is then inferred using this data[19]. Crop monitoring can be made more dynamic and proactive with the help of sensor-based techniques. Numerous metrics, including soil moisture, temperature, humidity, and other pertinent environmental conditions, may be measured using these sensors. Modifications to these factors may function as oblique markers of possible disease epidemics. Deviations from ideal conditions, for example, may indicate stress on the plants and warrant additional research into the potential existence of illnesses. Large agricultural fields can be continuously monitored non-invasively with this technology, providing a more timely and thorough insight of the crop health condition. However, issues including the reliance on environmental proxies and the restricted sensitivity of sensor data to specific diseases require cautious interpretation of the findings[20]. The incorporation of machine learning algorithms into sensor-based data might improve disease detection accuracy by facilitating the recognition of complex patterns and correlations that might not be discernible through manual examination. In precision agriculture, the integration of sensor-based methods and machine learning techniques is a synergistic approach that offers a promising path towards more focused and efficient crop disease management.

5.3 Data-driven Models

In the context of machine learning applications for crop disease identification, "Data-driven Models" refer to the employment of many approaches to extract relevant patterns from agricultural data. Below is a brief overview of the three data-driven model subcategories:

5.3.1 Supervised Learning:

In supervised learning, a model is trained using a labelled dataset in which every input sample has a matching output, or target label. This approach allows the model to learn the association between input characteristics (such spectral signatures or environmental factors) and the corresponding illness labels in the context of crop disease detection. Support vector machines (SVM), random forests, and convolutional neural networks (CNNs) are examples of popular supervised learning methods[21]. Precision agriculture techniques and focused interventions are made possible by supervised learning's superior performance in classification problems, which enable the precise detection of certain crop diseases based on historical data.

5.3.2 Unsupervised Learning:

Conversely, unsupervised learning uses datasets without pre-established labels. Without explicit instruction, the algorithm finds patterns, clusters, or abnormalities in the data. Unsupervised learning techniques like dimensionality reduction (e.g., Principal Component Analysis) and clustering (e.g., K-Means) can uncover latent structures in the data for crop disease identification. Unsupervised methods are useful for grouping related cases without prior knowledge of diseases, conducting exploratory research, and discovering unique disease patterns. They are also a useful tool for comprehending the intricacy of crop health dynamics.

5.3.3 Deep Learning:

A subtype of machine learning known as "deep learning" is distinguished by the application of multilayered neural networks, or "deep neural networks." Deep learning models, such Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs), show impressive performance in crop disease diagnosis when it comes to feature extraction and pattern recognition from complicated and high-dimensional data, like pictures or time-series[22]. With its exceptional ability to capture complex relationships in the data, deep learning provides crop disease identification with cutting-edge precision. Deep learning models are especially useful for situations where the underlying patterns may be deeply nested or abstract because of their capacity to automatically learn hierarchical representations.

To summarise, data-driven models for crop disease detection use deep learning for complex feature extraction and classification, unsupervised learning for uncovering patterns, and supervised learning for labelled datasets. These methods increase the accuracy and productivity of agricultural disease monitoring and management.

Proposed Architecture:

A versatile and effective machine learning architecture is necessary for an extensive investigation on crop disease extraction and detection in precision agriculture. Convolutional Neural Networks (CNNs) have shown promise in image-based tasks; hence, a well-designed CNN architecture would be appropriate for this use case. This is a recommended architecture:

Input Layer:

- Accept RGB images of crops as inputs.

Convolutional Layers:

For each convolutional layer:

- Convolve the input with an increasing number of filters to capture hierarchical features.
- Apply batch normalization to normalize the activations.
- Apply Rectified Linear Units (ReLU) for non-linearity.

Pooling Layers:

For each pooling layer:

- Apply max pooling to reduce spatial dimensions, preserving important features.

Additional Convolutional Blocks:

For each convolutional block:

- Repeat the convolutional layers with batch normalization and ReLU activation.
- Utilize dropout layers to prevent overfitting.

