



CLASSIFICATION OF PLANT LEAF DISEASES USING IMAGE PROCESSING AND MACHINE LEARNING ON REAL-WORLD IMAGES

Merve OZKAN 

Department of Computer Engineering, Ankara University, Ankara, Türkiye, merveozkan@ankara.edu.tr

Article Info

Received: October 17, 2025

Revised: November 17, 2025

Accepted: January 2, 2026

Keywords

*Plant diseases,
Image processing,
Machine learning,
Deep learning,
Leaf classification,
PlantDoc.*

ABSTRACT

Plant diseases pose a serious threat to global food security by directly affecting agricultural production. Traditional expert observation-based diagnosis processes are time-consuming, subjective, and error-prone, making early and accurate diagnosis difficult. This has necessitated the development of image processing and artificial intelligence-based systems that can automatically recognize disease symptoms from leaf images. This study aims to automatically classify plant leaf diseases using the PlantDoc dataset, which consists of images collected under real-world conditions. First, various image processing steps, such as denoising, color space transformations, segmentation, and contour detection, were applied to the leaf images to extract color, texture, and geometry-based features. The resulting features were classified using Support Vector Machines, Random Forests, and k-Nearest Neighbors, and the performance of these models was compared. Furthermore, a deep learning-based MobileNetV2 model was trained using transfer learning and data augmentation techniques and compared with classical methods. Experimental results show that the Random Forests model achieved the highest accuracy rate among classical methods, at 81.5%, while the MobileNetV2 model outperformed all other methods, with an accuracy rate of 86.9%. These findings demonstrate that deep learning-based approaches have higher generalization capabilities on complex, multi-class real-world data. Furthermore, classical methods, thanks to their interpretability and low computational cost, can be a good alternative in resource-limited systems.

1. INTRODUCTION

Plant diseases are one of the major problems that directly impact agricultural productivity and threaten global food security. The Food and Agriculture Organization of the United Nations (FAO) reports that approximately 20% of global agricultural production is lost to diseases each year [1]. Especially in developing countries, early and accurate disease diagnosis is often made through visual inspections based on expert knowledge, leading to time-consuming, subjective, and error-prone processes. These limitations have created a growing need for automated and digitally supported diagnostic systems.

Image processing and artificial intelligence-based classification systems offer effective solutions for automatically detecting disease symptoms on plant leaves. Image processing techniques allow the detection of visual cues such as discoloration, staining, and abnormalities in vein structure. Machine learning and deep learning algorithms can classify these cues to enable highly accurate disease diagnosis [2]. Such systems not only accelerate the diagnostic process but also contribute to reducing pesticide use, preventing economic losses, and promoting environmental sustainability.

Many studies in this area in the literature have been conducted on homogeneous images, such as the PlantVillage dataset, created under controlled laboratory conditions [3]. Although the PlantVillage dataset contains 38 classes and 54,303 examples of healthy and diseased leaves, it has been criticized for its structure, which is far removed from the real world, such as plain-colored backgrounds and ideal lighting conditions. Similarly, the Digipathos dataset focuses on plant diseases in Brazil and contains a

limited number of examples, mostly captured with professional cameras. While it has high visual quality, it contains a limited number of images (approximately 1,500) [4]. The dataset usually belongs to a single plant species (e.g., banana or soybean), and the class diversity is limited. The AI Challenger Agricultural Disease dataset is another widely used dataset in the literature. Developed in China, it contains approximately 61,486 labeled images covering 61 classes across 10 different crop species [5]. However, inconsistencies in label accuracy have been reported, and the ensemble-based labeling process has been criticized for insufficient quality control. Therefore, the generalization performance of models trained on such datasets under field conditions may be limited. To address this shortcoming, the PlantDoc dataset was created by collecting images of diseased leaves from various plant species in natural light, complex backgrounds, and diverse environmental conditions [6]. The dataset consists of 2,598 images covering 13 different plant species and 17 disease classes, providing a test environment closer to field reality. These features allow PlantDoc to enable the development of more reliable and generalizable models.

This study proposes a system for the automatic diagnosis of plant leaf diseases using the PlantDoc dataset. In the first phase, various image processing techniques (denoising, color space transformations, segmentation) were applied to leaf images to extract basic visual features (color, texture, shape). In the second stage, classification was performed using various machine learning algorithms (SVM, Random Forest, K-NN) based on these features. In the third stage, the performance of these methods was compared with a CNN-based deep learning approach. Thus, the accuracy, stability, and generalization capabilities of classical learning techniques and deep learning-based systems on field data were comparatively analyzed. The main contribution of this study is the integration of classical image processing methods and modern machine learning approaches on a dataset reflecting real-world conditions and the demonstration of performance differences between the methods. The obtained findings are intended to guide the development of digital diagnostic systems in agriculture and the design of cost-effective decision support tools for producers in rural areas.

The remaining sections of the study are structured as follows. In the second section, the structure of the PlantDoc dataset, its classes, and the basic properties of the images are introduced; the applied image processing steps and visual feature extraction are detailed. In the third section, the training and testing processes of traditional machine learning algorithms and deep learning-based models are explained, and their classification performances are compared. In the fourth section, experimental findings are presented, and the methods are evaluated in terms of accuracy, generalization ability, and computational efficiency. Finally, the general conclusions of the study are summarized and recommendations for future work are made.

2. LITERATURE REVIEW

In recent years, intensive research has been conducted on computer vision-based automated systems for the early and accurate detection of plant diseases in agricultural production. These systems aim to replace traditional processes based on expert observation, providing faster, scalable, and objective solutions. Classification systems based on leaf images, in particular, have gained significant ground in both academic studies and industrial applications of smart agricultural technologies. Analysis of visual symptoms such as color changes, mottling, and drying marks on the leaf surface is used as a key determinant in disease diagnosis.

