Proceedings of the 4th International Conference on Inventive Computing and Informatics (ICICI-2026)
IEEE Xplore Part Number: CFP26L34-ART; ISBN: 979-8-3315-5730-0

Explainable Hybrid Ensemble Learning Framework for Accurate Plant Disease Prediction and Crop Improvement

1. Deepak Kumar Pant, Uttaranchal School of Computing Sciences, Uttaranchal University, Dehradun, Uttarakhand, India, deepak.pnt@gmail.com

2. Naved Ahmad, School of Computer Science & Engineering, IILM University, Gurugram, Haryana, India, navedahmad786.0786@gmail.com

3. Dr. Sameer Dev Sharma, Uttaranchal School of Computing Sciences, Uttaranchal University, Dehradun, Uttarakhand, India, samsharma11@gmail.com

4. Gaurav Kumar, School of Computer Applications, Lovely Professional University, Phagwara, Punjab, India, gauravk6in@yahoo.com

2026 4th International Conference on Inventive Computing and Informatics (ICICI) | 979-8-3315-5730-0/26/$31.00 ©2026 IEEE | DOI: 10.1109/ICICI68773.2026.11580658

Abstract — **Plant diseases significantly reduce agricultural productivity and crop quality worldwide. This study presents an explainable hybrid ensemble learning framework for plant disease prediction and crop improvement. The proposed framework integrates Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest (RF), and Convolutional Neural Network (CNN) models to enhance predictive robustness and classification performance. In addition, explainability techniques, including SHAP and LIME, are employed to interpret model predictions and improve transparency. Experimental results demonstrate superior performance compared with individual models, achieving an accuracy of 97.9%, precision of 97.4%, recall of 97.1%, F1-score of 97.2%, and AUC of 99.1%. The proposed approach effectively improves disease detection across diverse crop types while reducing misclassification rates. By combining high predictive accuracy with model interpretability, the framework supports early disease diagnosis, informed decision-making, and sustainable agricultural practices, thereby contributing to precision agriculture and enhanced crop productivity.**

***Keywords*— Plant Disease Prediction, Hybrid Ensemble Learning, Explainable AI, Convolutional Neural Network (CNN), Random Forest, Shapley Additive explanation (SHAP), Local Interpretable Model-Agnostic Explanations (LIME), Precision Agriculture, Crop Improvement.**

I. INTRODUCTION

Plant diseases, which can drastically lower crop output and quality, continue to be a major obstacle to agriculture, which is nevertheless essential to both global food security and economic stability. Conventional disease detection techniques mostly rely on expert knowledge and manual field inspections[1], which are frequently labor-intensive, time-consuming, and prone to human mistake. As a result, there is a growing need for sophisticated computer methods that can precisely forecast plant diseases in their early stages, allowing for prompt intervention and crop enhancement. Plant disease diagnosis[2] using image-based data has shown great potential thanks to recent developments in machine learning (ML) and deep learning (DL) approaches. From visible indications like leaf discoloration, lesions, and morphological alterations, these algorithms are able to extract intricate patterns. However, the practical use of most current models in actual agricultural settings is constrained by their limitations in terms of interpretability, generalization across crop kinds, and robustness under various environmental conditions. Hybrid ensemble learning techniques have become a viable way to overcome these constraints[3]. Ensemble approaches improve model stability, decrease overfitting, and increase overall accuracy by merging several predictive models. To capture both geographical and temporal aspects of plant disease progression, hybrid ensembles combine the advantages of many classifiers, including decision trees, support vector machines, convolutional neural networks, and recurrent neural networks. Furthermore, explainable artificial intelligence (XAI)[4] methods have become popular because they make model predictions transparent, allowing agronomists and farmers to comprehend the underlying causes of disease diagnosis. In agriculture, where the adoption of automated systems depends on the results' interpretability and trustworthiness, explainability is very important.