Flatten Layer:

Flatten the output of the last convolutional block for input to fully connected layers.

Fully Connected Layers:

For each fully connected layer:

- Connect the flattened output to densely connected layers.
- Apply ReLU activation for non-linearity.
- Apply dropout for better generalization.

Output Layer:

- Apply a softmax activation for multi-class classification.
- Each class in the output layer represents a specific crop disease.

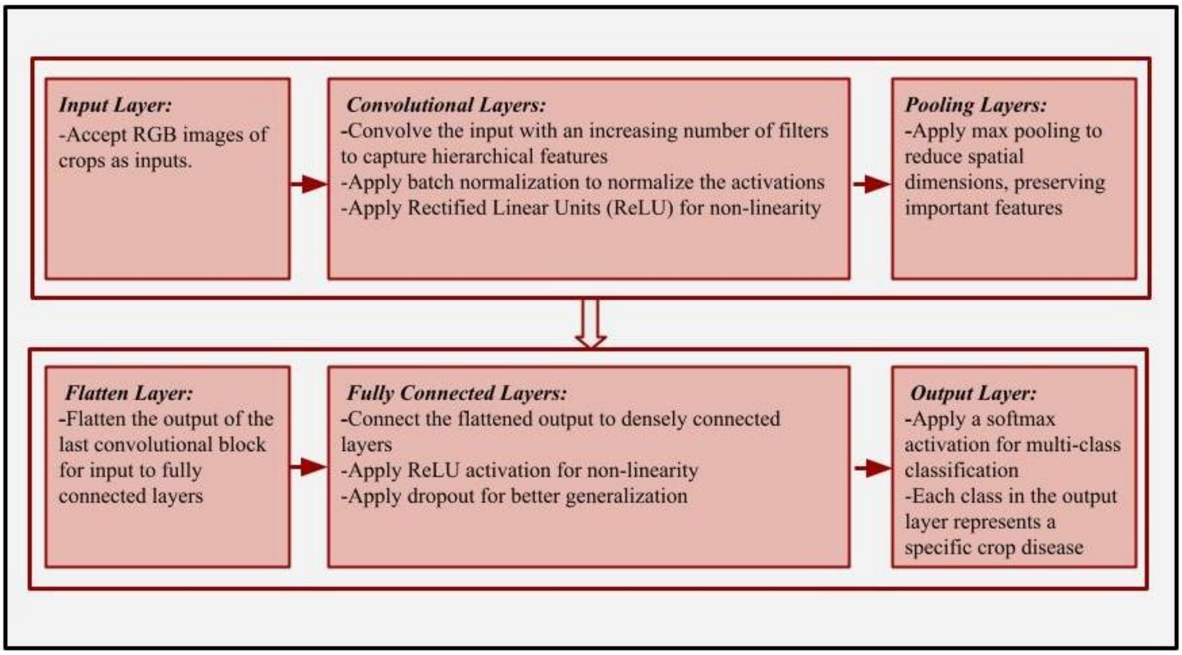

Fig 1: System Architecture for Crop Disease detection using Machine Learning

Key Considerations:

- Train the model using an appropriate optimisation technique (e.g., Adam, RMSProp).
- Use data augmentation methods to make the dataset appear more diverse than it is.
- To maximise model performance, play about with hyperparameters like dropout rates, number of filters, and learning rate.
- When data is scarce, think about combining transfer learning with pre-trained models to achieve better results.

The suggested CNN architecture for crop disease detection in precision agriculture can be implemented in an organised manner with the help of this step-by-step algorithm.

