

# Plant Leaf Disease Detection Using Machine Learning And Deep Learning

V. Kavitha<sup>1</sup>, S.Jayananthini ME.<sup>2</sup>

<sup>1</sup>Assistant professor, Department of Computer Science and Engineering, Vivekananda College of engineering for women, Elayampalayam, Tiruchengode, Namakkal - 637205, TamilNadu, India

<sup>2</sup>Department of Computer Science and Engineering, Vivekananda College of engineering for women, Elayampalayam, Tiruchengode, Namakkal - 637205

**Abstract**—Plant diseases significantly reduce agricultural productivity, leading to lower crop quality and major economic losses worldwide. Early and accurate identification of leaf diseases is essential for effective crop management and timely intervention. Traditional manual inspection is slow, labor-intensive, and prone to errors, creating a strong need for automated and intelligent disease detection systems. Advances in machine learning and deep learning now enable fast and reliable image-based diagnosis for precision agriculture.

This study proposes an automated plant leaf disease detection framework using both classical machine learning algorithms and deep learning techniques. A curated dataset of leaf images is preprocessed through resizing, noise removal, augmentation, and segmentation to improve model performance. Classical models such as SVM, Random Forest, k-NN, and Logistic Regression are trained using handcrafted features derived from color, texture, and shape descriptors.

To achieve higher accuracy, a Convolutional Neural Network (CNN) is incorporated to automatically learn hierarchical image features. The CNN consists of convolutional, pooling, and fully connected layers, optimized with dropout, batch normalization, the Adam optimizer, and learning-rate scheduling. Comparative analysis shows that the CNN significantly outperforms traditional models in both accuracy and generalization.

Extensive experiments using accuracy, precision, recall, F1-score, and confusion matrices confirm the effectiveness of the proposed approach. The CNN accurately classifies common plant diseases—including bacterial blight, leaf spot, mildew, and rust—across various species. Cross-validation and testing on unseen images further demonstrate the model's robustness and deployment readiness.

Overall, the system provides a fast, reliable, and scalable tool for farmers and agronomists. With support for real-time mobile or web-based deployment, it offers a practical solution for large-scale agricultural monitoring. The study highlights that combining classical machine learning with deep learning,

especially CNNs, greatly enhances plant disease detection, enabling smarter and more sustainable crop management.

**Keywords**—Deep Learning, Leaf Disease, Plant Disease Detection, Image Processing, Neural Classification Network, ENCN, Support Vector Machine, SVM

## I. INTRODUCTION

Plant disease detection is an essentially significant field of study using Deep Learning for the detection and diagnosis of plant diseases [1]. Since they may seriously affect agricultural production, it is very important to diagnose these diseases in time and to take precautions against them to guarantee food security and healthy crops. Firstly, the process would have required a lot of time and money from the specialists who had to work out, on a laborious and error-prone way, the identification of plant diseases. The following figure, Fig-1 represents the dataset image samples.



Fig.1 Dataset Image Samples

However, with the advent of AI and ML, one can now automate the identification of plant diseases with high accuracy and speed in modern agriculture, which is very promising [2][3]. One of the common techniques that many researchers employ is using supervised learning to build machine learning models for plant disease diagnosis. The algorithms are trained over labeled datasets, wherein the plant images are divided into either healthy or diseased types of images [4]. The model learns to discriminate between healthy and diseased plants by analysis of millions of images and considered variables such as color, texture, shape and some presence of unique patterns characteristic to several diseases. The advantage of proposed model ENCN lies in their ability to process complex visual information with the built-in automatic extraction of relevant features from images. That is exactly why they are used to mark subtler signs of plant diseases in situations where conventional methods might not work very well. The performance of the detection system relies chiefly upon the quality of the dataset on which the machine-learning models are trained. The datasets contain both healthy and diseased plants, including quality images of plants grown under managed conditions. Datasets also have labels to indicate the presence or absence of diseases. Several of such datasets contain examples of three highly frequent agricultural diseases: leaf rust, powdery mildew, and blight; in training models to detect signs of disease, these diseases are of outstanding importance in biological and agricultural research. A robust model would require a large, diverse and balanced dataset, which is not very easy to establish [5][6]. This dataset should include a range of plant species and diseases.