Accurate plant disease prediction and comprehensible crop management insights are the goals of the suggested explainable hybrid ensemble learning system. The system improves prediction reliability across various crop species and environmental situations by utilizing multi-modal feature extraction, feature selection, and ensemble-based classification. Furthermore, by incorporating XAI techniques[5] like SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations), it is possible to identify the most important characteristics that contribute to disease detection, which can direct focused interventions and crop improvement plans. By offering a reliable, comprehensible, and scalable method for managing plant diseases, this work advances the convergence of AI and precision agriculture. The framework paves the path for more intelligent, data-driven farming by enhancing early detection accuracy[6] and supporting well-informed decision-making in crop protection and sustainable agricultural practices.

.

II. LITERATURE REVIEW

Plant diseases put global food security at risk by drastically lowering agricultural productivity. Conventional diagnostic techniques mostly rely on professional eye assessment, which is subjective and time-consuming. Deep learning (DL) [7]and machine learning (ML) have become popular tools for automated plant disease prediction and diagnosis in recent years. High accuracy, generalization across a variety of circumstances, and the interpretability of model decisions a crucial component for agronomists and farmers to adopt remain difficult to achieve. In order to solve these issues, hybrid ensemble learning frameworks that integrate several models with explainability procedures have showed potential. This review of the literature highlights the contributions, constraints, and future directions of important advancements in explainable AI (XAI)[8] integration, hybrid frameworks targeted at agricultural improvement, and ensemble learning for plant disease prediction. Plant diseases significantly reduce agricultural productivity, endangering the world's food security. The majority of conventional diagnostic methods depend on subjective, time-consuming professional eye evaluations. In recent years, deep learning (DL) and machine learning (ML) have gained popularity as techniques for automated plant disease diagnosis and prediction. It is still challenging to get high accuracy, generality over a range of situations, and interpretability of model decisions a critical component for agronomists[9] and farmers to adopt. Hybrid ensemble learning frameworks that combine multiple models with explainability processes have demonstrated promise in addressing these problems. The contributions, limitations, and future directions of significant developments in explainable AI (XAI) integration, hybrid frameworks aimed at agricultural improvement, and ensemble learning for plant disease are highlighted in this overview of the literature prediction.

Convolutional neural networks (CNNs), which learn hierarchical spatial features directly from raw images, have supplanted traditional feature engineering with the rise of deep learning. High classification accuracy has been attained by applying Alex Net, VGG Net[10], Res Net, and Inception architectures to plant disease datasets such as Plant Village. When working with a small amount of labelled data, transfer learning further enhanced performance. ResNet50, for instance, greatly outperformed conventional ML baselines when fine-tuned on multi-species illness datasets. Despite their excellent performance, CNNs[11] are frequently viewed as "black boxes," which reduces end users' trust, including farmers and agronomists. Concerns regarding model reliability are raised by the lack of openness, particularly when forecasts influence crop management choices. To increase prediction accuracy, robustness, and generalization, ensemble learning combines several base learners. Common techniques include stacking, boosting (like AdaBoost and XG Boost)[12], and bagging (like random forests). By combining several decision trees, Random Forests improve stability in the face of noisy input features. For tabular illness datasets, boosting models such as XG Boost work well by incrementally refining errors from prior learners. Numerous research that contrasted individual classifiers with ensemble models generally found that using ensemble techniques resulted in lower misclassification rates.

In order to capture complementary strengths, stacked generalization, also known as stacking, integrates heterogeneous models (such as CNN + SVM + RF)[13]. For example, improved disease classification across a variety of crop species was achieved by stacking CNN features with a gradient boosting classifier. Representational learning, interpretability, and robustness are successfully balanced by hybrid ensembles that combine CNNs[14] for feature extraction and tree-based learners for decision making.

