10.48047/jocaaa.2025.34.05.45

Plant health diagnosis: AI-based classification on plant health issues for agricultural optimization

L. Anantha Lakshmi[1*], Guttikonda Rithwik Reddy[2], Gaddam Adhikeshava Reddy[2], Eeru Vivek[2], Yekkaluri Nagarjun Reddy[2], Boga Dheeraj Palhani[2]

[1]Assistant Professor, [2]UG Student, [1,2]Department of Computer Science and Engineering (AI&ML), [1,2]Malla Reddy Engineering College and Management Sciences, Kistapur, Medchal, 501401, Telangana, India

*Corresponding author: ananthalakshmi@mrem.ac.in

Received: 02.03.2025 Revised: 15.04.2025 Accepted: 25.05.2025

Abstract

Plant disease classification plays a vital role in improving agricultural productivity and sustainability. Traditional disease identification methods, such as manual inspection and laboratory analysis, are often time-consuming, subjective, and costly, making them unsuitable for large-scale farming systems. These limitations frequently result in delayed disease detection and substantial crop losses. To address these challenges, this study proposes an automated plant disease classification framework based on machine learning techniques applied to sparse and categorical Internet of Things (IoT) data. The system integrates robust data preprocessing strategies, including missing value imputation, label encoding, and the Synthetic Minority Oversampling Technique (SMOTE), to enhance data quality and resolve class imbalance issues. Multiple machine learning models Gaussian Naive Bayes, Support Vector Machines, K-Nearest Neighbors, and a Decision Tree Classifier are implemented and evaluated within a multi-class classification pipeline. Experimental results indicate that the Decision Tree model outperforms other classifiers, achieving an accuracy of 99.07% along with precision, recall, and F1-scores exceeding 98%. These results demonstrate the effectiveness and reliability of the proposed approach. The system enables early and accurate disease diagnosis, supports targeted pesticide application, and reduces unnecessary chemical usage. Overall, this research presents a scalable, efficient, and data-driven solution for plant disease management, contributing to precision agriculture and sustainable farming practices.

Keywords: Plant Disease Classification, Machine Learning, IoT Data, Decision Tree, SMOTE, Precision Agriculture, Crop Disease Detection

1. Introduction

Agricultural biodiversity is essential for providing humans with food and raw materials and is an essential component of human civilization [1, 2]. The disease can occur when pathogenic organisms such as fungi, bacteria, and nematodes; soil PH; temperature extremes; changes in the quantity of moisture and humidity in the air; and other elements continuously harm a plant. Plant diseases can have an impact on the growth, function, and structure of plants and crops, affecting the people that rely on them. Plant diseases can have a severe impact on crop health, leading to substantial damage, reduced yields, and financial losses for farmers. The majority of farmers still use manual methods to detect and classify plant ailments because it is difficult to do so early on, and this reduces productivity. Agriculture's productivity is a significant economic factor. As a result, disease identification and classification in plants are critical in agricultural industries [3]. If proper precautions are not taken, it can have serious consequences for plants by reducing the quality, quantity, or productivity of the corresponding products or services. Automatic disease detection and classification recognize symptoms

10.48047/jocaaa.2025.34.05.45

at an early stage, i.e., when they first appear on plant leaves, lowering the amount of labor necessary to monitor large farms of crops. According to [4] Plant leaf disease is a major issue in rice production, and the disease has the potential to harm the crop, resulting in a drop in products. Farmers have a difficult time detecting and classifying plant leaf diseases. The timely and precise detection of plant diseases plays a vital role in implementing effective disease management and prevention strategies.

2. Literature Survey

This section offers an extensive overview of prior works in the context of this research, encompassing machine learning models with the goal of identifying their strengths, limitations, and research gaps, thus forming the basis for the proposed method. An extensive review by Li et al. [5] presents the recent research progress in using deep learning technology for crop leaf disease identification. The authors discuss the current trends, challenges, and unresolved issues pertaining to plant leaf disease detection using deep learning and advanced imaging techniques are discussed, with the aim of providing a valuable resource for researchers in the field.