As research progresses and various datasets become more widely available, this will become a useful tool for farmers to safeguard crops and, therefore, increase global food security by providing enhancements in the speed, accuracy, and scalability of machine-learning algorithms to identify instances of plant diseases [7]. First, one has to collect a large amount of datasets containing images of good and poor conditions of plants as a step in kernel research of machine learning for plant disease detection. Images of plants collected to be labelled to show the absence or presence of certain diseases have been captured by devices like mobile phones, drones, or digital cameras. Next, these images undergo preprocessing to enhanced form features for

appearance. This includes resizing, normalization, and augmentation (such as rotating, flipping, or scaling) in order to achieve data heterogeneity, thus preventing overfitting. Image preprocessing permits the machine learning model to concentrate on the main features such as shape, color, and texture, which are critical in the detection of disease. An important aspect of autonomous feature extraction is deep learning models such as CNNs, which enable the system to internally learn complex patterns of input images [8][9]. After training on a labeled dataset, a model's performance is evaluated using various metrics such as accuracy, precision, and recall. In this case, a validation set is used for this purpose. The model is capable of being adjusted as necessary, in order to improve its performance. Once the model is accurate enough, it will be used in real-time applications, just like online platforms or mobile applications, so that farmers can identify disharmony in his/her crops by uploading pictures. In order to help guide farmers toward the rapid detection of possible illnesses, the software analyses these images and provides predictions. In addition, the system may improve itself over time in terms of different plant species and climatic conditions, since it learns from new data. There is great potential for machine-learning-powered plant disease detection to revolutionize the existing agricultural sector by reducing losses on crops and increasing food security.

The process of automating and improving the accurate and timely diagnosis of plant disease can be achieved through machine learning, especially via convolution neural networks. Convolutional neural networks, unlike any other machine learning algorithm, learn salient features from raw image data, such as pattern, texture, and shape specific to the plant disease. Training on a large dataset of labeled images allows convolutional neural networks to detect some early signs of diseases that might not be perceivable by the human eye. This allows prompt diagnosis, enabling farmers to take corrective or preventive measures before the disease spreads and causes large crop losses [10]. The ultimate goal is to implement an easy-to-use living system to allow real-time diseases identification in crops using different platforms such as mobile application; this will allow farmers better management of crop health and hence increased agricultural yield.

(i) To enrich the model and avoid overfitting, image synthesis is used to create additional images of infested plants and hence augment the dataset.

(ii) In general, such approaches produce artificial images of sick plants by training their convolutional neural networks with the data definitions provided through GANs. Where there is the absence of adequate labeled datasets, this can prove useful.

(iii) Early prediction and curing of plant diseases and minimizing losses to crops are made easier through the use of synthetic images produced by image synthesis, which increases the accuracy of disease detection models.

(iv) Synthetic images of infected plants can be used to fine-tune pre-trained CNN models. This allows efficient transfer learning and more accurate disease diagnosis.

## II. RELATED WORKS

Plants form the basis of the world's food supply; nonetheless, plant diseases cause considerable losses in the output of crops, related to many environmental conditions [11]. Nevertheless, Tropical plant disease identification undertaken by people becomes a long and tedious task. It is not always very reliable as a tool in Plant Disease detection and control. One way to tackle these difficulties is by implementing modern technologies like Deep Learning (DL) and Machine Learning (ML). These will allow for the early detection of plant illnesses. This study delves into the latest developments in plant disease diagnosis using ML and DL approaches. The trials included in this paper show that these methods can improve the efficiency and accuracy of plant disease detection, and the research focuses on publications from 2015 to 2022. In addition to identifying plant diseases, this paper discusses the difficulties and limits of ML and DL in this field, including problems with data availability, image quality, and distinguishing between healthy and sick plants. The research offers a thorough overview of the current research in plant disease detection, discusses the advantages and disadvantages of these methods, and suggests ways to overcome the obstacles that may arise during their implementation. As a result, it is valuable for researchers, practitioners, and industry professionals in this field.

Scientists are now adopting intelligent agriculture-a means of greatly enhancing production with the incorporation of the use of Artificial Intelligence

(AI)-to deal with various problems within agricultural fields [12]. There are a lot of illnesses that harm crop yields, and there are a lot of plants in the globe, thus finding and classifying plant diseases isn't an easy task. Any AI-based system aims to accurately classify plant illnesses and detect them early. In order to greatly enhance the accuracy of plant leaf disease categorization, this research suggests a hybrid architecture. The power of Convolutional Neural Networks (CNNs) and Vision Transformers (ViT) is utilized in this suggested model. Robust global features are extracted using an ensemble model that includes the popular CNN designs VGG16, Inception-V3, and DenseNet20. The next step in plant disease detection with high accuracy is applying a ViT model for local feature extraction. Under testing in the apple and corn public datasets, there are four classes per each dataset. The apple dataset has an accuracy of 99.24% while the corn dataset has 98%. This hybrid model will efficiently ascertain and classify multi-class plant leaf diseases in reference to other similar published models.