Some of the pioneering studies in this field are summarized in Table 1, providing a comparative framework of the methods, datasets, and success rates used. For example, Mohanty et al. (2016) [3] reported an accuracy rate of over 99% across 38 different classes in their study using Convolutional Neural Network (CNN)-based deep learning architectures on the PlantVillage dataset. This study demonstrated the potential of AI-based plant disease classification with such high performance for the first time and has been recognized as a milestone in the literature. However, the fact that the images used were collected in controlled environments, with a homogeneous background and constant lighting conditions limits the model's generalizability to field conditions.

Sladojevic et al. (2016) [7] achieved over 96% accuracy by testing different CNN architectures on a specialized dataset containing 13 disease classes. However, the dataset used in the study was both limited

in size and contained class imbalance, indicating that the model was vulnerable to overfitting. In a study conducted during the same period, Brahim et al. (2017) [8] increased the diversity of the training data using data augmentation techniques, improving both accuracy and the model's ability to discriminate between classes. This study, which particularly contributed to reducing the class imbalance problem, emphasizes the impact of data preprocessing strategies on model success.

Table 1. Summaries of the related studies.

Paper	Year	Method	Dataset	Result
Mohanty et al.	2016	CNN (PlantVillage)	PlantVillage	%99.35
Sladojevic et al.	2016	CNN(Special architecture)	Special Data	%96.30
Brahimi et al.	2017	CNN + Visualization	Tomato Disease	%91.20
Zhang et al.	2018	SVM, RF, SRC	Cucumber Disease	%85.30
Singh et al.	2020	CNN (MobileNet)	PlantDoc	%81.00
Zhai	2020	Case-Based Reasoning	Smart Agriculture	-(decision support-based work)
Khan et al.	2023	TomFormer (CNN + Transformer)	PlantDoc (Tomatoes)	%81.00 mAP (object detection)
Shah et al.	2023	YOLOv5, YOLOv7 (object detection)	PlantDoc	mAP: %62 (YOLOv5), %42 (YOLOv7)
Bera et al.	2024	PND-Net (GCN + Xception)	PlantDoc	%84.30
Zubair et al.	2025	Ensemble (MobileNetV2 + InceptionResNetV2 + EfficientNetB3)	PlantVillage, PlantDoc	~%60 (PlantDoc)
Duhan et al.	2025	Lightweight MobileNetV2 variant	PlantDoc	%64.7
Hussein	2025	MobileViT + SE/ECA/Triples Attention	PlantDoc	~%97

The potential of traditional machine learning methods in this area was examined by Zhang et al. (2018) [9]. In the study, color histogram, Haralick texture features, and geometry-based features were extracted from leaf images; These features were classified using algorithms such as SVM, Random Forest, and Sparse Representation Classifier (SRC). The accuracy rate of around 85% obtained demonstrates that classical methods can still provide an effective solution, particularly in low-resource systems. Furthermore, the low hardware dependency of such methods provides advantages in mobile or embedded systems with low processing power.

Most of these studies rely on datasets created in controlled environments, offering limited representation for real-world applications. The PlantDoc dataset, developed to address this deficiency, was introduced to the literature by Singh et al. (2020) [6]. This dataset offers a significant innovation in that it includes real-world factors such as natural light conditions, complex backgrounds, dust on leaves, shadows, and cuts. Singh and his colleagues tested the generalization capabilities of various CNN models by training them on PlantDoc and showed that models that achieved high accuracy in controlled environments experienced significant performance degradation under field conditions. This suggests that models intended for agricultural applications must be designed to account for site-specific conditions. adequate generalization

Recently, several studies have revisited the PlantDoc dataset using lightweight CNNs, MobileNet variants, ensemble architectures and transformer-enhanced models. Zubair et al. [10] proposed a high-capacity ensemble architecture combining MobileNetV2, InceptionResNetV2, and EfficientNetB3 models and evaluated its performance under both laboratory and field conditions. The model achieved a high accuracy of 99.69% on the PlantVillage dataset, but its accuracy dropped to approximately 60% on PlantDoc. This significant drop reveals a significant generalization gap between datasets with homogeneous backgrounds and controlled real-world images. The study highlights that PlantDoc is a challenging benchmark for deep learning models due to lighting, angle, background complexity, and class imbalance. Therefore, the limited performance of even powerful architectures like ensemble on PlantDoc makes the robust results of the simple MobileNetV2 model used in our current study more meaningful.

Duhan et al. [11] developed RTR_Lite_MobileNetV2, a lightweight CNN model optimized for mobile devices and embedded systems. This architecture aims to increase energy efficiency and hardware compatibility by reducing the number of parameters and FLOPs of MobileNetV2. The study also tested the PlantDoc dataset, and the model achieved only 64.73% accuracy, demonstrating limited success in adapting to real-world conditions. The authors note that background clutter, leaf deformation, and unbalanced class distribution pose significant challenges for lightweight models. These findings indicate that although low-cost lightweight models have practical advantages, more advanced preprocessing and training strategies are required to achieve satisfactory generalization.

Hussein [12] proposed a hybrid model that aims to be both lightweight and highly accurate by enhancing the MobileViT architecture with advanced attention mechanisms such as SE, ECA, and Triplet Attention. This method combines the local feature extraction power of CNNs with the global information flow based on transformers to provide comprehensive representation learning. The model demonstrated high performance on PlantDoc, achieving results in the 97–99% accuracy range; however, this success is associated with the additional computational burden introduced by complex attention mechanisms. The authors note that such hybrid models can better capture the variance in field images, but their use in practical applications is not always feasible due to energy and speed constraints.