6. Datasets for Crop Disease Detection

6.1 Overview of Available Datasets

Table 1: Summary of available dataset for crop disease detection

Dataset Name	Description	Crop(s)	Image Type	Labels/Classes	Size	Availability
PlantVillage Dataset	Large collection of images covering various crops	Multiple crops	RGB images	Various diseases	54,000 images	Publicly available
Tomato Disease Detection	Dataset focused on tomato plant diseases	Tomato		Bacterial Spot, Early Blight, Late Blight, Healthy	100 images for training	Publicly available
Cassava Leaf Disease	Dataset for detecting diseases in cassava leaves	Cassava	RGB images	| Cassava Mosaic Disease, Cassava Bacterial Blight, Healthy	5,800 images	Publicly available
Paddy Crop Disease Dataset	Dataset for detecting diseases in paddy crops	Paddy	RGB images	Brown Spot, Hispa, Leaf Blast, Healthy	1,500 images	Publicly available
Grape Leaf Diseases Dataset	Dataset focusing on diseases affecting grape leaves	Grape	RGB images	Black Rot, Black Measles, Healthy	2,000 images	Publicly available
Apple Diseases Dataset	Dataset specifically for detecting apple diseases	Apple	RGB images	Apple Scab, Cedar Apple Rust, Healthy	2,000 images	Publicly available
Cotton Disease Dataset	Dataset for detecting diseases in cotton crops	Cotton	RGB images	Various cotton diseases	3,000 images	Publicly available
Potato Disease Dataset	Dataset for monitoring diseases in potato crops	Potato	RGB images	Late Blight, Early Blight, Healthy	1,200 images	Publicly available
Maize Disease Dataset	Dataset focusing on diseases affecting maize crops	Maize	RGB images	Rust, Northern Leaf Blight, Healthy	2,500 images	Publicly available

Table 2: Summary of available dataset for cotton crop disease detection

Dataset Name	**Cotton Disease Dataset**
Description	A curated collection of images focused on cotton crops to facilitate research and development in the field of cotton disease detection. The dataset encompasses various cotton diseases, capturing diverse symptoms and manifestations across different growth stages of the plants.
Crop(s)	Cotton
Image Type	RGB images
Labels/Classes	Multiple cotton diseases, including but not limited to Verticillium Wilt, Fusarium Wilt, Cotton Leaf Curl Virus, and Bacterial Blight. Additionally, a category for healthy cotton plants.

Size	Approximately 3,000 labeled images, representing a mix of disease-infected and healthy cotton plants.
Availability	Publicly available for research purposes. Researchers and practitioners can access the dataset for developing and evaluating machine learning models for cotton disease detection.
Dataset Name	Cotton Disease Dataset

6.2 Challenges in Dataset Collection and Annotation

Many obstacles must be overcome to gather and annotate agricultural disease datasets, which will affect the calibre and efficiency of machine learning models that are developed later. First of all, because agricultural techniques, crop varieties, and weather circumstances can vary greatly, gathering representative and varied datasets might be difficult[23]. Complicating matters further is the geographical and seasonal distribution of diseases, since some may only be common areas or at specific seasons. The problem is exacerbated by inconsistent data gathering procedures used by various sources, which may result in biases and restrict the generalizability of the dataset. Another set of difficulties arises from the process of annotation, which involves labelling images with ground truth data. Accurate disease identification requires expertise, and there might not be enough qualified annotators with subject-matter expertise. A comprehensive understanding is necessary to interpret subtle disease symptoms or changes in distinct crop growth stages[24]. Furthermore, variability may be introduced by the subjective character of some annotations, which could affect the dataset's dependability. The dynamic aspect of crop health poses additional difficulties in modelling real-world situations, such as the co-occurrence of many illnesses on a single plant or the evolution of diseases over time[25]. It is crucial to address these issues for reliable and useful machine learning models in agricultural contexts since creating extensive annotations while adhering to financial and temporal restrictions complicates the process of creating datasets.

6.3 Challenges and Opportunities

6.3.1 Challenges in Machine Learning-based Crop Disease Detection

Crop disease identification using machine learning has great potential to transform agricultural methods, but it is not without its difficulties. The requirement for solid and varied datasets is one of the main barriers. Acquiring substantial and representative datasets covering diverse crops, illnesses, and environmental circumstances is essential for efficiently training machine learning models[26]. Lack of well-annotated datasets can impede models' ability to generalise, which could result in biases and poor performance in practical settings. Additionally, because changes in soil types, temperatures, and farming techniques can dramatically influence disease manifestation, maintaining data integrity across various geographic regions and agricultural practices offers a challenging task.