To protect agricultural crop output and guarantee food security, early and precise identification of plant leaf diseases is of the utmost importance [13]. Bacteria, fungus, weather, and other environmental variables are among the many causes of leaf diseases that plants experience during their life cycles. By combining the best features of the enhanced Vision Transformer with ResNet9, the authors of this study provide a new hierarchical residual vision transformer that can help with the early diagnosis of leaf illnesses. By lowering the number of trainable parameters and using fewer calculations, the suggested model is able to extract more relevant and discriminating features. Tests using 13, 38, and 51 distinct leaf disease classes are conducted on the Local Crop dataset, the Plant Village dataset, and the Extended Plant Village Dataset, respectively, to assess the efficacy of the suggested approach. Using ResNet 9 for feature classification and the optimal trail parameters from Improved Vision Transformer, the suggested model is trained. When tested on the aforementioned datasets, the suggested model beat competitors like InceptionV3, MobileNetV2, and ResNet50 across a variety of metrics.

Agriculture is an essential need and their primary source of domestic income for many countries [14]. Plant diseases effected by more than one pathogen

(as in bacteria, fungus and viruses) are so common that agricultural corporations lose big bucks worldwide. It is critical to monitor plant diseases in order to ensure the quantity and quality of harvests. This highlights the critical nature of plant disease detection. Symptoms of the plant disease syndrome manifest in certain plant tissues. Still, individual plant leaves are usually the first to show signs of infection. Several researchers have used computer vision, deep learning, few-shot learning, and soft computing approaches to automatically detect plant diseases from leaf images. Quick and suitable efforts to prevent a decrease in crop quality and quantity can also be achieved by farmers using these strategies. By using these methods to illness recognition, we may speed up technology and research while avoiding the drawbacks of origin by avoiding factious feature selection and extraction. Additionally, specific molecular methods have been developed to forestall or lessen the impact of the infectious danger. Thus, this research assists the researcher in designing automated plant disease identification systems using deep learning, machine learning, and few shots of learning. It also gives specific diagnostic methods to prevent disease. We also discuss some of the next steps in illness categorization.

Reducing economic repercussions and optimizing agricultural output require precise and timely detection of plant leaf diseases [15]. The problem with precisely identifying certain illnesses is that farmers rely on traditional manual approaches, which makes it difficult. Applying the YOLOv4 algorithm to the problem of plant leaf disease detection and identification is the focus of this study. The big Plant Village Dataset is composed of more than 50,000 pictures of healthy and diseased plant leaves from fourteen different species; the study prepares the very general advanced systems for prediction of agricultural diseases. To enhance the dataset and fortify the model's durability, data augmentation techniques such horizontal flip and histogram equalization were employed. We compared the YOLOv4 algorithm's performance against that of other well-known target recognition methods, such as Densenet, Alexanet, and neural networks, as part of our thorough evaluation. An astounding 99.99% accuracy was attained by YOLOv4 when applied to the Plant Village dataset. The proposed method was verified based on consistently very high values, with 0.99 scores for all the following metrics: accuracy,

precision, recall, and F1-score. The findings of this study highlight the remarkable progress made in plant disease detection and highlight the potential of YOLOv4 as an advanced tool for precise disease prediction. Because they increase our ability to manage diseases and safeguard crops, these innovations are incredibly important for researchers, farmers, and everyone else working in the agricultural sector. After the model is developed, it has to be trained using a large number of pictures that have the necessary objects tagged. Keep in mind that the EfficientNetV2B1 model can only learn to differentiate between the classes if the input is balanced. After that, you need to feed the data into the EfficientNetV2B1 model. Dataset size determines whether this is best done in batches or in one continuous run. Next, a suitable optimizer, such as Adam or SGD, has to be used to train the model. In order for the model to learn to identify different objects in the images, its weights are modified continually during the training phase. After then, the testing set may be used to assess the model's correctness. The model's performance may be evaluated using a variety of measures, including recall, precision, and F1 score. We may measure the model's performance by keeping track of how many images are appropriately classified.