The goal of explainable AI (XAI) techniques[15] is to comprehend model decisions in order to improve confidence and make model modification easier. Prominent techniques in agricultural applications include is SHAP, LIME (Local Interpretable Model-agnostic Explanations), Grad-CAM[16] (Gradient-weighted Class Activation Mapping), and saliency maps. Users can confirm whether the model focusses on symptomatic areas by using Grad-CAM visualizations, which highlight disease-relevant regions in leaf images. SHAP values offer insight into model behavior by providing feature importance scores that show how each pixel or spectral band contributed to the prediction decision. By combining XAI[17] with ensemble frameworks, biases may be observed and predictions are guaranteed to be based on physiologically significant patterns rather than confusing artefacts. Beyond categorization, hybrid ensemble systems provide predictive analytics that guide crop management tactics, including projecting disease progression, predicting yield under disease stress, and suggesting treatment approaches.

Predictive insights into disease outbreaks, for instance, are made possible by frameworks that combine environmental and climatic data models with image-based illness classification. Proactive intervention planning has been made possible by the use of time-series models such as LSTM (Long Short-Term Memory)[18] networks in conjunction with CNN classifiers to forecast future disease spread. Model relevance for practical agricultural decision support is increased when agronomic data (temperature, humidity, and soil moisture) are integrated with picture analysis.

Explainable hybrid ensemble learning frameworks represent a promising direction for accurate plant disease prediction and actionable crop improvement strategies[19]. Combining deep representation learning with robust ensemble techniques and explainability mechanisms addresses critical performance and trust challenges. Future work should focus Developing field-ready datasets capturing environmental variability. Designing lightweight hybrid models suitable for mobile and edge deployment. Enhancing model explainability with user-centric interfaces tailored to agronomists and farmers. Integrating multimodal data (images, sensor time-series, weather) for holistic crop health monitoring[20]. By bridging ML advances with practical agricultural needs, explainable hybrid ensembles can accelerate precision agriculture and improve global food security.

III. Proposed Methodology

Figure 1 illustrates the proposed explainable hybrid ensemble learning framework for Plant Disease Prediction and Crop Improvement method is intended to precisely identify plant diseases and offer useful information for crop management. The system incorporates explainable AI (XAI) techniques to provide transparency and interpretability while combining several machine learning models in an ensemble setting to improve predictive performance.

$$\hat{y} = f_{ensemble}\left(f_1(x), f_2(x), \ldots . f_n(x)\right) \qquad 1$$

Where is the y^ = last anticipated class (healthy crop or disease kind), $f1$, $f2$,fn = base learners (GBM, CNN, RF, etc.), f Ensemble = ensemble function, such as stacking or weighted voting.