The interpretability of machine learning models in the context of agriculture is another significant difficulty. The "black-box" character of some complicated models, especially deep learning architectures, might make it difficult for farmers to trust and comprehend how decisions are made. It is common for farmers to need interpretable models that offer insights into the characteristics impacting disease predictions. Successful adoption requires bridging the gap between the model's decision and practical insights for farmers. Moreover, the dynamic characteristics of agricultural diseases, which can change over time and show a variety of symptoms, complicate the process of creating machine learning models that are precise and flexible. To overcome these obstacles, researchers, agricultural specialists, and tech developers must work together to improve the quality of datasets, the interpretability of models, and the general applicability of machine learning in crop disease detection.

6.3.2 Emerging Technologies and Future Trends

Examining new technology and potential developments in machine learning-based crop disease diagnosis reveals a dynamic environment that could lead to major improvements in farming methods. The integration of sophisticated sensing technologies, including drones, satellite data, and hyperspectral and multispectral photography, is becoming more and more important as technology advances. These technologies provide a comprehensive view of crop health metrics by enabling more accurate and rapid data collecting. When combined with machine learning methods, these abundant datasets enable models to recognise faint spectral patterns that may be signs of illnesses, leading to improved precision in both detection and classification.

The development of interpretable and explainable models is another indicator of the direction that machine learning-based crop disease detection is taking. Models that can offer transparent insights into the decision-making process in addition to precise predictions are becoming more and more necessary as the agriculture sector depends more and more on data-driven decision-making. It is anticipated that explainable AI techniques—like model-agnostic methodologies and explainable neural networks—will become more popular. This will help stakeholders and farmers understand the reasoning behind the predictions and build confidence in the technology. Furthermore, it is projected that the cooperative integration of edge computing and Internet of Things (IoT) devices in agriculture will expedite real-time data processing, enabling prompt reactions to new disease threats and supporting the adoption of precision agriculture tactics. Fundamentally, the potential for revolutionising the detection, management, and mitigation of crop diseases through the convergence of cutting-edge technologies and machine learning might usher in a new era of agricultural efficiency and sustainability.

6.3.3 Research Gaps and Opportunities for Improvement

Although there has been a lot of development in using machine learning for this, the authors probably point out some difficulties that indicate knowledge gaps. These could include the restricted generalizability of models trained on crops or areas, as well as the need for more representative and diverse datasets that portray the intricacies of real-world agricultural circumstances. Refinement of interpretability and explain ability of the model is probably where there is the most room for improvement, as these are critical components that farmers and other agricultural stakeholders need to trust and use. To improve the robustness and accuracy of illness detection models, the authors may also discuss the necessity of integrating multi-modal data sources, such as merging satellite images with data from on-field sensors. Further investigation into the possibilities of cutting-edge technologies such as edge computing for on-device processing and in-the-moment decision-making could prove to be a beneficial subject of study. The next wave of developments in machine learning applications for crop disease detection will be guided by the identification of these research gaps and opportunities, opening the door to more useful and practical precision agricultural solutions.

7. Comparative Analysis

Comparison with Traditional Methods vs Comparative Analysis of Machine Learning Models

In the field of machine learning-based agricultural disease detection, evaluating the improvements over conventional techniques requires a comparative analysis. First off, machine learning models show a notable improvement in efficiency and accuracy over more conventional methods like visual inspection and lab testing. Conventional techniques are subjective and heavily dependent on human labour; in contrast, machine learning algorithms—especially those that use computer vision—can process large volumes of data quickly and objectively. This makes it possible to identify crop diseases more quickly and precisely, allowing for prompt actions to lessen their negative effects on yields.

