

Feature Selection Random Forest Model for Crop Disease Detection in Greenhouse Vegetable Images

Alexandra Marin^{1,*}, Mariana Constantinescu¹ and Walter Rentea¹

¹ Faculty of Information Engineering, University of Turda, Turda, 401100, Romania

*Corresponding author: alexandra.ma@info.uturda.ro

Abstract. The disease-detection system model for greenhouse vegetables must be very accurate, simple to understand, and compatible with a low-cost monitoring apparatus. The majority of deep learning techniques necessitate large-scale annotated images and provide less insight into the visual symptoms that influenced the classification choice. Make a selection of explainable features. This study used a Random Forest algorithm to detect crop diseases in photos of greenhouse vegetables. Using a method, extract lesion color, texture, form, boundary, and local contrast features from illness candidate locations. Then, use a relevance-redundancy-importance selection strategy to create a small set of symptom-sensitive features. The chosen features are used to train a Random Forest classifier, and category-level illness difference is interpreted using feature significance analysis. Experimentally evaluate the computational efficiency, illumination robustness, confusion among related diseases, feature reduction ratio, and detection accuracy. With a general accuracy of 94.3%, a macro-F1 score of 93.6%, a feature reduction rate of 61.8%, and an average inference speed of 6.4 ms per image, all of the metrics in the suggested model have produced excellent results. As a result, the selected symptom descriptors will perform well in categorization and offer more comprehensible features at the representation level.

Keywords: *Explainable Disease Detection, Greenhouse Vegetable Images, Symptom Feature Selection, Random Forest*

Received on 27 September 2022, Accepted on 15 January 2023, Published on 23 January 2023

Copyright © 2023 Author, licensed to JIIC. This is an open access article distributed under the terms of the CC BY-NC-SA 4.0, which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

Introduction

Greenhouse vegetable production is essential for promoting high-quality, intensive-agricultural-management greenhouses and ensuring food security. Greenhouse systems have more control over temperature, humidity, irrigation, and nutrient conditions than open-field production, but they are also enclosed spaces, and if early signs of bacterial, viral, or fungal diseases are not addressed, they will spread swiftly. As a result, image-based disease detection has been employed in the real world to increase the accuracy of plant protection, and cameras in permanent monitoring stations or mobile inspection equipment can be used to continuously capture leaf images without causing harm to the plants [1]. The model crops of cucumber, tomato, pepper, lettuce, and eggplant were chosen, and it was examined whether early disease detection might lower the use of pesticides while also determining what type of early intervention methods were required for local treatment [2]. Automatic disease identification in greenhouse monitoring systems is also feasible due to the increasing affordability of low-cost cameras and edge devices [3].

Nevertheless, it is difficult to categorize photos of greenhouse vegetable diseases. The color distribution of the lesions is being impacted by variations in leaf shade, additional bulbs, and uneven lighting brought on the plastic film. Superfluous visual aspects unrelated to the disease include leaf overlap, water droplets, dirt background, stem occlusion, and variations in camera distance [4]. Although deep convolutional networks have improved the identification of plant diseases, they often require a large number of tagged photos and may be computationally demanding for inexpensive greenhouse terminals [5]. The performance of traditional machine learning models

based on handcrafted features is largely dependent on whether the feature set can capture information like lesion color, texture, shape, and boundaries without retaining too many background descriptors, even though these models are still reasonable choices for datasets with moderate sizes and when interpretability is required [6]. Although Random Forest offers a measure of feature relevance and is comparatively insensitive to non-linear interactions among features, duplicated or correlated image features may nevertheless compromise classification stability [7].

In this paper, develop an explainable feature selection Random Forest model for crop disease diagnosis in photos of greenhouse vegetables. Features are assessed for relevance and redundancy using a method to extract symptom-oriented descriptors from disease-candidate regions. Tree-based importance then chooses which features to include in a reduced set for Random Forest classification. Reducing feature redundancy without compromising disease recognition accuracy, improving discrimination between visually similar disease categories, and providing feature-level interpretability appropriate for greenhouse plant protection decisions are the three practical issues addressed in this study. The system aims to provide high classification accuracy, cheap inference costs, stable performance with variations in greenhouse imaging circumstances, and strong model transparency.

Related Work

Image-Based Plant Disease Diagnosis

From color thresholding and texture descriptors to deep learning and mobile identification systems, image-based plant disease diagnosis has advanced. In the past, lesion segmentation, color moments, gray-level co-occurrence matrices, local binary patterns, and shape characteristics were employed to distinguish between healthy and diseased leaves. These methods are nevertheless appropriate for minimal training data and visually distinct illness symptoms [8]. Deep models are currently widely used for various recognition tasks due to the expansion of public plant disease databases and the use of convolutional neural networks. Research on apple, rice, tomato, and palm diseases demonstrates that deep learning can enhance classification accuracy in a controlled setting by directly learning discriminative lesion features from photos [9]. By using features pre-trained on general image recognition or other related crop disease datasets, transfer learning and domain adaptation lessen the need for large-scale labelled datasets [10].

The detection of greenhouse diseases differs from various benchmark datasets. The production environment's images feature a variable background, overlapping leaves, inconsistent lighting, and early-stage lesions with weak contrast. Models trained under one imaging setting may perform poorly when the illness presentation or acquisition equipment are different, according to open-set and cross-domain research [11]. Simultaneously, a sample's diagnosis of powdery mildew, leaf spot, downy mildew, or nutritional stress-like disease must be explained for practical usage. Heatmaps can be produced by a single deep learning model, but they cannot be directly connected to agronomic properties like lesion color, edge sharpness, spot circularity, and texture roughness [12]. In order to preserve visual feature engineering and improve the accuracy of feature selection and classification, hybrid models have been put forth.

Feature Selection in Agricultural Image Analysis

In this work on plant disease identification, feature selection is also necessary for the manually created image descriptors. Color, texture, shape, border, and contrast descriptors can produce hundreds of candidate variables; many of these are redundant or illumination-sensitive. The classifier may learn background patterns, leaf variety characteristics, or imaging artifacts rather than illness symptoms if descriptors are the only ones employed [13]. Wrapper approaches look for subsets based on classifier performance, filter methods utilize correlation, mutual information, or separability indices to establish the relevance of features, and embedding methods use model-specific importance during training [14]. By selecting symptom-sensitive descriptors associated with disease categories, feature selection in agricultural image analysis minimizes computation and enhances interpretability [15].

Images of greenhouse vegetables should have features that are both non-redundant and capable of discrimination. Boundary features indicate if the lesion edge is diffuse or clearly separated from healthy tissue,

texture features are utilized to evaluate fungal growth or spot density, and lesion color features can distinguish between necrosis and chlorosis [16]. There is a strong correlation between descriptors from the same texture operator or color space. The two selection strategies have distinct complements and do not overlap. For this reason, Random Forest classification is also appropriate; nevertheless, the significance of correlated variables may still be dispersed and therefore less consistent, even if it can handle mixed-type features [17].