Shah et al. [13] examined the performance of YOLOv5 and YOLOv7 models performing real-time object detection on the PlantDoc dataset, presenting an alternative approach that focuses directly on object detection rather than classification. The results show that YOLOv5 achieved approximately 62% mAP, while YOLOv7, despite being newer, achieved only 42% mAP, demonstrating that PlantDoc also poses a significant challenge for object detection. The study reports that detection performance is limited, particularly due to the partial appearance of leaves, the complexity of the background, and the spread of diseased areas across small areas. These results demonstrate that PlantDoc is a challenging and realistic benchmark not only for classification but also for object detection tasks. Therefore, this study demonstrates that PlantDoc should be carefully evaluated by researchers as it tests its generalization capacity as a dataset representative of real-world field conditions.

Beyond classification-focused approaches, a study conducted by Zhai (2020) [14] focused on the integration of artificial intelligence with decision support systems in agriculture. In this study, flexible and interpretable systems capable of providing recommendations to farmers were developed using case-based reasoning. This approach aims not only to diagnose but also to provide transparency to the user by providing recommendations based on past case information. Thus, a structure has been developed that facilitates interaction between agricultural experts and system outputs and builds trust. In this respect, the study contributes not only to accuracy rates but also to system reliability and user acceptance.

Recently developed new methods have further advanced the application of deep learning-based systems to this field. For example, the TomFormer model developed by Khan et al. (2023) [15] combined CNN and Transformer architectures to perform object-based classification for the detection of tomato leaf diseases. The model attracted attention with its 81.00% mAP accuracy on the PlantDoc dataset. TomFormer, which demonstrated particularly effective performance in distinguishing diseases with similar symptoms, was able to simultaneously evaluate both local and global features thanks to its attention mechanism. However, the study's limitation of only tomato leaves and its failure to cover the entire dataset is considered a major limitation that reduces the model's generalizability. Furthermore, because the mAP metric does not provide a direct class-based measure of accuracy, more detailed metrics for class discrimination are necessary.

Similarly, the PND-Net model developed by Bera et al. (2024) [16] achieved 84.30% accuracy in the classification of plant diseases by combining a graph convolutional network (GCN) with the Xception architecture. GCN enables more effective learning of inter-class transitions by modeling structural similarities between classes at the topological level. However, training this architecture is quite complex and requires high computational resources. Furthermore, it has been reported that the model tends to experience performance degradation as the number of classes increases. Therefore, future studies could integrate such GCN-based structures with lighter-weight models to optimize them for mobile agricultural systems.

Generally, the literature is dominated by models that either focus solely on CNN-based approaches or remain limited to traditional methods. This study, however, compares both classical image processing techniques and deep learning architectures on the same dataset, making a unique methodological contribution to the literature. Furthermore, using a dataset such as PlantDoc, which reflects field conditions, fills a significant gap in the literature in terms of assessing the applicability of the developed systems in real-world situations. In addition to deep CNN-based architectures, recent studies have explored metaheuristic optimization techniques to enhance classification performance. Methods such as Genetic Algorithms (GA), Particle Swarm Optimization (PSO), Grey Wolf Optimizer (GWO), and Harmony Search have been employed for hyperparameter optimization, feature subset selection, and convergence acceleration. Several works combining MobileNet variants with PSO or GA-based fine-tuning have reported improvements ranging from 3% to 5% on challenging datasets, indicating that metaheuristic-driven optimization can further strengthen the generalization capability of lightweight models.

3. THE PROPOSED METHODOLOGY

This section presents in detail the dataset, preprocessing steps, feature extraction methods, and applied machine learning and deep learning algorithms used in this study. A flowchart of the applied methodology is presented in Figure 1.

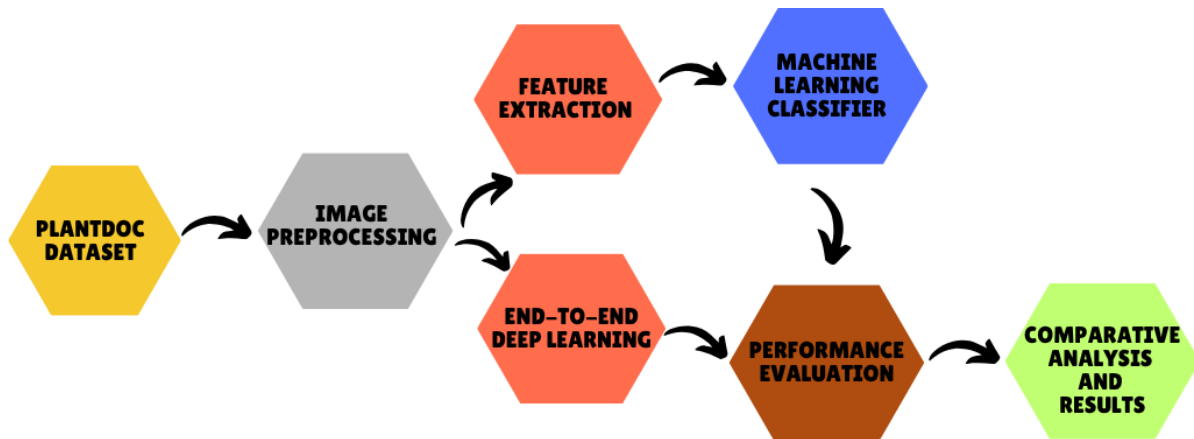


Figure 1. Flowchart of the proposed methodology.

According to the proposed methodology, model training and performance evaluation processes are systematically explained and the experimental framework of the research is presented.