The assessment of several machine learning models used in crop disease detection constitutes the second component of the comparative analysis. Support Vector Machines and Convolutional Neural Networks, two examples of supervised learning models, have proven their abilities in classification tasks by exhibiting excellent accuracy in detecting diseases based on labelled datasets. Without predetermined labels, unsupervised learning techniques such as clustering algorithms provide insights into patterns and anomalies within the data. Deep learning models have shown impressive effectiveness, particularly in image-based disease identification, because to their capacity to automatically learn complex information. The goal of the comparative analysis is to draw attention to the advantages and disadvantages of each strategy, assisting scholars and industry professionals in choosing the best models for their agricultural applications. Precision agriculture practices will continue to evolve based on this mix of traditional and modern methods for crop disease identification.

8. Conclusion

In conclusion, this study delves into the application of machine learning for crop disease detection and extraction in precision agriculture, with a particular focus on understanding disease patterns. Leveraging Convolutional Neural Networks (CNNs) and other advanced techniques, the research offers a comprehensive review of machine learning's evolution in agriculture. The comparative analysis with traditional methods emphasizes the advancements made, guiding future research. The study's exploration of transfer learning, data augmentation, and interpretable techniques contributes to improving model robustness and interpretability. Overall, this research provides valuable insights for sustainable and optimized crop management, contributing to the ongoing transformation of precision agriculture.

References

[1] A. Bouguettaya, H. Zarzour, A. Kechida, and A. M. Taberkit, “A survey on deep learning-based identification of plant and crop diseases from UAV-based aerial images,” *Cluster Comput.*, vol. 26, no. 2, pp. 1297–1317, 2023, doi: 10.1007/s10586-022-03627-x.

[2] A.-K. Mahlein, “Present and Future Trends in Plant Disease Detection,” *Plant Dis.*, vol. 100, no. 2, pp. 1–11, 2016, doi: 10.1007/s13398-014-0173-7.2.

[3] B. V. Gokulnath and G. Usha Devi, “A survey on plant disease prediction using machine learning and

deep learning techniques,” *Intel. Artif.*, vol. 23, no. 65, pp. 136–154, 2020, doi: 10.4114/intartif.vol23iss65pp136-154.

[4] J. Liu and X. Wang, “Plant diseases and pests detection based on deep learning: a review,” *Plant Methods*, vol. 17, no. 1, pp. 1–18, 2021, doi: 10.1186/s13007-021-00722-9.

[5] W. B. Demilie, “Plant disease detection and classification techniques: a comparative study of the performances,” *J. Big Data*, vol. 11, no. 1, 2024, doi: 10.1186/s40537-023-00863-9.

[6] B. J. Gorad and D. S. Kotrappa, “Comparative Study and Development of Disease Prediction System for Indian Crops,” *SSRN Electron. J.*, 2019, doi: 10.2139/ssrn.3350329.

[7] I. M. Nasir *et al.*, “Deep learning-based classification of fruit diseases: An application for precision agriculture,” *Comput. Mater. Contin.*, vol. 66, no. 2, pp. 1949–1962, 2020, doi: 10.32604/cmc.2020.012945.

[8] D. M. Kabala, A. Hafiane, and L. Bobelin, “Image - based crop disease detection with federated learning,” *Sci. Rep.*, pp. 1–19, 2023, doi: 10.1038/s41598-023-46218-5.

[9] S. Lee and C. M. Yun, “A deep learning model for predicting risks of crop pests and diseases from sequential environmental data,” *Plant Methods*, pp. 1–8, 2023, doi: 10.1186/s13007-023-01122-x.

[10] I. Ahmed and P. K. Yadav, “A systematic analysis of machine learning and deep learning based approaches for identifying and diagnosing plant diseases,” *Sustain. Oper. Comput.*, vol. 4, no. September 2022, pp. 96–104, 2023, doi: 10.1016/j.susoc.2023.03.001.

[11] C. Jackulin and S. Murugavalli, “A comprehensive review on detection of plant disease using machine learning and deep learning approaches,” *Meas. Sensors*, vol. 24, no. September, p. 100441, 2022, doi: 10.1016/j.measen.2022.100441.