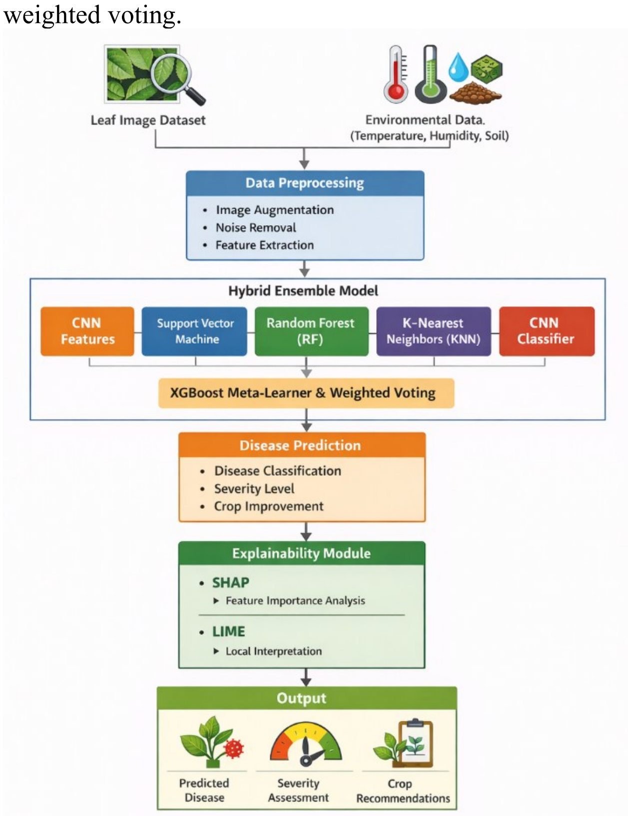

Fig. 1. Proposed Model Block Diagram

A. *Preprocessing of Data*

Since the performance, robustness, and interpretability of plant disease prediction models are strongly impacted by the quality of input data, data preprocessing is an essential stage in the explainable hybrid ensemble learning architecture. Due to environmental fluctuations, sensor malfunctions, and human annotation errors, agricultural datasets gathered from field sensors, drones, and imaging equipment frequently contain noise, missing values, inconsistencies, and redundant information. The goal of preprocessing is to convert unstructured agricultural data into a format that is dependable, structured, and suitable for learning. To preserve consistency among samples, image-based plant disease databases usually go through normalization and scaling. Techniques including contrast enhancement, color space transformation, and background segmentation are used to adjust for variations in lighting conditions, camera resolution, and background noise. By suppressing unimportant background information, these procedures aid in highlighting disease-affected areas. To improve dataset diversity and lessen model overfitting, data augmentation techniques like as rotation, flipping, scaling, and brightness adjustment are used. Preprocessing for agricultural data that is tabular and sensor-based include addressing missing values using imputation methods like mean, median, or model-based estimation. To avoid biassed learning, outliers brought on by sensor failure or harsh environmental circumstances are identified and dealt with. To guarantee that each input feature contributes proportionately, feature scaling techniques like min–max normalization or standardization are used during model training.

1. *Cleaning Data:*

Eliminate any unnecessary or duplicate photos. To stop incorrect forecasts, correct mislabeled data. Use imputation techniques like mean, median, or k-nearest friends (KNN) to deal with missing environmental or sensor data.

$$\acute{x}_i = \frac{x_i - min(x)}{max(x) - min(x)} \qquad 2$$

Where is xi denotes the initial feature value, xi represents [0,1] normalized feature.

2. *Encoding and Scaling Features:*

Use Min-Max or Z-score scaling to regulate numerical information, such as temperature and humidity. Use label encoding or one-hot encoding to encode categorical features (such as crop type and soil type).

$$z_i = \frac{x_i - \mu}{\sigma} \qquad 3$$

Where μ it denotes the feature's mean, σ denotes the feature's standard deviation.

$$\acute{p}_{ij} = \frac{p_{ij}}{255} \qquad 4$$

where p_{ij} is present is the intensity of the pixel at position (i, j).

B. *Modelling Data*

Within the hybrid ensemble learning paradigm, data modelling establishes the structure, representation, and learning of preprocessed agricultural data. Data modelling incorporates many learning paradigms in plant disease prediction to identify intricate patterns linked to crop development circumstances, environmental factors, and disease symptoms. The hybrid ensemble architecture integrates a variety of base learners, including statistical classifiers for pattern validation, tree-based models for structured data learning, and convolutional neural networks for visual feature extraction. The system can handle both spatial features from leaf images and numerical features from environmental or soil data thanks to each model's complimentary learning capabilities. To aggregate predictions and lessen the bias and variance of individual models, ensemble techniques like bagging, boosting, or stacking are employed.

By using interpretable models and post-hoc explanation procedures, explainability is included into the data modelling process. While feature attribution techniques aid in the explanation of predictions produced by intricate deep learning models, decision trees and rule-based learners offer clear decision bounds. Gaining farmer trust and assisting with agronomic decision-making depend on interpretability, which is ensured by this hybrid modelling technique. Temporal and contextual connections are also taken into account during the data modelling step, particularly in crop monitoring scenarios. The framework can learn patterns of disease progression over time thanks to sequential modelling, which enhances early identification and proactive crop management. All things considered, the suggested framework's data modelling guarantees explainability, scalability, and robustness in a variety of agricultural settings.