3.1. Used Dataset: PlantDoc

In this study, the PlantDoc dataset, which was obtained under real field conditions and includes diseases from various plant species, was used to classify plant leaf diseases [6]. Unlike homogeneous datasets created in a laboratory environment, PlantDoc reflects real-world scenarios with images obtained from heterogeneous, natural environments. The dataset contains a total of 2,598 leaf images from 13 different plant species (e.g., tomato, corn, grape, pepper, cotton, peach, etc.) and 17 different disease classes (e.g., leaf spot, rust, downy mildew, bacterial spot, mosaic virus, etc.).

The images were captured from different angles, under varying lighting conditions, complex backgrounds, and various environmental factors. This provides an important testing ground for the generalizability of both traditional image processing approaches and deep learning-based methods. The inclusion of natural images, particularly those taken in the field, allows for a more realistic performance evaluation, considering the challenges that AI-based disease detection systems will face during their transition to field applications. In addition, PlantDoc contains more complex classification problems compared to controlled datasets such as PlantVillage, which are frequently used in the literature due to the diversity and difficulty of the data, and has become one of the important benchmark datasets preferred to test the suitability of the developed models for practical use. Figure 2 shows example images of healthy and diseased leaf in the dataset.



Figure 2. Healthy and Diseased Leaf Images in the PlantDoc Dataset.

The images are in JPEG format, and each image is labeled according to its plant species and disease class. During the model training process, the dataset was randomly divided into two subsets, using a common practice of 80% training and 20% testing. This allowed for objective assessment of both the model's learning capacity and its performance on previously unseen data. Care was taken to maintain a balance between classes during data separation, ensuring a representative representation of each class in the training and test sets. Various data augmentation techniques such as random rotation, horizontal/vertical flipping, zooming, contrast/brightness adjustment, and small amounts of noise addition were applied to the training set to prevent overfitting and improve generalization. This made the model more robust to image variations and enabled accurate predictions under different environmental conditions. In this context, the implemented strategy not only increased the model's accuracy but also contributed to the system's robustness against uncertainties that may be encountered in real-world applications.

3.2. Image Processing

Because the PlantDoc dataset was obtained under real-world conditions, the raw images contain a wide range of resolutions, complex background structures, and varying lighting conditions. This diversity necessitates extensive preprocessing before proceeding directly to model training. This stage aims to both adapt the images to artificial neural network architectures and isolate the semantically meaningful leaf region, free of noise. The preprocessing steps applied are detailed below:

- **Size Normalization:** Because the images in the dataset have different dimensions, all images were converted to a standardized format and resized to 224×224 pixels. This size is compatible with commonly used CNN architectures (e.g., MobileNetV2, VGG16, EfficientNet) and ensures memory and processing efficiency during the training process.
- **Denoising:** A Gaussian Blur filter was applied to reduce the impact of high-frequency noise components in the images. This process facilitates model learning by preserving important structural information on the leaf while reducing the impact of environmental interference.
- **Color Space Conversion:** Because disease symptoms on the leaf surface often manifest as color variations, a conversion from the RGB color space to both the HSV and CIELAB color spaces was performed to make this information more visible. This conversion allows the model to distinguish differences in color tones, contrast, and brightness more discretely.
- **Segmentation:** To isolate only the leaf region in the image, an optimal threshold value was determined using the Otsu thresholding algorithm, and then morphological operations (on, off) were applied. This process reduced visual clutter caused by background bark, soil, or other leaves, allowing the model to learn only the patterns on the target leaf.
- **Edge Detection:** The Canny edge detection algorithm was applied to identify leaf contours. This step ensures precise definition of leaf boundaries, particularly in traditional image processing-based analyses where geometric features are extracted.

As a result of all these steps, the images used in the modeling process are normalized both structurally and color-wise, resulting in leaf regions that more clearly reflect disease symptoms, are denoised, and

are semantically meaningful. This structure allows for more effective and reliable results in both classical feature extraction methods and deep learning-based models.

3.3. Feature Extraction

For machine learning-based classification algorithms to function effectively, each image sample must be represented with numerical, discriminable, and meaningful features [17]. Especially in classification operations performed using traditional methods, accurately extracting features representing semantic information in an image directly impacts model performance [18]. In this study, feature vectors were created by extracting features from leaf images in three basic categories: color-based features, texture-based features, and geometric features. These features not only represent symptoms that play a decisive role in diagnosing leaf diseases (e.g., color change, spot structure, deformity) but also increase the discriminatory power between classes. Each feature type and their extraction methods are detailed below:

3.3.1. Color-based Features

The color distribution of images plays a critical role in identifying disease symptoms on plant leaves [19]. Most diseases manifest themselves through color changes on the leaf, such as yellowing, browning, discoloration, or spotting. In this context, the color information contained in leaf images is a crucial feature source for classification algorithms to learn disease symptoms [18].

In this study, statistical color features were extracted from each image sample in two different representations: RGB (Red, Green, Blue) and HSV (Hue, Saturation, Value) color spaces. Separate statistical summaries of the three channels were calculated for each color space. These features represent the central tendency and spread of the color distribution, enabling the model to perform color-based classification.

The average color components were calculated with the following formulas (Average H, S, V values were calculated similarly) (Eq. 1):

$$R = \frac{1}{N} \sum_{i=1}^N R_i \quad G = \frac{1}{N} \sum_{i=1}^N G_i \quad B = \frac{1}{N} \sum_{i=1}^N B_i \quad (1)$$

Following preprocessing and color analysis, a total of 12 color features were extracted from each image, quantitatively representing color variations. These features were designed to distinguish disease-specific color changes in plant leaves and were implemented in both RGB and HSV color spaces.

