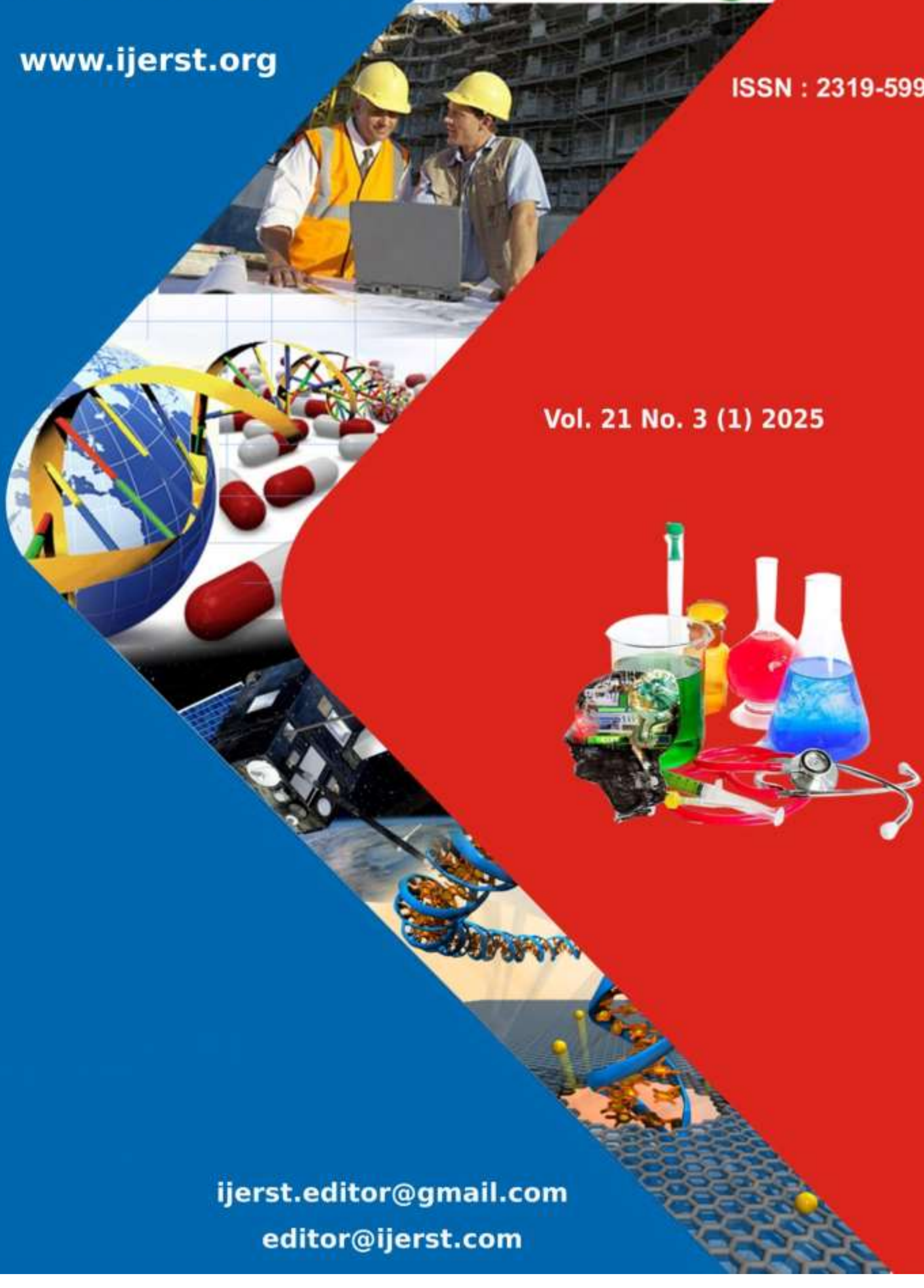
IJERST

International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991



Vol. 21 No. 3 (1) 2025

ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 21, No. 3(1), 2025

Research Paper

AI-POWERED PLANT HEALTH ASSESSMENT: AUTOMATED CLASSIFICATION FOR ENHANCED CROP MONITORING AND PRODUCTIVITY

G. Divya, Banoth Mahender, Gowlikar Mahesh, Amgothu Pavan Kalyan
Department of Computer Science and Engineering (AI&ML), Kommuri Pratap Reddy Institute of Technology, Ghatkesar, Medchal, 500088.

