Identification of Plant Diseases Using *Imaging Processing and Machine Learning*

1st Amitabh Chaurasia
School of Computer Science and Engineering
Lovely Professional University
Phagwara, India
amitabh.12004548@lpu.in

2nd Baljinder Kaur
School of Computer Science and Engineering
Lovely Professional University
Phagwara, India
baljinder.neha@gmail.com

Abstract— **This comprehensive review explores the most current developments in the use of image processing and machine learning methods for the diagnosis of plant diseases. After a thorough review of the literature, we critically assess the different frameworks and methods used to classify disorders affecting plants. Our study highlights the advantages and disadvantages of each approach while focusing on how accurate it is in diagnosing a wide variety of illnesses. In addition, we investigate new developments and trends in this quickly developing industry. We conclude by talking about the ongoing difficulties and suggesting some directions for further study to improve the effectiveness of plant disease detection systems.**

Keywords— CNN, VGG16, image processing, classification, neural networks, and machine learning

I. INTRODUCTION

A large proportion of the world's population depends heavily on agriculture for their nutrition, and many livelihoods, especially in places like India, are closely related to agricultural pursuits. However, farmers are still plagued by the possibility of plant diseases, which seriously jeopardize agricultural output. These illnesses, which mostly affect plant leaves and fruits and are frequently caused by bacteria, viruses, or other pathogens, cause significant losses in agricultural output.

Combining machine learning and image processing methods has become a viable approach to address these issues in agriculture in recent years. Through the use of large-picture datasets that comprise both healthy and sick plant samples, scientists hope to create effective tools for the identification and classification of different plant diseases[1]. Notably, one of the most important places to start when identifying a disease is when symptoms appear on the surface of leaves or fruits[2].

In the field of plant disease diagnosis, several approaches have received substantial attention:

1. Advanced image processing techniques
2. Supervised learning algorithms
3. Unsupervised learning approaches

Preprocessing, segmentation, and feature extraction are important steps in these approaches that help precisely identify disease-affected regions on plant surfaces. In particular, by identifying critical characteristics retrieved from pictures, supervised learning approaches are critical in the prediction of illness. By utilizing machine learning models like VGG16 and Convolutional Neural Networks (CNNs), researchers may efficiently diagnose illnesses by training on a variety of feature datasets that are derived from photos of plants.

Unsupervised learning approaches have also generated interest because of their capacity to find patterns in data on their own, which makes it easier to find new features of diseases. Deep Convolutional Neural Networks (CNNs) are a notable method in this field that efficiently discriminate between healthy plant instances and those that are ill by utilizing several convolutional layers to extract important information.

Many deep learning network topologies that draw inspiration from biological neural networks are being studied in this period of study. Automatic learning techniques enable models such as CNN and VGG16 to identify unique characteristics in training photos and correctly categorize test images into appropriate illness groups.

This study begins with a thorough examination of the literature, concentrating on up to eight publications that use these approaches to categorize plant diseases. This study attempts to offer insights into the effectiveness of various techniques in plant disease detection by closely examining datasets, accuracy levels, and research procedures. Furthermore, investigating ensemble modeling appears to be a viable way to improve the consistency and accuracy of categorization. By combining predictions from many models, such as CNNs and VGG16, ensemble modeling aims to minimize the drawbacks of individual models while optimizing performance overall.

This research aims to further the current discussion on plant disease diagnosis by demonstrating how machine learning and image-processing approaches might transform agricultural practices.

II. LITERATURE REVIEW

Plant disease monitoring and detection have seen a major transformation in recent years because of the incorporation of cutting-edge technology like machine learning and image processing. Advancements have been achieved in the development of automated systems that can identify plant diseases in their early stages, facilitating prompt intervention and reducing crop losses.

To forecast disease signs across agricultural areas, Radha et al. [3] (2021) proposed an automated system for early-stage plant disease identification that makes use of convolutional neural networks (CNNs). Their method, which was trained on a collection of pictures of both healthy and sick leaves, showed encouraging outcomes, with an accuracy of up to 85%.