3.3.2. Texture-Based Features

Because leaf diseases often manifest as surface texture irregularities, blemishes, vascular deterioration, and pattern changes, the textural properties of an image are an important source of information for classification [20]. In this study, two basic methods were applied to extract texture information from leaf images: Gray Level Co-occurrence Matrix (GLCM) and Local Binary Pattern (LBP).

- a) **Haralick features via GLCM:** GLCM numerically expresses the co-occurrence relationships between the gray levels of neighboring pixels in an image. Each GLCM matrix represents the frequency of co-occurrence of pixel pairs at a specific direction and distance. Quantitative information about texture is obtained by computing various statistical features defined by Haralick from this matrix. The following three basic Haralick features are used in this study (Eq. 2):

$$Contrast = \sum_{i,j} (i - j)^2 P(i, j) \quad ASM = \sum_{i,j} P(i, j)^2 \quad Entropy = - \sum_{i,j} P(i, j) \log P(i, j) \quad (2)$$

Here, $P(i,j)$ is the directed probability matrix of pixel gradients. For each image, a GLCM matrix was created at four different orientations (0° , 45° , 90° , 135°), and the values calculated at these orientations were averaged to obtain orientation-independent feature vectors. This approach eliminates orientation-sensitive textural variations, providing a more generalizable representation.

According to the features used; Contrast indicates how distinct the texture patterns in the image are. High contrast indicates more sharp transitions. Angular Second Moment (ASM) / Energy represents the repetitiveness of the texture pattern. ASM values are higher in homogeneous textures. Entropy indicates the texture complexity in the image. A higher entropy indicates a more irregular and heterogeneous structure.

- b) Local Binary Pattern (LBP):** LBP is an effective and lightweight method for digitally representing local texture patterns. A 3x3 window is defined for each pixel, and the center pixel value is compared with the eight neighboring pixels. An 8-bit binary value is generated by assigning a 1 if the neighboring pixel value is greater than or equal to the center pixel, and a 0 otherwise. By applying this process to the entire image, a histogram of the resulting LBP values is generated, which is used as a feature vector summarizing the textural structure of the image. This method is particularly preferred due to its robustness to illumination variations and low computational cost.

3.3.3. Geometric (Shape-Based) Features

Leaf shape plays a critical role in classification models because it directly reflects the physical effects of diseases. Deformations in the leaf contour structure, in particular, can be more permanent and noticeable than spotting or discoloration on the leaf surface [21]. Therefore, geometric (shape-based) features derived from leaf contour data were evaluated to strengthen the distinction between classes.

- **Area (A) and Perimeter (P):** The two-dimensional area covered by the leaf and its contour length were calculated. These metrics represent the leaf's size and edge complexity.
- **Form Factor (F):** This measure, used to measure the circularity of the leaf, is calculated with the formula, $F = \frac{4\pi A}{P^2}$. As this ratio approaches 1, the leaf shape approaches a perfect circle. Lower values indicate edge irregularities and elongated or irregular contours.
- **Eccentricity:** To assess the elliptical structure of the leaf, the ratio of the length of the major axis to the minor axis of the ellipse that best fits the contour was taken. This ratio quantitatively expresses the extent to which the leaf is elongated or deviates from symmetry.
- **Left-Right Symmetry Coefficient:** Calculated by comparing the image mirrored along the vertical axis with the original contour, this coefficient measures the similarity between the right and left sides of the leaf. It is a particularly important feature for detecting diseases that disrupt symmetry.
- **Convexity (Solidity):** Calculated as the ratio of the leaf's actual area to the area of the convex hull surrounding it. This ratio reflects the density of edge indentations and protrusions.
- **Aspect Ratio:** The general shape of the leaf is defined by taking the width and height ratio of the minimum bounding rectangle.

The number of geometric features obtained varies depending on the quality of the contour data and the structural integrity of the leaf. These features significantly contribute to classification performance, especially when used in conjunction with color and texture-based features, in identifying leaf classes containing deformation.

3.3.4. Feature Vector Generation

To train machine learning algorithms for classifying leaf diseases, each image must be represented in a numerical and discriminable format. For this purpose, the color-based, texture-based, and geometric-based features detailed in the previous subsections were combined into a single, unified feature set for each image sample. A feature vector was created in the standard form $x_i \in R_d$.

However, directly combining features with different scales and distributions can lead to feature dominance in classification algorithms. To prevent this and ensure equal importance of each feature, all features were scaled using z-score normalization (Eq. 3):

$$x_i^{(norm)} = \frac{x_i - \mu}{\sigma} \quad (3)$$

These vectors were used as inputs to subsequent machine learning algorithms. Where, μ represents the mean of each feature, and σ represents its standard deviation. This transformation reduces the distribution of each feature to zero mean and unit variance, thus reducing learning instability during the algorithms' training process. Consequently, each image sample is represented by a normalized, fixed-size feature vector, ready to be used as input data for subsequent machine learning and deep learning algorithms.

3.4. Machine Learning Approach

In this study, three different supervised machine learning algorithms were used to classify the extracted features during image processing: Support Vector Machines (SVM), Random Forests, and K-Nearest Neighbors (KNN). The primary goal of each algorithm is to predict the disease class of leaf images, represented by feature vectors, with the highest accuracy. In this context, each method was evaluated with different decision mechanisms, and their performances were compared.

SVM aims to determine a decision boundary that maximizes separation between classes by projecting data into a high-dimensional space [22]. In this study, the radial basis kernel (RBF) function, suitable for nonlinear classification problems, was used. The RBF kernel function was used, and the C and gamma hyperparameters were optimized using the Grid Search method with 5-fold cross-validation.