Deep learning techniques were initially applied to plant image recognition with a focus on analyzing leaf vein patterns. A notable study employed a CNN architecture with 3–6 layers to successfully classify three leguminous plant species: white bean, red bean, and soybean [6]. The CNN model extracted and analyzed intricate features of leaf vein structures unique to each species, capturing low-level features such as edges and textures and progressing to more complex patterns. This hierarchical feature extraction enabled effective species differentiation based on leaf vein characteristics. The study demonstrated the potential of CNNs in plant species classification, highlighting the effectiveness of deep learning in capturing subtle biological variations for accurate identification, and laying the groundwork for further applications in plant image recognition.

Subsequently, Mohanty et al. [7] employed a classifier model to accurately identify 14 different crop species and 26 crop diseases, achieving an impressive accuracy of 99.35%. Building upon this work, Kawasaki et al. [8] proposed a CNN-based system specifically designed for the precise recognition of diseases affecting cucumber leaves, achieving a commendable accuracy rate of 94.9%. In an investigation by Ma et al. [9], a deep CNN was employed for the identification and recognition of four diseases in cucumber plants, including downy mildew, anthracnose, powdery mildew, and target leaf spots, based on their distinct symptoms. This study reported a recognition accuracy of 93.4%.

Impressive strides have been made in plant leaf disease recognition; however, the availability of diverse datasets remains limited. CNN training necessitates extensive and diverse datasets, which are currently scarce in the field of plant leaf disease recognition. Consequently, transfer learning has emerged as a viable solution to enhance the robustness of CNN classifiers. By adapting pre-trained CNNs and retraining them with smaller datasets showcasing different distributions, transfer learning has proven to be an effective approach in implementing deep learning models [10]. Promising results have been demonstrated by leveraging pre-trained CNN models from the ImageNet dataset and retraining them for leaf disease recognition [11].

Recently, MTL models have been explored in the field of plant disease detection, enabling the simultaneous execution of multiple tasks. These models can effectively handle tasks, such as severity estimation of leaf disease, identification of plant species, and classification of plant disease concurrently [12]. To address the limitations in information utilization and scalability of existing models, Lee et al. [13] introduced a novel Conditional Multi-Task Learning (CMTL) model in this context. This method employs an MTL mechanism with a conditional scheme that links host species and disease

10.48047/jocaaa.2025.34.05.45

representations, mimicking how plant pathologists infer diseases from species information. This improves plant disease identification performance compared to traditional joint species–disease pair modeling. It efficiently utilizes available information and is scalable, not requiring distinctions between similar species or diseases, thus enhancing accuracy and suitability for real-world applications.

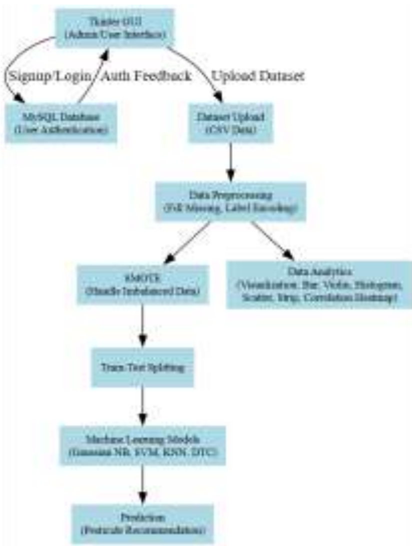
However, ViTs offer advantages over MTL frameworks in plant disease detection. They capture both local and global context information, enabling accurate analysis of plant images. ViTs also generalize well with limited labeled data through pre-training on large-scale datasets like ImageNet. Additionally, they provide flexibility and scalability by adapting to different disease detection tasks with a single model, reducing complexity and computational requirements.

In their work, Thakur et al. [14] introduced a hybrid model, named PlantViT, for automating plant disease detection, which integrates the strengths of a CNN and a ViT, incorporating a multi-head attention module. The model was evaluated on two extensive plant disease detection datasets, namely, PlantVillage and Embrapa. Experimental results showed that the model achieves detection accuracies of 98.61% and 87.87% on the PlantVillage and the Embrapa datasets, respectively. These results outperform the current state-of-the-art methods, highlighting the effectiveness of PlantViT in plant disease detection.