With a focus on color and texture traits, Madilwalar and Wyawahare [4] (2021) investigated a variety of image-processing algorithms for the identification of plant diseases.

Their research, which examined an RGB picture dataset, demonstrated how well grey-level co-occurrence matrix (GLCM) features work to identify normal leaves and the color characteristics of anthracnose-affected leaves. With a support vector machine classifier, they were able to get an astounding 83.34% accuracy.

Convolutional neural networks were utilized by Shrestha et al.[5] (2020) to identify a variety of plant diseases with an accuracy rating of 88.80%. Their results demonstrated the promise of deep learning techniques for the diagnosis of plant diseases, despite the computing difficulties involved.

Apart from deep learning methodologies, conventional machine learning algorithms have also played a significant role in the identification of diseases. Bhange et al.[6] (2015) created a web-based application that uses characteristics including color, morphology, and CCV to identify fruit diseases. With the use of support vector machines and clustering techniques, they were able to diagnose pomegranate disorders with 82% accuracy.

Additionally, the problem of unbalanced datasets in illness prediction has been addressed by developments in resampling approaches. To forecast plant diseases using unbalanced datasets, Bhatia et al.[10] (2019) suggested applying Extreme Learning Machine (ELM) classifiers. They enhanced the ELM classifier's performance by using strategies like Random Over Sampling (ROS) and Synthetic Minority Over-Sampling Technique (SMOTE), producing noteworthy improvements in classification accuracy and area under the curve.

To improve the precision and effectiveness of plant disease diagnosis systems, there is an increasing focus on the integration of various approaches and techniques, from deep learning architectures to conventional machine learning algorithms, as the area continues to develop.[7] Furthermore, tackling issues like dataset imbalance and computational complexity continues to be a priority for upcoming research projects.

Plant disease diagnosis with machine learning and image processing techniques has made encouraging strides, but there are still several obstacles and restrictions to overcome. Variability in plant traits and environmental circumstances is a major obstacle that can affect the precision and applicability of disease detection algorithms. Furthermore, there are issues with training dataset quality and availability, especially when it comes to uncommon or recently discovered illnesses.[8] Moreover, deep learning models' computational complexity can make it difficult for them to scale and be used practically in environments with limited resources. The possibility of bias in datasets, which can result in skewed predictions and decreased model resilience, is another drawback. Furthermore, it is still challenging to interpret complicated models like convolutional neural networks, which makes it challenging to comprehend the underlying decision-making process.

The literature review summarizes the noteworthy advancements in plant disease detection by machine learning and image processing methods. Research has indicated that several algorithms and approaches exhibit efficacy in identifying illnesses in diverse crop types and environmental settings.[9] Still, there are holes in several areas. First, to enable reliable model training and assessment, standardized datasets covering a wide variety of plant diseases and environmental circumstances are required. Furthermore, studies concentrating on the interpretability and explainability of machine learning models might improve end users' acceptance and confidence in them. Research on the scalability and practical use of disease detection systems is still lacking, especially in low-resource agricultural contexts.

To sum up, the integration of machine learning and image processing methodologies signifies a noteworthy progression in the domain of plant disease detection and tracking. The effectiveness of different algorithms and approaches in identifying illnesses in a range of crops and environmental situations has been demonstrated by the literature study. Even with the encouraging advancements, several obstacles and restrictions still exist.

The diversity of plant traits and environmental conditions, which can affect the precision and applicability of disease detection models, is one of the main obstacles. Furthermore, there are still issues with training dataset quality and availability, particularly about uncommon or developing disorders. Another obstacle to deep learning models' scalability and practical use is their computational complexity, which is especially problematic in agricultural contexts with limited resources.

Also, there are concerns about things like dataset bias and how interpretable sophisticated models like convolutional neural networks are. To overcome these obstacles, policymakers, researchers, and other stakeholders in agriculture must work together.

Standardized datasets covering a wide variety of plant diseases and environmental variables are urgently needed going the future. Furthermore, end-user acceptability and trust may be fostered by research aimed at improving the interpretability and explainability of machine learning models.[11] Moreover, efforts have to be focused on creating workable, scalable disease detection technologies that are available to farmers everywhere.