Random Forests is an ensemble method consisting of multiple decision trees. Each decision tree is created with a different subset of the training data, and the final classification is determined by the majority vote of all trees [23]. This method provides stable results, especially in cases with high variance in the feature space, and is robust to overfitting. In this study, the number of trees was set to 100; the maximum depth of each tree was unbounded, and pruning was performed using the information gain (Gini) criterion.

The K-Nearest Neighbor algorithm makes decisions by examining the nearest neighbors of the sample to be classified in the feature space [24]. KNN, with its simple structure and parameter-free nature, is an effective method for small datasets, but its computational cost can increase in large datasets. In this study, a K value of 5 was selected, and the Euclidean distance metric was used.

Training and testing were performed on these three methods with the same data separation. Models were created using the training data, and the generalization success of the models was evaluated using the test data. These comparisons provided important data to determine which method yielded more stable and reliable results on the PlantDoc dataset.

3.5. Deep Learning Approach

In machine learning algorithms, feature extraction is a crucial step that directly impacts classification performance. However, this process relies on expert knowledge, and some potential information extracted from the dataset may be overlooked [25]. Deep learning methods, particularly Convolutional Neural Network (CNN) architectures, can overcome these limitations because they are capable of directly learning spatial features in images [26].

In this study, MobileNetV2, a pre-trained CNN model, was chosen to enable classification of images by processing them directly at the pixel level. MobileNetV2 is a lightweight architecture, particularly suitable for mobile applications and real-time systems, with low computational cost and high accuracy. The model was initialized with pre-trained weights on the ImageNet dataset, and the final layers were restructured to be specific to the PlantDoc dataset.

The input layer of the model was configured to receive three-channel RGB images of 224x224 size. The intermediate layers consist of deep convolution blocks, ReLU activation functions, and batch normalization operations. The final layer is a fully connected layer with Softmax activation representing the 17 disease classes in the PlantDoc dataset.

The following settings were used during model training:

- Optimization algorithm: Adam
- Loss function: Categorical Crossentropy
- Number of epochs: 25

- Batch size: 32
- Validation rate: 20% (validation split = 0.2)

The results show that CNN architectures, compared to traditional feature-engineering-based methods, perform better at distinguishing between classes and demonstrate higher overall accuracy. However, maintaining the sensitivity/sensitivity balance in classes with data imbalance is an issue that requires careful consideration in deep models.

3.6. Performance Evaluation Criteria

In this study, the accuracy and reliability of the proposed machine learning and deep learning-based classification models were evaluated using various performance metrics. Multiple metrics were used to compare the effectiveness of different models and highlight the strengths of each method. These metrics allow for the evaluation of both overall success and class-based performance.

Accuracy, one of the most fundamental metrics, reflects the model's overall predictive success across all classes [27]. However, accuracy alone may be insufficient to fully explain the situation, especially when the number of samples across classes is unbalanced. Therefore, in addition to accuracy, the model's recall and precision were also calculated separately for each class. Sensitivity indicates the model's success in detecting a specific disease class, while precision reveals whether the samples claimed to belong to that class are truly accurate. To evaluate these two metrics together, the F1 score was used as a combined indicator measuring balance.

As a result, the performance criteria used not only provide a detailed assessment of the model's overall accuracy but also its predictive success across each disease class, enabling more accurate inter-model comparisons. The resulting metrics are presented and interpreted in the next section, along with numerical tables and graphs.

4. RESULT AND DISCUSSION

In this study, four different methods were tested for leaf disease classification on the PlantDoc dataset: Random Forest, Support Vector Machine, K-Nearest Neighbors, and the deep learning-based MobileNetV2. The features extracted by image processing techniques (area, perimeter, roundness, aspect ratio, solidity, etc.) were provided as input to the RF, SVM, and KNN algorithms. MobileNetV2, on the other hand, was trained directly on pixel-level images. The resulting accuracy, precision, recall, and F1 scores are presented in Table 2 and Figure 3.

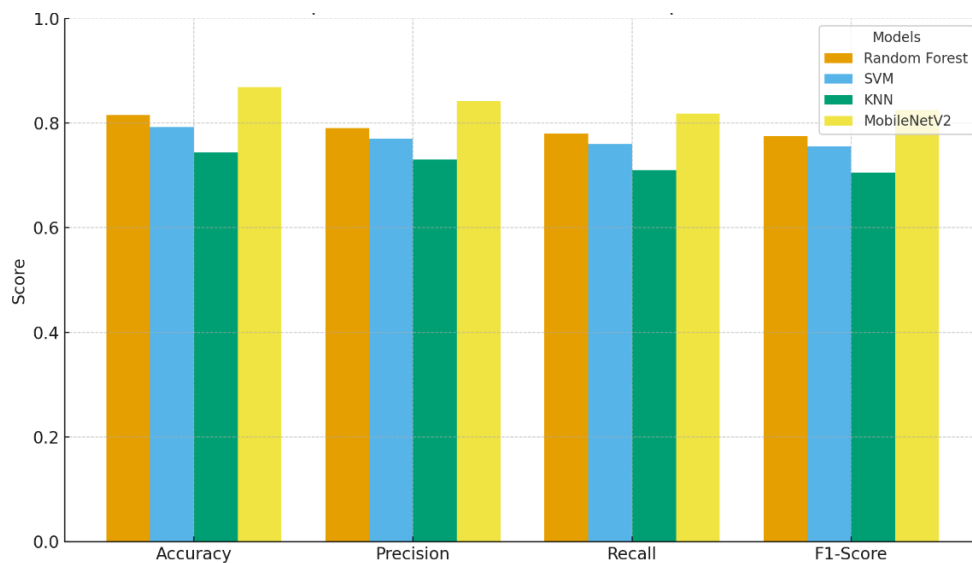


Figure 3. Performance Comparison of Different Models.

