

Leaf Disease Detection and Pesticide Recommendation Using Pretrained CNN Models in Keras on the Augmented New Plant Diseases Dataset

Vijayalakshmi S. Abbigeri

School of Computing and Information Technology, REVA University, Bengaluru, India
vijayalaxmisa@gmail.com (corresponding author)

Geetha D. Devanagavi

School of Computing and Information Technology, REVA University, Bengaluru, India
dgeetha@reva.edu.in

Received: 10 October 2025 | Revised: 5 November 2025 | Accepted: 10 November 2025

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.15443>

ABSTRACT

Timely and precise identification of diseases in plant leaves is crucial for protecting crops and promoting sustainable agricultural practices. In this work, a custom Convolutional Neural Network (CNN) model built with Keras is presented to identify 38 different classes of leaf diseases from the Augmented New Plant Diseases Dataset. The model architecture consists of several convolutional layers with increasing filter sizes (from 32 to 512), ReLU activation functions, and max pooling for spatial downsampling. It concludes with fully connected layers and includes dropout to reduce the risk of overfitting. Training is performed using the Adam optimizer with a learning rate of 0.0001, and the loss is calculated using the sparse categorical cross-entropy function. To improve the model's robustness and adaptability to diverse lighting and environmental conditions, data augmentation methods are applied. The results confirm the model's high accuracy, with a training accuracy of 98.37% and a test accuracy of 96.29%, supporting its application for real-time detection of plant diseases at scale. Furthermore, a rule-based system is incorporated to recommend suitable pesticides based on the detected disease category. This approach emphasizes the value of deep learning in advancing smart agriculture and providing automated support for disease identification and treatment decisions.

Keywords-plant leaf disease; deep Convolutional Neural Networks (CNNs); data augmentation; Keras; augmented new plant diseases dataset

I. INTRODUCTION

Agricultural productivity is a cornerstone of food security and economic stability, particularly in regions where farming is the primary livelihood. However, plant diseases pose a persistent threat to crop health, often resulting in substantial yield losses and economic damage [1]. Recent estimates indicate that plant pathogens and pests account for a significant portion of agricultural losses, often ranging from 20% to 40% of global crop yields annually, highlighting the critical need for effective disease management strategies. Detecting plant diseases manually can be slow, requires significant effort, and is often susceptible to mistakes due to human limitations, especially in early disease stages or when multiple symptoms are visually similar [2]. Traditional methods relying on laboratory tests are often costly, require specialized expertise, and lack the scalability required for monitoring large agricultural areas. Furthermore, non-deep learning computer

vision approaches are limited by their reliance on manual feature extraction, which is time-consuming and often fails to generalize across diverse environmental conditions. These challenges underscore the need for automated, scalable, and accurate plant disease diagnosis systems.

Recent progress in computer vision and deep learning has enabled the development of more intelligent and efficient systems for monitoring crop health [3]. Convolutional Neural Networks (CNNs) have shown strong performance in image classification, which makes them well-suited for disease identification in plant foliage [4]. CNNs are capable of learning hierarchical image features automatically, removing the reliance on manual feature extraction and allowing reliable plant disease classification across varying environmental conditions [5].

This study introduces a framework powered by deep learning techniques for plant leaf disease identification and

pesticide recommendation using a custom-designed CNN implemented in Keras [6]. The model is trained and evaluated on the Augmented New Plant Diseases Dataset [7], which comprises over 50,000 labeled images across 38 classes, including healthy and diseased plant leaves. Extensive data augmentation techniques are applied during training to improve the model's ability to generalize and minimize the risk of overfitting.

Our CNN architecture is composed of several convolutional layers with increasing filter depths, interleaved with max-pooling operations to progressively reduce spatial dimensions. The final dense layers incorporate dropout for regularization, followed by a softmax output layer for classification into multiple categories. The model is optimized using the Adam optimizer and trained using sparse categorical cross-entropy loss.

Furthermore, a rule-based module is integrated into the framework to provide pesticide recommendations based on the predicted disease class. This hybrid approach not only enables high-accuracy disease classification but also supports actionable agronomic decision-making. The experimental results validate the efficiency of our method, showing high classification accuracy and strong generalization across multiple disease types.