Bridging current gaps and turning research discoveries into workable solutions will be made possible through interdisciplinary cooperation and knowledge exchange. We can create more precise, effective, and easily accessible plant disease detection technologies by taking on these obstacles head-on. The ultimate goal of these developments is to ensure that agricultural systems remain resilient in the face of changing challenges by promoting sustainable farming practices and global food security.

Beyond tackling current issues, future study projects have to concentrate on utilizing cutting-edge technology like Internet of Things (IoT) gadgets and remote sensing methods for monitoring diseases in real-time. By combining these technologies with machine learning algorithms, farmers will be able to detect disease outbreaks quickly and implement proactive disease control techniques. [12]Additionally, there is a need for easily navigable mobile applications that offer decision support tools and actionable information to farmers through the usage of disease prediction algorithms. Giving farmers access to these technologies can help them be better able to allocate resources more efficiently and make timely adjustments, which would eventually improve crop productivity and agriculture.

Table 1: Comparison of Techniques.

Study	Methodology	Dataset	Accuracy
N. Radha et al.	CNN training, automated technique	Mixed (healthy and diseased)	85%
Shiroop Madilwalar and Medha Wyawahare	Image processing with colour and texture features, SVM classification	110 RGB images	83.34%
Garima Shrestha et al.	CNN with three blocks, high-resolution RGB images	3000 images	88.80%
M. Bhange et al.	Web-based tool, colour, morphology, CCV, k-means clustering, SVM	Mixed (fruit images)	82%
J.D Pujari et al.	Segmentation (crop-specific), texture features, various classifiers	Various crops	83.72% - 90.72% Vary-from data to data
Anand. H. Kulkarni and Ashwin Patil R. K	Image processing with Gabor filter, colour information, ANN classification	Raw plant images	91%
Maniyath	Comparative analysis of SVM, KNN, RF, Naïve Bayes, using HOG for feature extraction	Diseased leaves	70.14%
Bhatia et al.	ELM classifier with resampling techniques (ROS, SMOTE, RUS, IMPS)	TPMD dataset	88.57%

III. RESEARCH METHODOLOGY

The basis of this research was a wide collection of leaf photos that were carefully tagged with particular disease-plant pairings. The New Plant Diseases Dataset served as the main source of data, including a large number of cases of different plants and illnesses (Table 2).

Table 2:Types of Diseases in Dataset.

1: Apple_Apple_scab
2: Apple_Black_rot
3: Apple_Cedar_apple_rust
4: Apple_healthy
5: Blueberry_healthy
6: Cherry_(including_sour)_healthy
7: Cherry_(including_sour)_Powdery_mildew
8:Corn_(maize)_Cercospora_leaf_spot Gray_leaf_spot
9: Corn_(maize)_Common_rust
10: Corn_(maize)_healthy
11: Corn_(maize)_Northern_Leaf_Blight
12: Grape_Black_rot
13: Grape_Esca_(Black_Measles)
14: Grape_healthy
15: Grape_Leaf_blight_(Isariopsis_Leaf_Spot)
16: Orange_Haunglongbing_(Citrus_greening)
17: Peach_Bacterial_spot
18: Peach_healthy
19: Pepper_bell_Bacterial_spot
20: Pepper,_bell___healthy
21: Potato___Early_blight
22: Potato___healthy
23: Potato___Late_blight
24: Raspberry___healthy
25: Soybean___healthy
26: Squash___Powdery_mildew
27: Strawberry___healthy
28: Strawberry___Leaf_scorch
29: Tomato___Bacterial_spot
30: Tomato_Early_blight
31: Tomato_healthy_image
32: Tomato_Late_blight_image
33: Tomato_Leaf_Mold
34: Tomato_Septoria_leaf_spot
35: Tomato_Spider_mites_Two-spotted_spider_mite
36: Tomato_Target_Spot_image
37: Tomato_Tomato_mosaic_virus
38: Tomato_Tomato_Yellow_Leaf_Curl_Virus

All possible combinations of plant and disease are included in the New Plant Diseases Dataset, which makes thorough study and testing easier. To guarantee uniformity among all techniques, images were standardized to 224 by 224 pixels. Throughout the stages of model optimization and prediction, this consistent resizing procedure was upheld.