In [15], a lightweight deep learning approach leveraging the ViT architecture is proposed for real-time automated plant disease classification. The study rigorously compares various models, including ViT, traditional CNNs, and a hybrid of CNN and ViT. Extensive training and evaluation across multiple datasets reveal that while attention blocks significantly enhance accuracy, they also introduce latency in predictions. However, integrating attention blocks with CNN blocks strikes an optimal balance between accuracy and speed, offering a robust solution for real-time plant disease classification.

3. Proposed Methodology

This project is a comprehensive Python application designed to diagnose plant diseases using machine learning. It integrates data preprocessing, visualization, machine learning, and a user-friendly interface to facilitate the diagnosis of plant diseases using IoT data. It offers a robust platform for both data scientists (Admins) and end-users, supporting the entire pipeline from data exploration to model training and real-time predictions with actionable recommendations. Here's an overview of its main components



10.48047/jocaaa.2025.34.05.45

Fig. 1: System architecture of proposed plant disease classification with pesticide suggestion from categorical IoT data.

Decision Tree model

A Decision Tree is a flowchart-like structure where internal nodes represent tests on features, branches represent the outcome of these tests, and leaf nodes represent class labels. The tree is constructed by recursively splitting the dataset based on feature values that provide the maximum information gain or the best reduction in impurity (e.g., Gini impurity or entropy).

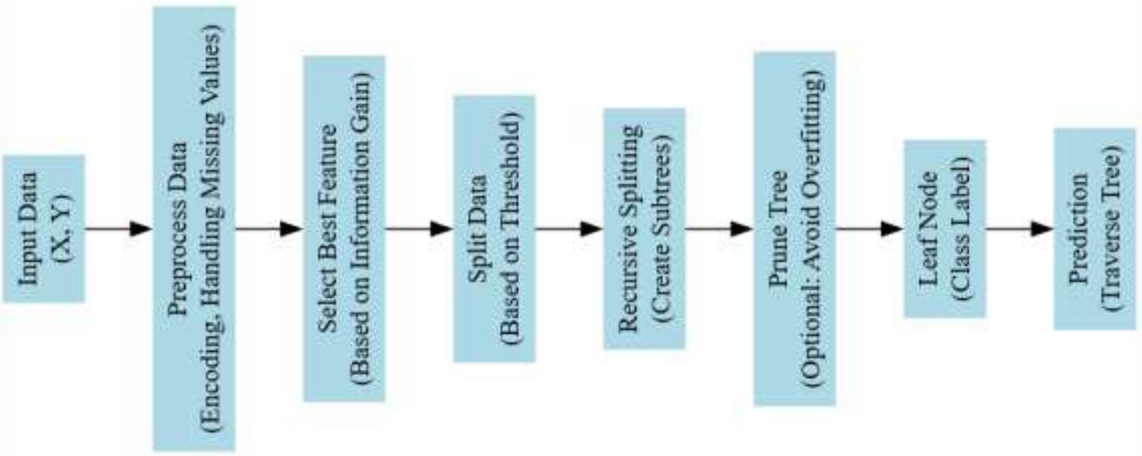


Fig. 2: DTC model workflow.

Key Points:

- **Interpretable Model:** The decision-making process is easy to visualize and interpret.
- **Recursive Splitting:** The tree divides the dataset into smaller subsets based on the most informative features.
- **Prone to Overfitting:** Without proper pruning, decision trees can create overly complex trees that do not generalize well.