Received: 05-6-2025 Accepted: 03-7-2025 Published: 10-7-2025

ABSTRACT

Plant disease classification plays a vital role in advancing modern agriculture, transitioning from traditional manual diagnosis to intelligent, automated systems powered by machine learning. Historically, identification of plant diseases relied on visual inspections, expert advice, and lab tests—methods that were accurate for small-scale use but often subjective, slow, and inconsistent. These limitations resulted in delayed treatment and substantial crop losses, highlighting the inefficiency and high cost of conventional approaches, especially at scale. To address this, the proposed system introduces an innovative machine learning-based solution capable of accurately classifying plant diseases using sparse and categorical IoT data. It incorporates comprehensive data preprocessing techniques, including handling missing values, label encoding, and class imbalance correction using the Synthetic Minority Oversampling Technique (SMOTE), ensuring a high-quality dataset for model training. The classification pipeline integrates multiple models—Gaussian Naive Bayes, Support Vector Machines, K-Nearest Neighbors, and a novel Decision Tree Classifier. Among these, the Decision Tree model demonstrated superior performance, achieving an accuracy of 99.07% with precision, recall, and F1-scores consistently exceeding 98%, confirming its robustness and reliability. This research is significant in offering real-time, data-driven diagnostics that enable early disease detection and precise pesticide recommendations. It not only improves crop yield and reduces financial losses but also promotes environmentally sustainable agriculture by limiting excessive chemical usage. By overcoming the limitations of traditional methods—such as subjectivity, delay, and lack of scalability—this system presents a transformative approach to plant disease management through advanced machine learning, marking a pivotal shift toward precision agriculture.

Keywords: Plant Disease Classification, Precision Agriculture, Early Disease Detection, Decision Tree Classifier.

1. INTRODUCTION

Agricultural biodiversity is essential for providing humans with food and raw materials and is an essential component of human civilization. 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. 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 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 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 play 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%.

3. PROPOSED METHODOLOGY

The 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.

International Journal of Engineering Research and Science & Technology

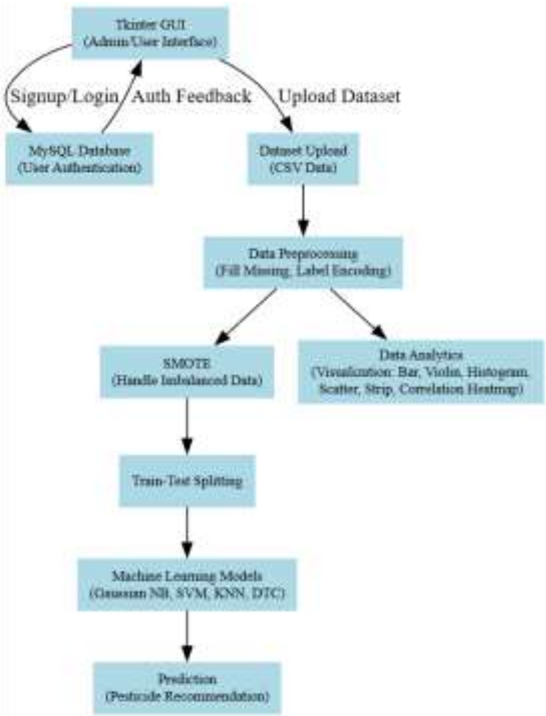


Fig. 1: System architecture of proposed plant disease classification with pesticide suggestion from categorical IoT data.

Decision Tree Classifier 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).

Key Points:

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

Comparative Discussion

After training multiple models such as GNB classifier, SVM classifier, KNN classifier, and DTC model, a comprehensive evaluation of performance metrics such as accuracy, precision, recall, and F1-score revealed that the DTC model achieved the best overall performance. The superior performance of the DTC model can be attributed to its ability to capture complex, non-linear relationships inherent in the plant disease dataset. Unlike probabilistic models like GNB classifier, which assume independence among features, the DTC approach effectively identifies and utilizes the most informative features through recursive splitting. This inherent feature selection mechanism allows the model to handle the diversity of sensor data and the subtle variations in plant disease symptoms more adeptly.