Tree Ensemble Models for Interpretable Classification

Tree ensemble models are frequently employed in applied classification tasks without considerable feature scaling because of their ability to handle non-linear relationships, mixed-scale features, and limited amounts of data very well. In order to lower the variance of a single decision tree, Random Forest creates numerous decision trees using bootstrap sampling and random feature subspace selection [18]. In addition to providing class-wise feature importance, Random Forest can be used to combine lesion color, texture, and shape information for image-based disease identification. Although Gradient Boosting and XGBoost often have higher accuracy, they are more challenging to implement on low-cost devices and are somewhat susceptible to hyperparameter adjustment [19].

Another factor in selecting a feature-selection Random Forest model for greenhouse disease monitoring is interpretability. Agronomists can ascertain whether a classifier is based on physiologically significant symptoms or unrelated background elements by using Feature Importance. Feature contribution analysis needs to be more visible when the outcomes of a model will direct real-world scenarios, according to research on explainable artificial intelligence [20]. When diagnosing plant diseases, interpretable categorization can determine which symptom descriptors differentiate visually similar diseases and which descriptors are sensitive to light changes [21]. In order to balance accuracy, robustness, and field application, a compact and explainable Random Forest model has been chosen. For displaying the non-linear threshold effect of color, texture, or lesion area descriptors, tree-based interpretation is also practical [22]. Additionally, local-to-global explanation techniques aid in expanding the use of individual picture diagnostic to knowledge at the category level [23]. Instead, then giving explanations as independent importance scores, the majority of explainable AI research focuses on connecting explanations to domain knowledge [24]. Agricultural data analysis also requires interpretability in machine learning, i.e., the ability to ascertain whether a given prediction has biological or environmental significance [25]. By lowering the number of potential split variables and preventing unstable significance distributions among redundant descriptors, feature selection can further improve the interpretability of tree ensembles [26]. To reduce the inference cost while retaining non-linear discrimination capacity, a smaller collection of features is selected for Random Forest image classification [27]. Reducing redundant leaf descriptors increases classification stability, according to research on rice disease using feature selection [28]. Based on the findings, it is suggested that feature selection, Random Forest classification, and explanation analysis be developed as an integrated approach rather than separate modules for the identification of greenhouse vegetable diseases [29]. Therefore, a highly complex model with opaque decision evidence is not as appropriate for the low-cost agricultural monitoring application as a compact and interpretable classifier [30].

Explainable Feature Selection Random Forest Model

Disease Candidate Region Extraction and Symptom Feature Encoding

The proposed method begins with greenhouse vegetable image preprocessing. Each RGB image is resized to 512×512 pixels, and a color constancy correction is applied to reduce the effect of plastic-film reflection and supplemental lighting. The leaf region is separated from background by combining excess green response and adaptive saturation thresholding. Disease candidate regions are then enhanced by emphasizing local deviations from the surrounding healthy tissue, because early lesions are often small and have weak contrast.

$$I_n(x, y) = \frac{\kappa I(x, y)}{\mu_\Omega(x, y) + \varepsilon} \quad \text{Eq.(1)}$$

Here, $I_n(x, y)$ is the illumination-normalized pixel value, $\mu_\Omega(x, y)$ denotes the local neighborhood mean, κ is a scale coefficient, and ε prevents unstable division in dark regions. This operation keeps lesion contrast while suppressing slow illumination drift across the leaf surface.

$$S(x, y) = \alpha_1 D_{\text{color}}(x, y) + \alpha_2 T_{\text{lbp}}(x, y) + \alpha_3 E_{\text{edge}}(x, y) \quad \text{Eq.(2)}$$

The terms Decolor, Telp, and Edge describe color deviation, local binary texture variation, and boundary response. In the implementation, α_1 , α_2 , and α_3 are selected on the validation set, and the final candidate mask is obtained from the top 18% lesion-score pixels within the leaf region.

$$r_k = \frac{A_k}{A_{\text{leaf}}}, c_k = \frac{4\pi A_k}{P_k^2} \quad \text{Eq.(3)}$$

The ratio r_k measures lesion occupation in the leaf region, while c_k measures compactness. Circular spots, irregular blight regions, and diffuse mildew areas can therefore be separated by morphology-sensitive descriptors rather than by color alone.

$$f_i = \frac{d_i - \text{median}(d_i)}{\text{MAD}(d_i) + \varepsilon} \quad \text{Eq.(4)}$$

Figure 1 shows the disease candidate extraction and symptom feature encoding workflow. The figure is placed after feature construction because it connects image preprocessing, lesion localization, and descriptor generation into a unified pipeline.



Figure 1. Explainable symptom feature selection workflow for greenhouse vegetable disease images

Relevance-Redundancy-Importance Feature Selection

The initial feature pool contains 186 descriptors, including 54 color features, 72 texture features, 38 shape and boundary features, and 22 local contrast features. Directly using this pool increases computation and may reduce interpretability because correlated descriptors distribute importance across multiple variables.

$$\text{Rel}(f_i) = \frac{I(f_i; y)}{H(y)} \quad \text{Eq.(5)}$$

The relevance score is normalized mutual information between feature f_i and disease label y . It estimates whether the feature contains class-discriminative information independent of its numerical scale.

$$\text{Red}(f_i, F_s) = \max_{j \in F_s} |\text{corr}(f_i, f_j)| \quad \text{Eq.(6)}$$

Redundancy is measured by the maximum absolute correlation between the candidate feature and the already selected feature subset F_s . This maximum form avoids retaining several almost identical color or texture descriptors.

$$\text{Imp}(f_i) = \frac{1}{T} \sum_t \sum_{v \in N_t(i)} \Delta G_v \quad \text{Eq.(7)}$$

Here, T is the number of trees, $N_t(i)$ is the set of nodes split by feature f_i in tree t , and ΔG_v is the Gini impurity reduction at node v . Features repeatedly selected near upper tree levels tend to obtain higher values.

$$\text{Score}(f_i) = \lambda_1 \text{Rel}(f_i) + \lambda_2 \text{Imp}(f_i) - \lambda_3 \text{Red}(f_i, F_s) \quad \text{Eq.(8)}$$

The weights λ_1 , λ_2 , and λ_3 are tuned on the validation set. In the experiment, $\lambda_1 = 0.42$, $\lambda_2 = 0.38$, and $\lambda_3 = 0.20$ provide the best balance between accuracy and feature compactness.

$$\Delta Q_k = Q(F_s \cup \{f_k\}) - Q(F_s) \quad \text{Eq.(9)}$$

If ΔQ_k is lower than 0.002 for five consecutive candidate features, the selection process stops. This prevents the model from adding features that only increase dimensionality without improving disease discrimination.

$$B(f_i) = \frac{\min_c R_c(F_s \cup \{f_i\})}{\max_c R_c(F_s \cup \{f_i\})} \quad \text{Eq.(10)}$$

The final feature utility is sensitive to both global accuracy and class-level balance, which is important when early blight, leaf spot, and nutrient stress-like symptoms appear visually similar.

