International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 14 Issue: 2
***Article Received:** 25 April 2026 **Revised:** 12 May 2026 **Accepted:** 30 June 2026*

Orange Disease Classification Using Hybrid Deep Feature Extraction and Machine Learning Classification

C.Mahalakshmi[1],
Research Scholar, Department of Computer and Information Science,
Annamalai University, Annamalai Nagar Chidambaram, Tamil Nadu, India, mahacisau@gmail.com

Dr. S. Saravanan[2],
Assistant Professor/Programmer, Department of Computer and Information Science,
Annamalai University, Annamalai Nagar, Chidambaram, Tamil Nadu, India, aucissaran@gmail.com

Abstract— Orange cultivation is a major crop worldwide. However, citrus plants are vulnerable to many diseases, including Greening, Black Spot and Canker, which lead to crop quality deterioration and yield loss. Traditionally, detection of disease relies on human scouting that is hardly repeatable, so there is a growing interest in automated techniques to address the problem. The present work suggests a new hybrid method for orange disease detection that integrates deep feature extraction with backbones EfficientNetB3, and classical machine learning classifiers - K-Nearest Neighbors (KNN), Support Vector Machine (SVM) and Random Forest (RF). This proposed experiment utilized deep features in conjunction with traditional classifiers to gain a tradeoff between good performance, low computing power demand, and model transparency thus making the solution feasible for developing countries or rural areas. Testing was done with a publicly available dataset on Kaggle which comprises four classes: Greening, Canker, Black Spot, and Healthy oranges. The experimental results revealed that SVM recorded more overall accurateness of 94.6% while the instances causing the misclassification were mostly between visually very similar Canker and Black Spot samples. Hence, the proposed method is a practical example of combining deep learning and classical machine learning approaches for an efficient, correct and deployable orange disease identification tool that can support precision agriculture.

Keywords— *Orange disease classification, Citrus disease detection, Hybrid deep learning, EfficientNetB3, SVM classifier, Image-based diagnosis*

I. INTRODUCTION

Orange cultivation represents a significant sector of global agriculture contributing to food supply, and rural economies [1]. However, orange and citrus crops are at risk from a variety of diseases that negatively impact fruit quality, yield, and economic value. Some common diseases in orange crops are citrus canker, citrus greening (Huanglongbing), black spot, anthracnose, scab, melanose, and sooty mold. If necessary measures are not taken at an early stage, these diseases may get out of control and cause huge commercial losses [2].

Conventional approaches for disease identification mostly depend on agricultural experts visually inspecting plants, which is an approach that suffers from subjecexafor lar SA Developments in artificial intelligence and computer vision have made it possible to automatically detect diseases through image-based analysis. Deep learning, especially Convolutional Neural Networks (CNNs), have demonstrated a great potential in extracting highly discriminative features from fruit images [3]. Transfer learning allows making the most of pretrained networks to further improve the models, thus lessening both training complexity and the amount of required data. On the other hand, such end-to-end deep learning models may have vigorous requirements for computational resources, and this could pose a problem if they were to be deployed in real agricultural environments that have limited resources[4].

In order to overcome these drawbacks, methods that integrate deep learning with traditional machine learning techniques have been gaining attention[5]. In such systems, deep learning architectures are used for automated feature extraction, while it is the task of machine learning classifiers to perform the final classification. This symbiosis offers increased flexibility, better interpretability, and computational efficiency thus making it suitable for real-world agricultural applications[6].

This study introduces a hybrid machine learning and deep learning pipeline for classifying orange disorders utilising photos sourced from a publicly accessible Kaggle resource. The dataset contains four classes: Greening, Canker, Black Spot, and Fresh (healthy) oranges. Among these disease types, those presenting the highest frequency, the severest economic impact, and the most evident visual characteristics were chosen. Greening is considered as one of the devastating citrus diseases globally, whereas canker and black spot are fruit appearance-related diseases that affect their marketability. Healthy oranges inclusion is critical to obtain a reliable diseased versus healthy discrimination. Figure 1

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 14 Issue: 2
***Article Received:** 25 April 2026 **Revised:** 12 May 2026 **Accepted:** 30 June 2026*

displays sample images for the respective classes of the orange disease dataset, namely (a) Fresh, (b) Greening, (c) Canker and (d) Black Spot. Through targeting these major disease categories only, the proposed method is expected to facilitate early disease diagnosis and thus, enable better decision-making in precision agriculture. The Fig. 1 shows the sample images from the orange disease dataset.

a) Greening b) Fresh c) Canker d) Blackspot

Fig. 1 Sample images from the orange disease dataset for (a) Fresh,
(b) Greening, (c) Canker and (d) Black Spot

II. LITERATURE REVIEW

Deep learning has recently advanced to a point where it can automatically and accurately detect diseases in citrus and orange crops. These methods also diminish the need for manual inspection and facilitate the implementation of timely disease control measures in farming.

