Black Pepper Leaf Disease Detection and Crop Yield Prediction

Prof. Nithya B P[1], Prof. Nivin K S[2], Devika S Nambiar[3], Pallavi M P[4], Nandhana M Nambiar[5], Reshikesh[6]
[1,2,3,4,5,6]*Department of Artificial Intelligence and Machine Learning, Srinivas Institute of Technology, Visveswaraya Technological University, Valachil, Mangaluru, India*
doi.org/10.64643/IJIRTV12I12-206674-459

***Abstract*—Black pepper is an important commercial crop in tropical regions, but its production is often affected by leaf diseases and unstable yield conditions. Traditional methods used for identifying these issues are usually slow and depend heavily on manual observation, which may not always be accurate. This can lead to delayed treatment and reduced productivity. This study presents a smart system that combines image-based disease detection with crop yield prediction. A Convolutional Neural Network (CNN) is used to examine leaf images and identify diseases based on visible features such as color variation, texture changes, and spots. At the same time, a regression-based machine learning model predicts crop yield using soil nutrient values, mainly nitrogen (N), phosphorus (P), and potassium (K).**

***Index Terms*—Black Pepper, Leaf Disease Detection, Crop Yield Prediction, Machine Learning, CNN, Image Processing**

I. INTRODUCTION

Black pepper is widely cultivated and serves as a major source of income for farmers in many regions. However, its cultivation is highly sensitive to plant diseases and soil nutrient imbalance. Leaf diseases usually develop slowly, making them difficult to detect at an early stage through simple observation. In addition, crop yield depends greatly on soil nutrients such as nitrogen, phosphorus, and potassium.

Traditional farming practices rely heavily on experience-based assessment, which can lead to elayed intervention and inconsistent yield outcomes.

With the growth of digital technologies, there is an opportunity to improve agricultural practices through automation. Image-based systems can help identify diseases by analyzing leaf patterns, while machine learning models can estimate crop yield using soil data. However, most existing systems handle these tasks separately.

Image-based disease detection has shown promise in identifying plant health issues by analyzing leaf surface patterns, while data-driven yield estimation models provide insight into productivity trends based on nutrient availability. However, many existing approaches address these problems independently, limiting their effectiveness in real-world agricultural scenarios. With the growth of digital technologies, there is an opportunity to improve agricultural practices through automation. Image-based systems can help identify diseases by analyzing leaf patterns, while machine learning models can estimate crop yield using soil data. However, most existing systems handle these tasks separately. This project aims to combine both approaches into a single system. By using leaf images along with nutrient data, the system provides a more accurate and practical solution for monitoring plant health and predicting productivity.

II. LITERATURE REVIEW

Earlier approaches to plant disease detection mainly relied on manual inspection and laboratory analysis. Farmers or experts would visually examine plants for symptoms such as discoloration, spots, or wilting. In some cases, plant samples were tested in laboratories for detailed analysis. In this approach, farmers or agricultural experts visually examined crops to identify symptoms like discoloration, spots, or wilting. Recent developments in machine learning and deep learning have introduced automated solutions. Image processing techniques are now widely used to detect plant diseases, while data-driven models are used to predict crop yield. Combining these methods can improve both accuracy and efficiency, making them more suitable for real-world agricultural use. Although these methods were reliable, they were highly timeconsuming, labor-intensive, and required significant domain expertise. effectively.

III. SYSTEM FEATURES AND FUNCTIONALITY

The proposed system is developed to assist farmers by offering automated evaluation of black pepper plant health. It enables users to upload leaf images for identifying diseases and enter soil nutrient values for estimating crop yield. The system handles these inputs using trained machine learning models and presents the output in a simple, easy-to-understand format. Key features include automatic image preprocessing, classification of diseases based on visible symptoms, analysis of nutrient data, and quick generation of results. The system focuses on ease of use, allowing individuals with limited technical background to operate it without difficulty. By combining disease identification and yield estimation in one platform, the system helps farmers make better and more informed agricultural decisions.