Random Forest Disease Decision and Feature Explanation

The selected feature subset is used to train a Random Forest classifier. Each tree is trained on a bootstrap sample, and each split considers a random subset of selected descriptors. This mechanism reduces variance and improves robustness under greenhouse image variability.

$$h_t(x) = \arg \max_c P_t(c | x_F) \quad \text{Eq.(11)}$$

The prediction of one tree is determined by the class posterior estimated at the leaf node reached by the selected feature vector off. The forest prediction is obtained by aggregating all tree decisions.

$$P(c | x) = \frac{1}{T} \sum_t \mathbf{1}(h_t(x) = c) \quad \text{Eq.(12)}$$

The final disease category is the class with the largest aggregated probability. In the implementation, the forest contains 220 trees, the maximum depth is limited to 18, and the minimum leaf size is set to 4.

$$G(v) = 1 - \sum_c p_{v,c}^2 \quad \text{Eq.(13)}$$

A split is selected when it maximizes impurity reduction while satisfying the minimum sample constraint. This favors features that separate disease classes consistently across image subsets.

$$\Phi(f_i, c) = \text{mean}_{x \in D_c} |P(c | x) - P(c | x_{-i})| \quad \text{Eq.(14)}$$

Figure 2 presents the Random Forest decision aggregation and feature explanation mechanism. It highlights that the model not only outputs a disease label but also generates category-level symptom evidence through selected feature importance.

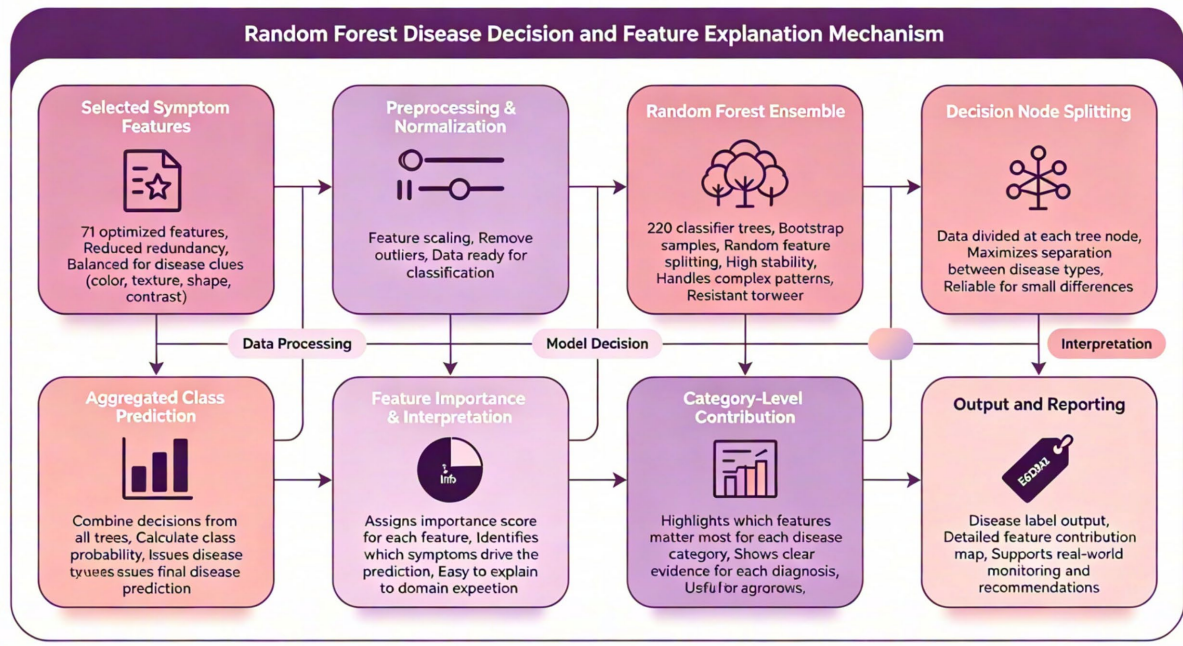


Figure 2. Random Forest decision aggregation and feature explanation mechanism

$$J = F1_{\text{macro}} - \beta_1 \frac{|F|}{|F_{\text{all}}|} - \beta_2 \tau_{\text{inf}} + \beta_3 S_{\text{exp}} \quad \text{Eq.(15)}$$

The final model keeps 71 features from the original 186 descriptors. This subset preserves disease-sensitive color and texture features while removing many redundant background and weak-shape descriptors.

Experimental Results and Comparative Discussion

Disease Detection Performance Across Models

The experimental dataset contains 8,640 photos of greenhouse vegetables from tomato, cucumber, pepper, and lettuce producing locations. Healthy leaves, powdery mildew, downy mildew, early blight, bacterial spot, and leaf mold are the six varieties. To avoid image-neighborhood leaking, split the dataset's acquired batches into 70% for training, 15% for validation, and 15% for testing. KNN, SVM, Decision Tree, Full-feature Random Forest, XGBoost, Lightweight CNN, and MobileNetV2 are the baseline models. There are 220 trees and 71 selected features in the feature-selection Random Forest that is suggested here.

The overall detection results are shown in Figure 3. Accuracy, macro-precision, macro-recall, and macro-F1 are compared in Figure 3(a). The suggested model has 94.3% accuracy and 93.6% macro-F1. The accuracy of the full-feature Random Forest is 0.918, whereas the accuracy of SVM and XGBoost is 0.887 and 0.924, respectively. The class-level recall is shown in Figure 3(b); powdery mildew and leaf mold have recall rates of 96.1% and 94.8%, respectively, while bacterial spot continues to be the most challenging class with a recall rate of 90.5%. The proposed model has a speed of 6.4 ms per image, which is significantly quicker than MobileNetV2 at 18.7 ms and full-feature Random Forest at 9.8 ms [31]. Figure 3(c) illustrates the inference time contrast.

It is incorrect to think of feature selection as a decrease in the number of features. Eliminating superfluous descriptors enhances the quality of tree splits and lessens the possibility that the initial splits will be led by backdrop or illumination-sensitive descriptors. When compared to a full-feature Random Forest, the suggested model has a higher macro-F1 score and 61.8% less features. When it comes to early-stage lesions with little sick patches, a lightweight CNN is not particularly stable, but it works well for clear-cut symptoms. Because the chosen color-texture descriptors are still sensitive to local fungal texture and poor lesion contrast, the suggested model has retained good accuracy [32].