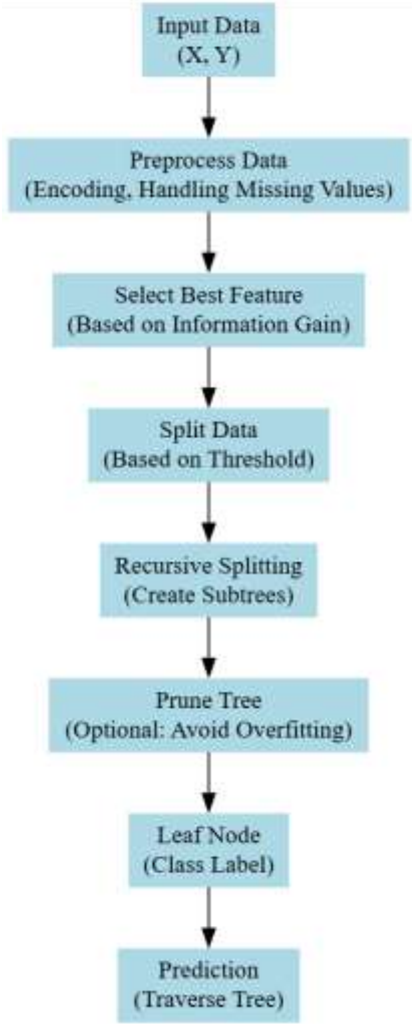


Fig. 2: DTC model workflow.

Additionally, the interpretability of the DTC model provides valuable insights into the decision-making process. The tree structure clearly outlines the sequence of feature-based decisions that lead to a final classification, making it easier to understand which factors most significantly contribute to the diagnosis. This transparency not only aids in model validation but also facilitates further refinement of feature engineering strategies. Moreover, while models like SVM and KNN offer robust performance in many contexts, they may struggle with high-dimensional, sparse, and categorical data unless extensively tuned or transformed. The DTC model, on the other hand, naturally accommodates such complexities without the need for extensive preprocessing beyond the standard data normalization and encoding steps.

4. RESULTS AND DISCUSSION

Fig. 3 in this figure, the admin signup interface is displayed. It shows a dedicated window where new administrative users can enter their username and password. The design focuses on simplicity and clarity, ensuring that admins can quickly register their credentials before accessing the system's management features. Similar to the admin signup.

Fig. 3: Admin signup interface.

Fig. 4 presents the signup interface for regular users. It allows users to create an account by providing necessary credentials. The interface is user-friendly, ensuring that even non-technical users can sign up with ease.

Fig. 4: User signup interface.

Fig. 5: Admin login interface.

Fig. 5 illustrates the login screen designed for administrators. It features fields for entering the username and password, along with clear prompts and buttons that guide the admin through the authentication process, ensuring secure access to the advanced functionalities of the system.

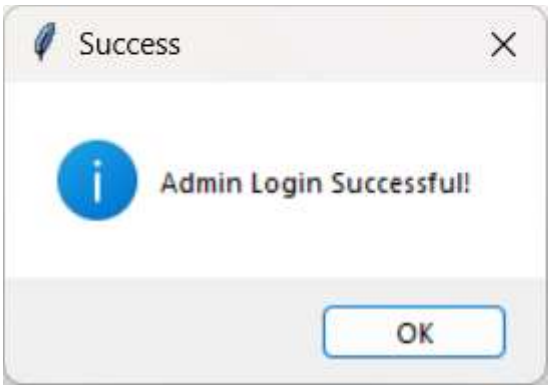


Fig. 6: Successful login interface.

Fig. 6 demonstrate after a successful login, this interface confirms that the credentials are valid, and access is granted. It often shows a welcome message or transition screen that confirms the user has been authenticated, providing immediate feedback to the admin.

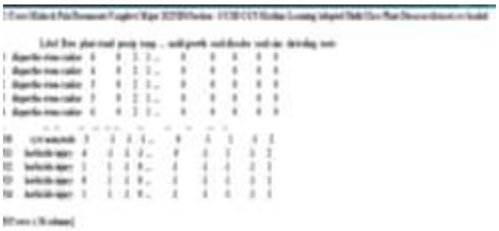
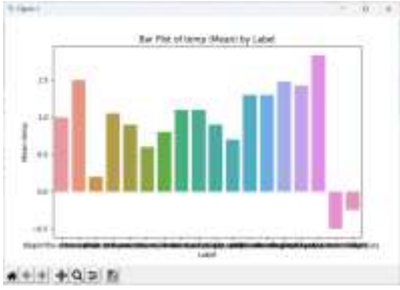


Fig. 7: Illustration of GUI application after uploading the dataset.