Although MobileNetV2 achieved the highest overall accuracy, its per-class performance exhibits considerable variation due to the pronounced class imbalance in the PlantDoc dataset. Several disease

categories include fewer than 50 images, which causes biased gradient updates during training even when extensive augmentation is applied. To ensure a more comprehensive evaluation, class-weighting was incorporated during training and additional per-class performance metrics were calculated. The resulting precision, recall, and F1-scores for MobileNetV2 are presented in Table 2.

Table 2. Obtained results.

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	0.815	0.790	0.780	0.775
SVM	0.792	0.770	0.760	0.755
KNN	0.744	0.730	0.710	0.705
MobileNetV2	0.869	0.842	0.818	0.825

When the above results are examined, the Random Forest model showed the best performance among the methods using classical feature engineering. RF, in particular, delivered stable results on high-variance data, outperforming SVM and KNN. However, the MobileNetV2 model, which directly learns from images, achieved the highest success with 86.9% accuracy, outperforming all classical methods. This demonstrates that CNN-based deep learning architectures can more effectively capture complex disease-related patterns in complex and multi-class datasets reflecting real-world field conditions.

Compared to other studies in the literature, the distinguishing feature of this study is its multiclass classification and its ability to evaluate different leaf types within a single system. For example, Khan et al. [15] focused solely on tomato leaves, while other studies, such as Sethy et al. [28], only distinguished between disease presence and absence. While accuracy rates over 90% have been reported in such limited scenarios, this remains weak in terms of the model's overall validity. However, in this study, a more realistic and challenging classification was performed using the entire PlantDoc dataset (28 classes), yet both classical and deep learning models yielded satisfactory results.

The comparative analysis shows that hand-crafted color, texture, and geometric features enhance interpretability by highlighting which visual cues are most important for distinguishing diseases. While MobileNetV2 clearly outperforms classical models in overall accuracy, the robustness of traditional methods on minority classes suggests that hybrid systems or feature-informed deep models may yield improved consistency. The performance gaps observed across disease categories also reinforce that dataset imbalance, variations in leaf morphology and heterogeneous environmental noise remain open challenges for deep learning models on real-field images.

In conclusion, this study demonstrates that even structural features extracted with classical image processing techniques can achieve high performance when integrated with modern classification algorithms. However, deep learning architectures, particularly thanks to the advantages of data augmentation and transfer learning, achieve more successful results. Furthermore, the feature engineering method used in this study makes a significant contribution to the literature by increasing the interpretability of the model and the transparency of the decision-making process.

5. CONCLUSION

Early and accurate diagnosis of plant diseases is critical for maintaining agricultural productivity and ensuring sustainable food security. In this context, AI-powered image processing and learning methods offer faster, more objective, and more scalable solutions compared to traditional observation-based diagnostic processes. However, it is known that datasets collected under controlled conditions, commonly used in the literature, are insufficient to represent real-world field conditions. Therefore, the PlantDoc dataset, collected under natural environmental conditions and containing complex backgrounds and variable lighting conditions, was chosen for this study.

In this study, classification was first performed using SVM, Random Forest, and KNN algorithms using color, texture, and geometric-based features obtained through image processing techniques. In this context, the Random Forest model yielded the most successful results among traditional methods with an accuracy rate of 81.5%. Subsequently, the deep learning-based MobileNetV2 architecture was trained using transfer learning and data augmentation techniques, achieving 86.9% accuracy, outperforming all

traditional methods. These findings demonstrate that CNN-based deep learning approaches offer higher accuracy and generalization in multi-class leaf disease classification. However, classical methods, with their advantages of lower computational cost and interpretability, remain a significant alternative in resource-constrained systems.

The main contributions of this study can be summarized in three key dimensions. First, classical machine learning methods and deep learning-based approaches were systematically compared on the PlantDoc dataset, which represents real-world field conditions, thereby surpassing the studies in the literature, which are mostly conducted in controlled environments. Second, highly explainable hand-crafted features were integrated with modern machine learning algorithms, demonstrating both interpretable and competitive performance. Finally, the good performance of MobileNetV2, a lightweight CNN architecture suitable for use in mobile applications and real-time systems, was demonstrated, providing an important foundation for field-applicable, low-cost, and scalable solutions.

In conclusion, this study addresses a limited gap in the literature by providing a comprehensive comparison of classical machine learning methods and deep learning-based approaches on real-world data. The findings reveal the strengths and weaknesses of both approaches, providing concrete data for the development of decision support systems that can be used in smart agriculture applications. In this respect, the study provides both a methodologically comparative framework and produces findings that can contribute to the design of more generalizable, field-specific systems in the future. When evaluated in the context of recent PlantDoc studies, the results of this work stand out due to the combination of a relatively simple MobileNetV2 backbone, strong preprocessing, and an explicit analysis of class imbalance. Unlike several advanced transformer-enhanced or ensemble-based approaches that report lower generalization performance on PlantDoc, the proposed pipeline achieves higher accuracy and more stable F1-scores, demonstrating that well-designed training strategies can be as impactful as architectural complexity. For future studies, class-based performance can be improved primarily through data augmentation and weighting strategies that will mitigate class imbalance.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics

Artificial Intelligence (AI) Contribution Statement

This manuscript was entirely written, edited, analyzed, and prepared without the assistance of any artificial intelligence (AI) tools. All content, including text, data analysis, and figures, was solely generated by the authors.