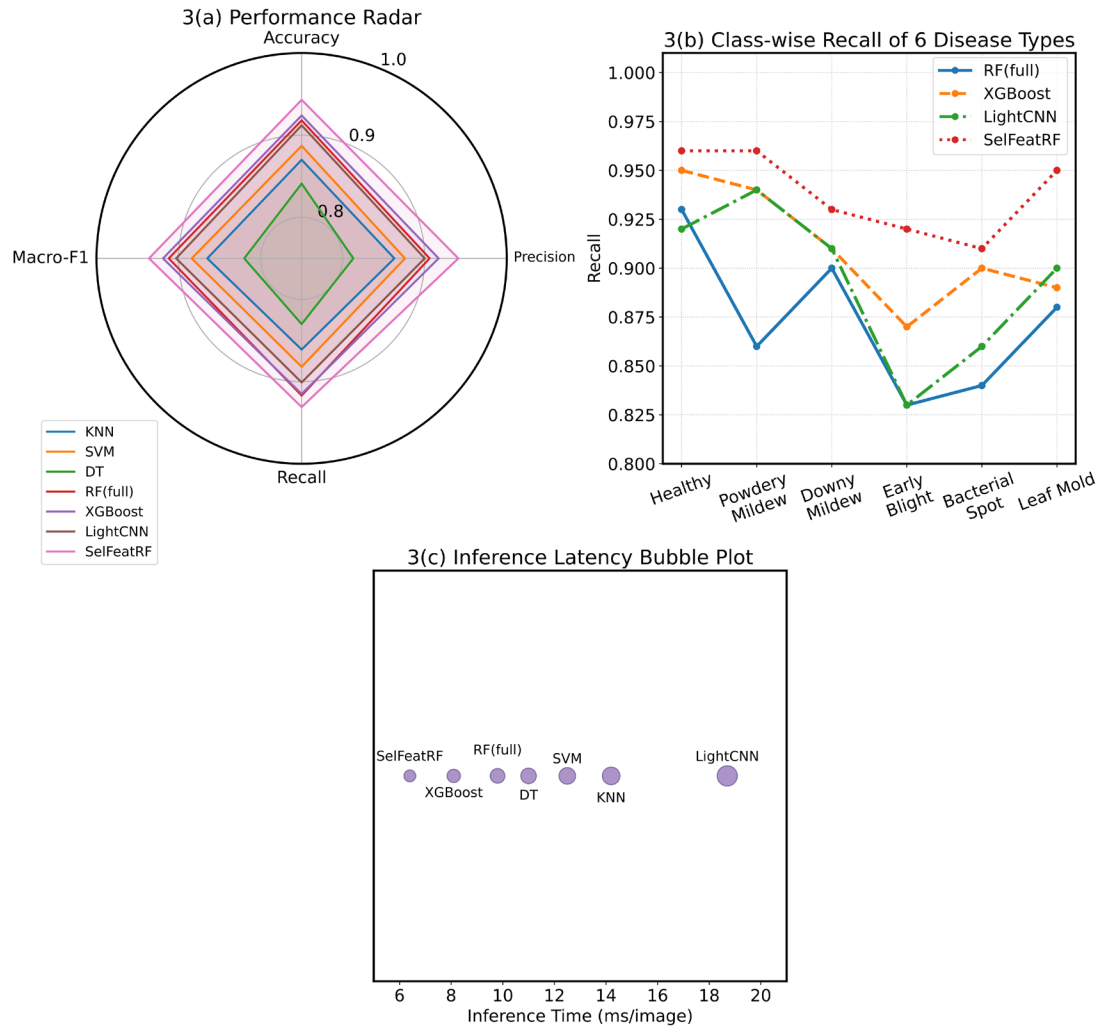


Figure 3. Disease detection performance comparison across classification models. (a) Overall accuracy, precision, recall, and macro-F1 comparison. (b) Class-level recall of six greenhouse vegetable disease categories. (c) Inference latency comparison among machine learning and lightweight deep learning models.

Feature Selection Effect and Category-Level Interpretation

The number of features, descriptor type, importance distribution, and class-level explanations are used to assess the efficacy of feature selection. The validation accuracy curve as a function of the number of chosen features is shown in Figure 4(a). Before 45 features, accuracy increases quickly; at 71 features, it approaches 94.3%; beyond that, it is nearly constant. Descriptor composition is shown in Figure 4(b): 26 color descriptors, 29 texture descriptors, 9 shape descriptors, and 7 contrast descriptors were chosen. The top 15 features in the chosen set include Lab-a mean, HSV saturation variance, LBP uniformity, GLCM contrast, lesion compactness, and edge irregularity (Figure 4(c)). The average absolute feature correlation following selection is 0.19, as seen in Figure 4(d), which is less than the initial 0.42.

The chosen characteristics align with the greenhouse's apparent illness symptoms. A large local brightness variance and a fine texture density are typically indicative of powdery mildew [33]. Downy mildew has a low-saturation lesion area and a greater shift in yellow-green color. The first kind of blight is rough, ringed, and dark in color. A bacterial spot is a tiny, irregularly bordered, high-contrast region. These feature patterns demonstrate that the Random Forest classifier uses symptom-level descriptions along with findings from plant pathology in addition to a generic leaf color [34].

The category-level explanation is shown in Figure 5. The class-wise feature contribution matrix is displayed in Figure 5(a), with texture consistency accounting for 18.6% of the categorization of powdery mildew and edge irregularity for 16.2% of the detection of bacterial spots. Shape descriptors are required to differentiate between

the confused bacterial spot and early blight samples because they have similar dark lesion colors but somewhat different spot boundary distributions [35]. Figure 5(b) illustrates the proper categorization and confusion of samples.