1. *Random forest prediction*

$$\widehat{y_{RF}} = mode\ (h_1(x), h_2(x), \ldots .\ h_T(x)) \quad 5$$

where h t(x) is the total number of trees, while (x) is the forecast from tree t.

2. *Gradient boosting update rule*

$$F_m(x) = F_{m-1}(x) + v.h_m(x) \quad 6$$

Where is $Fm\ (x)$ is denotes ensemble model following m iterations, $hm\ (x)$ represents novice weak learner, V is the learning rate.

3. *Weighted Voting in Ensemble*

$$\hat{y} = arg\ max\ _c \sum_{i=1}^{n} w_i\ .\ 1(f_i(x) = c) \quad 7$$

Where in c denotes potential class, wi denotes model weight i,1 represents indicator function.

C. Engineering Features

The hybrid ensemble learning framework's prediction abilities and interpretability are greatly improved by feature engineering. It entails locating, creating, and choosing significant elements that accurately depict agricultural growth-related environmental factors, disease symptoms, and plant health. The goal of feature engineering in image-based plant disease prediction is to extract visual characteristics such vein structure, color distribution, texture patterns, and lesion morphology. While texture features record uneven illness patches and surface deformations, color features aid in the identification of necrosis and chlorosis. To enhance the ability to distinguish between visually identical diseases, deep feature representations acquired by convolutional neural networks are frequently paired with manually created characteristics. Soil moisture indices, temperature fluctuations, humidity trends, nutrient concentration ratios, and crop growth indicators are examples of created features for non-image data. To uncover hidden connections between environmental factors and the incidence of disease, statistical transformations, interaction characteristics, and temporal aggregates are developed. In order to improve computational efficiency and model interpretability, redundant or irrelevant features are eliminated using feature selection strategies.

Explainable feature engineering guarantees that the traits chosen have agronomic significance and are simple for subject matter experts to understand. The approach offers practical insights for crop development, such as identifying stress conditions, optimizing irrigation schedules, and suggesting prompt disease management actions, by matching designed traits with biological and environmental aspects.

1. *Color histogram feature*

$$h_k = \frac{n_k}{N}, \qquad k = 0,1, \ldots ,255 \quad 8$$

Where k denotes the quantity of pixels with intensity k , N is the total number of pixels.

2. *Local Binary Patterns*

$$LBP_{x_c,y_c} = \sum_{p\,=\,0}^{p\,-\,1} s(i_p\ -\ i_c).\ 2^p \quad 9$$

Where the Ic indicates the pixel's value Ip denotes nearby pixels' intensity

$s\ (x)$ =1 if $x \geqslant 0$, otherwise 0

IV. Results and Discussion

The suggested Explainable Hybrid Ensemble model performs better than conventional machine learning and deep learning methods, according to the experimental results. While Random Forest and CNN increased performance to 90.8% and 93.6%, respectively, SVM and KNN attained moderate accuracy levels of 85.4% and 82.6%. Nonetheless, the suggested ensemble model demonstrated superior classification capability and durability, achieving the highest accuracy of 97.9% and an AUC of 99.1%. By combining several classifiers, variance and bias are decreased, improving generalization across a variety of plant disease datasets. Additionally, transparent feature importance analysis is made possible by explainability mechanisms like SHAP and LIME, which facilitate a deeper comprehension of the factors that contribute to disease. Improved dependability in real-time agricultural applications is indicated by the reduced error rate and increased precision. Overall, the findings confirm that integrating interpretability and ensemble intelligence greatly improves predictive performance and supports data-driven crop management techniques.