Study [7] proposed an architecture of deep learning-based Orange crop disease identification using DenseNet-121 such as black spot and greening. Transfer learning along with image enhancement helped to better features extraction from a small sample of dataset. The paper demonstrated the use of DenseNet architecture for disease identification in farming images.

Study [8] presented a CNN-based labeling method through the use of pretrained deep learning models, mainly ResNet-50, combined with the Context Data Fusion method and Faster-CNN deployed for real-time processing on an edge computing platform. The algorithm separated the orange pictures in fresh and diseased categories, thus proving the pretrained CNN models can be effectively used for automated identification of plant diseases in real time.

Likewise, [9] integrated a Faster-CNN model on the edge computing system to identify and classify six citrus diseases i.e. canker, black spot, greening, scab, melanose, and healthy fruits. A Context Data Fusion approach was used to merge RGB and near-infrared (NIFR) images, which further improved robustness of disease detection by transfer learning.

By means of [10], a comprehensive study of several transfer learning-based deep learning architectures such as ResNet50, InceptionV3, MobileNetV2, and EfficientNetB3 was facilitated by the Citrus Plant Dataset. The paper showed modern convolutional architectures to be highly effective in citrus disease identification.

Research [11] was about the identification and categorization of orange maladies by adopting the MobileNetV2 model. The author utilized transfer learning from the pretrained ImageNet model, data augmentation, and a stratified division of the dataset to boost the model generalization capabilities. The study results suggested that compact deep learning models could be a good fit for image-based recognition of orange diseases. As per [12], a range of machine learning models was tested and in particular, CNN-based models utilizing MobileNetV2 were given more emphasis for the classification of the four categories of orange fruit condition: Black spot, Canker, Fresh, and Greening. The authors found that CNN-based methods are capable of differentiating among many disease classes in citrus fruits accurately.

Research [13] detailed a MobileNetV2-based deep learning framework that was trained upon a publicly accessible dataset from Kaggle. The dataset was partitioned into training and validation sets. The author employed transfer learning, dropout, and average pooling layers to enhance the model's generalisation capacity, demonstrating that pretrained lightweight models for classifying citrus diseases possess significant potential.
The hybrid categorizing technique in [14] was done by incorporating a Convolutional Neural Network for extracting features with a Support Vector Machine (SVM) for the classification of diseases. The authors discussed how important the lesion properties such as the shape, size, and texture are for differentiating among various orange diseases.

In conclusion, [15] revealed a transfer learning-based system with deep CNN models including InceptionV3, ResNet50, VGG16, and VGG19, supported with the data augmentation techniques. The authors proved that transfer learning can reduce the computational intensity of the task while retaining a high level of robustness in disease classification, which thus makes this technique highly usable for agricultural practices.

Research Gap and Motivation

Deep learning and transfer learning-based methods have been the main focus for detecting diseases in oranges and citrus and consumption of such treatment products. However, most of the studies depend entirely on comprehensive neural networks for classification and feature extraction without considering other options. Such methods normally need high-performance computing and large datasets, which unintentionally can be the reasons for limitations in their efficiency, interpretability, and deployment in the agricultural environment with scarce resources. Besides, traditional machine learning methods that are known for their robustness and less computational complexity, have been scarcely combined with deep learning-based feature extraction approaches of citrus diseases classification. Those few attempts that explore hybrid models are not only narrow and limited but also they do not inquire into into the mutual benefits of machine learning and deep learning classical models.

Therefore, an integrated method is needed which merges deep learning for automatic and highly discriminative machine learning algorithms and feature extraction for fast and accurate classification. In response to these shortcomings, a hybrid machine learning–deep learning architecture is put forward by this research work which is capable of not only improving the classification accuracy but also lessening the computational load, thus, making orange disease detection more convenient and scalable for real-world agricultural scenarios. The Fig 2. Shows Workflow of the proposed hybrid orange disease classification.