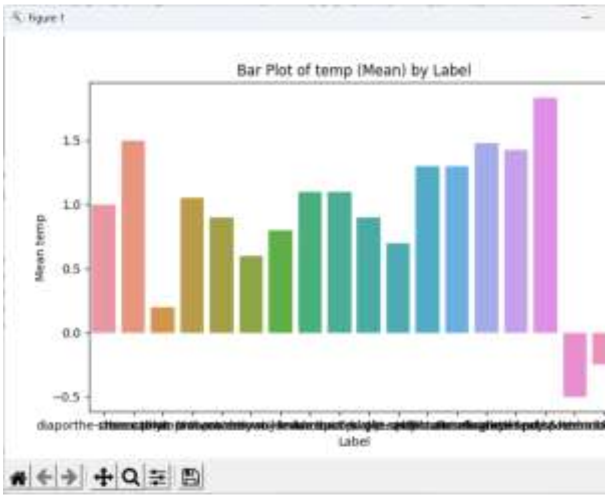
Database Integration

The database integration in this project is centered around managing user authentication for different roles, specifically Admin and User. This integration is implemented using a MySQL database, and the Python module pymysql is employed to establish a connection between the application and the database. Overall, the database integration in this project effectively supports user management by:

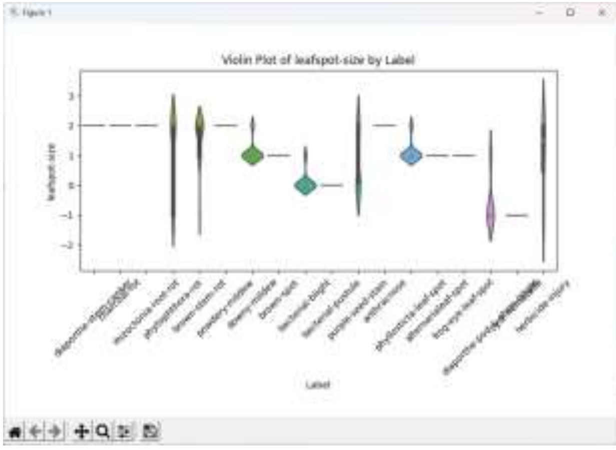
This integration is essential for maintaining the integrity and security of the application while allowing it to scale and manage multiple users with different levels of access.

4. Results Analysis

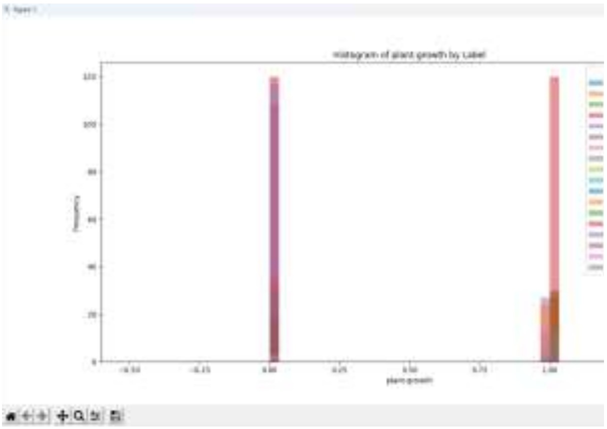
10.48047/jocaaa.2025.34.05.45



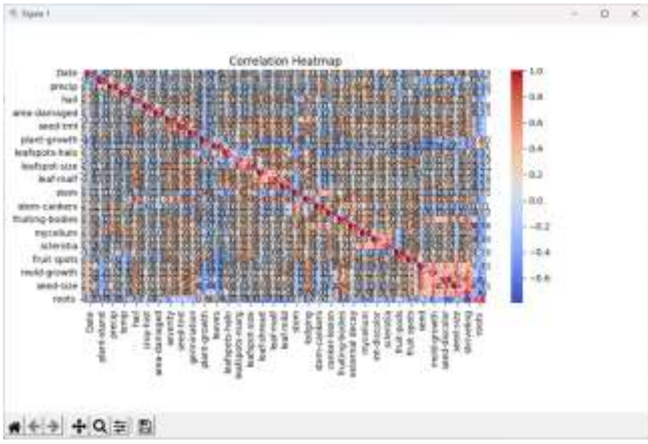
(a)



(b)



(c)



(d)

Fig. 3: Displaying the data visualization with various plots. (a) Bar plot. (b) Violin plot. (c) histogram plot. (d) correlation heat map.