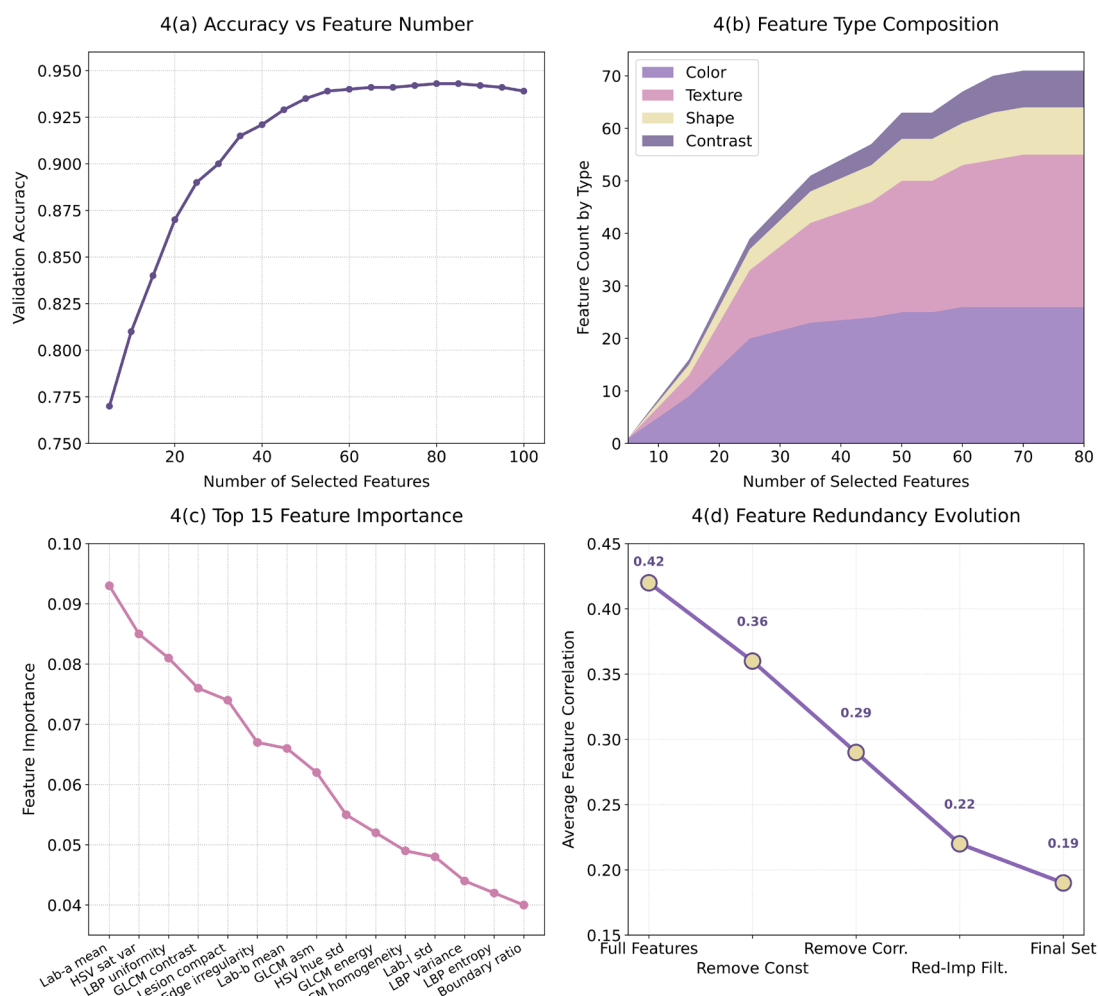


Figure 4. Feature selection effectiveness and symptom feature importance. (a) Validation accuracy under different selected feature numbers. (b) Composition of selected color, texture, shape, and contrast descriptors. (c) Top-ranked symptom features based on composite selection score. (d) Feature redundancy reduction before and after selection.

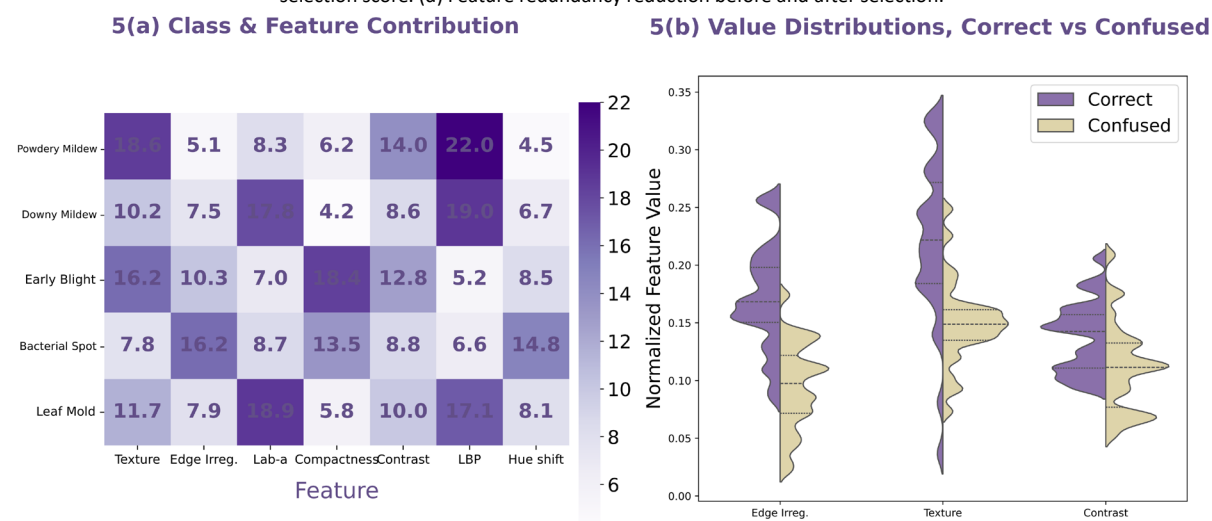


Figure 5. Category-level disease interpretation based on selected features. (a) Class-wise contribution matrix of selected symptom descriptors. (b) Comparison of dominant feature responses in correctly classified and confused disease samples.

The results of the interpretation also demonstrate how feature selection enhances model stability. The full-feature model includes a number of illumination-sensitive RGB descriptors at higher tree levels, which causes the classifier to give brightness variations an excessive amount of weight. Choose one to more uniformly distribute the texture descriptors throughout the tree splits by substituting lab and HSV descriptors for the raw RGB features. Normalization of local textural elements and color spaces makes them more resilient to variations in greenhouse lighting.

Robustness and Deployment Efficiency

For robustness tests, three greenhouse imaging disturbances are introduced: backdrop complexity, leaf overlap, and uneven illumination. The accuracy for -30% and +30% illumination changes is displayed in Figure 6(a); under extreme underexposure, SVM drops to 84.1%, while the suggested model stays above 90.8%. By adding areas of soil, support wire, and irrigation pipe close to the leaf boundary, Figure 6(b) evaluates background disturbance; the accuracy loss for the suggested model is only 2.7 percentage points, while full-feature Random Forest loses 5.9 points. The performance with Gaussian picture noise is shown in Figure 6(c), and at a noise variance of 0.03, the macro-F1 remains 91.2% [36].