III. METHODOLOGY

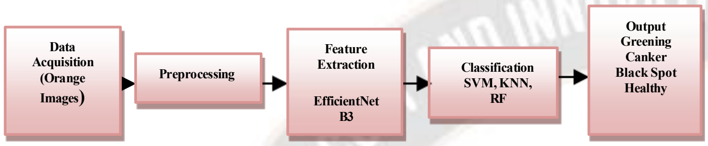

Fig 2. Workflow of the proposed hybrid orange disease classification

Dataset Description

The Orange Diseases Dataset on Kaggle provided the data used in this study, which features photos of orange fruits and leaves that are divided into four categories: Greening (Huanglongbing), Citrus Canker, Black Spot, and Healthy (Fresh). The pictures have different backgrounds, lighting conditions, and disease severities, thus they realistically represent the patterns of orange diseases for training and testing models.

There were 347 images of Greening, 281 images of Healthy, 179 images of Canker, and 184 images of Black Spot for training. The testing set included 22 images of Greening, 33 images of Healthy, 22 images of Canker, and 22 images of Black Spot. To align with the EfficientNet architecture, Pixel values were normalised to the [0,1] range and all images were scaled to 224 × 224 pixels. Additionally, a number of data augmentation methods, including rotation, flipping, zooming and brightness adjustments were employed to improve the model's capacity for generalisation.

EfficientNetB3

Tan and Le (2019) created it as a component of the EfficientNet model family. It is a convolutional neural network that uses a compound scaling technique to simultaneously modify the network's depth, breadth and resolution. This architecture allows it to deliver a balance among accuracy, computational cost and size of the model, thus it can be used efficiently on small-to-medium datasets like the orange disease dataset in this work. The model first receives four 300 × 300 × 3 images as input. A 3 × 3 convolutional layer with 32 filters and a stride of 2 is then applied to the input for feature mapping at the low-level (edges, color gradients, textures, etc.). The resulting feature maps are then passed through several Squeeze-and-Excitation (SE) modules combined with Mobile Inverted Bottleneck Convolution (MBConv) blocks which eliminate channel-wise feature responses in an adaptive way.

These MBConv blocks differ in filter numbers, kernel sizes, expansion rates, and stride thus they progressively acquire more complex and abstract features. Moreover, early MBConv layers can understand simple shapes, whereas the deeper ones are capable of identifying disease spots, lesions as well as color changes on orange leaves and fruits.

In addition to that, strided convolutions lower the spatial size of the feature maps, which allows hierarchical feature extraction and simultaneously, decreases the computation price. The MBConv blocks are followed by a 1 × 1 convolutional layer with 1280 filters. Global average pooling is then used to obtain a fixed-length 1280-dimensional features vector that locates the image's main information both spatially and chromatically.

Transfer learning is done by using pretrained weights from ImageNet, and the obtained feature vectors serve as input to conventional machine-learning classifiers like KNN, SVM and Random Forest to accurately make a distinction among four classes: Greening, Canker, Black Spot, and Healthy.

When compared to bigger EfficientNet models like B7, EfficientNetB3 has much fewer parameters which not only reduces the danger of overfitting but also keeps the ability of feature extraction in a good balance.

EfficientNetB3 thus blends deep feature extraction with classical classifiers to give high representational power with a relatively low computational cost, and for this reason, it is extremely successful for small-to-medium-sized datasets.

Machine Learning Classifiers

In this proposed work, the feature vectors obtained from EfficientNetB3 are employed as the input of three standard machine learning classifiers: K-Nearest Neighbors (KNN), Support Vector Machine (SVM) and Random Forest (RF) [15]. SVM is a supervised learning approach that can handle nonlinear interactions by using kernel functions to identify the best hyperplane to divide the classes in the feature space of potentially very high dimensions. KNN is an instance-based, non-parametric learning technique that gives the sample to be classified the label of most of the k nearest neighbours in the feature space with Euclidean distance being the most used measure of similarity. Random Forest is an ensemble learning technique for regression and classification that works by building a large number of decision trees during training and producing the class that is the mean prediction (for regression) or the mode of the classes (for classification) of the individual trees., thus reducing variance, yielding improved prediction accuracy, and thus also reduces overfitting. By leveraging the deep feature extraction power of EfficientNetB3 with the above classifiers, the system ultimately enables accurately.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 14 Issue: 2
***Article Received:** 25 April 2026 **Revised:** 12 May 2026 **Accepted:** 30 June 2026*