This research contributes to the development of intelligent, automated plant disease monitoring systems and supports the broader objective of precision agriculture through Artificial Intelligence (AI)-driven diagnosis and decision support.

We present a robust performance analysis using a custom-designed CNN model, achieving a high-accuracy benchmark of 98.25% on the large Augmented New Plant Diseases Dataset for 38 distinct disease classes. We also develop an end-to-end system that successfully integrates high-accuracy disease detection with a practical rule-based pesticide recommendation system, bridging the crucial gap between diagnosis and immediate intervention for farmers. A comprehensive evaluation using detailed metrics, including precision (PPV) and recall (sensitivity), ensures a complete and transparent assessment of the model's diagnostic reliability.

II. RELATED WORK

The use of machine learning (ML) techniques, especially deep learning, has revolutionized plant disease detection and classification [8]. Various ML approaches have been utilized for identifying and classifying plant diseases; however, with the emergence of deep learning—a specialized branch of ML—this field has seen significant improvements in accuracy and overall potential. Common plant diseases encompass root rot, powdery mildew, mosaic virus, leaf spots, and fruit rot [9]. To address these challenges, deep learning provides high classification accuracy and greater processing speed. CNNs consist of multiple layers that can automatically detect and classify diseases present in plant images. Many studies have demonstrated the superiority of CNNs over traditional ML methods for plant disease identification [10].

For example, several CNN-based models have been trained on diverse plant disease datasets, consistently delivering high

accuracy in detecting a wide range of diseases across multiple plant types. These findings demonstrate CNNs' capability to automatically extract relevant features from raw images, eliminating the need for handcrafted feature engineering. In one notable study, a mobile application was developed that could analyze plant images captured in real-world agricultural settings and return results in under a second [11]. In many of these works, there is extensive utilization of data augmentation techniques, such as flipping, rotation, and scaling [12].

However, several challenges remain in the practical application of deep learning models for plant disease diagnosis. The greatest limitation of existing non-deep learning methods is their reliance on manual feature extraction, which is time-consuming, prone to human error, and lacks the necessary scalability for large farming operations. Variations in image quality due to lighting conditions, camera angles, and background noise can affect the model's performance [13]. Moreover, the computational demands of certain CNN architectures can pose challenges for deployment on devices with limited resources, such as smartphones or embedded systems.

Authors in [14] conducted a seminal study using pre-trained deep learning models to recognize 26 diseases across 14 crop species, utilizing the PlantVillage dataset. Their approach involved fine-tuning AlexNet and GoogleNet architectures, achieving high accuracy in classifying healthy and diseased leaves. Authors in [15] presented Caffenet, a single GPU variation of Alexnet, for classifying 13 different classes of leaf diseases using a dataset of around 34,000 images.

Other research efforts have concentrated on creating efficient, lightweight CNN architectures designed for deployment on mobile devices, allowing for real-time disease detection in agricultural settings [12]. Most studies in the current literature primarily use the PlantVillage dataset, often focusing on diagnosing diseases in individual plant species rather than classifying all plant diseases present in the dataset [16]. Recent research also indicates a trend towards using deep learning techniques that offer faster convergence rates for plant disease recognition [17]. Transfer learning approaches, leveraging pre-trained models on large datasets like ImageNet, have also gained popularity in plant disease detection [18].

Building on prior studies, our research focuses on enhancing current methodologies by introducing an effective and streamlined CNN model for identifying plant diseases and recommending appropriate pesticides, utilizing a thoroughly augmented dataset for training [19].

III. METHODOLOGY

A. Dataset Overview

The proposed deep learning framework is developed and evaluated using the Augmented New Plant Diseases Dataset, a publicly available and extensively used benchmark dataset for plant disease classification tasks. This dataset consists of high-resolution images of healthy and diseased plant leaves across a diverse range of plant species.