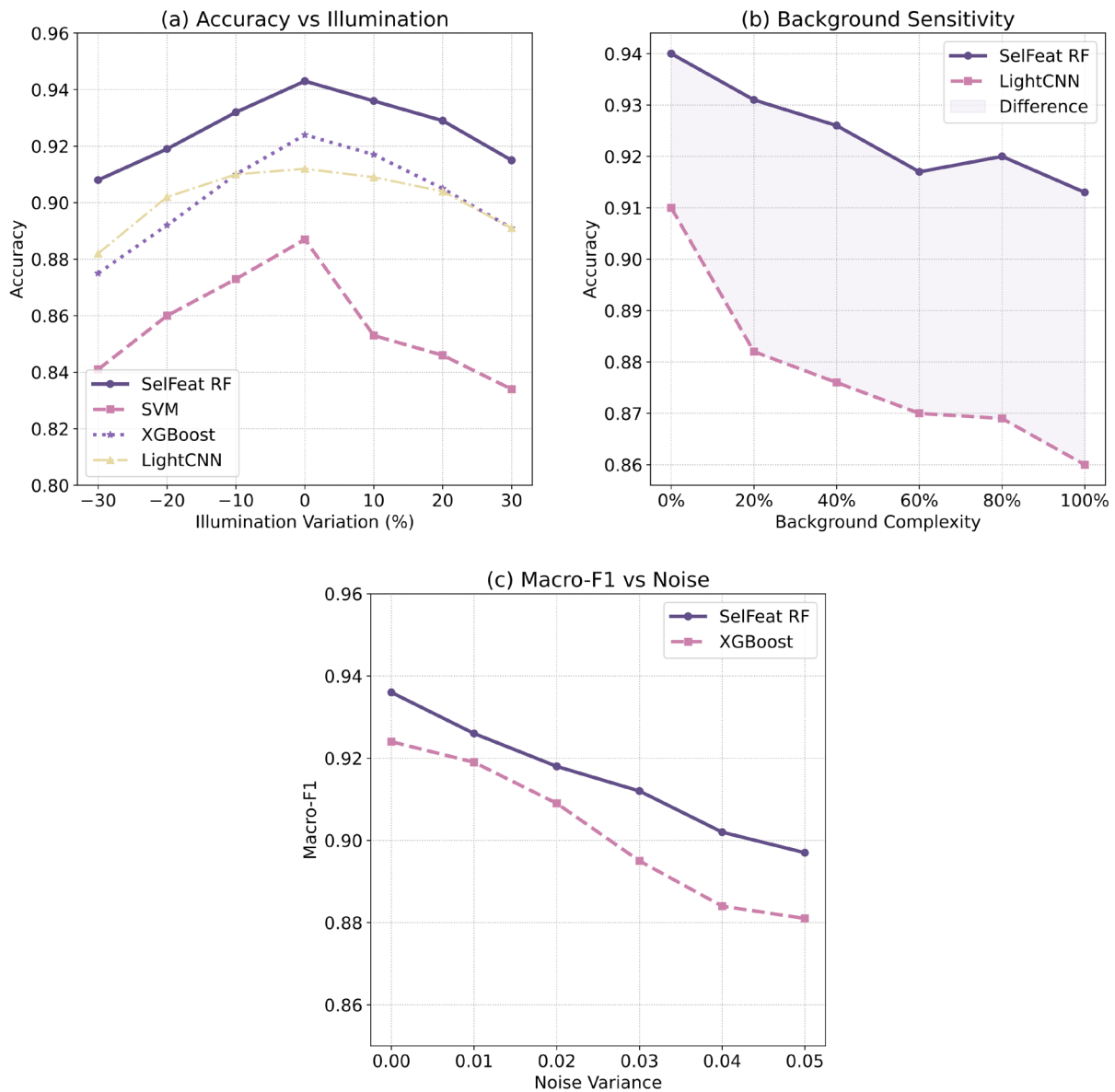


Figure 6. Robustness under greenhouse image disturbance. (a) Disease detection accuracy under illumination intensity variation. (b) Performance degradation under complex greenhouse background. (c) Macro-F1 variation under image noise perturbation.

These are the two categories of robustness. First, to lessen the influence of background pixels and restrict the descriptor's computation region to symptomatic areas. Second, several descriptors that are susceptible to leaf border artifacts or global brightness are eliminated via feature selection. The accuracy is 91.6% rather than 94.3% in the absence of color constancy preprocessing. The chosen subset grows to 113 features when redundancy suppression is turned off, but the accuracy falls to 92.7%; hence, increasing the number of features does not always increase the differentiating capability for illness diagnosis [37].

Compactness and deployment efficiency are contrasted in Figure 7. The suggested model is 3.8 MB, MobileNetV2 is 13.6 MB, and the full-feature Random Forest is 6.1 MB, as shown in Figure 7(a). The CPU inference speed for a batch size of one is displayed in Figure 7(b); the suggested model can process 156 photos per second on a desktop CPU and 41 images per second on a low-power edge board. The accuracy-efficiency trade-off is shown in Figure 7(c), and when compared to other non-deep and lightweight deep baselines, the suggested model is nearer the high-accuracy and low-latency region [38].