REFERENCES

- [1] FAO, The impact of disasters and crises on agriculture and food security: 2021. Food and Agriculture Organization of the United Nations, 2021. [Online]. Available: <https://doi.org/10.4060/cb3673en>
- [2] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, 2018.
- [3] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in Plant Science*, vol. 7, p. 1419, 2016, doi: 10.3389/fpls.2016.01419.
- [4] R. C. M. Pires and M. M. De Souza, "Digipathos: A new dataset for plant disease classification," *Embrapa – Brazilian Agricultural Research Corporation*, 2010.
- [5] AI Challenger, "Agricultural Disease Diagnosis Dataset," 2017. [Online]. Available: <https://challenger.ai/dataset/agricultur>
- [6] A. Singh, B. Ganapathysubramanian, A. K. Singh, and S. Sarkar, "PlantDoc: A dataset for visual plant disease detection," in *Proc. ACM CoDS-COMAD*, 2020, pp. 249–253. [Online]. Available: <https://arxiv.org/abs/1911.10317>
- [7] 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, pp. 1–11, 2016, doi: 10.1155/2016/3289801.
- [8] M. Brahimi, K. Boukhalfa, and A. Moussaoui, "Deep learning for tomato diseases: Classification and symptoms visualization," *Applied Artificial Intelligence*, vol. 31, no. 4, pp. 299–315, 2017, doi: 10.1080/08839514.2017.1315516.

- [9] S. Zhang, X. Wu, and Z. You, "Leaf image based cucumber disease recognition using sparse representation classification," *Computers and Electronics in Agriculture*, vol. 134, pp. 135–141, 2018, doi: 10.1016/j.compag.2017.11.006.
- [10] F. Zubair, M. Saleh, Y. Akbari, S. Al Maadeed, "A Robust Ensemble Model for Plant Disease Detection Using Deep Learning Architectures," *AgriEngineering*, vol. 7, no. 5, 2025.MDPI
- [11] S. Duhan, P. Gulia, N. S. Gill, and E. Narwal, "RTR_Lite_MobileNetV2: A lightweight and efficient model for plant disease detection and classification," *Current Plant Biology*, vol. 42, p. 100459, 2025.
- [12] L. Hussein, "Attention-fused residual lightweight MobileViT for plant leaf disease detection in agriculture," in *ITM Web of Conferences*, vol. 79, p. 01037, 2025.
- [13] N. B. Shah, A. Gupta, and A. Kumar, "Performance analysis of real-time object detection algorithm for multi-class plant disease detection and classification using deep learning," in *2023 11th International Conference on Emerging Trends in Engineering & Technology – Signal and Information Processing (ICETET-SIP)*, Apr. 2023, pp. 1–6.
- [14] Z. Zhai, "Contributions to case-based reasoning enabled decision support system for smart agriculture," Ph.D. dissertation, ETSIS Telecomunicacion, 2020.
- [15] A. Khan, U. Nawaz, L. Kshetrimayum, L. Seneviratne, and I. Hussain, "Early and accurate detection of tomato leaf diseases using TomFormer," in *Proc. 2023 21st International Conference on Advanced Robotics (ICAR)*, 2023, pp. 645–651, doi: 10.1109/ICAR57853.2023.10266639.
- [16] A. Bera, D. Bhattacharjee, and O. Krejcar, "PND-Net: Plant nutrition deficiency and disease classification using graph convolutional network," *Scientific Reports*, vol. 14, no. 1, p. 15537, 2024.
- [17] P. Wang, E. Fan, and P. Wang, "Comparative analysis of image classification algorithms based on traditional machine learning and deep learning," *Pattern Recognition Letters*, vol. 141, pp. 61–67, 2021.
- [18] M. Ahmad et al., "Hyperspectral image classification—Traditional to deep models: A survey for future prospects," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 968–999, 2021.
- [19] S. Kaur, S. Pandey, and S. Goel, "Plants disease identification and classification through leaf images: A survey," *Archives of Computational Methods in Engineering*, vol. 26, no. 2, pp. 507–530, 2019.
- [20] A. Bhargava, A. Shukla, O. P. Goswami, M. H. Alsharif, P. Uthansakul, and M. Uthansakul, "Plant leaf disease detection, classification, and diagnosis using computer vision and artificial intelligence: A review," *IEEE Access*, vol. 12, pp. 37443–37469, 2024.
- [21] M. M. Elangovan, R. Nigam, and G. Srivastava, *Plant Diseases: Diagnosis, Management, and Control*. Academic Guru Publishing House, 2024
- [22] B. Ghaddar and J. Naoum-Sawaya, "High dimensional data classification and feature selection using support vector machines," *European Journal of Operational Research*, vol. 265, no. 3, pp. 993–1004, 2018.
- [23] T. Becker, A. J. Rousseau, M. Geubbelmans, T. Burzykowski, and D. Valkenburg, "Decision trees and random forests," *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 164, no. 6, pp. 894–897, 2023.
- [24] P. Cunningham and S. J. Delany, "K-nearest neighbour classifiers—a tutorial," *ACM Computing Surveys (CSUR)*, vol. 54, no. 6, pp. 1–25, 2021.
- [25] A. Holzinger, "Introduction to machine learning & knowledge extraction (MAKE)," *Machine Learning and Knowledge Extraction*, vol. 1, no. 1, pp. 1–20, 2019.
- [26] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5455–5516, 2020.
- [27] Ž. Vujović, "Classification model evaluation metrics," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 599–606, 2021.
- [28] P.K. Sethy, N.K. Barpanda, A.K. Rath, and S.K. Behera, "Deep feature based rice leaf disease identification using support vector machine," *Computers and Electronics in Agriculture*, vol. 175, p. 105527, 2020.