A.Leaf Disease Detection

The leaf disease detection module is created to help users identify plant health conditions in a simple and efficient manner using image-based analysis. In this feature, users can upload images of plant leaves through the system interface using an easy upload option. Once the image is provided, it is processed by a trained Convolutional Neural Network (CNN) model, which automatically examines different visual characteristics such as color changes, texture variations, spots, and deformities found on the leaf surface. Unlike traditional approaches, the CNN does not depend on manual feature extraction, as it learns important patterns directly from the dataset during training After processing the image, the system classifies the leaf into groups such as healthy or diseased, and in certain cases, it can also determine the exact type of disease affecting the plant. The results are produced quickly, enabling users to take prompt action and reduce the spread of infection. This feature greatly lowers the reliance on agricultural experts and allows farmers to monitor crop health more effectively using digital technology

B.Crop Yield Prediction

The crop yield prediction feature is designed to estimate expected agricultural output using important soil and environmental factors. In this module, users are asked to provide numerical values representing soil nutrients, specifically Nitrogen (N), Phosphorus (P), and Potassium (K), which play a key role in plant growth. These inputs are then processed by a machine learning regression model that has been trained using past agricultural datasets. The model studies the relationship between nutrient levels and crop productivity to generate an estimated yield value. This allows farmers to better understand how current soil conditions can influence production and helps them make informed choices regarding fertilizer application and crop planning. By offering predictive insights, this feature improves resource management and supports the optimization of agricultural output while minimizing unnecessary expenses. Unnecessary costs.

C. Multi-Modal Data Prediction

By combining these two types of inputs, the system delivers more accurate and meaningful results. This multi-modal approach ensures that both visible signs and underlying factors are included in the analysis. The first method of data collection involves gathering images of black pepper leaves. These images act as the main input for the disease detection module. Leaf images are captured under different lighting conditions and backgrounds to represent real field situations. The dataset contains both healthy and diseased leaves, enabling the model to learn visual differences such as color variations, texture changes, and disease symptoms. This visual information forms the base for training and evaluating the convolutional neural network used for disease classification

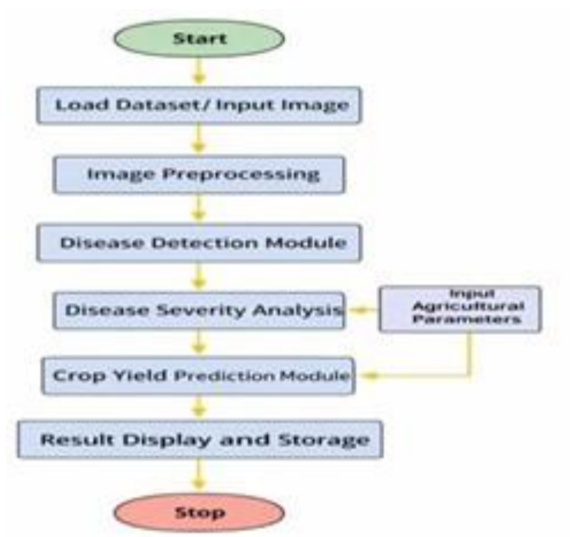

Fig 3.1: DFD Diagram

D. Crop Yield Prediction Using NPK Values

Along with disease detection, the system also offers crop yield prediction based on soil nutrient values. Users can enter Nitrogen (N), Phosphorus (P), and Potassium (K) levels, which are important indicators of soil fertility. These inputs are processed by a machine learning regression model to estimate the expected crop yield. This feature supports better nutrient management planning and helps farmers make informed decisions about fertilizer application. Along with disease detection, the system also offers crop yield prediction based on soil nutrient values. Users can enter Nitrogen (N), Phosphorus (P), and Potassium (K) levels, which are important indicators of soil fertility. A machine learning regression model processes these inputs to estimate expected crop yield. This feature supports nutrient planning and helps farmers make informed decisions about fertilizer usage

IV. METHODOLOGY

The system follows a well-defined methodology that includes several stages, starting from data collection and ending with result generation. The workflow begins with acquiring data and continues through processing to final output. Leaf images are preprocessed to remove noise and standardize their size. Feature extraction methods are used to identify important visual patterns related to common black pepper diseases. These extracted features are then classified using a trained machine learning model. At the same time, nutrient values provided by the user are normalized and passed into a yield prediction model. The outputs from both models are combined to present disease condition and estimated crop yield. This approach ensures that both visual symptoms and soil fertility factors are analyzed together, improving overall system reliability

A.Data Collection.