To improve the diagnosis of plant diseases using leaf photographs, creative research methods were used. This work adopted ensemble learning techniques, departing from traditional single-model designs. Convolutional neural networks (CNNs), which make up the ensemble, and the VGG16 architecture were picked for their ability to complement one another and operate well together on image classification tasks.

Plant disease identification has undergone a paradigm shift with the introduction of the ensemble model paradigm. Single-model approaches are still useful in some situations, but they might not be able to adequately convey the subtleties seen in leaf imagery. The ensemble seeks to get beyond these restrictions and provide a more reliable and accurate classification system by utilizing the combined intelligence of several models.

CNNs are the foundation of the ensemble model and are essential to modern image classification applications. These networks extract complicated patterns and spatial correlations necessary for the diagnosis of plant diseases by using convolutional layers to discern spatial hierarchies in images.

By identifying complex patterns in pictures, the VGG16 design enhances the CNN component. VGG16's deep design, which consists of 16 layers, makes it an excellent tool for interpreting complicated structures. This helps the ensemble identify small changes that are linked to different plant illnesses.

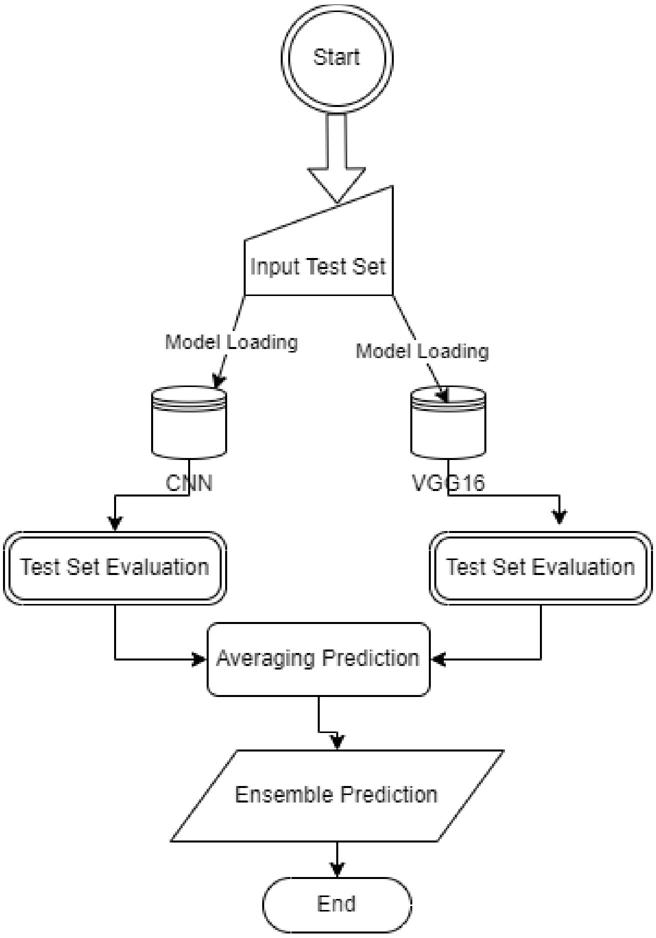
By combining the advantages of both the CNN and VGG16 architectures, a mutually beneficial connection is created. Although CNNs are quite good at extracting features and identifying spatial hierarchies, the deeper architecture of VGG16 improves the ensemble's ability to extract complex patterns and hierarchical representations.

The efficiency of the ensemble model technique is demonstrated by many tests carried out on a variety of datasets. When compared to single-model baselines, the ensemble model shows better classification accuracy, resilience, and generalization capacity. Furthermore, a broad range of plant species and disease combinations tested extensively highlight the ensemble framework's adaptability and applicability in real-world applications.