Fig. 3 is divided into sub-figures:

- **(a) Bar Plot:** Shows mean values of a selected numerical feature per disease label.
- **(b) Violin Plot:** Displays the distribution and density of a feature across disease labels.
- **(c) Histogram Plot:** Illustrates the frequency distribution of feature values.
- **(d) Scatter Plot:** Reveals relationships between two features with color-coded disease labels.
- **(e) Strip Plot:** Plots individual data points with jitter to show the spread within groups.
- **(f) Correlation Heat Map:** Visualizes pairwise correlations among numerical features using a color gradient.
- **(g) Another Scatter Plot:** Provides an additional perspective on the interaction between different variables.

10.48047/jocaaa.2025.34.05.45

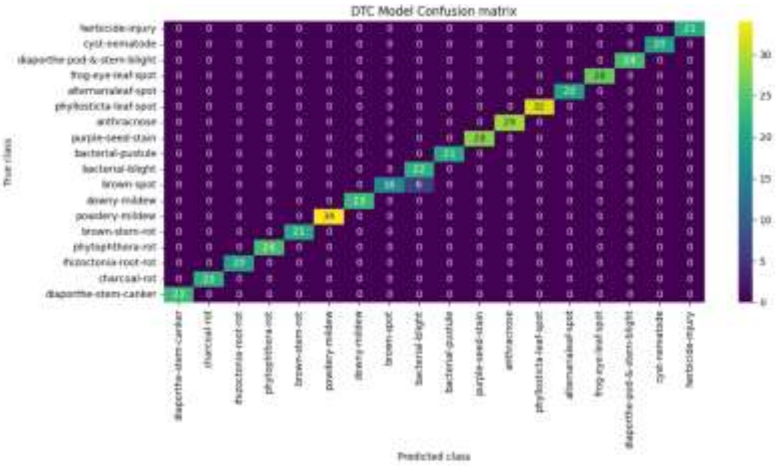


Fig. 4: Confusion matrices obtained using DTC model.

Fig. 4 includes confusion matrices for Showing its classification performance. Illustrating how well the SVM model differentiates between disease classes. Visualizing the performance of the K-Nearest Neighbors algorithm. Depicting the confusion matrix of the decision tree model, which may show superior performance compared to the others.

Fig. 5: Sample prediction and pesticide suggestions on test data.

In the Fig. 5, the application displays the outcome of a prediction on a test dataset. It shows the classified disease and the corresponding pesticide recommendation, thereby completing the full cycle from data input to actionable output for disease management.

Table 1: Performance comparison of existing GNBC, SVM, KNN, and proposed DTC models.

Model/Metric	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
GNBC model	72.68	72.58	72.72	68.17
SVM classifier	92.36	92.92	92.56	92.09
KNN classifier	97.91	97.62	97.56	97.42

10.48047/jocaaa.2025.34.05.45

Proposed DTC model	99.07	98.88	98.76	98.86

Table 1 provides a performance comparison among four different models used for plant disease classification: the Gaussian Naive Bayes Classifier (GNBC), the Support Vector Machine (SVM) classifier, the K-Nearest Neighbors (KNN) classifier, and the proposed Decision Tree Classifier (DTC). The table lists four key metrics Accuracy, Precision, Recall, and F1-score in percentage terms to evaluate how well each model performs. Demonstrates that as we move from simpler to more advanced models, the performance metrics improve considerably. The proposed DTC model, with the highest scores across all metrics, represents the most effective approach in this project for achieving accurate and reliable plant disease classification.

5. Conclusion

In conclusion, the project successfully demonstrates the power of machine learning in revolutionizing plant disease diagnosis through the development of a comprehensive system that integrates data ingestion, preprocessing, model training, and prediction with an intuitive GUI. By leveraging sparse and categorical IoT data, the system effectively addresses the limitations of traditional, manual methods of disease identification, providing rapid, accurate, and scalable diagnosis. Advanced preprocessing techniques, including label encoding and the application of SMOTE, ensure that the data is optimally balanced and suitable for training. Among the various models tested, the proposed Decision Tree Classifier outperformed others in terms of accuracy, precision, recall, and F1-score, underscoring its suitability for the complexities of multi-class classification in agricultural contexts. The robust performance of the system, coupled with actionable pesticide recommendations, positions this approach as a significant step forward in precision agriculture. Ultimately, this innovative framework not only enhances early detection and intervention but also contributes to cost-effective, sustainable, and scalable disease management, paving the way for further advancements in agricultural technology and data-driven decision-making.