IV EXPERIMENTAL RESULT

A dataset of images from four classes—Greening, Canker, Black Spot, and Healthy—was gathered via Kaggle and used to assess the suggested orange disease classification framework. Training and testing sets were created from the dataset using the following distribution: 347 Greening, 281 Healthy, 179 Canker, and 184 Black Spot images for training, and 22 Greening, 33 Healthy, 22 Canker and 22
Black Spot images for testing, resulting in an approximate training-to-testing ratio of 80:20.
The feature extractor used was EfficientNetB3, generating 1280-dimensional feature vectors for each image. These features were then classified using K-Nearest Neighbors (KNN), Support Vector Machine (SVM) and Random Forest (RF) classifiers. Accuracy, precision, recall, and F1-score were used to assess performance. The findings show that SVM attained the greatest overall performance, followed by Random Forest and KNN. Misclassifications mostly occurred between Canker and Black Spot, which share similar visual characteristics, while Greening and Healthy classes were classified with high accuracy. These results demonstrate that using deep features from EfficientNetB3 significantly improves classification performance compared to training traditional classifiers on raw images. The Fig 3 shows Confusion matrices showing the classification performance of SVM, KNN, and Random Forst classifiers using features extracted from EfficientNetB3.

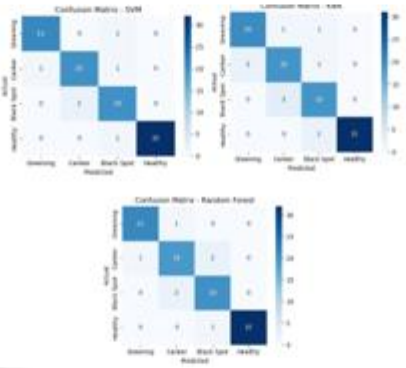

Fig 3. Confusion matrices showing the classification performance of SVM, KNN, and Random Forest classifiers using features extracted from EfficientNetB3.

The Table 1 shows the Classification performance of SVM, KNN and Random Forest classifiers using features extracted from EfficientNetB3 for the orange disease dataset. The Fig 4 shows the accuracy comparison of SVM, KNN and Random Forest classifiers using features extracted from EfficientNetB3.

Class	Precision (%)			Recall (%)			F1-Score (%)		
	SVM	KNN	RF	SVM	KNN	RF	SVM	KNN	RF
Greening	95.5	90.9	95.5	95.5	90.9	95.5	95.5	90.9	95.5
Canker	90.9	86.4	86.4	90.9	86.4	86.4	90.9	86.4	86.4
Black Spot	90.9	86.4	90.9	90.9	86.4	90.9	90.9	86.4	90.9
Healthy	97	93.9	97	97	93.9	97	97	93.9	97
Overall	94.6	90.9	92.7	94.6	90.9	92.7	94.6	90.9	92.7

Table 1. Classification performance of SVM, KNN and Random Forest classifiers using features extracted from EfficientNetB3 for the orange disease dataset.

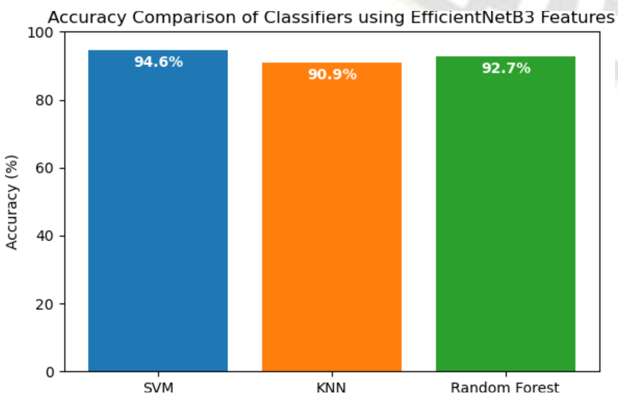

Fig 4. Accuracy comparison of SVM, KNN and Random Forest classifiers using features extracted from

EfficientNetB3.

