Plant Leaf Disease Prediction Using Machine Learning: A Review

Ms. Rani Sahu[1], Mrs. Swati Tiwari[2]

[1] *Research Scholar, Department of Computer Science and Engineering, Columbia Institute of Engineering and Technology, Raipur (C.G.), India*

[2]*Assistant Professor and HOD, Department of Computer Science & Engineering, Columbia Institute of Engineering and Technology, Raipur, (C.G.), India*

***Abstract*—Agriculture remains one of the most essential pillars of the economy, especially in countries like India where a majority of the population depends on farming for their livelihood. One of the biggest challenges in modern agriculture is the rapid spread of plant diseases, especially those affecting leaves, which can significantly reduce crop quality and yield. If these diseases are not identified and treated in time, they can lead to large-scale losses and affect food security. Traditionally, the identification of plant leaf diseases has relied on human observation and expertise. However, this process can be slow, inaccurate, and not feasible for large-scale farming. With the advancement of technology, particularly in artificial intelligence and machine learning (ML), it has become possible to develop systems that can automatically detect diseases from leaf images with high accuracy. This project focuses on creating a machine learning-based model to predict and classify different types of leaf diseases in vegetable and fruit crops.**

***Index Terms*—Plant Leaf Disease, Machine Learning Models, Convolutional Neural Network (CNN), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree, Naive bias, Agricultural Technology, Image Processing, Plant Health Monitoring.**

I. INTRODUCTION

Agriculture plays a vital role in the economy and food security of many countries, particularly in developing nations like India, where a large portion of the population depends on farming for their livelihood. With the rising global population, there is a growing demand for agricultural products such as fruits, vegetables, and grains. However, one of the major challenges that farmers face today is the occurrence of plant diseases, which significantly reduce crop yield and quality. Among various plant parts, leaves are often the first to exhibit symptoms of disease, making them critical indicators of plant health. Early and accurate detection of leaf diseases is essential for effective crop management and protection. Traditional methods of disease detection rely heavily on human expertise, including visual inspection by farmers or agricultural officers. These methods are often time-consuming, prone to human error, and may not be practical for large-scale farms. Additionally, farmers in remote or rural areas may lack access to trained specialists who can diagnose and suggest appropriate treatments. With advancements in the fields of computer science and artificial intelligence (AI), particularly in machine learning (ML) and image processing, it is now possible to automate the detection of plant diseases through analysis of leaf images. These technologies offer scalable, fast, and highly accurate solutions that can assist farmers in identifying diseases early and taking timely action. In this thesis, a machine learning-based approach is proposed for the prediction and classification of plant leaf diseases using image data. The aim is to develop a robust system that can identify diseases in commonly grown vegetables and fruits, such as tomato, potato, chili, brinjal, and papaya, based on the visual symptoms appearing on leaves. By using image datasets collected from online resources and real-world farms, the system is trained and evaluated using various machine learning models such as Convolutional Neural Network (CNN), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naive Bayes, and Decision Tree algorithms.

The proposed system includes several phases: image acquisition, preprocessing, feature extraction, model training, disease classification, and result output. Among the models tested, the CNN-based model achieved the highest accuracy, demonstrating its ability to learn complex patterns in leaf images. The system can potentially be deployed in the form of a web or mobile application, enabling users to capture and upload leaf images and receive disease predictions instantly. The objective of this research is not only to build an accurate disease prediction model but also to contribute to smart and sustainable farming practices. Through automation and technology, farmers can reduce crop losses, minimize pesticide usage, and increase agricultural productivity. This work serves as a stepping stone toward more advanced agricultural technologies such as precision farming and AI-based crop health monitoring systems.

II. LITRETURE REVIEW

the detection and classification of plant diseases have become a critical area of research due to their direct impact on agricultural productivity and food security. Traditional methods of plant disease identification have heavily relied on the expertise of agricultural specialists who visually inspect plant leaves and diagnose potential diseases. However, such manual inspection methods are time-consuming, labor-intensive, and prone to human error. This has led to an increased interest in the development of automated systems for plant disease detection using computer vision and machine learning techniques. Earlier research efforts focused primarily on classical image processing methods where hand-crafted features such as color, texture, and shape were extracted manually from plant leaf images. These features were then used to classify diseases using basic classifiers like decision trees or rule-based logic. While these techniques showed some level of accuracy, they were limited in their generalization capability. Their performance often deteriorated in uncontrolled environments with variable lighting, complex backgrounds, and overlapping symptoms among different diseases. The dependency on manually selected features made the models rigid and less adaptable to new datasets or different plant species.