Subsequent investigations will delve into innovative group configurations that include sophisticated methods including adversarial training, transfer learning, and attention processes. The goal is to push the envelope of innovation to spark groundbreaking discoveries in the detection of plant diseases, which will ultimately advance sustainable farming methods and global food security. Exploration will go beyond traditional methods. We will look into novel group designs that use state-of-the-art methods such as transfer learning, adversarial training, and attention mechanisms. These initiatives seek to promote sustainable agricultural methods, global food security, and groundbreaking breakthroughs in the detection of plant diseases.

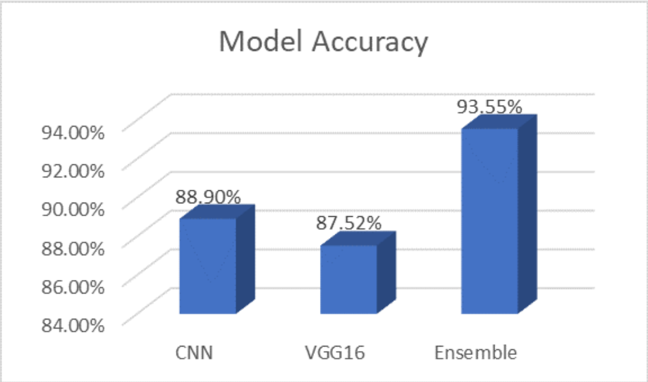
The trained models' accuracy results are displayed in Figure 2, where the CNN and VGG16 exhibit 88.90% and 87.52% accuracy, respectively. The CNN and VGG16 models were combined using an ensemble technique after both models had been trained to a high degree of accuracy. Following the loading of both models and their application to the test set, predictions were generated for the CNN and VGG16 separately. After that, the forecasts from both models were averaged to provide an ensemble prediction. Based on the examination, the CNN scored 88.90%, the VGG16 scored 87.52%, and the ensemble model outperformed the individual models with a much higher accuracy of 93.55%.

Figure 1: Ensemble Flowchart



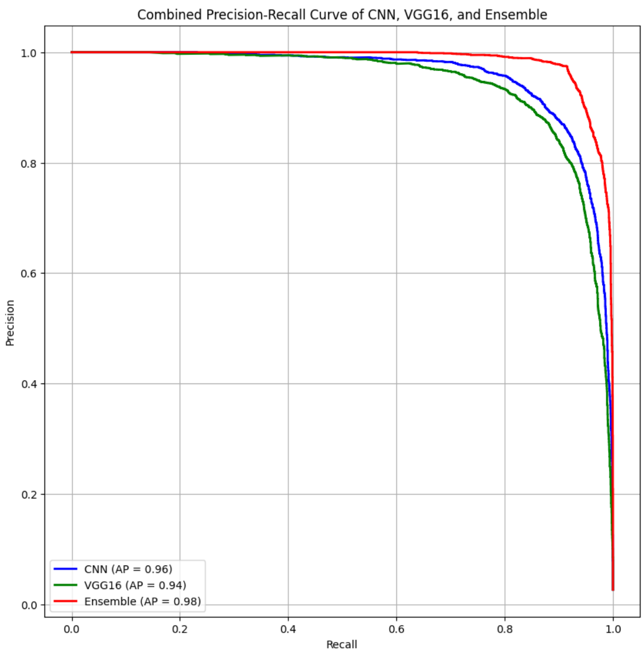
IV. RESULTS AND DISCUSSION

Figure 2: Model Accuracy



As seen in Figure 2, the ensemble model outperformed the individual models (CNN: 88.90%, VGG16: 87.52%) in terms of accuracy (93.55%). Combining the feature extraction capabilities of VGG16 with CNN's predictive capacity, the ensemble highlights the effectiveness of ensemble learning in the identification of plant diseases.