References

[1] Tirkey D, Singh KK, Tripathi S. Performance analysis of AI-based solutions for crop disease identification detection, and classification. Smart Agric Technol. 2023. https://doi.org/10.1016/j.atech.2023.100238.

[2] Ramanjot, et al. Plant disease detection and classification: a systematic literature review". Sensors. 2023. https://doi.org/10.3390/s23104769.

[3] Krishnan VG, Deepa J, Rao PV, Divya V, Kaviarasan S. An automated segmentation and classification model for banana leaf disease detection. J Appl Biol Biotechnol. 2022;10(1):213–20. https://doi.org/10.7324/JABB.2021.100126.

[4] S Mathulaprangsan K Lanthong S Patarapuwadol. 2020. Rice Diseases Recognition Using Effective Deep Learning Models. Telecommun. Eng Media Technol with ECTI North Sect Conf Electr Electron Jt Int Conf Digit Arts Comput. https://doi.org/10.1109/ECTIDAMTNCON48261.2020.9090709

[5] Li, L., Zhang, S., Wang, B.: Plant disease detection and classification by deep learning—a review. IEEE Access **9**, 56683–56698 (2021). https://doi.org/10.1109/ACCESS.2021.3077026

[6] Grinblat, G.L., Uzal, L.C., Larese, M.G., Granitto, P.M.: Deep learning for plant identification using vein morphological patterns. Comput. Electron. Agric. **127**, 418–424 (2016). https://doi.org/10.1016/j.compag.2016.07.003

10.48047/jocaaa.2025.34.05.45

[7] Mohanty, S.P., Hughes, D.P., Salathé, M.: Using deep learning for image-based plant disease detection. Front. Plant Sci. **7**, 1419 (2016). https://doi.org/10.3389/fpls.2016.01419

[8] Kawasaki Y, Uga H, Kagiwada S, Iyatomi H Basic study of automated diagnosis of viral plant diseases using convolutional neural networks. In: Advances in visual computing: 11th international symposium ISVC 2015 Las Vegas NV USA December 14–16 2015 Proceedings Part II 11, 638–645. Springer International Publishing (2015)

[9] Ma, J., Du, K., Zheng, F., Zhang, L., Gong, Z., Sun, Z.: A recognition method for cucumber diseases using leaf symptom images based on deep convolutional neural network. Comput. Electron. Agric. **154**, 18–24 (2018). https://doi.org/10.1016/j.compag.2018.08.048

[10] Mukti IZ, Biswas D (2019) Transfer learning based plant diseases detection using ResNet50. In: 2019 4th International conference on electrical information and communication technology (EICT), 1–6. IEEE (2019, December)

[11] Chen, J., Chen, J., Zhang, D., Sun, Y., Nanehkaran, Y.A.: Using deep transfer learning for image-based plant disease identification. Comput. Electron. Agric. **173**, 105393 (2020). https://doi.org/10.1016/j.compag.2020.105393

[12] Keceli, A.S., Kaya, A., Catal, C., Tekinerdogan, B.: Deep learning-based multi-task prediction system for plant disease and species detection. Eco. Inform. **69**, 101679 (2022). https://doi.org/10.1016/j.ecoinf.2022.101679

[13] Lee SH, Goëau, H., Bonnet, P., & Joly, A.: Conditional multi-task learning for plant disease identification. In: 2020 25th international conference on pattern recognition (ICPR), 3320–3327. IEEE (2021, January). https://doi.org/10.1109/ICPR48806.2021.9412180

[14] Thakur PS, Khanna P, Sheorey T, Ojha A Vision transformer for plant disease detection: PlantViT. In: International Conference on Computer Vision and Image Processing, 501–511. Cham: Springer International Publishing (2021, December)

[15] Boukabouya RA, Moussaoui A, Berrimi M: Vision Transformer based models for plant disease detection and diagnosis. In: 2022 5th international symposium on informatics and its applications (ISIA), 1–6. IEEE (2022, November). https://doi.org/10.1109/ISIA.2022.9923797