

Agritech-Plant Disease Detection and Classification

Anushri Awari¹, Vaishnavi Bhokare², Harshada Daundkar³, Swarupa Dhas⁴ Dr.

Praveen Blessington Thummalakunta⁵

1,2,3,4,5 Department of Computer Engineering & Zeal College of Engineering and Research, Pune, India

_____***

Abstract - The Plant Disease Detection and Classification project is designed to help farmers identify crop diseases at an early stage using images of plant leaves. By applying advanced machine learning (ML) and deep learning (DL) techniques, this system aims to automatically detect and classify various plant diseases with high accuracy. Early identification of plant diseases plays a vital role in improving crop yield, reducing unnecessary pesticide usage, and encouraging sustainable farming practices. Once a disease is detected, the system not only identifies it but also provides a brief description and recommends appropriate prevention and treatment methods. To enhance performance, the project uses data augmentation and normalization techniques-helping the model handle variations in lighting, angles, and environmental conditions. The system is trained on a diverse dataset containing images of both healthy and diseased leaves from different plant species, categorized into 38 different classes.

The project incorporates various machine learning algorithms such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Decision Trees, and Random Forests, as well as a Convolutional Neural Network (CNN) for deep learning. Among these, CNNs have shown outstanding performance due to their ability to learn spatial features from images effectively. While existing research has shown promising results, real-world challenges such as varying image quality, lighting conditions, and background clutter still affect accuracy. This project aims to overcome these issues by training on large, diverse datasets and fine-tuning model parameters. Ultimately, this project hopes to empower farmers with an intelligent tool to monitor crop health, make informed decisions, and contribute to food security through precision agriculture.

Keywords:

Plant Disease Detection, Automated Disease Detection System, Machine Learning, Deep Learning, Data Augmentation, Data Normalization, Support Vector Machine, KNN Classifier, Convolutional Neural Network, Decision Trees, Random Forest

1.INTRODUCTION

Plant diseases continue to pose a major threat to global agriculture, contributing to significant crop damage and financial losses each year. Detecting these diseases early is essential for effective management and timely treatment. However, conventional diagnosis methods often depend on expert analysis and lab-based testing, which can be time-consuming, costly, and not always accessible especially in rural or low-resource farming areas.

To overcome these limitations, our Plant Disease Detection and Classification project focuses on building an intelligent, automated system powered by modern machine learning and deep learning techniques. The system is designed to analyze techniques to artificially increase the size and diversities of th images of plant leaves and accurately determine the presence and type of disease. Along with the diagnosis, it will also provide detailed insights into the symptoms, possible causes, and recommended treatments, helping farmers take timely and informed action.

At the core of this system are powerful deep learning architectures like Convolutional Neural Networks (CNNs), which are particularly effective in handling image data. These models will be trained to extract and learn key features from leaf images, enabling them to distinguish between healthy and diseased plants, and further classify the disease type. To enhance the system's performance and reduce overfitting, we will apply data augmentation techniques—generating varied versions of existing images to increase dataset diversity.

In our system, we have implemented the Convolutional Neural Network (CNN) algorithm for effective image classification, as it excels at recognizing patterns and features within visual data. CNN enables the model to automatically learn the distinguishing characteristics of different plant diseases from leaf images. Additionally, we have integrated OpenCV, a powerful computer vision library, for preprocessing tasks such as resizing, filtering, and segmenting images to enhance model performance. Beyond detection, our user interface not only allows farmers to upload images for diagnosis but also provides tailored supplements and treatment recommendations based on the identified disease. This holistic approach ensures that users receive both accurate diagnoses and practical remedies, promoting healthier crops and sustainable farming practices.

A key component of our approach is the creation of a highquality, diverse dataset. This dataset will consist of numerous images from different plant species affected by various diseases. Each image will be carefully pre-processed to improve clarity, adjust lighting, and isolate relevant features to aid in more accurate learning.

The system's performance will be thoroughly evaluated using key metrics including accuracy, precision, recall, and F1-score. Validation will be performed on a wide-ranging test dataset to ensure strong generalization capabilities. Once the models demonstrate dependable results, we will integrate them into a user-friendly application. Through this platform, farmers will be able to simply upload images of affected plants and receive immediate feedback on potential diseases and suggested remedies.