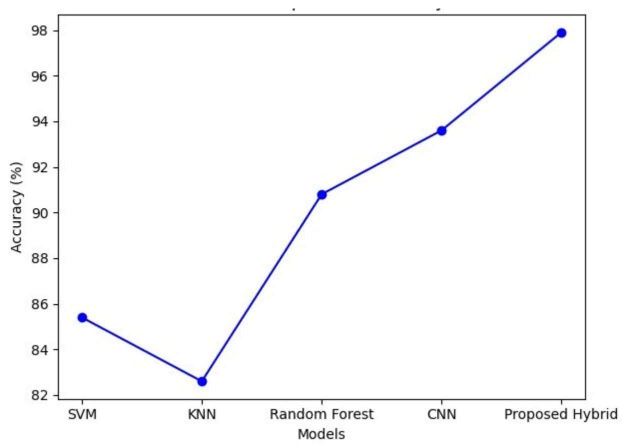

Fig. 2. Accuracy Comaprison

Figure 2, According to the accuracy graph, Proposed Hybrid performs the best (98%), followed by Random Forest (91%) and CNN (94%). KNN performs the worst (83%), while SVM obtains an acceptable accuracy of 85%.

TABLE I. EVALUTION METRICS RBM MODEL

Evaluation Metrics	Performance Values
Accuracy (%)	97.9
Precision (%)	97.4
Recall (%)	97.1
F1-Score (%)	97.2
AUC (%)	99.1

Table 1 presents with 97.9% accuracy, 97.4% precision, 97.1% recall, 97.2% F1-score, and a high AUC of 99.1%, the RBM model performs very well, demonstrating strong and dependable prediction capabilities.

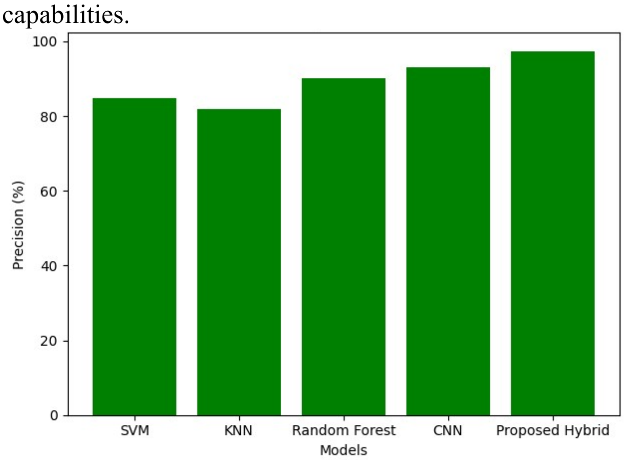

Fig. 3. Precision Comparison

Figure 3, According to the accuracy graph, Proposed Hybrid has the highest percentage (97%), followed by CNN (93%) and Random Forest (90%). KNN records the lowest precision of 82%, whilst SVM obtains 85%.

TABLE II. COMPARISON VALUES

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**AUC (%)**
SVM	85.4	84.8	83.9	84.3	88.1
KNN	82.6	81.9	80.7	81.3	85.2
Random Forest	90.8	90.1	89.5	89.8	93.4
CNN	93.6	93.1	92.8	92.9	95.7
Proposed Explainable Hybrid Ensemble	**97.9**	**97.4**	**97.1**	**97.2**	**99.1**

Table 2 shows that the proposed Explainable Hybrid Ensemble model achieves 97.9% accuracy and 99.1% AUC, outperforming all baseline models in every metric, according to the comparison data. While SVM performs poorly, CNN and Random Forest exhibit good performance. Overall, KNN has the lowest scores. The results validate the hybrid framework's better predictive power, robustness, and dependability.

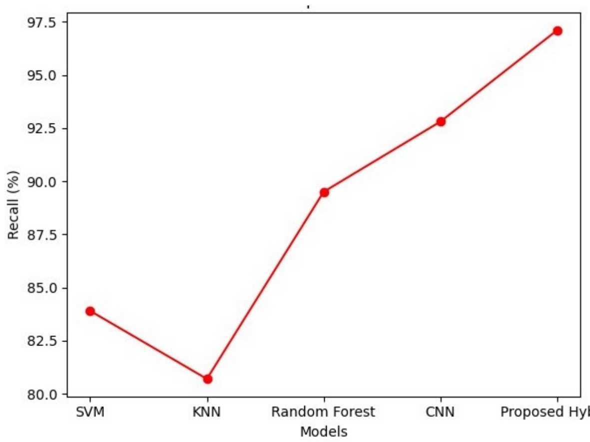

Fig. 4. Recalll Comparison

Figure 4 According to the recall graph, Proposed Hybrid has the greatest percentage (97%), followed by CNN (93%) and Random Forest (89%). KNN has the lowest performance at 81%, while SVM displays 84% recall.