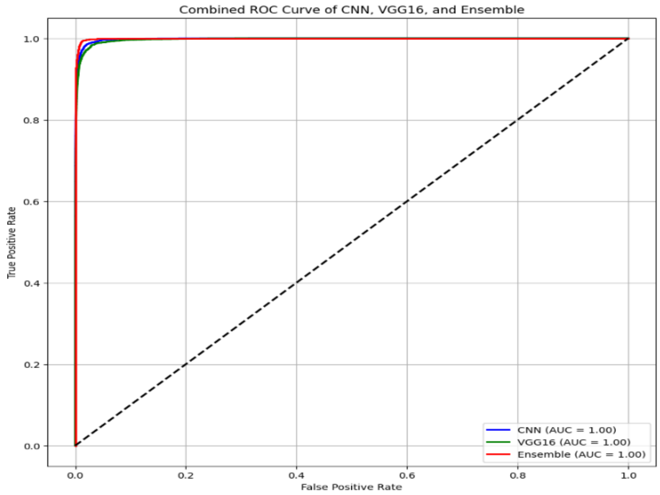
Figure 3: Precision Graph



The precision values attained by each model in recognizing plant diseases are highlighted in the precision graph, which is shown in Figure 3. While VGG16 obtained a precision of 0.94, CNN displayed a precision value of 0.96. Remarkably, the ensemble model achieved an accuracy value of 0.98, outperforming both individual models.

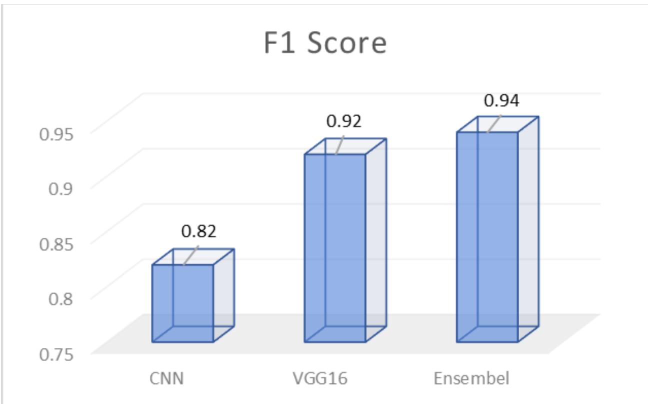
These findings demonstrate the ensemble's higher accuracy in recognizing ill plants. By combining the advantages of VGG16 and CNN, the ensemble model shows improved illness classification accuracy and dependability. The ensemble's ability to reduce false positives and enhance the overall performance of plant disease diagnosis systems is demonstrated by the greater precision value.

Figure 4: ROC Curve



The Area Under the Curve (AUC) values for CNN, VGG16, and the ensemble model are shown by the ROC curve in Figure 4, which all achieve a flawless AUC value of 1. This demonstrates the models' remarkable precision in differentiating between sick and healthy plants. These results demonstrate how well each model performs in accurately detecting plant diseases while reducing false positives and false negatives. The models' accuracy and dependability in identifying diseases are shown by their flawless AUC values, which also highlight their potential as useful instruments for agricultural applications.

Figure 5:F1 Score



The accuracy and recall of CNN, VGG16, and the ensemble model are displayed in Figure 5 along with their F1 Score values. With an F1 Score of 0.82, CNN demonstrated a balance between recall and precision in the diagnosis of diseases. With an F1 Score of 0.92, VGG16 performed better, demonstrating its ability to recognize intricate patterns in plant photos. The ensemble's improved F1 Score highlights how well it performed in correctly identifying plant illnesses, underscoring the importance of ensemble learning in improving disease detection systems.

V. CONCLUSION

Using sophisticated machine learning algorithms and an emphasis on ensemble learning, this study explored the field of plant disease diagnosis. By utilizing cutting-edge algorithms and a varied dataset, our goal was to improve the precision and dependability of disease detection systems used in agricultural settings. We carefully gathered and pre-processed the data to create a comprehensive dataset that included a wide range of plant species and disease combinations. The robust performance of our models was ensured across many circumstances by using this dataset as the basis for training and assessment.

The importance of ensemble learning in the detection of plant diseases was demonstrated by our research into machine learning methods. We illustrated the synergistic advantages of mixing different models by integrating the VGG16 architecture with Convolutional Neural Networks (CNNs). When compared to individual models, the ensemble model demonstrated higher accuracy, precision, and F1 Score, highlighting the effectiveness of ensemble learning in improving illness detection skills.