The dataset is organized into training and validation subsets, located in separate directories. It includes 38 distinct classes, covering various plant species and disease types such as Apple Scab, Grape Black Rot, Tomato Mosaic Virus, Potato Early Blight, and several healthy leaf conditions. Each class represents either a specific plant disease or a healthy leaf category, allowing for fine-grained multi-class classification.

The data are loaded using the `image_dataset_from_directory` method from TensorFlow, which automatically labels the images based on the subdirectory names and resizes them to a uniform input shape of 256×256 pixels. This facilitates efficient batch loading and preprocessing during model training. The entire dataset was split into a ratio of 80% for the training set and 20% for the validation/testing set to ensure robust and unbiased evaluation. A sample listing of the class labels is shown below in Figure 1 to highlight the diversity of disease types covered.

```
0: Apple__Apple_scab
1: Apple__Black_rot
2: Apple__Cedar_apple_rust
3: Apple__healthy
4: Blueberry__healthy
5: Cherry_(including_sour)__Powdery_mildew
6: Cherry_(including_sour)__healthy
7: Corn_(maize)__Cercospora_leaf_spot Gray_leaf_spot
8: Corn_(maize)__Common_rust_
9: Corn_(maize)__Northern_Leaf_Blight
```

Fig. 1. Disease classes included in the Augmented New Plant Diseases Dataset.

The training and validation directories are structured as follows:

- Training set: Contains the majority of the samples for each class, used to train the model.
- Validation set: Used for evaluating the model's performance on unseen data during training to monitor generalization.

The dataset is augmented to increase variability and robustness. This includes transformations such as random rotation, flipping, zooming, and color jittering, which improve the model's ability to handle real-world image distortions.

B. Data Preprocessing and Training Data Enrichment

To improve the model's robustness and ability to generalize across diverse inputs, systematic data preprocessing and augmentation techniques were applied before training the CNN [20]. These steps ensured input quality and consistency while preventing overfitting across the different leaf disease categories.

This study utilizes the Augmented New Plant Diseases Dataset, comprising color images of both healthy and diseased plant leaves across various species. The images were loaded from structured directories using TensorFlow's `image_dataset_from_directory()` function, which automatically infers class labels from the directory structure and creates a `tf.data.Dataset` object suitable for model training.

All images were uniformly resized to 256×256 pixels to match the input dimensions required by the CNN architecture. Subsequently, a Rescaling operation was applied to normalize the pixel values. Originally ranging from 0 to 255, the pixel intensities were scaled to the range $[0, 1]$ using a Rescaling $(1.0/255)$ transformation. This normalization ensures stable and efficient gradient updates during the optimization process. The normalization was applied to both training and validation datasets using the `map()` function of the TensorFlow dataset pipeline.

To improve the robustness and generalization of the proposed CNN model, image transformation techniques were integrated into the preprocessing pipeline. Data augmentation plays a vital role in preventing overfitting by synthetically increasing the diversity of training samples through various transformations while preserving the semantic content of the images.

In this study, the Augmented New Plant Diseases Dataset was already pre-augmented to some extent. However, additional on-the-fly data augmentation techniques were employed using TensorFlow and Keras to simulate real-world variations in environmental conditions such as lighting, orientation, and scale. The following augmentation techniques were applied:

- Random horizontal and vertical flipping: Helps the model become invariant to leaf orientation.
- Random rotation (up to 20°): Simulates varied leaf angles and perspectives.
- Random zoom (range: 0.8–1.2): Allows the model to focus on both global and local leaf features.
- Width and height shift: Introduces minor translations to simulate camera misalignment.
- Brightness adjustment: Makes the model robust to varying lighting conditions.

Crucially, all augmentation parameters were carefully selected to ensure the core pathological features (e.g., spots, lesions, discoloration patterns) remained clearly visible and relevant to the true disease class, thereby preserving the distribution and representativeness of the original dataset. These augmentations were implemented using the `tf.keras.preprocessing.image.ImageDataGenerator` and integrated into the training pipeline. The augmented images were generated dynamically during training, ensuring memory efficiency and maximum variability across epochs. This augmentation strategy significantly contributed to the model's ability to generalize well across different leaf types and disease manifestations, especially when evaluated on the validation dataset.