The dataset contains images of black pepper leaves, including both healthy and diseased samples, collected under different environmental conditions. In addition, soil nutrient data consisting of nitrogen, phosphorus, and potassium values is gathered. The quality and variation of the data play a key role in determining system performance. To achieve effective learning and realistic prediction results, the project uses two main data sources: leaf image data and soil nutrient data

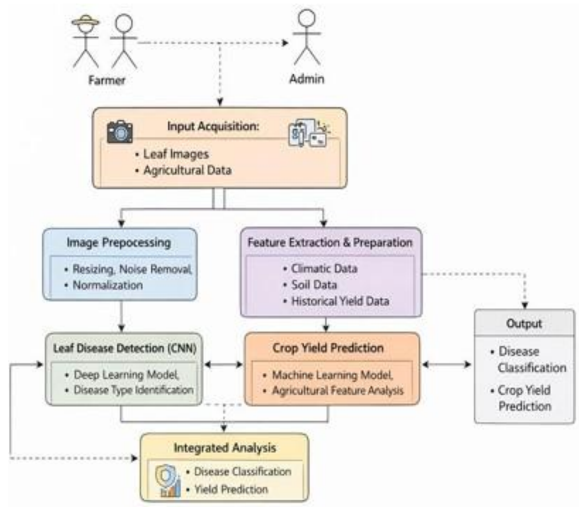

Fig4.1: Dataset Collection

B.Data Preprocessing.

Before providing the data to the models, preprocessing is carried out to maintain consistency and improve accuracy. Image preprocessing involves resizing images to a fixed size, normalizing pixel values, and reducing noise. Data augmentation methods such as rotation and flipping are used to increase dataset diversity. For numerical data, preprocessing includes scaling values and handling missing entries. For leaf image data, preprocessing starts by resizing all images to a consistent dimension. This step ensures compatibility with the convolutional neural network and reduces computational load. The images are then normalized by adjusting pixel intensity values to a standard range, which helps stabilize training and speeds up convergence. Noise reduction methods are applied to reduce the effect of unwanted background elements, shadows, or lighting changes that could affect disease recognition.

C. Modal Training.

The disease detection module relies on a convolutional neural network that is trained using labeled leaf images. The model learns to recognize patterns related to different plant diseases. For crop yield prediction, a regression model is trained with nutrient data to understand the connection between soil fertility and crop output. To improve the robustness of the disease detection model, image augmentation methods such as rotation, flipping, and zooming are applied during the training phase. These techniques artificially expand the dataset and help the model learn stable features, which reduces overfitting and improves its ability to generalize to new, unseen images. Augmentation is especially useful in agricultural datasets, where collecting large amounts of labeled image data can be difficult

D. Hybrid Analysis.

The results from both models are combined to produce a final analysis. This integrated method ensures that both disease conditions and nutrient levels are taken into account, leading to more dependable predictions, as it brings together outputs from different analytical modules into a single decision-making framework. Rather than handling disease detection and yield prediction as separate tasks, this hybrid approach merges insights from both visual and numerical data to offer a more accurate and meaningful assessment of overall crop condition.

V. SYSTEM REQUIREMENTS

A.Hardware Requirements

The system can be implemented on a standard computer or laptop with moderate processing capability. A system with at least adequate RAM and a modern processor is recommended for the efficient execution of machine learning models. The project is based on clearly defined hardware and software requirements. Since it involves image processing, machine learning model execution, and data analysis, the system should be able to handle moderate computational workloads while still being accessible for both academic and practical use. The requirements described in this section ensure smooth model training, accurate prediction, and consistent system performance.

B.Software requirements

The system is developed using open-source tools and technologies. Python is used as the main programming language because of its strong support for machine learning and data analysis. Libraries such as TensorFlow, OpenCV, NumPy, Pandas, and Scikit-learn are used for building models and processing data. Development environments like Jupyter Notebook and Visual Studio Code are used for coding and testing. These environments, including integrated development tools and notebooks, help in experimentation, debugging, and visualization of results

VI. EXERIMENTAL RESULTS

A. Crop Yield Prediction Results

The regression model was able to predict crop yield effectively using nutrient values. The results indicated that well-balanced nutrient levels lead to improved yield, while deficiencies or excessive levels reduce productivity. The crop yield prediction module generated consistent and reliable estimates based on Nitrogen, Phosphorus, and Potassium (NPK) values provided as input. The regression model successfully learned the relationship between nutrient levels and expected crop output, showing that proper nutrient balance results in higher predicted yield, whereas deficiency or excess has a negative effect on productivity. Healthy leaves were recognized by their uniform green color and smooth texture, while diseased leaves displayed signs such as discoloration, spots, and uneven patterns. The model showed stable performance across various test samples.

B. Disease Detection Results

The CNN model achieved strong accuracy in classifying leaf images. Healthy leaves were identified by their consistent green color and smooth surface, while diseased leaves showed visible symptoms such as discoloration, spots, and irregular patterns. The model maintained consistent performance across different test samples.