2. LITERATURE REVIEW

Plant Disease Detection Using Machine Learning by Anamika Jain, Anagha Langhe, Harsh Choudhary, and Ashutosh Mishra (2024, IEEE) explores the use of machine learning to detect plant diseases early. The authors aim to reduce pesticide use, enhance crop yields, promote sustainability, and make disease detection more accessible to farmers. They employed



techniques such as data augmentation, feature engineering, and a CNN architecture optimized with a tuned learning rate and batch size (32) to achieve faster convergence, using TensorFlow's Data Pipeline. However, the effectiveness of this system depends heavily on data quality and environmental factors, which can vary. The approach may struggle with symptom variations and is also less accessible in remote areas or on diverse devices.[1]

In their 2024 paper, Analysis of Formal Concepts for verification of pests and diseases of crops using ML, authors Jamalbek Tussupoy, Moldir Yessenova, Gulzira Abdikerimova, Aidyn Aimbetov, and Kazbek Baktybekoy (IEEE) focus on formalizing concepts for verifying crop pests and diseases with machine learning. Their approach combines spectral-space data with machine learning methods, including logistic regression, Vanilla CNN, and XGBoost, to detect and classify diseases. However, their methodology is currently limited to spectral-space data, which restricts its application to a narrow data type. Broader applications using diverse datasets and alternative techniques could improve the system's adaptability.[2]

In the 2024 paper, Improving Plant Disease Classification With Deep-Learning-Based Prediction Model Using Explainable Artificial Intelligence, Natasha Nigar, Hafiz Muhammad Faisal, Muhammad Umer, Olukayode Oki, and Jose Manappattukunnel Lukose focus on early plant disease detection using a deep-learning model. Their goal is to reduce pesticide usage, enhance crop yields, and support sustainable farming practices. The authors apply data augmentation, feature engineering, and a CNN architecture with a specifically tuned learning rate and batch size (32) for improved convergence speed, utilizing TensorFlow's Data Pipeline to optimize performance.While the model shows promise, its accuracy depends greatly on data quality and environmental conditions, which may vary widely and affect detection reliability. Additionally, the model faces challenges with symptom variations and may be less accessible in remote areas or across different devices, pointing to areas for further enhancement.[3]

In their 2024 research paper titled "An Approach Toward Classifying Plant-Leaf Diseases and Comparisons With the Conventional Classification," Anita Shrotriya and her coauthors delve into the use of machine learning for identifying plant diseases and pests. The study places particular emphasis on the Support Vector Machine (SVM) algorithm, a widely used supervised learning model known for its effectiveness in classification tasks. The researchers highlight that while SVM performs well, its accuracy heavily depends on the clarity and quality of the leaf images being analyzed. In other words, better-quality images lead to more accurate disease detection, underlining the importance of high-resolution data in machine learning-based agricultural solutions.[4]

In the 2024 paper Plant Disease Detection and Classification Techniques: A Comparative Study of the Performances, Wubetu Barud Demilie conducts a review of machine learning (ML) and deep learning (DL) techniques for detecting and classifying plant diseases. The study provides a comparative analysis of recent ML and DL methods, highlighting their effectiveness in this area. However, the author notes that there is a need for improved dataset robustness and the implementation of automated parameter searches for weather data to enhance the overall performance of these techniques.[5]

A Systematic Literature Review on Plant Disease Detection by Wasswa Shafik, Ali Tufail, Abdallah Namoun, Liyanage Chandratilak De Silva, and Rosyzie Anna Apong (2023, IEEE) examines various machine learning and deep learning approaches, such as CNNs, for plant disease detection. This review focuses on models like pre-trained CNNs, SVM classifiers, and ensemble learning techniques to provide an overview of the existing methodologies. The authors note limitations in relying on pre-trained models, which may struggle with generalization. They suggest that future research should test other classifiers and explore different model combinations to address these challenges.[6]