### III. METHODOLOGY

In our proposed model image processing method is used for the construction of system through which leaf disorder is detected if any distorted picture is supplied with in very short time. As a result a farmer without sufficient sense disease detection knowledge, modern techniques and software can be effortlessly applied this system. The dataset which is used as input is mixed of healthy and distorted images and after completing the action of input dataset the system output provides the affected and healthy leaves. A chart is introduced below as the proposed methodology. Figure 1 is that flowchart.

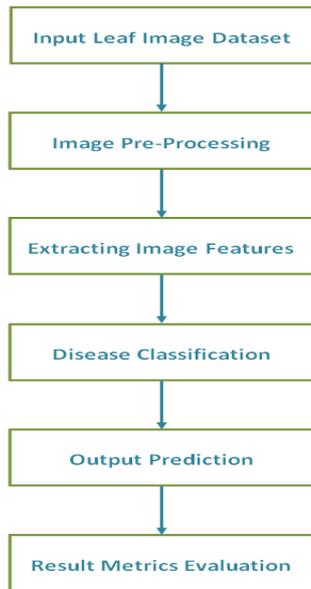


Fig.2 System Flow Diagram

## CONVOLUTIONAL NEURAL NETWORKS

One variant of deep neural networks is called as convolutional neural networks(CNN).A CNN combines well-read features with input data, and then it uses 2D convolution layers, and hence makes this architecture more suitable for processing 2D data, like images,CNNs abolish the demand for manual feature removal and extraction for the classification of the images. The CNN model of its own extracts features straight from images. the features that are extracted aren't pre-trained; they are well-read while the network is trained on few groups of images. The CNN model has numerous of layers which execute the processing of image in convolutional layers include-Input layer, Output layer, convo Layer, Fully,Soft-max layer, Connected layer, Pooling layer.

(i) Data Collection: It is challenging to guarantee the efficacy of transfer learning because the initial MobileNet pre-trained model was trained on the ImageNet dataset, which does not only include the pictures that are required. So, in order to train the model, we require a dataset that includes trash photos. There is currently no fixed dataset that is universally used for trash categorization jobs. The TrashNet dataset is used for rubbish classification, however it doesn't adequately reflect the real situation of residential waste categorization in India since it contains too few categories. Consequently, this article builds a dataset specifically for visual trash sorting using network retrieval and real-world

scene imaging in the lab, covering both single-object and multiple-scene scenarios.

(ii) Image Processing: To augment and normalize images of plants so that accuracy in identification of plant disease is improved, an Image Preprocessing obtains images. Depending on the kind of input image it is processing, it resizes images into fixed resolution, normalizes pixel values, incorporates some augmentation techniques, removes noise or other irrelevant qualities, and enhances image quality with brightening, contrast, and saturation enhancing. The whole model is thus reinforced, has less influence from external variables, and improves the accuracy in diagnosis.

(iii) User Interface: Dataset was divided into three subsets: validation, training, and testing. User Explanation provides a user-friendly interface for farmers and agricultural experts to receive disease diagnosis and recommendations for treatment. It explains the detected disease, its symptoms, causes, and prevention methods, enabling users to take informed decisions about crop management and disease control. Also, the process will provide users with personal advice on pesticide application, fertilizer application, and irrigation management, allowing users to optimize their farm management systems. It also provides information on how the disease may potentially affect crop yield and quality, enabling users to plan accordingly. Likewise, a feature on the process will track the disease's history and give alerts for possible outbreaks of the disease, allowing users to implement some of their proactive measures to avoid the spread of any disease.

(iv) CNN Architecture: The proposed Convolutional Neural Network (CNN) architecture consists of a sequence of convolutional and pooling layers designed to automatically learn hierarchical features from plant leaf images. It begins with an input layer followed by three convolutional blocks, each containing Conv2D layers with ReLU activation, batch normalization, and max pooling to extract spatial patterns while reducing dimensionality. Increasing filter sizes (32, 64, and 128) help capture both low-level and high-level features, while dropout is applied to prevent overfitting. The extracted features are then flattened and passed through fully connected layers, including a dense layer with 256 neurons and a final softmax layer for multi-class disease classification. This architecture

provides a robust, efficient, and scalable solution for accurate plant leaf disease detection..