[12] M. Lamba, Y. Gigras, and A. Dhull, “Classification of plant diseases using machine and deep learning,” *Open Comput. Sci.*, vol. 11, no. 1, pp. 491–508, 2021, doi: 10.1515/comp-2020-0122.

[13] F. Ahmed, “Machine Learning-Based Tea Leaf Disease Detection : A Comprehensive Review Introduction :”

[14] T. Domingues, T. Brandão, and J. C. Ferreira, “Machine Learning for Detection and Prediction of Crop Diseases and Pests: A Comprehensive Survey,” *Agric.*, vol. 12, no. 9, pp. 1–23, 2022, doi: 10.3390/agriculture12091350.

[15] A. Sharma, A. Jain, P. Gupta, and V. Chowdary, “Machine Learning Applications for Precision Agriculture: A Comprehensive Review,” *IEEE Access*, vol. 9, pp. 4843–4873, 2021, doi: 10.1109/ACCESS.2020.3048415.

[16] S. P. Mohanty, D. P. Hughes, and M. Salathé, “Using Deep Learning for Image-Based Plant Disease Detection,” vol. 7, no. September, pp. 1–10, 2016, doi: 10.3389/fpls.2016.01419.

[17] P. Kulkarni, A. Karwande, T. Kolhe, and S. Kamble, “Plant Disease Detection Using Image Processing and Machine Learning.”

[18] S. S. Harakannanavar, J. M. Rudagi, V. I. Puranikmath, and A. Siddiqua, “Plant leaf disease detection using computer vision and machine learning algorithms,” vol. 3, no. April, pp. 305–310, 2022, doi: 10.1016/j.gltp.2022.03.016.

[19] A. Haque, S. Marwaha, C. K. Deb, and S. Nigam, “Deep learning - based approach for identification of diseases of maize crop,” *Sci. Rep.*, pp. 1–14, 2022, doi: 10.1038/s41598-022-10140-z.

[20] C. Jackulin and S. Murugavalli, “Measurement : Sensors A comprehensive review on detection of plant disease using machine learning and deep learning approaches,” *Meas. Sensors*, vol. 24, no. September, p. 100441, 2022, doi: 10.1016/j.measen.2022.100441.

[21] P. Chandre, P. Mahalle, and G. Shinde, “Intrusion prevention system using convolutional neural network for wireless sensor network,” *IAES Int. J. Artif. Intell.*, vol. 11, no. 2, pp. 504–515, 2022, doi: 10.11591/ijai.v11.i2.pp504-515.

[22] P. R. Chandre, P. N. Mahalle, and G. R. Shinde, “Machine Learning Based Novel Approach for Intrusion Detection and Prevention System: A Tool Based Verification,” in *2018 IEEE Global Conference on Wireless Computing and Networking (GCWCN)*, Nov. 2018, pp. 135–140, doi: 10.1109/GCWCN.2018.8668618.

[23] M. Xu, J. Eun, J. Lee, J. Yang, and S. Yoon, “Plant Disease Recognition Datasets in the Age of Deep Learning : Challenges and Opportunities.”

[24] A. Chug, A. P. Singh, D. Singh, and D. L. Based, “Retraction Retracted: A Systematic Analysis of Machine Learning and Deep Learning Based Approaches for Plant Leaf Disease Classi fi cation: A Review,” vol. 2022, 2023.

[25] P. Kwabena *et al.*, “CCMT : Dataset for crop pest and disease detection,” *Data Br.*, vol. 49, p. 109306, 2023, doi: 10.1016/j.dib.2023.109306.

[26] C. Sarkar, D. Gupta, U. Gupta, and B. B. Hazarika, “Leaf disease detection using machine learning and deep learning: Review and challenges,” *Appl. Soft Comput.*, vol. 145, p. 110534, 2023, doi: 10.1016/j.asoc.2023.110534.