In, Real-Time Plant Disease Dataset Development and Detection Using Deep Learning (2023), Diana Susan Joseph, Pranav M. Pawar, and Kaustubh Chakradeo explore developing datasets to enhance real-time disease detection in crops like rice, wheat, and maize. The authors create specialized datasets that capture various stages of disease progression, using data augmentation to enrich the model's learning process. However, their work primarily relies on annotated images, and they suggest that future improvements could focus on integrating these datasets into object detection models to evaluate disease severity, making the system more applicable for real-world agricultural needs.[7]

In the 2023 paper, Machine Learning and Deep Learning for Plant Disease Classification and Detection, Vasileios Balafas, Emmanouil Karantoumanis, Malamati Louta, and Nikolaos Ploskas Provide an analysis that compares machine learning and deep learning techniques for identifying plant diseases.. The study evaluates five object detection and eighteen classification algorithms across existing datasets, providing insights into the effectiveness of different approaches. The authors highlight a need for datasets with greater diversity and realism, suggesting that adding non-image data could enhance the robustness of ML and DL models, making them more effective in varied agricultural environments.[8]

In A Comparative Study on Disease Detection of Plants Using Machine Learning Techniques (2021), Vishnu S. Babu, R. Satheesh Kumar, and R. Sunder investigate various machine learning methods to detect plant diseases such as bacterial blight and sheath rot. The authors use a five-step approach: dataset creation, preprocessing, data augmentation, feature selection, and classification. While their study provides valuable insights, it primarily focuses on specific plants like cardamom, and they suggest that future research should expand to a broader range of crops to increase the model's versatility.[9]

Study of Machine Learning Techniques for Plant Disease Recognition in Agriculture (2021), Pallavi Dwivedi, Sumit Kumar, Surbhi Vijh, and Yatender Chaturvedi did research on machine learning and image processing methods and algorithms for identifying illnesses in plant leaves. The authors note that future research could benefit from hybrid deep learning models that incorporate fuzzy logic, as well as the development of automated systems for detecting healthy versus diseased leaves, which could significantly improve the effectiveness of agricultural disease management.[10]



3. PROPOSED METHODOLOGY



Fig -1: Proposed Methodology

This section outlines the methodology adopted for the plant disease detection and classification system, as illustrated in Fig. 1. The approach harnesses deep learning techniques to build a robust and scalable model. The process follows a systematic pipeline comprising data preprocessing, feature extraction through transfer learning, followed by model training, validation, and performance evaluation. The implementation is carried out using Python and TensorFlow within a Jupyter Notebook environment. For experimentation, the publicly accessible PlantVillage dataset is used, which includes a wide range of leaf images depicting both healthy and diseased conditions. To ensure effective learning, the dataset is first balanced and augmented to handle class imbalances and improve generalization. Techniques such as rotation, flipping, zooming, and brightness adjustment are applied to artificially expand the dataset and expose the model to various orientations and lighting conditions. A pre-trained Convolutional Neural Network (CNN) model is then fine-tuned using transfer learning to extract meaningful features from the leaf images. This significantly reduces training time and improves accuracy by leveraging knowledge from large-scale image datasets.

1.Image Acquisition

The process of image acquisition serves as the initial stage in the image processing workflow, involving the capture of realworld visuals using devices such as digital cameras, satellite sensors, medical scanners, and other imaging tools. These devices convert visual scenes into digital images, which may be in grayscale—depicting intensity through a single channel or in color formats like RGB or CMYK that use multiple channels. The clarity and detail of the captured image can be affected by factors such as lighting, resolution, and the quality of the imaging sensor. A major concern during this phase is image noise, which can degrade image quality and negatively impact subsequent processing tasks. The primary objective of this stage is to obtain a clear and accurate representation of the target object or scene to enable effective further analysis.

2.Image Preprocessing

After capturing the image, the next step involves preprocessing, which aims to refine image quality and prepare it for further analysis tasks like segmentation and feature extraction. This phase typically involves techniques to minimize noise, standardize pixel intensity, and improve image contrast. Noise can be reduced using methods such as median filtering or Gaussian smoothing, while contrast enhancement like histogram equalization—helps highlight essential features within the image. Additionally, operations like resizing, cropping, or rotating may be performed to ensure consistent alignment and scale across all images. Pre-processing helps ensure the images are clean, consistent, and ready for effective analysis in the later stages of the pipeline.