(v) Model Training: Following model creation, it is necessary to train the model using a large number of photographs annotated with the necessary items. If you want the EfficientNetV2B1 model to learn to differentiate between the different classes accurately, you must ensure that the data is balanced. Once the data is available, it has to be loaded into the EfficientNetV2B1 model. Depending on the dataset size, this can be done in batches or all at once. The next step is to train the model with a suitable optimizer, such as Adam or SGD. Throughout the training phase, the model's weights are consistently adjusted to help it learn to identify different objects in the images. Next, the model may be run on the testing set to see how well it performs. There are a number of ways to evaluate the model's efficacy, including recall, accuracy, and F1 score. One way to measure the model's performance is to see how many photos it accurately labels. To further assess the model's generalizability, it can be run with data that has never been seen before. The model's performance with unobserved data will be shown by this.

#### IV. RESULTS AND DISCUSSION

In our research work, the proposed model is provided expected output after completing the training, testing and validation using various dataset. The detail description about this model output is given below.

##### A. Analytical Analysis

The training and validation accuracy is obtained through this model 77.99% and 34.17% respectively. With time, after each run the model experienced trained and the exactness of the result improved as well. After completing 10 runs the training and validation accuracy become 89.56% and 61.91% respectively and the learning rate decreased at .0005. When the successful run is counted 30 the training and validation accuracy reached to 93.8% and 95.8% respectively and the learning rate decreased at 3.124e-05. After completing final run the height accuracy is obtained and the value for training and validation accuracy is 94.29% and 96.28% respectively.

##### B. Accuracy Graph

The term over fitting is used for describing and referring a model and this appear during capturing noise of the data. So the indication is that the perfect fitting of model or algorithmic data occur over fitting. The term under fitting is introduced during showing the lacking of capturing insufficient underlying data. The model which we introduced here shows absence of over or under fitting. The graph for training and validation accuracy and loss of our model is shown in Fig 3.

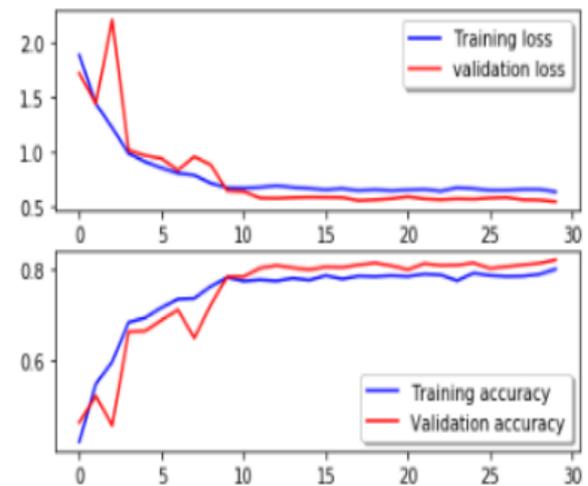


Fig.3 Training and validation

##### C. Confusion Matrix

The performance of the model is shown by the specific error table known as confusion or error matrix. Among total 67 images true image is 67 and false image is 0 only for Alternaria blight, for Anthracnose true images is 41 and false image is 45 among total approximate image 86, The Downy mildew shows 91 true and 3 false image among 94, For Powdery mildew consider total 80 images where 44 images are true and the remain 36 is false. In Healthy, among 70 image 59 detected as true and 11 is false. Finally for Black rot, among total 80 images 58 shows true and 22 is false. The values introduced in diagonal position of the confusion matrix are bigger than others. The values in diagonal position used (4x4) shape is showing their best performance comparing to the other position and this part maximizing the data and that's why its color is deep blue. Fig.4. Shows confusion matrix and Table 1 shows classification report of our model.