Furthermore, research demonstrated the flexibility and adaptability of machine learning models in solving real-world problems. pushed the limits of innovation in plant disease diagnosis by combining state-of-the-art approaches and methodologies, opening the door for sustainable agricultural practices and global food security.

In conclusion, this work offers important new information and techniques to improve agricultural operations and raise crop output and quality. It also signifies a substantial advancement in the field of plant disease diagnosis. By pursuing ongoing research and innovation, we can fully utilize machine learning to address intricate agricultural problems and advance global food security.

VI. REFERENCES

1. Detection of Cercospora Leaf Spot in Sugar Beet through robust template matching for disease identification. Zhou et al., 2014 Published in Computers and Electronics in Agriculture, Volume 108, Pages 58-70.
2. Nazir, Binzy, and Shayini R. authored, A Review on Plant Disease Detection Using Image Processing Tools and Machine Learning Techniques, published in the International Research Journal of Engineering and Technology (IRJET), volume 7, issue 11, in 2020.
3. At the International Conference on Artificial Intelligence and Smart Systems (ICAIS 2021) Radha, N., and Swathika, R., presented a paper titled Polyhouse: Plant Monitoring and Disease Detection using CNN. The article with the following DOI is available: 10.1109/ICAIS50930.2021.9395847. The results were presented in the conference proceedings.
4. Madiwalar, S. C., and Wyawahare, M. V., presented a study titled is A Comparative Study on Plant Disease Identification at the 2017 International Conference on Data Management, Analytics, and Innovation (ICDMAI), where they discussed the identification of plant diseases. The paper was published in the conference proceedings, and it can be accessed via DOI: 10.1109/ICDMAI.2017.8073478.
5. Shrestha, G., Deepsikha, Das, M., and Dey, N., presented a paper titled Plant Disease Detection Using CNN at the 2020 IEEE Applied Signal Processing Conference (ASPCON). DOI: 10.1109/ASPCON49795.2020.9276722.
6. Bhange, M., and Hingoliwala, H.A., presented a paper titled Pomegranate Disease Detection Using Image Processing at the Second International Symposium on Computer Vision and the Internet in 2015. Their research focused on the implementation of smart farming techniques for the detection of diseases in pomegranate plants. The paper can be found in Volume 58 of the symposium proceedings, spanning pages 280-288.
7. Pujari, J.D., Yakkundimath, R., and Byadgi, A.S., discussed the detection of fungal diseases in plants using image processing in their paper presented at the International Conference on Information and Communication Technologies in 2015. Their study utilized image-processing techniques to identify fungal diseases affecting plants. The paper is included in Volume 46 of the conference proceedings, covering pages 1802-1808.
8. Anand H, Kulkarni, and Patil, R. K. A. presented research on plant disease detection through image processing in the International Journal of Modern Engineering Research (IJMER), Volume 2, Issue 5, PP. 3661-3664, 2012 (ISSN: 2249-6645).
9. Maniyath, S. R. et al. presented research on plant disease detection utilizing machine learning techniques at the 2018 International Conference on Design Innovation in Communication, Computer, and Control (ICDI3C 2018). The paper, published in the conference proceedings, can be accessed via DOI: 10.1109/ICDI3C.2018.00017.
10. Bhatia, A., Chug, A., & Prakash Singh, A. (2020). Plant disease prediction using an extreme learning machine on a highly unbalanced dataset. Journal of Statistical Management Systems, 23(6), 1059-1068.DOI:10.1080/09720510.2020.1799504
11. Meghana Govardhan et al, "Diagnosis of Tomato Plant Diseases using Random Forest",2019, Global Conference for Advancement in Technology (GCAT)
12. Arnawa et al (2019), Food Security program towards community food consumption. Journal of Advanced Research in Dynamical and Control Systems, 11(2), 1198-1210.
13. Sunil S et al, "Plant leaf disease detection using computer vision and machine learning algorithms" Global Transition Processing 3 (1) (2022) 305-310.
14. V. Ananthi, Fused segmentation algorithm for the detection of nutrient deficiency in crops using SAR images, Artificial Intelligence Techniques for Satellite Image Analysis, Springer,2020, pp. 137-159.