(a)

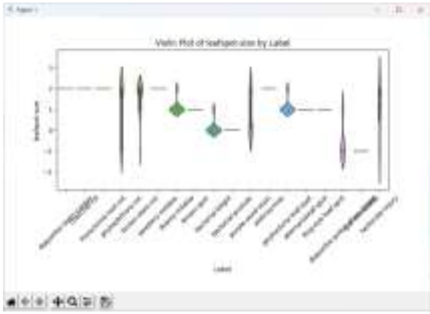


Fig. 8: Displaying the data visualization with various plots. (a) Bar plot. (b) Violin plot. (c) histogram plot.

Fig. 8 is divided into sub-figures:Shows mean values of a selected numerical feature per disease label. Displays the distribution and density of a feature across disease labels. Illustrates the frequency distribution of feature values.

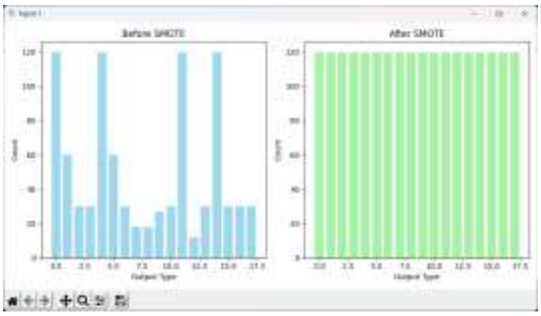
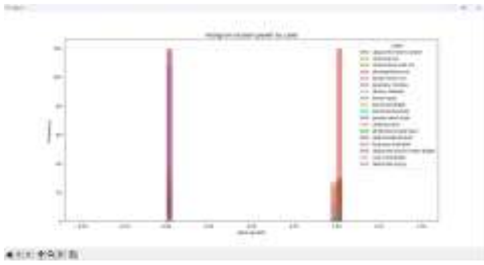


Fig. 9: Bar plot of count versus output before and after applying SMOTE algorithm.

Fig. 9 contrasts the class distribution before and after the application of the SMOTE

Fig. 7 captures the state of the GUI immediately after an admin uploads a dataset. The file path is shown along with a preview or log of the dataset content in a text widget. It demonstrates how the system provides feedback about the dataset being loaded and ready for further processing.

(b)



(c)

algorithm. Two bar plots are shown side-by-side: one displaying the imbalanced counts of each disease class, and the other illustrating the balanced class

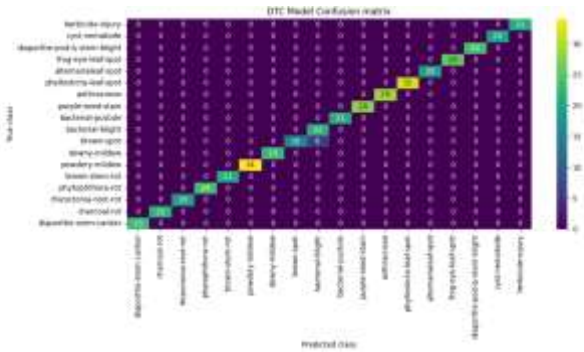


Fig. 10: Confusion matrices obtained using DTC model.

Fig. 10 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. 11: User login interface.

Fig. 11 shows the login interface tailored for regular users. It mirrors the admin login interface but is designed for user-specific access, ensuring secure entry into the user functionalities of the application.

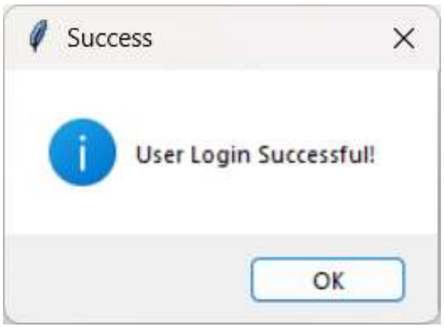


Fig. 12: Successful login of user.

Upon successful authentication, this interface (as shown in Fig. 12) confirms that the user has been logged in. It might display a welcome message or transition to a new screen dedicated to prediction tasks, confirming that access has been granted.

Fig. 13: Sample prediction and pesticide suggestions on test data.

In the Fig. 13 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
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. Overall summary, 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

[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