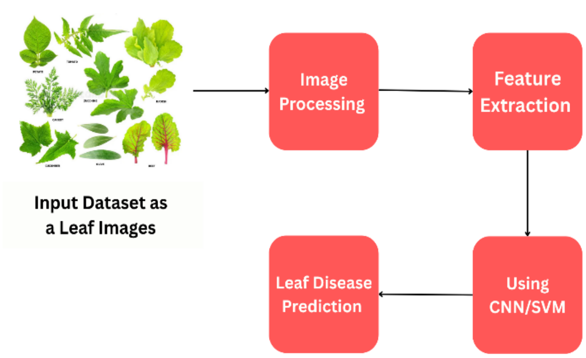
The advancement of machine learning has significantly improved the potential of automated plant disease detection systems. Traditional machine learning algorithms like Support Vector Machines (SVM), k-Nearest Neighbors (KNN), and Naive Bayes (NB) have been applied to various agricultural datasets to detect and classify leaf diseases. These models, when trained on well-preprocessed and labeled image data, can effectively learn the patterns and symptoms associated with specific diseases. For instance, SVM models have demonstrated strong performance in binary and multi-class classification tasks, especially when used in combination with texture and color-based features. KNN has been employed for its simplicity and effectiveness in small datasets, although it is computationally expensive during prediction. Naive Bayes, despite its assumption of feature independence, has shown decent results in scenarios with limited training data.

The evolution of deep learning has marked a significant turning point in the field of plant disease detection. Convolutional Neural Networks (CNNs), in particular, have emerged as powerful tools due to their ability to automatically learn relevant features from raw image data without requiring manual feature extraction. Unlike traditional methods, CNNs can identify complex patterns and hierarchical features in plant leaf images, which significantly improves classification accuracy. Studies have demonstrated that deep learning models like AlexNet, VGGNet, InceptionNet, and ResNet achieve remarkably high accuracy, often exceeding **95%,** on datasets containing multiple plant diseases. These models not only outperform traditional machine learning algorithms but also offer the advantage of scalability and transferability across different domains. Among the pioneering works in this domain, researchers have developed CNN-based systems that are capable of classifying multiple diseases in crops like tomato, potato, and grape with exceptional accuracy.

The introduction of transfer learning, where pre-trained models such as ResNet50 are fine-tuned on specific plant disease datasets, has further enhanced the performance and efficiency of these systems. Transfer learning significantly reduces the computational cost and training time, while still achieving high generalization on unseen data. Researchers have successfully leveraged these pre-trained networks, demonstrating that they can be adapted effectively to various agricultural applications even when the available data is limited. Another significant advancement in the field has been the integration of data augmentation techniques. Techniques such as rotation, flipping, brightness adjustment, and zooming help in artificially increasing the size and diversity of training datasets. This reduces overfitting and improves the robustness of the model in dealing with real-world variations. Additionally, preprocessing steps such as grayscale conversion, histogram equalization, and noise filtering play a crucial role in enhancing image quality and enabling better feature extraction by the neural network.

Recent works have also explored hybrid models that combine deep learning with traditional classifiers. For example, features extracted using CNNs can be classified using SVM or Random Forest classifiers for improved accuracy. Ensemble methods, where multiple models are combined to make final predictions, have also been shown to boost the performance and reliability of disease detection systems. Furthermore, object detection frameworks such as YOLO (You Only Look Once) and Faster R-CNN have been utilized for locating and classifying diseased regions in real-time, opening new possibilities for mobile and drone-based agricultural monitoring systems. Despite these advancements, several challenges and research gaps remain. Most studies are conducted using controlled datasets like PlantVillage, which lack the complexity and variability of real-world farming environments. As a result, models trained on such datasets may not perform well when deployed in the field.

III. PRAPOSED METHEDOLOGY



Data Collection

The primary objective of data collection in this research is to obtain high-quality, labeled images of plant leaves exhibiting different disease conditions as well as healthy states. These images serve as the input for training, validating, and testing the deep learning model. A robust dataset ensures that the model learns to recognize subtle differences in leaf patterns, discolorations, textures, and shapes associated with specific diseases.