V FUTURE ENHANCEMENT

Future research can primarily focus on the orange disease identification system, which could be boosted by using bigger and more varied datasets for model generalization. To generate artificial samples of the rare disease categories, advanced data augmentation methods such as generative adversarial networks (GANs) could be utilised. Additionally, to improve feature extraction and classification accuracy, attention-based architectures or ensemble deep learning models should be explored. Automating the operation of real-time detection on edge devices or mobile can also make the system more suitable for use in the field of the orchards.

VI CONCLUSION

In order to classify orange diseases, this study combined classic machine learning classifiers including SVM, KNN and Random Forest with EfficientNetB3-based feature extraction. The system was evaluated on a Kaggle dataset containing four classes: Greening, Canker, Black Spot, and Healthy. EfficientNetB3 effectively captured deep features representing color, texture, and shape, which were then enhanced and fed into the classifiers. SVM demonstrated the efficacy of integrating deep feature extraction with traditional machine learning techniques by achieving the greatest overall accuracy of 94.6% among the classifiers. The proposed approach provides an accurate and efficient system for detecting orange diseases and future enhancements can further improve generalization and enable real-time deployment in orchard environments.

REFERENCES

[1] Ayobami IO, Oluwabusayo I, Olajide AO. Classification and detection of citrus disease using feature extraction and support vector machine (SVM). *Int J Comput Appl*. 2019;177(17):17–25. https:// doi. org/ 10. 5120/ ijca2 01991 9582.

[2] Farooq MS, Mehboob A. Prediction of citrus diseases using machine learning and deep learning: classifier, Models SLR. Cornell University. 2023. https:// doi. org/ 10. 48550/ arxiv. 2306. 01816.

[3] Sharif M, Khan MA, Iqbal Z, Azam M, Lali MIU, Javed MY. Detection and classification of citrus diseases in agriculture based on optimized weighted segmentation and feature selection. *Comput Electron Agric.* 2018;150:220–34. https:// doi. org/ 10. 1016/j. compag. 2018. 04. 023.

[4] Kaur G, Sharma N, Malhotra S, Devliyal S, Kumar BV. Citrus Leaf Disease Classification Using VGG16 Convolutional Neural Network Architecture. *In 2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*. 2024, April; 1–6. IEEE.

[5] Tanwar V, Anand V, Aggarwal P, Kumar M. Demonstration of a Highly Developed CNN-SVM Model to Accurately Assess the Degree of Brown Rot in Orange Leaf. *2024 Fourth International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), Bhilai, India*. 2024; 1–6, https:// doi. org/ 10. 1109/ ICAEC T60202. 2024. 10468 934.

[6] R. Indhumathi and A. Geetha, “Machine learning approach for head gesture classification,” *Journal of Emerging Technologies and Innovative Research (JETIR)*, vol. 5, no. 12, pp. 175–183, Dec. 2018.

[7] Banerjee D, Kukreja V, Chamoli S, Hariharan S, Thapliyal S. Enhanced Citrus Fruit Disease Detection: Combining CNN and Random Forest Models. *In 2024 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC)*. 2024, January; 1–7. IEEE.

[8] A. Goyal and K. Lakhwani, "Integrating advanced deep learning techniques for enhanced detection and classification of citrus leaf and fruit diseases," *Scientific Reports*, vol. 15, p. 12659, 2025.

[9] Z. Huang, X. Jiang, S. Huang, S. Qin, and S. Yang, "An efficient convolutional neural network-based diagnosis system for citrus fruit diseases," *Frontiers in Genetics*, 2023.

[10] K. N. Arifin, S. A. Rupa, M. M. Anwar, and I. Jahan, "Lemon and orange disease classification using CNN-extracted features and machine learning classifier," *arXiv preprint arXiv:2408.14206*, 2024.

[11] E. Ahmed, M. F. Hassan, and A. K. Jaber, "Classification of Citrus Diseases Using Optimization Deep Learning Approach," *PubMed*, 2022. M. Sharma, A. Patel, and R. Singh, "Plant leaf disease detection and classification using convolution neural networks model: A review," *Artificial Intelligence Review*, 2025.

[12] S. Verma and P. Gupta, "Fruit Disease Detection using AI: A Review of Classical and Deep Learning Approaches," *Indian Journal of Agricultural Research*, 2026.

[13] Z. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Machine Learning (ICML)*, Long Beach, CA, USA, 2019, pp. 6105–6114.

[14] R. Indhumathi and A. Geetha and, “Real time interface for assistive e-learning,” *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 8, no. 8, pp. 2434–2439, Jun. 2019.