C. CNN Architecture and Training

The CNN framework is based on established models like LeNet to leverage their feature extraction and classification capabilities [21]. The model was customized for plant disease detection by optimizing layer configurations and

hyperparameters [22]. Transfer learning was also adopted to enhance feature extraction efficiency.

The proposed CNN model was implemented using the Keras Sequential API, which facilitates layer-by-layer stacking of neural network components. The model is designed to extract rich hierarchical features from input leaf images and perform multi-class classification over 38 plant disease categories. The system architecture is shown in Figure 2.

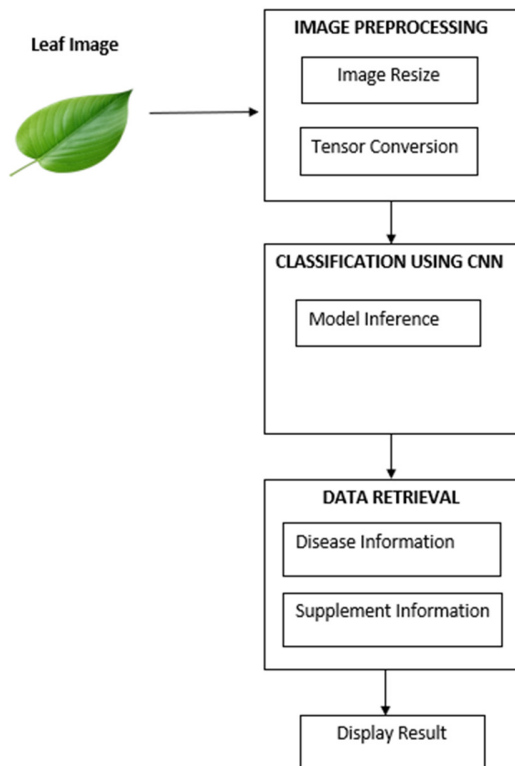


Fig. 2. CNN-based system architecture for plant leaf disease classification.

The network begins with a series of convolutional layers to extract spatial and textural features:

- Block 1: Two convolutional layers with 32 filters of size 3×3 , activated by the ReLU function and using 'same' padding to preserve spatial dimensions. This is followed by a MaxPooling layer with pool size 3×3 , reducing the spatial resolution while retaining salient features.
- Block 2: Two convolutional layers with 64 filters, each of size 3×3 , again followed by a MaxPooling layer with the same configuration.
- Block 3: This block repeats the same pattern using 128 filters, capturing more complex and abstract features.
- Block 4: The number of filters increases to 256, and no pooling is applied after this block, allowing for a deeper feature representation.
- Block 5: Two high-capacity convolutional layers with 512 filters and a larger kernel size of 5×5 are used to extract

high-level features. These layers are also followed by 'same' padding.

The convolutional feature maps are flattened and fed into a fully connected Dense layer with 1,568 units and ReLU activation, allowing the model to capture intricate non-linear combinations of features. To mitigate overfitting, a Dropout layer with a rate of 0.5 is incorporated, randomly deactivating neurons during training. The output layer contains 38 neurons, each representing a class, with a Softmax activation generating a probability distribution across disease categories.

The model was trained using the Adam optimizer with a learning rate of 0.0001, enabling adaptive updates during training. Sparse categorical crossentropy was employed as the loss function, suitable for multi-class classification with integer-labeled targets. Accuracy served as the primary evaluation metric. Training was conducted for 10 epochs using the `fit_generator()` function to enable efficient batch-wise processing of the large dataset.

Model performance was continuously monitored on the validation dataset, and consistent convergence across epochs indicated stable learning and reliable generalization on the Augmented New Plant Diseases Dataset.

D. Pesticide Recommendation Module

In addition to disease classification, the proposed system integrates a pesticide recommendation module that assists farmers or agricultural practitioners in selecting appropriate chemical treatments based on the identified plant disease. This module bridges the gap between diagnosis and actionable intervention by mapping each predicted disease class to a corresponding pesticide or management practice.