Data Image Preprocessing

Image processing plays a crucial role in the field of plant disease detection. Before a machine learning or deep learning model can be trained to recognize and classify leaf diseases, the raw image data must be preprocessed and transformed into a form that enhances relevant features while minimizing noise. This chapter explains the various stages of image processing applied to the dataset and how it improves the accuracy and efficiency of the prediction model.

Image processing is a vital step in the pipeline of plant disease prediction. Through preprocessing and augmentation, the quality and quantity of the data were significantly improved. This ensures that the deep learning model receives clean, well-structured, and varied inputs, allowing it to learn more effectively. The combination of resizing, normalization, augmentation, and optional segmentation leads to more accurate, stable, and generalizable predictions. The entire image processing workflow contributes greatly to the overall success of the model and its potential for real-world agricultural applications.

Data Annotation

Proper labeling is essential for supervised learning. Each image was annotated with a unique label corresponding to the disease type or "healthy" status. In cases where the labels were not provided (especially in field data), annotation was performed manually by agricultural domain experts using online tools like Labeling or via spreadsheet documentation. This ensured the dataset remained organized and accurate.

Data Augmentation

The core purpose of data augmentation in this research is to generate a diverse set of training images by simulating real-world variations in plant leaves without having to collect additional physical images. It allows the model to learn invariant features and become robust to environmental factors such as orientation, brightness, background noise, and camera angle. This is especially important in agricultural applications, where plant leaves may be captured under different weather conditions, camera settings, or angles, and still need to be correctly classified by the system.

Data augmentation is an essential strategy in the success of the plant leaf disease prediction model. By simulating real-world variations through image transformation techniques, the dataset becomes more robust and diverse. This not only improves the model's accuracy but also ensures its practical applicability in the field, where perfect image capture conditions are rarely guaranteed. In conclusion, augmentation plays a vital role in enhancing the learning process and making the model ready for real-life agricultural use cases.

Feature Extraction

Feature extraction is a fundamental step in any image classification task, especially in plant disease detection systems. It involves identifying and extracting the most informative patterns or characteristics from input images that can effectively represent the underlying class of the image, such as healthy or diseased leaf. In machine learning and deep learning, features serve as the input for the model to learn and make predictions. The quality and relevance of extracted features directly affect the accuracy and robustness of the final model.

In this research, feature extraction is crucial for converting raw pixel data into meaningful information that can be used by the classifier to distinguish between different diseases and healthy leaves. This process allows the model to focus on important patterns such as shape, color variation, vein structure, lesion texture, and leaf surface abnormalities.

IV. MACHINE LEARNING MODELS

In the context of plant leaf disease prediction, ML models are trained using labeled images of plant leaves to identify and classify various diseases based on patterns, textures, shapes, and colors present in those images. These models help in developing an automated decision support system that assists farmers and agricultural experts in early and accurate disease detection, thereby increasing crop yield and reducing economic losses.

Several machine learning algorithms have been developed and applied in this domain. Each algorithm has its strengths and is chosen based on the dataset size, complexity of the classification task, and feature space. In this research, the following machine learning models were studied and evaluated: Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naive Bayes (NB), Decision Tree (DT), and Convolutional Neural Network (CNN). Machine learning models play a central role in automating the detection and classification of plant leaf diseases. Each algorithm has its advantages and ideal use cases. While models like SVM and KNN offer interpretability and work well on structured feature data, CNN offers end-to-end learning from raw images and achieves higher accuracy and robustness. This thesis leverages the strengths of deep learning through CNNs, specifically ResNet50, to build a scalable and practical solution for plant disease prediction in agricultural environments.

V. CONCLUSION

The accurate and timely detection of plant leaf diseases is essential for improving agricultural productivity, ensuring food security, and minimizing economic losses for farmers. This research aimed to develop an efficient and reliable system for the automatic classification of plant leaf diseases using machine learning and deep learning techniques. By leveraging image processing, feature extraction, and powerful classification models, particularly Convolutional Neural Networks (CNNs), this study demonstrated the effectiveness of automated disease diagnosis systems in the agricultural domain.

REFERENCE

[1] KaziNaimur Rahman, Sajal Chandra Banik, Raihan Islam, Arafath Al Fahim (Volume 4, Issue 1, February 2025, 100092) A real time monitoring system for accurate plant leaves disease detection using deep learning " ELSEVIER ".

[2] Karamjeet Kaur 1, Priyanshi sharma2, Ankit Chaubey3, Aman Kumar Mishra4, Priyanka Pathak5 (© 2023 IJNRD | Volume 8, Issue 5 May 2023 | ISSN: 2456-4184 | IJNRD.ORG) LEAF DISEASE PREDICTION "IJNRD ".