3.Image Segmentation

Image segmentation is a vital step in the image analysis pipeline that involves dividing an image into distinct regions, each corresponding to meaningful components such as objects, textures, or boundaries. This division simplifies further analysis by isolating specific areas of interest within the image. Depending on the characteristics of the image and the desired outcome, various segmentation approaches can be employed. Basic methods like thresholding classify pixels based on intensity values, while edge detection techniques such as Canny or Sobel help outline object boundaries. Region-based techniques group together pixels with similar properties, and more advanced algorithms like the watershed method utilize topological features to delineate segments. Through this process, segmentation enhances the focus on relevant areas, enabling more accurate feature extraction and classification in subsequent stages.



Fig -2: Randomly selected plant images before preprocessing **4.Feature Extraction**

Feature extraction is the process of identifying and isolating key patterns or attributes from a segmented image, which are crucial for subsequent tasks like detection and classification. These attributes can be categorized into several types: edge-based, texture-based, shape-based, and color-based features. Edge features focus on identifying the boundaries of objects within the image, while texture features analyze the spatial distribution of pixel intensities using techniques like the Gray-Level Cooccurrence Matrix (GLCM). Shape features, such as Hu Moments or Fourier transforms, capture the geometric properties of objects, and color features examine the color distribution across the image. The goal of this step is to condense the image into a set of significant features that retain essential information for recognition or classification. This reduction in complexity facilitates more efficient analysis while preserving the key details needed for the next stages of the system.

5.Detection and Classification

Detection and classification represent the final stages of the image processing pipeline, where the focus shifts to identifying



and categorizing objects within the image. After detecting the objects, they are assigned to predefined categories using machine learning models such as Support Vector Machines (SVM), Decision Trees, or deep learning techniques like Convolutional Neural Networks (CNNs). This step enables the system to make informed decisions based on the recognized objects, including identifying specific objects, classifying images by content, or detecting anomalies and patterns. Detection and classification play a crucial role in various fields such as medical imaging, facial recognition, and object tracking in video analytics



Fig -3: Randomly selected plant images after preprocessing

A.MODEL TRAINING AND VALIDATION Model Performance Metrics

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
CNN (Custom)	98.9	98.4	98.5	98.45

Table 1.CNN Performance

These values are derived from the final evaluation in model:

- Training Accuracy: 96.7%
- Validation Accuracy: 98.7%
- Test Accuracy: 98.9%

B. PREDICTION EXPLANIBILITY

In the Plant Disease Detection and Classification project, prediction explainability focuses on understanding and interpreting how the trained model arrives at its decisions for each input image. To achieve this, the system analyzes the influence of different regions of the image by slightly altering or masking portions of the leaf and observing how these changes affect the model's output. By examining these variations, it becomes possible to identify which parts of the image contribute most to the final classification. The results of these experiments are then used to build a simple, interpretable model that mimics the behavior of the complex deep learning model on a local scale. This approach helps in visualizing and interpreting the decision-making process of the model. it easier to trust and making validate the predictions, especially when diagnosing specific plant diseases from leaf images.

4. EXPERIMENTAL ANALYSIS A.EXPERIMENTAL SETUP

1)DATASET

In this paper, the dataset named as 'New Plant Diseases' is taken from Kaggle. This dataset a rich collection of 87,000 images, providing a diverse range of plant diseases 38 distinct classes across 14 plant species. Determining the economic relevance of diseases per plant depends on various factors such as the crop's economic importance, the severity and prevalence of the disease, the cost of control measures, and the potential yield losses. However, some insights are provided for diseases which are most economically relevant, per plant as shown in 'red' color in Table 1 and healthy instances are shown in 'green' color. Moreover, this dataset is imbalanced. It means that number of images in each class are not equal.