The mapping between plant diseases and recommended pesticides was manually curated based on agricultural best practices and publicly available agricultural resources. The system uses a dictionary structure (`disease_to_pesticide`) to associate each disease class (as predicted by the CNN) with the most effective chemical control agent(s), such as Copper-based sprays, Mancozeb, Metalaxyl, or Neem oil. For healthy plants, the recommendation appropriately states "No pesticide needed."

Once the disease is predicted using the trained CNN model, the predicted class label is passed as a key to the dictionary to fetch the relevant pesticide(s). This process is encapsulated in a utility function that accepts an input image, preprocesses it, performs inference, and overlays the predicted disease and corresponding pesticide information directly onto the image using `matplotlib`.

The integration of this module demonstrates the potential of AI not only for accurate disease identification but also for supporting real-time, context-aware agricultural decision-making, which is essential for enhancing productivity and minimizing unnecessary pesticide use.

The model accurately identifies crops and their associated diseases across 38 distinct categories, highlighting its potential for large-scale deployment in mobile-based plant disease diagnosis. Effective detection of plant diseases plays a crucial

role in enhancing both the yield and quality of agricultural produce while minimizing the excessive use of chemical pesticides [23, 24].

IV. RESULTS AND DISCUSSION

The proposed CNN model was trained and evaluated using the Augmented New Plant Diseases Dataset. The performance of the model was assessed using key metrics such as accuracy, precision, and recall ensuring a comprehensive understanding of its classification capability.

The model achieved an impressive training accuracy of 98.37% and a test accuracy of 96.29%, indicating strong generalization to unseen data. This narrow gap between training and test accuracy suggests that the model effectively learned the distinguishing features of different plant diseases while avoiding overfitting. Furthermore, both the precision and recall scores were recorded at 96.29%, reflecting a balanced performance across all classes. High precision signifies that most of the diseases predicted by the model were correctly identified, whereas high recall indicates the model's effectiveness in detecting actual disease cases without omission.

Figure 3 illustrates the training and validation accuracy of the model over 10 epochs. Initially, both accuracies increase rapidly, with validation accuracy slightly higher than training accuracy during the first few epochs. As training progresses, the training accuracy steadily improves, reaching approximately 98% by the ninth epoch. The validation accuracy also stabilizes between 95% and 96%, showing minor fluctuations but no significant drop, which indicates strong generalization ability. The small gap between the training and validation curves suggests that the model is not overfitting and is capable of maintaining high predictive performance on unseen data. Overall, the model demonstrates efficient learning behavior and robust validation performance.

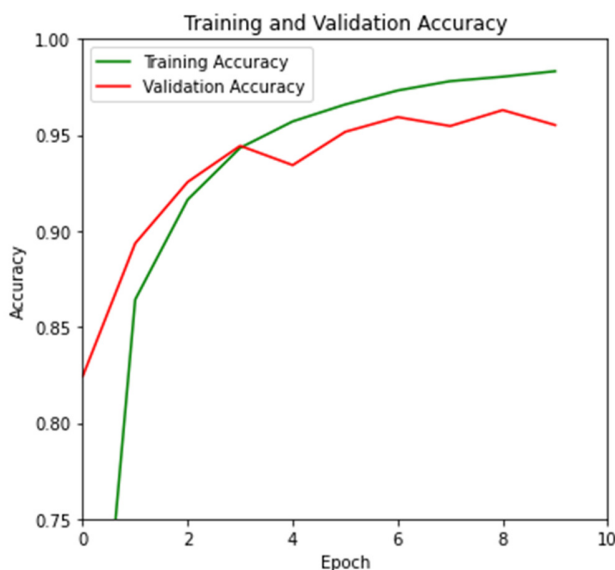


Fig. 3. Plot of training and validation accuracy across epochs, showing high performance and minimal overfitting, with validation accuracy stabilizing above 95%.

Figure 4 shows the training and validation loss curves over 10 epochs. The training loss decreases sharply during the initial epochs and continues to decline steadily, reaching a low value close to zero by the ninth epoch. Similarly, the validation loss decreases significantly at the beginning and then stabilizes, showing minor fluctuations but remaining low throughout the training process. The consistent downward trend of both training and validation loss, along with the small gap between them, indicates that the model is learning effectively without significant overfitting. Overall, the loss curves demonstrate that the model achieves efficient convergence and maintains good generalization to the validation data.