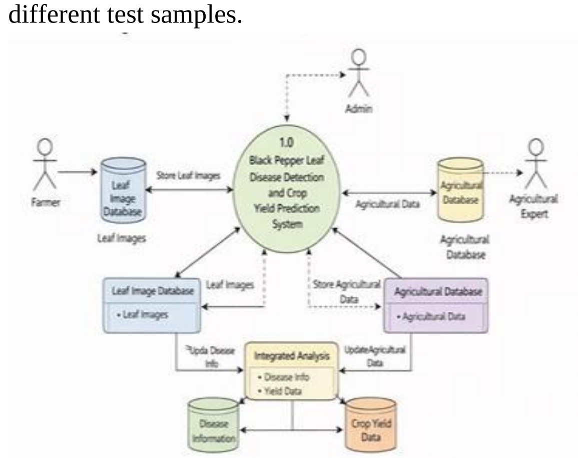

Fig 6.1: Confusion Matrix

During experimentation, leaf images representing different health conditions were processed through the trained CNN model. Healthy leaves were correctly identified based on their uniform green coloration and smooth texture, while diseased leaves exhibited noticeable visual deviations such as yellowing, dark spots, and uneven surface patterns. The model successfully learned these distinguishing features during training, enabling accurate classification when tested on unseen images.

C. Hybrid Analysis Results

The combined analysis offered more practical and realistic insights. For instance, even when leaves appeared healthy, poor nutrient levels led to lower yield predictions. In the same way, diseased plants showed reduced yield even when nutrients were sufficient. This highlights the importance of combining both types of analysis. In situations where leaves looked healthy but nutrient levels were not optimal, the system predicted lower yield, showing the hidden effect of soil conditions. On the other hand, even with proper nutrient levels, the presence of disease resulted in reduced yield predictions. This integrated analysis reflects real agricultural conditions more accurately than separate evaluation and improves the reliability of decision making and enhances decision-making reliability.

D. Performance Evaluation

The system showed quick response time and stable performance. It was capable of processing multiple inputs efficiently without noticeable delays. The results remained consistent and dependable, making the system suitable for real-world applications.

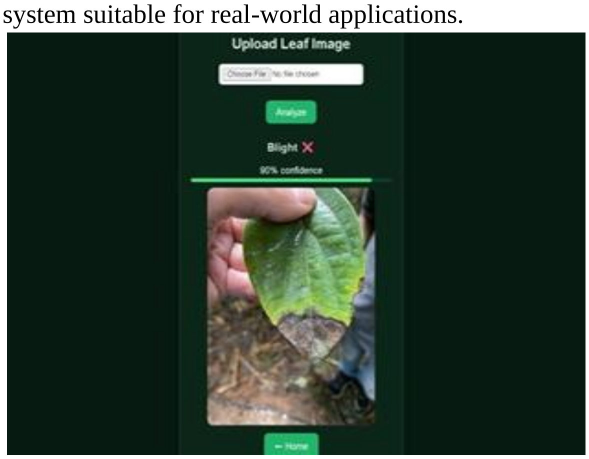

Fig 6.2: Outcome

The system demonstrated efficient performance by generating disease detection and yield prediction results within a short response time. The predictions remained consistent across repeated inputs, indicating stability and reliability. The results also show that the system can manage multiple inputs without performance issues, making it appropriate for practical deployment.

Fig 6.3: User Interface

VII. CONCLUSION

This project presents an integrated system for detecting black pepper leaf diseases and predicting crop yield using machine learning techniques. By combining image analysis with nutrient data, the system offers a complete evaluation of plant health and overall productivity.

The results indicate that the proposed method improves accuracy and supports better decision-making in agriculture. Early detection of diseases helps reduce crop loss, while reliable yield prediction assists in efficient management of resources.

Future improvements may include expanding the dataset, adding environmental factors such as temperature and humidity, and developing a mobile application for easier accessibility. Overall, the system contributes to the advancement of smart and sustainable agricultural practices.

The disease detection module effectively distinguishes between healthy and diseased black pepper leaves by analyzing visual features such as color differences, texture patterns, and visible symptoms. Early identification of leaf diseases allows timely action, reducing the spread of infection and minimizing yield loss. The crop yield prediction module further supports decision-making by estimating expected output based on Nitrogen, Phosphorus, and Potassium values, highlighting the importance of soil nutrients in plant growth and productivity.