2)PERFORMANCE MEASURE

This section presents the quantitative metrics employed to assess classifier performance. The classifier is attempting to identify what kind of disease it is given plant species. It can also identify the leaf as healthy leaf which is the case in Blueberry, Raspberry, and Soyabean. In classification problems where results are categorized into positive or negative classes, the evaluation involves four potential states, often referred to as the confusion matrix.

i)True positive (TP): Correctly identifying instances of the positive class

ii)True negative (TN): Correctly identifying instances of the negative class

iii)False positive (FP): Incorrectly classifying instances as belonging to the positive class

iv)False negative (FN): Incorrectly classifying instances as belonging to the negative class

The performance of the results is assessed through accuracy, precision, and recall, computed as follows:

Accuracy =	TP + TN	(1)
1100001000	$\overline{FP+TN+TP+FN}$	
Precision =	TP	(2)
	TP + FP	
Recall =	TP	(3)
	TP + FN	

B.PARAMETER SETTINGS

The hyperparameter settings employed for the classification of plant diseases are outlined in Table2.

Parameters	Values
Input Size	224*224
Batch Size	.32
Error Function	Categorical Cross Entropy
Activation	Relu
Optimizer	ADAM
Learning Rate	0.0001
Dropout	0.5
Train/Validation	80% / 20%

Table 2.Parameters

L



C.RESULTS AND DISCUSSION

1.COMPARISON WITH EXSISTING SYSTEMS

Traditional methods for plant disease detection primarily depend on manual inspection by farmers or agricultural experts. These approaches, while based on experience and observational skills, typically offer an accuracy of about 70–80%. However, they are prone to human error and subjectivity, which can lead to inconsistent or delayed diagnoses. Moreover, such methods lack scalability, making them inefficient when dealing with large-scale agricultural data or widespread monitoring.

In contrast, machine learning (ML)-based methods have significantly improved the accuracy and efficiency of plant disease classification, achieving accuracy rates between 87% and 100%. These techniques leverage advanced algorithms capable of processing and analyzing large volumes of data with speed and precision. By reducing human intervention, ML systems minimize errors and enhance the objectivity of disease detection. Additionally, these methods are highly scalable and well-suited for real-time monitoring and decision-making in modern agricultural systems, making them a more robust and reliable solution for plant disease management.

2.UML DIAGRAM



3.USER INTERFACE

The user interface (UI) of the Plant Disease Detection and Classification project is designed to be intuitive, responsive, and user-friendly, catering to both technical and non-technical users. It allows users to easily upload images of plant leaves affected by potential diseases. Once the image is uploaded, the system processes it through a trained machine learning model and displays the predicted disease along with its confidence level. The interface also provides additional information such as symptoms, preventive measures, and suggested treatments for the identified disease. A clean layout with clearly labeled sections ensures smooth navigation, while visual cues and feedback messages enhance the user experience. The UI is optimized for both desktop and mobile devices, ensuring accessibility in field conditions where quick decision-making is crucial. This interactive and informative interface bridges the gap between advanced technology and practical agricultural needs, empowering users to take timely and informed action.



Fig5. Front Page



Fig6. Diagnosis Section



Fig7. Image Upload section

Fig -4: UML Diagram

L



Volume: 09 Issue: 04 | April - 2025

agricultural professionals with accessible, real-time solutions. With further improvements and real-world deployment, such ML-driven systems have the potential to revolutionize crop

6.REFERENCES

farming at a global scale.

[1] Anamika Jain, Anagha Langhe, Harsh Choudhary, Ashutosh Mishra, "Plant Disease detection using Machine Learning", 2024

management practices, increase yields, and support sustainable

[2] Jamalbek Tussupov, Moldir Yessenova, Gulzira Abdikerimova, Aidyn Aimbetov, Kazbek Baktybekov, "Analysis of formal concepts for verification of pests and diseases of crops using Machine Learning", 2024

Muhammad [3] Natasha nigar,Hafiz Faisal, Muhammad Umer, Olukayode Oki, and Jose Manappattukunnel lukose, "Improving Plant Disease Classification With Deep-Learning-Based Prediction Model Using Explainable Artificial Intelligence",2024

[4] Anita Shrotriya, Akhilesh kumar sharma, Shrikanth Prabhu, and Amit Kumar Bharwa, "An Approach Toward Classifying Plant-Leaf Diseases and Comparisons With the Conventional Classification", 2024

[5] Wubetu Barud Demilie1, "Plant disease detection and classification techniques: a comparative study of the performances", 2024

[6] Diana Susan Joseph, Pranav M. Pawar, Kaustubh Chakradeo, "Real-Time Plant Disease Dataset Development and Detection Using Deep Learning", 2023