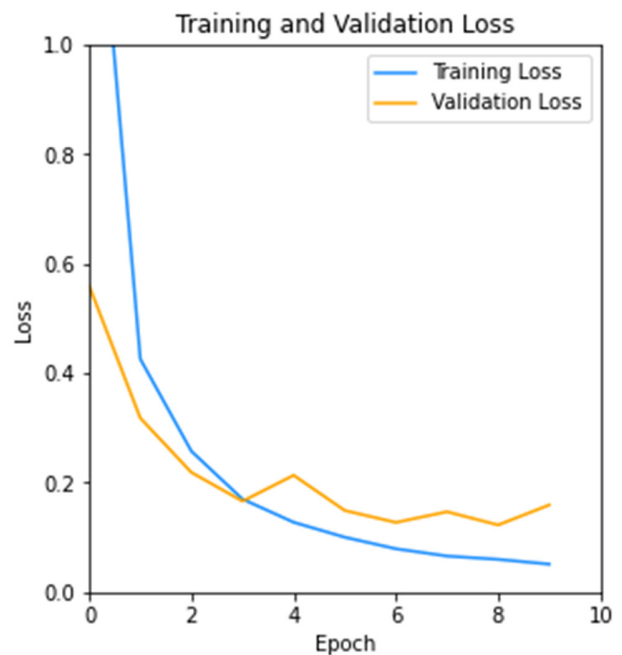


Fig. 4. Training and validation loss plots showing effective convergence and good generalization performance of the model.

Figure 5 illustrates the confusion matrix of the classification model evaluated across 38 distinct classes. In the confusion matrix, rows correspond to the true class labels, and columns represent the predicted labels. The diagonal elements indicate correct classifications for each class, whereas the off-diagonal entries reflect misclassified instances. A strong concentration of values along the diagonal suggests that the model successfully classified the majority of the input samples. Although minor misclassifications are observed for a few classes, they are relatively sparse and low in number, indicating a high overall classification accuracy. This performance suggests that the model effectively distinguishes between different classes with minimal confusion, validating its robustness and reliability for the plant disease classification task.

Figure 6 displays examples of the model's predictions and corresponding pesticide recommendations for test data. The first image shows a corn (maize) leaf affected by common rust, with the model correctly identifying the disease and suggesting the use of the pesticide Propiconazole for treatment. The