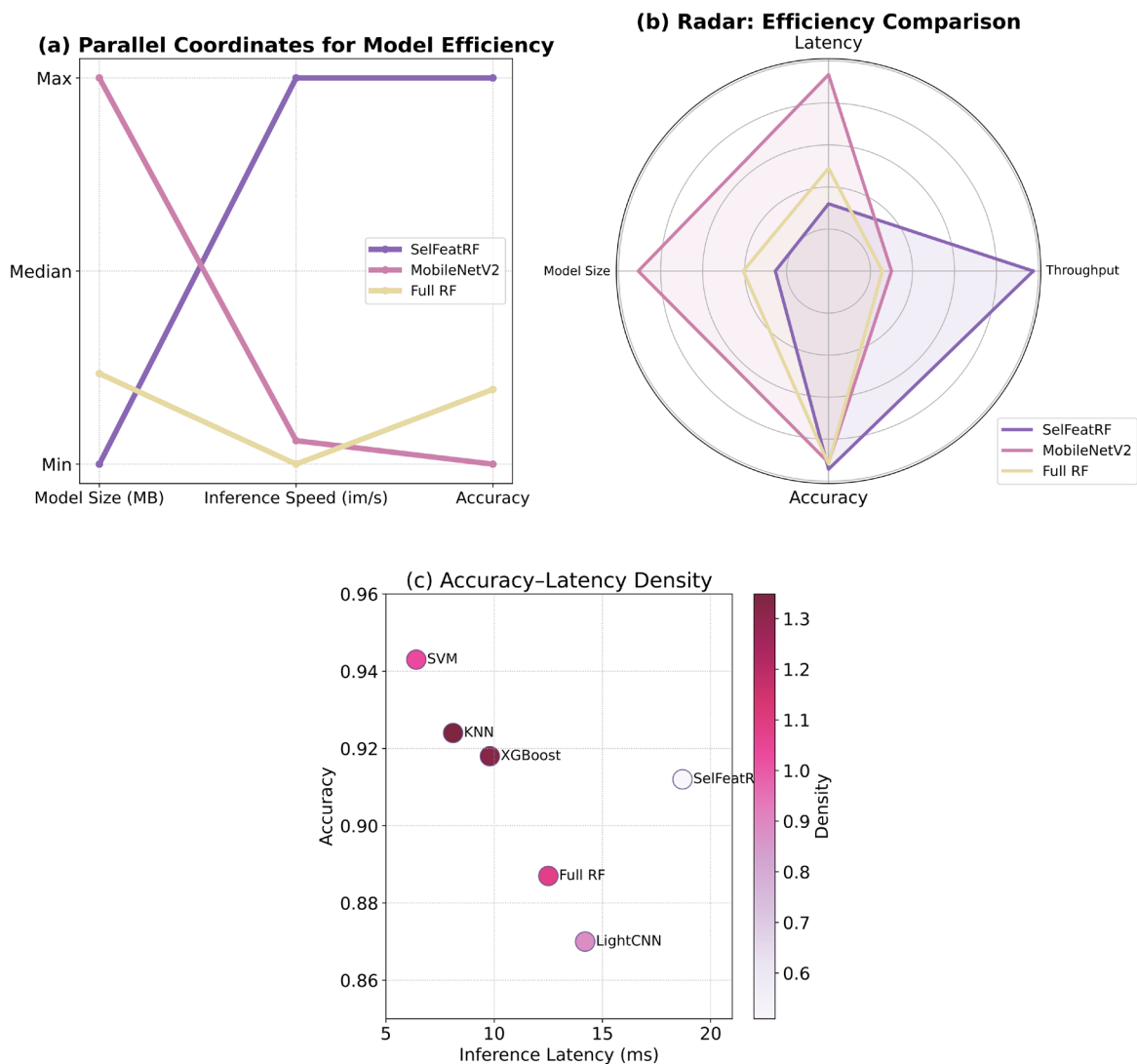


Figure 7. Model compactness and deployment efficiency comparison. (a) Model size comparison across classifiers. (b) CPU and edge-board inference throughput. (c) Accuracy-efficiency trade-off under greenhouse deployment conditions.

The contributions of particular modules are also identified using ablation studies. Because many descriptors are tainted by background and healthy leaves, excluding the extraction of lesion candidates lowers the accuracy to 91.9%. The macro-F1 drops to 91.8% when relevance ranking is removed and to 92.4% when redundancy suppression is removed. The stability of category-level explanations is reduced by 23.7% when Random Forest

significance input is eliminated. The suggested approach operates as an integrated pipeline rather than a single classifier replacement in light of the findings [39].

Under different imaging settings, a model can be used to keep an eye on a greenhouse with a few diseases. Compared to deep CNNs, the new method requires less training data and provides more understandable feature-level explanations. When compared to SVMs and single decision trees, it exhibits good non-linear separation and robustness. The primary shortcoming is that these handcrafted elements need to be adjusted when new crops or unidentified illness symptoms emerge. For better adaptability under prolonged greenhouse operation, future deployments may integrate the chosen-feature Random Forest with a lightweight lesion segmentation network or incremental feature update [40].

Conclusion

In order to detect crop diseases in photos of greenhouse vegetables, this research suggests an explainable feature selection Random Forest model. After extracting lesion-oriented color, texture, shape, border, and contrast descriptors, the model uses relevance, redundancy, and tree-based importance to choose symptom-sensitive features. The chosen feature subset is appropriate for an inexpensive greenhouse monitoring device since it increases classification accuracy while lowering the number of features and inference cost.

To put it briefly, feature selection can be used to both improve the discrimination and interpretability of disease data while reducing the number of dimensions. The suggested approach gives lesion color, texture density, edge irregularity, and compactness attributes more weight while suppressing redundant and illumination-sensitive descriptors. The category-level classification for outwardly comparable illnesses, including bacterial spot, early blight, downy mildew, and leaf mold, is thus explained more intuitively.

As a kind of deep learning technique, Feature Selection Random Forest may not be appropriate in all situations because it requires a bigger labelled dataset and is more computationally expensive than other methods. In the upcoming work, a lightweight lesion segmentation module will be introduced, the multi-view greenhouse image acquisition approach will be improved for better performance in difficult production scenarios, and incremental learning will be used to support new disease categories.

Author Contributions

Alexandra Marin contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, project administration, and funding acquisition. Mariana Constantinescu contributes to conceptualization, methodology, software, validation. Walter Rentea contributes to methodology, software. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

- [1] Janarthan, R., Thuseethan, S., Rajasegarar, S., & Yearwood, J. (2022). P2OP-Plant Pathology on Palms: A deep learning-based mobile solution for in-field plant disease detection. *Computers and Electronics in Agriculture*, 199, 107371. <https://doi.org/10.1016/j.compag.2022.107371>
- [2] Kerkech, M., Hafiane, A., & Canals, R. (2020). Vine disease detection in UAV multispectral images using optimized image registration and deep learning segmentation approach. *Computers and Electronics in Agriculture*, 174, 105446. <https://doi.org/10.1016/j.compag.2020.105446>
- [3] Zhong, Y., & Zhao, M. (2020). Research on deep learning in apple leaf disease recognition. *Computers and Electronics in Agriculture*, 168, 105146. <https://doi.org/10.1016/j.compag.2019.105146>

- [4] Jiang, P., Dong, Q., Jiang, B., Yang, Y., & Xu, X. (2021). Recognition of rice leaf diseases and wheat leaf diseases based on multi-task deep transfer learning. *Computers and Electronics in Agriculture*, 186, 106184. <https://doi.org/10.1016/j.compag.2021.106184>
- [5] Khan, M. A., Quadri, S. A., Bandy, S., & Shah, J. L. (2022). Deep diagnosis: A real-time apple leaf disease detection system based on deep learning. *Computers and Electronics in Agriculture*, 198, 107093. <https://doi.org/10.1016/j.compag.2022.107093>
- [6] Fuentes, A., Yoon, S., Kim, S. C., & Park, D. S. (2021). Open set self and across domain adaptation for tomato disease recognition with deep learning techniques. *Frontiers in Plant Science*, 12, 758027. <https://doi.org/10.3389/fpls.2021.758027>
- [7] Bedi, P., & Gole, P. (2021). Plant disease detection using hybrid model based on convolutional autoencoder and convolutional neural network. *Artificial Intelligence in Agriculture*, 5, 90-101. <https://doi.org/10.1016/j.aiia.2021.05.002>
- [8] Shrivastava, V. K., & Pradhan, M. K. (2021). Rice plant disease classification using color features: A machine learning paradigm. *Journal of Plant Pathology*, 103, 17-26. <https://doi.org/10.1007/s42161-020-00683-3>
- [9] Keivani, N., Mazloun, J., Sedaghatfar, E., & Tavakoli, H. (2020). Automated analysis of leaf shape, texture, and color features for plant classification. *Traitement du Signal*, 37(1), 17-25. <https://doi.org/10.18280/ts.370103>
- [10] Hylmi, M. S., Wiharto, W., & Suryani, E. (2022). Detection of potato leaf disease using multi-class support vector machine based on texture, color, and shape features. *2022 International Conference on Electrical and Information Technology*, 218-223. <https://doi.org/10.1109/IEIT56384.2022.9967866>
- [11] Rahul, M., & Rajesh, R. (2020). Image processing based automatic plant disease detection and stem cutting robot. *2020 Third International Conference on Smart Systems and Inventive Technology*, 1052-1057. <https://doi.org/10.1109/ICSSIT48917.2020.9214257>
- [12] Katafuchi, R., & Tokunaga, T. (2021). Image-based plant disease diagnosis with unsupervised anomaly detection based on reconstructability of colors. *Proceedings of the International Conference on Image Processing and Vision Engineering*, 112-120. <https://doi.org/10.5220/0010463201120120>
- [13] Yang, C., Hong, J., & Luo, X. (2022). Identification and detection for plant disease based on image segmentation and deep learning. *2022 IEEE International Conference on Electrical Engineering, Big Data and Algorithms*, 132-136. <https://doi.org/10.1109/EEBDA53927.2022.9744979>
- [14] Vijayakumar, V., Santhi, B., & Gokila, S. (2022). A deep learning-based approach for tomato plant leaf disease detection and classification. *2022 3rd International Conference on Smart Electronics and Communication*, 539-544. <https://doi.org/10.1109/ICOSEC54921.2022.9952007>
- [15] Menon, P., Ashwin, K., & Deepa, P. (2021). Plant disease detection using CNN and transfer learning. *2021 International Conference on Communication, Control and Information Sciences*, 1-6. <https://doi.org/10.1109/ICCISc52257.2021.9484957>
- [16] Thakur, P., Chug, A., & Singh, A. P. (2021). Plant disease detection of bell pepper plant using transfer learning over different models. *2021 8th International Conference on Signal Processing and Integrated Networks*, 327-332. <https://doi.org/10.1109/SPIN52536.2021.9565945>
- [17] Andrianto, H., Suhardi, Faizal, A., & Armandika, F. (2020). Smartphone application for deep learning-based rice plant disease detection. *2020 International Conference on Information Technology Systems and Innovation*, 387-392. <https://doi.org/10.1109/ICITSI50517.2020.9264942>
- [18] Babu, G. R., Maravarman, M., & Pitchai, R. (2022). Detection of rice plant disease using deep learning techniques. *Journal of Mobile Multimedia*, 18(3), 717-734. <https://doi.org/10.13052/jmm1550-4646.18314>
- [19] Applalanaidu, V., & Kumaravelan, G. (2021). A review of machine learning approaches in plant leaf disease detection and classification. *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks*, 716-722. <https://doi.org/10.1109/ICICV50876.2021.9388488>
- [20] Thenmozhi, K., & Reddy, U. S. (2019). Crop pest classification based on deep convolutional neural network and transfer learning. *Computers and Electronics in Agriculture*, 164, 104906. <https://doi.org/10.1016/j.compag.2019.104906>
- [21] Apley, D. W., & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society: Series B*, 82(4), 1059-1086. <https://doi.org/10.1111/rssb.12377>