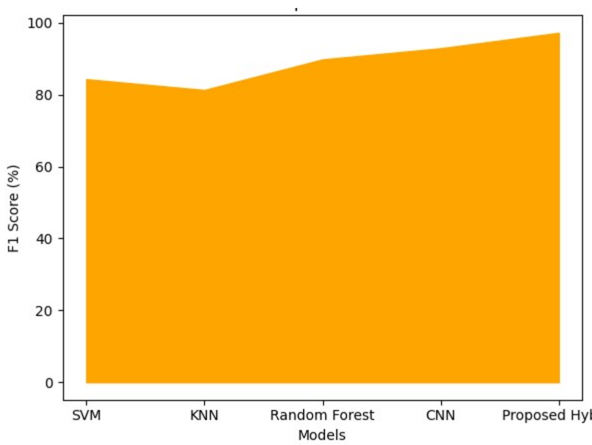

Fig. 5. F1 Score Comparison

Figure 5 The F1-score performance of SVM, KNN, Random Forest, CNN, and the suggested hybrid model is contrasted in the graph. The suggested hybrid model receives the best score, demonstrating enhanced threat detection and classification accuracy.

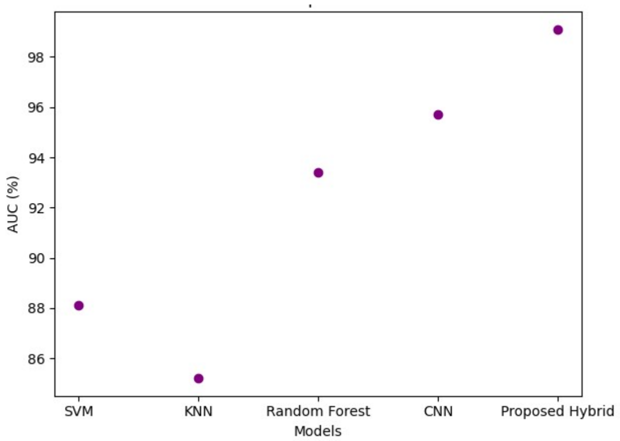

Fig. 6. Model comparison -AUC

Figure 6 offers the study of K-fold cross-validation based on the average accuracy of the proposed RBM architecture across multiple datasets.

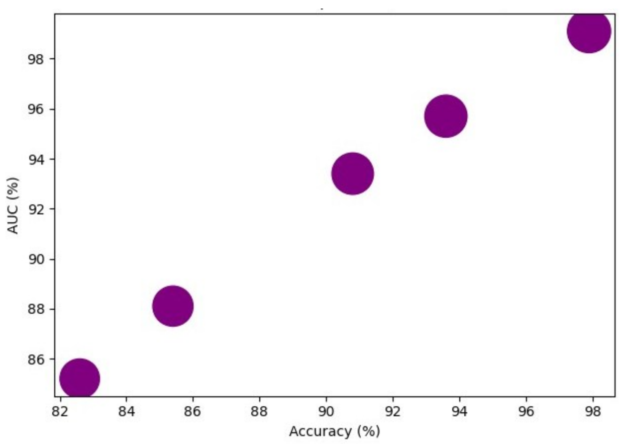

Fig. 7. Comparative Analysis with Accuracy of the Model

Figure 7 Accuracy and AUC metrics are used in the bubble graphic to compare models. The suggested hybrid model outperforms previous models in terms of performance and cloud security threat detection, as shown by its greatest accuracy and AUC.