[7] Vasileios Balafas, Emmanouil Karantoumanis, Malamati Louta, Nikolaos Ploskas, "Machine Learning and Deep Learning for Plant Disease Classification and Detection", 2023

[8] Wasswa Shafik, Ali Tufail, Abdallah Namoun, Liyanage Chandratilak De Silva, Rosyzie Anna Apong, "A Systematic Literature Review on Plant Disease Detection", 2023

[9] Vishnu S Babu, R. Satheesh Kumar, R. Sunder, "A Comparative Study on Disease Detection of Plants using Machine Learning Techniques", 2021

[10] Pallavi Dwivedi, Sumit Kumar, Surbhi Vijh, Yatender Chaturvedi, "Study of Machine Learning Techniques for Plant Disease Recognition in Agriculture", 2021

[11] Chunduri Madhurya, Emerson Ajith Jubilson, "YR2S: Efficient deep learning technique for detecting and classifying plant leaf diseases",2024

[12] Liu and X. Wang, "Plant diseases and pests detection based on deep learning: a review," Plant Methods, vol. 17, no. 22,2021, doi 10.1186/s13007-021-00722-9.

[13] Schwarz Schuler, Joao Paulo & Romaní, Santiago & Abdel-Nasser, Mohamed & Rashwan, Hatem & Puig, Domenec. (2021). Reliable Deep Learning Plant Leaf Disease Classification Based on Light-Chroma Separated Branches. 10.3233/FAIA210157.

[14] Kulkarni, Pranesh & Karwande, Atharva & Kolhe, Tejas & Kamble, Soham & Joshi, Akshay & Wyawahare, Medha. (2021). Plant Disease.

[15] Ramesh Maniyath, Shima & P V, Vinod & M, Niveditha & R, Pooja & N, Prasad & N, Shashank & Ram, Hebbar. (2018).Plant

[16] S. Bhattarai. New Plant Diseases Dataset. [Online]. Available:



Fig8. Prediction Section



Fig9. About Us Section



Fig10. Contact Section

5.CONCLUSION

The Plant Disease Detection and Classification system marks a significant advancement over traditional methods of disease diagnosis in agriculture. Traditional approaches, which rely heavily on human observation and expert knowledge, often suffer from subjectivity and limited scalability. These methods typically achieve an accuracy of around 70-80%, making them less reliable for large-scale or real-time applications, especially when early and precise disease identification is crucial for effective crop management.

In contrast, the machine learning-based system developed in this project demonstrates a notable improvement in both accuracy and efficiency. By leveraging advanced classification algorithms and deep learning models, the system has achieved an accuracy range of 87-100% in identifying and classifying various plant diseases from leaf images. This high level of precision significantly reduces human error, ensures consistent results, and enables rapid analysis of vast datasets, which is essential for timely decision-making and disease control.

Overall, the integration of intelligent algorithms with a user-

friendly interface provides a powerful tool for modern agriculture. The system not only enhances the accuracy and speed of plant disease detection but also empowers farmers and

L



https://www.kaggle.com/datasets/vipoooool/new-plant-diseasesdataset

[17] S. Ramesh, R. Hebbar, M. Niveditha, R. Pooja, N. Shashank, and P. V. Vinod, "Plant disease detection using machine learning," in Proc. Int. Conf. Design Innov. 3Cs Compute Communicate Control (ICDI3C), Apr. 2018, pp. 41–45.

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

[19] M. de Benito Fernández, D. L. Martínez, A. González-Briones, P. Chamoso, and E. S. Corchado, "Evaluation of XAI models for interpretation of deep learning techniques' results in automated plant disease diagnosis," in Proc. Sustain. Smart Cities Territories Int. Conf. Cham, Switzerland: Springer, 2023, pp. 417–428.

[20] M. Nahiduzzaman, M. E. H. Chowdhury, A. Salam, E. Nahid, F. Ahmed, N. Al-Emadi, M. A. Ayari, A. Khandakar, and J. Haider, "Explainable deep learning model for automatic mulberry leaf disease classification," Frontiers Plant Sci., vol. 14, Sep. 2023, Art. no. 1175515.