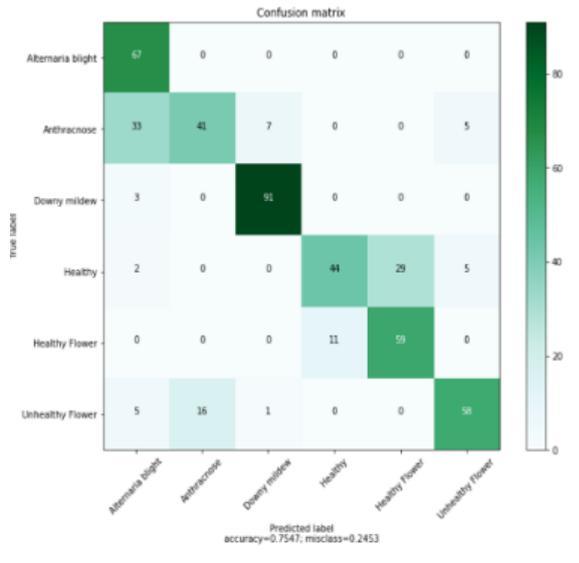


Fig.4 Confusion Matrix

PREDICTION

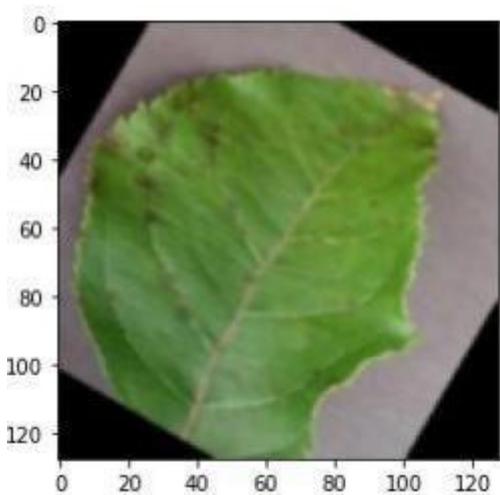


Fig. 5 Prediction

V. CONCLUSION

This work provides an authentic notion for detecting the attacked leaf ('Grape', 'Potato' and 'Strawberry') and the farmer who works for produce these fruits gets remedy so that they can enhance the production in agricultural industry. Specialist who works in agriculture department accepts quick disease detection process by image processing technique as a result Image Processing technology touch it's milestone within very short time. The transited portion of leaf easily segments and analyzes using CNN model and this model provides best possible result instantly. As a result the farmer who detects plant disease manually can save their time and diminish suspicion on possibilities of

wrong detection. Our future goal is to develop an open multimedia system and make software which automatically detect plant disease and provide their solution.

REFERENCES

- [1] Wubetu Barud Demilie, "Plant disease detection and classification techniques: a comparative study of the performances", Journal of Big Data, 2024.
- [2] Payal Trivedi, et al., "Plant Leaf Disease Detection and Classification Using Segmentation Encoder Techniques", Open Agriculture, 2024.
- [3] Manjunatha Shettigere Krishna, et al., "Plant Leaf Disease Detection Using Deep Learning: A Multi-Dataset Approach", J, 2025.
- [4] Priyanka Pradhan, et al., "Plant disease detection using leaf images and an involuntional neural network", Environment Conservation Journal, 2024.
- [5] B.V. Nikith, et al., "Leaf Disease Detection and Classification", Procedia Computer Science, 2023.
- [6] Cemal Ihsan SOFUOGLU, et al., "Potato Plant Leaf Disease Detection Using Deep Learning Method", Journal of Agricultural Sciences, 2024.
- [7] Alwan Fauzi, et al., "Development of a Mobile Application for Plant Disease Detection using Parameter Optimization Method in Convolutional Neural Networks Algorithm", International Journal of Engineering Technology, 2023.
- [8] Rashmi Ashtagi, et al., "Fusion of AI Techniques: A Hybrid Approach for Precise Plant Leaf Disease Classification", Journal of Electrical Systems, 2025.
- [9] Ashutosh Kumar Singh, et al., "Hybrid Feature-Based Disease Detection in Plant Leaf Using Convolutional Neural Network, Bayesian Optimized SVM, and Random Forest Classifier", Journal of Food Quality, 2022.
- [10] Kamaldeep Kaur, et al., "Enhancing Plant Disease Detection using Advanced Deep Learning Models", Indian Journal of Science and Technology, 2024.
- [11] Muhammad Shoab, et al., "An advanced deep learning models-based plant disease detection: A review of recent research Updated", Frontiers in Plant Science, 2023.

- [12] Sherihan Aboelenin, et al., "A hybrid Framework for plant leaf disease detection and classification using convolutional neural networks and vision transformer", *Complex Intell. Syst.*, 2025.
- [13] Sasikala Vallabhajosyula, et al., "A novel hierarchical framework for plant leaf disease detection using residual vision transformer", *Heliyon*, 2024.
- [14] Anuja Bhargava, et al., "Plant Leaf Disease Detection, Classification, and Diagnosis Using Computer Vision and Artificial Intelligence: A Review", *IEEE Access*, 2024.
- [15] Eman Abdullah Aldakheel, et al., "Detection and identification of plant leaf diseases using YOLOv4", *Frontiers in Plant Science*, 2024.