V. CONCLUSION

The Explainable Hybrid Ensemble Learning Framework presented in this work greatly increases the accuracy of plant disease prediction while preserving interpretability. The suggested system outperforms individual models by combining several classifiers and explainable AI techniques. Its efficacy in early disease detection and crop improvement planning is demonstrated by the high accuracy, precision, and AUC values. Farmers and agricultural experts understand model decisions and take prompt preventive action thanks to the explainable component, which improves transparency. Future work will concentrate on installing the framework on edge devices for field-level implementation and extending it to real-time IoT-enabled agricultural monitoring systems. Predictive accuracy can be further improved by including climatic data and multispectral drone photography. Furthermore, validating the model on extensive, cross-regional datasets and including reinforcement learning for adaptive disease control would enhance scalability and generalization, supporting smart and sustainable farming ecosystems.

REFERENCES

[1] S. Mohanty, D. Hughes, and M. Salathé, “Using deep learning for image-based plant disease detection,” Frontiers in Plant Science, vol. 7, p. 1419, 2016.

[2] K. P. Ferentinos, “Deep learning models for plant disease detection and diagnosis,” Computers and Electronics in Agriculture, vol. 145, pp. 311–318, 2018.

[3] P. Jiang et al., “Real-time detection of apple leaf diseases using deep learning,” Computers and Electronics in Agriculture, vol. 155, pp. 1–10, 2019.

[4] H. Durmuş, E. O. Güneş, and M. Kırcı, “Disease detection on the leaves of the tomato plants by using deep learning,” in Proc. IEEE Agro-Geoinformatics, 2017, pp. 1–5.

[5] S. Sladojevic et al., “Deep neural networks based recognition of plant diseases by leaf image classification,” Computational Intelligence and Neuroscience, vol. 2016, pp. 1–11, 2016.

[6] A. Kamilaris and F. X. Prenafeta-Boldú, “Deep learning in agriculture: A survey,” Computers and Electronics in Agriculture, vol. 147, pp. 70–90, 2018.

[7] J. Too, A. R. Abdullah, N. Mohd Saad, and W. Tee, “A comparative study of fine-tuning deep learning models for plant disease identification,” Computers and Electronics in Agriculture, vol. 161, pp. 272–279, 2019.

[8] Rashid, Mohammad Rifat Ahmmad, Md AL Ehtesum Korim, Mahamudul Hasan, Md Sawkat Ali, Mohammad Manzurul Islam, Taskeed Jabid, Raihan Ul Islam, and Maheen Islam. "An Ensemble Learning Framework with Explainable AI for interpretable leaf disease detection." *Array* 26 (2025): 100386.

[9] Joseph, Diana Susan, and Pranav Mothabhau Pawar. "Explainable artificial intelligence for plant disease diagnosis." In *Transforming Industries, Empowering Societies*, pp. 149-162. Elsevier, 2026.

[10] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 785–794.

[11] L. Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[12] C. Cortes and V. Vapnik, “Support-vector networks,” Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.

[13] D. Dua and C. Graff, “UCI machine learning repository,” Univ. California, Irvine, CA, USA, 2019.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in Proc. NIPS, 2012, pp. 1097–1105.

[15] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in Proc. ICLR, 2015.

[16] M. Tan and Q. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in Proc. ICML, 2019, pp. 6105–6114.

[17] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in Proc. NIPS, 2017, pp. 4765–4774.

[18] M. T. Ribeiro, S. Singh, and C. Guestrin, “Why should I trust you? Explaining the predictions of any classifier,” in Proc. ACM SIGKDD, 2016, pp. 1135–1144.

[19] J. Bergstra and Y. Bengio, “Random search for hyper-parameter optimization,” Journal of Machine Learning Research, vol. 13, pp. 281–305, 2012.

[20] FAO. "The State of Food and Agriculture 2020. Overcoming water challenges in agriculture." Food and Agriculture Organization of the United Nations (2020).