ACKNOWLEDGMENT

The authors would like to express their sincere appreciation to their project guide, Prof. Nithya B P, for her continuous support, valuable advice, and constant encouragement throughout the progress of this project. Her expertise and thoughtful feedback played an important role in shaping the direction and overall quality of this work. Her guidance helped the team overcome difficulties and improve the final outcome of the project.

The authors also extend their heartfelt thanks to the faculty members of the Department of Artificial Intelligence and Machine Learning for providing a strong academic base and technical knowledge. Their lectures, discussions, and support during the course of study contributed greatly to the successful completion of this research work.

The authors would like to acknowledge the Head of the Department and the management of Srinivas Institute of Technology for providing the required infrastructure, resources, and a supportive environment to carry out this project. Their encouragement and assistance were essential in completing this work efficiently.

The authors are also grateful to their friends and classmates for their cooperation, helpful suggestions, and moral support during different stages of the project. Their contributions, whether direct or indirect, helped improve the quality of the work and maintain consistency throughout the process.

Finally, the authors express their deepest gratitude to their parents and family members for their unconditional support, encouragement, and motivation. Their constant belief and understanding have been a strong source of inspiration, helping in the successful completion of this project.

REFERENCES

[1] S. P. Mohanty, D. P. Hughes, and M. Salathé, “Using Deep Learning for Image-Based Plant Disease Detection,” *Frontiers in Plant Science*, vol. 7, pp. 1419, 2016.

[2] K. P. Ferentinos, “Deep Learning Models for Plant Disease Diagnosis and Detection,” Computers and Electronics in Agriculture, vol. 145, pp. 311–318, 2018.

[3] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, “Deep Neural Networks for Recognition of Plant Diseases by Leaf Image Classification,” Computational Intelligence and Neuroscience, vol. 2016, Art. no. 3289801, 2016.

[4] A. Kamilaris and F. X. Prenafeta-Boldú, “Deep Learning in Agriculture: A Survey,” Computers and Electronics in Agriculture, vol. 147, pp. 70–90, 2018.

[5] D. Elavarasan and D. R. Vincent, “Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agriculture,” IEEE Access, vol. 8, pp. 86886–86901, 2020.

[6] J. H. Jeong, J. P. Resop, N. D. Mueller, et al., “Random Forests for Global and Regional Crop Yield Predictions,” PLOS ONE, vol. 11, no. 6, 2016.

[7] P. Kaur and R. Kumar, “Crop Yield Prediction Using Machine Learning Techniques,” Journal of Computational Science, vol. 13, pp. 1–10, 2019.

[8] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, “Machine Learning in Agriculture: A Review,” Sensors, vol. 18, no. 8, Art. no. 2674, 2018.

[9] A. Chlingaryan, S. Sukkarieh, and B. Whelan, “Machine Learning Approaches for Crop Yield Prediction and Nitrogen Status Estimation in Precision Agriculture: A Review,” Computers and Electronics in Agriculture, vol. 151, pp. 61–69, 2018.

[10] N. Islam, M. Rashid, and M. Hossain, “Leaf Disease Detection Using Image Processing and Machine Learning,” International Journal of Computer Applications, vol. 181, no. 43, pp. 1–6, 2018.

[11] A. Kumar and M. Vani, “Soil Nutrient-Based Crop Yield Prediction Using Machine Learning,” International Journal of Advanced Computer Science and Applications, vol. 10, no. 6, pp. 1–7, 2019.

[12] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, “A Comparative Study of Fine-Tuning Deep Learning Models for Plant Disease Identification,” Computers and Electronics in Agriculture, vol. 161, pp. 272–279, 2019.

[13] J. A. Pandian and R. Ramesh, “Precision Agriculture Using IoT and Machine Learning for Crop Monitoring,” International Journal of Engineering and Technology, vol. 7, pp. 140–145, 2018.

[14] M. Torky and A. E. Hassanein, “Integrating Deep Learning and Agricultural Data for Plant Disease Prediction,” Procedia Computer Science, vol. 163, pp. 456–463, 2019.

[15] S. Ramesh, R. Hebbar, and S. Nagesh, “Nutrient-Based Crop Productivity Analysis Using Data-Driven Models,” Journal of Agricultural Informatics, vol. 11, no. 2, pp. 15–25, 2020.
[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in Advances in Neural Information Processing Systems (NeurIPS), 2012, pp. 1097–1105.
[17] Y. LeCun, Y. Bengio, and G. Hinton, “Deep Learning,” Nature, vol. 521, no. 7553, pp. 436–444, 2015.
[18] R. Girshick, “Fast R-CNN,” in Proc. IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1440–1448.
[19] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.