- Engineering and Technology Management*, vol. 75, pp. 561–572, Jan. 2025.
- [6] V. Tiwari, R. C. Joshi, and M. K. Dutta, "Dense convolutional neural networks based multiclass plant disease detection and classification using leaf images," *Ecological Informatics*, vol. 63, July 2021, Art. no. 101289, <https://doi.org/10.1016/j.ecoinf.2021.101289>.
- [7] "New Plant Diseases Dataset." Kaggle. [Online]. Available: <https://www.kaggle.com/datasets/vipooooo/new-plant-diseases-dataset>.
- [8] A. Darwish, D. Ezzat, and A. E. Hassaniien, "An optimized model based on convolutional neural networks and orthogonal learning particle swarm optimization algorithm for plant diseases diagnosis," *Swarm and Evolutionary Computation*, vol. 52, Feb. 2020, Art. no. 100616, <https://doi.org/10.1016/j.swevo.2019.100616>.
- [9] O. Iparraguirre-Villanueva *et al.*, "Disease Identification in Crop Plants based on Convolutional Neural Networks," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 14, no. 3, pp. 519–528, Mar. 2023, <https://doi.org/10.14569/IJACSA.2023.0140360>.
- [10] K. Suwa, Q. H. Cap, R. Kotani, H. Uga, S. Kagiwada, and H. Iyatomi, "A comparable study: Intrinsic difficulties of practical plant diagnosis from wide-angle images," in *2019 IEEE International Conference on Big Data*, Los Angeles, CA, USA, 2019, pp. 5195–5201, <https://doi.org/10.1109/BigData47090.2019.9006556>.
- [11] A. A. Ahmed and G. H. Reddy, "A Mobile-Based System for Detecting Plant Leaf Diseases Using Deep Learning," *AgriEngineering*, vol. 3, no. 3, pp. 478–493, July 2021, <https://doi.org/10.3390/agriengineering3030032>.
- [12] T. S. Alam, C. B. Jowthi, and A. Pathak, "Comparing pre-trained models for efficient leaf disease detection: a study on custom CNN," *Journal of Electrical Systems and Information Technology*, vol. 11, no. 1, Feb. 2024, Art. no. 12, <https://doi.org/10.1186/s43067-024-00137-1>.
- [13] J. G. A. Barbedo, "Factors influencing the use of deep learning for plant disease recognition," *Biosystems Engineering*, vol. 172, pp. 84–91, Aug. 2018, <https://doi.org/10.1016/j.biosystemseng.2018.05.013>.
- [14] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using Deep Learning for Image-Based Plant Disease Detection," *Frontiers in Plant Science*, vol. 7, Sept. 2016, Art. no. 1419, <https://doi.org/10.3389/fpls.2016.01419>.
- [15] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification," *Computational Intelligence and Neuroscience*, vol. 2016, no. 1, June 2016, Art. no. 3289801, <https://doi.org/10.1155/2016/3289801>.
- [16] Ü. Atila, M. Uçar, K. Akyol, and E. Uçar, "Plant leaf disease classification using EfficientNet deep learning model," *Ecological Informatics*, vol. 61, Mar. 2021, Art. no. 101182, <https://doi.org/10.1016/j.ecoinf.2020.101182>.
- [17] S. M. Omer, K. Z. Ghafoor, and S. K. Askar, "Plant Disease Diagnosing Based on Deep Learning Techniques: A Survey and Research Challenges," *ARO-The Scientific Journal of Koya University*, vol. 11, no. 1, pp. 38–47, Feb. 2023, <https://doi.org/10.14500/aro.11080>.
- [18] K. R. Aravind, P. Raja, R. Anirudh, K. V. Mukesh, R. Ashwin, and G. Vikas, "Grape Crop Disease Classification Using Transfer Learning Approach," in *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering*, Palladam, India, 2018, pp. 1623–1633, https://doi.org/10.1007/978-3-030-00665-5_150.
- [19] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, Feb. 2018, <https://doi.org/10.1016/j.compag.2018.01.009>.
- [20] A. Fuentes, S. Yoon, S. C. Kim, and D. S. Park, "A Robust Deep-Learning-Based Detector for Real-Time Tomato Plant Diseases and Pests Recognition," *Sensors*, vol. 17, no. 9, Sept. 2017, Art. no. 2022, <https://doi.org/10.3390/s17092022>.
- [21] C. P. Lee, K. M. Lim, Y. X. Song, and A. Alqahtani, "Plant-CNN-ViT: Plant Classification with Ensemble of Convolutional Neural Networks and Vision Transformer," *Plants*, vol. 12, no. 14, July 2023, Art. no. 2642, <https://doi.org/10.3390/plants12142642>.
- [22] S. Prabavathi and P. Kanmani, "Plant Leaf Disease Detection and Classification using Optimized CNN Model," *International Journal of Recent Technology and Engineering*, vol. 9, no. 6, pp. 233–238, Mar. 2020, <https://doi.org/10.35940/ijrte.F5572.039621>.
- [23] M. H. Saleem, J. Potgieter, and K. M. Arif, "Plant Disease Classification: A Comparative Evaluation of Convolutional Neural Networks and Deep Learning Optimizers," *Plants*, vol. 9, no. 10, Oct. 2020, Art. no. 1319, <https://doi.org/10.3390/plants9101319>.
- [24] S. Alqethami, B. Almtanni, W. Alzhrani, and M. Alghamdi, "Disease Detection in Apple Leaves Using Image Processing Techniques," *Engineering, Technology & Applied Science Research*, vol. 12, no. 2, pp. 8335–8341, Apr. 2022, <https://doi.org/10.48084/etasr.4721>.