- [22] Lundberg, S. M., Erion, G. G., Lee, S. I., & others. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2, 56-67. <https://doi.org/10.1038/s42256-019-0138-9>
- [23] Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82-115. <https://doi.org/10.1016/j.inffus.2019.12.012>
- [24] Ryo, M. (2022). Explainable artificial intelligence and interpretable machine learning for agricultural data analysis. *Artificial Intelligence in Agriculture*, 6, 257-265. <https://doi.org/10.1016/j.aiia.2022.11.003>
- [25] Arora, P., & Kaur, H. (2020). A Bolasso based consistent feature selection enabled random forest classification algorithm: An application to credit risk assessment. *Applied Soft Computing*, 86, 105936. <https://doi.org/10.1016/j.asoc.2019.105936>
- [26] Narasimhulu, C. V. (2021). An automatic feature selection and classification framework for analyzing ultrasound kidney images using dragonfly algorithm and random forest classifier. *IET Image Processing*, 15(9), 1964-1976. <https://doi.org/10.1049/ipr2.12179>
- [27] Singh, V., Chourasia, A., & Raghuwanshi, M. M. (2021). BPSO based feature selection for rice plant leaf disease detection with random forest classifier. *International Journal of Engineering Trends and Technology*, 69(4), 39-44. <https://doi.org/10.14445/22315381/IJETT-V69I4P206>
- [28] Wan, R., Tian, Y., Zhang, Y., Liu, Z., & Wang, N. (2022). A multivariate temporal convolutional attention network for time-series forecasting. *Electronics*, 11(10), 1516. <https://doi.org/10.3390/electronics11101516>
- [29] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(12), 11106-11115. <https://doi.org/10.1609/aaai.v35i12.17325>
- [30] Khoder, J., & Dornaika, F. (2021). Ensemble learning via feature selection and multiple transformed subsets: Application to image classification. *Applied Soft Computing*, 113, 108006. <https://doi.org/10.1016/j.asoc.2021.108006>
- [31] Richey, T., & Shirvaikar, M. (2021). Deep learning based real-time detection of northern corn leaf blight crop disease using YOLOv4. *Real-Time Image Processing and Deep Learning 2021*, 11736, 1173604. <https://doi.org/10.1117/12.2587892>
- [32] Kolli, S., Vamsi, K., & Manikandan, V. M. (2021). Plant disease detection using convolutional neural network. *2021 IEEE Bombay Section Signature Conference*, 1-6. <https://doi.org/10.1109/IBSSC53889.2021.9673493>
- [33] Chanda, S., & Sarkar, S. (2021). Effective classification of plant disease using image processing and machine learning. *2021 Innovations in Power and Advanced Computing Technologies*, 1-6. <https://doi.org/10.1109/i-PACT52855.2021.9696947>
- [34] F. Danilevicz, M., & Bayer, P. E. (2022). Machine learning for image analysis: Leaf disease segmentation. *Methods in Molecular Biology*, 2464, 411-424. https://doi.org/10.1007/978-1-0716-2067-0_22
- [35] Kumar, A., & Jindal, R. (2022). Survey of plant disease detection techniques based on image processing and machine learning. *International Journal of Health Sciences*, 6(S4), 10898-10913. <https://doi.org/10.53730/ijhs.v6nS4.10096>
- [36] Sungheetha, A. (2022). State of art survey on plant leaf disease detection. *Journal of Innovative Image Processing*, 4(2), 111-124. <https://doi.org/10.36548/jiip.2022.2.004>
- [37] Kumar, A., Shukla, A., & Princee, P. (2022). Plant disease detection and crop recommendation using CNN and machine learning. *2022 International Mobile and Embedded Technology Conference*, 589-594. <https://doi.org/10.1109/MECON53876.2022.9752173>
- [38] Patel, V., Rathod, H., & Rathod, T. (2022). Detection of disease using machine learning. In *Neural Networks, Machine Learning, and Image Processing* (pp. 153-166). CRC Press. <https://doi.org/10.1201/9781003303053-14>
- [39] Aggarwal, M. (2022). Biological tomato leaf disease classification using deep learning framework. *International Journal of Biology and Biomedical Engineering*, 16, 270-277. <https://doi.org/10.46300/91011.2022.16.30>
- [40] Tjoa, E., & Guan, C. (2021). A survey on explainable artificial intelligence (XAI): Toward medical XAI. *IEEE Transactions on Neural Networks and Learning Systems*, 32(11), 4793-4813. <https://doi.org/10.1109/TNNLS.2020.3027314>