[3] Pushkar Deshwal1, Kaushal Sharma2, Mr. Suveg Moudgil3 (Volume 11 Issue V May 2023- Available at www.ijraset.com) Plant Leaf Disease Detectionusing Machine Learning "IJRASET ".

[4] C Jackulin, S. Murugavalli (Volume 24, December 2022, 100441)A comprehensive review on detection of plant disease using machine learning and deep learning approaches "ELSEVIER ".

[5] Nirmala Shinde & Asha Ambhaikar (Volume 18, article number 14, (2025) An efficient plant disease prediction model based on machine learning and deep learning classifiers " SPRINGER NATURE Link ".
[6] G. Geetha1, *, S. Samundeswari2, G.Saranya 3, K.Meenakshi 4 and M. Nithya5 (G. Geetha et al 2020 J. Phys.: Conf. Ser. 1712 012012) Plant Leaf Disease Classification and Detection System Using Machine Learning "PURPOSE- LED PUBLISHING ".
[7] Nivargi Anil Basavant1, Gururaj V2, Kavya N L3 (E-ISSN: 2582-2160 ● Website: www.ijfmr.com) Crop Prediction Analysis and Plant Disease Detection Using Machine Learning "IJFMR ".
[8] SunilS. Harakannanavar [a], Jayashri M. Rudagi [b], Veena I Puranikmath [b], Ayesha Siddiqua [a], R Pramodhini (Volume 3, Issue 1, June 2022, Pages 305-310) Plant leaf disease detection using computer vision and machine learning algorithms " KeAi ".
[9] Eman Abdullah Aldakheel 1 *, Mohammed Zakariah2 and Amira H. Alabdalall 3 (PUBLISHED 22 April 2024 DOI 10.3389/fpls.2024.1355941) "frontiers ".
[10] Shivappa M. Metagar[1], Gyanappa A. Walikar[2, *] (volume 44 issue 4 (august 2024): 591-602) Machine Learning Models for Plant Disease Prediction and Detection: A Review " **Arcc Journals** ".
[11] Sunku Rohan, Triveni S Pujar, Suma VR Amog Shetty, Rishabh F Tated (2019), CNN based Leaf Disease Identification and Remedy Recommendation System, IEEE conference paper.
[12] Bin Liu, Peng Jiang, Yuehan Chen, Dongjian He, Chunquan Liang, Real-Time Detection of Apple
[13] Karunya Rathan, Abirami Devaraj, K Indira and Sarvepalli Jaahnavi (2019), Identification of Plant Disease using Image Processing Technique, International Conference on Communication and Signal Processing, IEEE.
[14] Vellanki Krishna Vamsi, Velamakanni Sahithya, Brahmadevara Saivihari, Parvathreddy Sandeep Reddy and Karthigha Balamurugan (2019). GUI based Detection of Unhealthy Leaves using Image Processing Techniques, International Conference on Communication and Signal Processing
[15] Devi, S. S., & Santhi, V. (2018). Identification of apple leaf diseases using deep learning convolutional neural networks. International Journal of Pure and Applied Mathematics, 118(8), 373-380.
[16] R. Kumar, R. Yadav, M. Rathore, et al (2016). A Review on Detection of Plant Leaf Diseases using Image Processing Techniques, Procedia Computer Science.
[17] Md Nasim Adnan "Improving the Random Forest Algorithm by Randomly Varying the Size of the Bootstrap Samples"Adnan,(2014).
[18] Ziming Wu, Weiwei Lin, Zilong Zhang and Angzhan Wen "An Ensemble Random Forest Algorithm for Insurance Big Data Analysis"IEEE,(2017).
[19] Manjunath Badiger, Varuna kumara,Sachin CN shetty,Sudhir poojary "Leaf and skin disease detection using image processing" Global Transactions Proceedins,(2022).
[20] Niveditha M, Pooja R, Prasad Bhat N, shashank N, "Plant disease detection using machine learning" IEEE (2021).
[21] Nishant Shelar, Suraj shinde, Shubham sawant, Shreyas dhumal "Plant disease detection using CNN" Turkish Journal of Computer and Mathematics Education, (2021).
[22] Madhuri Devi Chodey, Dr.Noorilla Shariff C, Gauravi Shetty "Pest detection in crop using video and Image processing" IJRASET (2020).