International Journal of Robotics and Control Systems
Vol. 4, No. 3, 2024, pp. 1429-1445
ISSN 2775-2658
http://pubs2.ascee.org/index.php/ijrcs

Cucumber Disease Image Classification with A Model Combining LBP and VGG-16 Features

Miftahol Arifin [a,1,*], Anny Yuniarti [a,2], Nanik Suciati [a,3]

[a] Department of Informatics, Institut Teknologi Sepuluh Nopember, Sukolilo, Surabaya 60111, Indonesia
[1] miftaholarifin2017@gmail.com; [2] anny@if.its.ac.id; [3] nanik@if.its.ac.id
* Corresponding Author

ARTICLE INFO

Article history
Received June 19, 2024
Revised July 27, 2024
Accepted August 16, 2024

Keywords
Cucumber Disease Classification;
Hybrid Model;
Image Classification;
Local Binary Pattern (LBP);
Random Forest (RF);
Texture Features;
Visual Feature Extraction;
VGG-16

ABSTRACT

Cucumber (Cucumis sativus) is a significant horticultural crop worldwide, highly valued for both fresh consumption and processing. However, cucumber cultivation faces challenges due to diseases that can substantially reduce yield and quality. Diseases like leaf spots, stem wilt, and fruit rot are caused by pathogens including viruses, bacteria, and fungi. Traditionally, disease detection in cucumbers is performed manually, which is time-consuming and inefficient. Therefore, developing machine vision-based models using Deep Learning (DL) and Machine Learning (ML) for early disease detection through image analysis is crucial for assisting farmers. While many studies on plant disease classification using various DL and ML models show optimal results, research on cucumbers has mostly focused on leaf diseases. This study aims to optimize cucumber disease image classification by developing a model that combines Local Binary Pattern (LBP) texture features and VGG-16 convolutional features. The dataset used, Cucumber Disease Recognition Dataset consists of 8 classes of cucumber plant disease images covering leaves, stems, and fruits. This study classifies cucumber plant disease images using Random Forest (RF) combined with LBP texture features and VGG-16 visual features and compares its performance with models using VGG-16, LBP+RF, and VGG-16+RF on the same dataset. The results show that the proposed model achieved a precision of 84.7%, recall of 84%, F1-Score of 83.8%, and accuracy of 84%. These results outperform the comparative models, demonstrating the effectiveness of the combined approach in classifying cucumber plant diseases.

1. Introduction

Cucumber (Cucumis sativus) is one of the important horticultural crops in the world. It is highly suitable for consumption in both fresh and processed forms. Cucumber plants belong to the Cucurbitaceae botanical family, characterised by climbing growth and yellow flowers similar to melon plants (Cucumis melo) [1]. Cucumber is a diverse source of nutrients, including vitamins, iron, calcium, niacin, thiamine, fibre, and other minerals, making it a beneficial plant suitable for daily consumption [2]. Cultivating cucumber plants often faces challenges from diseases that attack, causing significant declines in yield and quality. Diseases such as leaf spots, fusarium wilt, and fruit rot are caused by various pathogens, including viruses, bacteria, and fungi [3]. Disease detection in cucumber plants is usually done manually by individuals, which is time-consuming and ineffective for both farmers and retailers. Therefore, it is crucial to develop vision-based models like Machine Learning (ML) and Deep Learning (DL) to minimize human effort, costs, and production time in the agricultural industry by identifying diseases in cucumber plants [4], [5].

Image analysis and classification of plant diseases is an effective method for identifying and distinguishing diseases in plants. Advances in artificial intelligence and image processing techniques provide opportunities to expand research in the agricultural field. Recently, Deep Learning (DL) has shown its advantages in various aspects of the agricultural sector, including plant disease diagnosis, plant growth prediction, insect recognition, and more [6] Deep learning, utilizing layered artificial neural networks (deep neural networks), facilitates feature extraction, pattern analysis, and automatic image data classification. This method employs layers of learning to recognize complex features in images, which are challenging to achieve with traditional image processing methods. Thus, deep learning becomes a powerful tool for improving accuracy and efficiency in plant disease detection and classification, aiding farmers in making informed decisions for disease management and ultimately enhancing agricultural productivity [7].

Research on the classification of diseases affecting cucumber plants using image datasets of diseases that attack various parts of the plant such as leaves, stems, and fruits, based on deep learning (DL) and machine learning (ML), is rarely conducted. Most previous studies relied on cucumber disease datasets that only focused on leaf disease [8]-[13]. Therefore, the authors aim to test the classification performance of the Cucumber Disease Recognition Dataset. This dataset not only includes diseases that affect the leaves but also diseases that generally harm farmers, such as those affecting leaves, stems, and fruits. Considering that previous studies on image classification with this dataset have not been conducted, we include related research that intersects with the authors' proposed methods.

Conventional and Convolutional Neural Network (CNN) image classification methods have been applied in various studies to identify diseases in plant images. This approach shows that integrating advanced technologies such as CNN in image analysis can enhance accuracy and efficiency in identifying plant diseases, encouraging further research combining complex datasets for a holistic understanding of diseases affecting various plant parts. Previous research on cucumber disease classification suggested the use of specific Convolutional Neural Network (CNN) algorithms designed to improve the accuracy of existing models. The dataset used in this research consists of 4,868 cucumber disease images, covering five types of cucumber leaf disease images and one healthy cucumber leaf image. Experiments were conducted by comparing the proposed CNN algorithm with several pre-trained classification models such as AlexNet, Inception-V3, and ResNet-50. The proposed CNN algorithm achieved recognition accuracy of 98.19% with augmented datasets and 100% accuracy with publicly available plant disease datasets. Experimental results show that the proposed CNN algorithm is more effective in recognizing cucumber leaf diseases compared to other transfer learning algorithms [14].

Previous research [15] focused on the intelligent identification and classification of cucumber plant diseases in greenhouses. The aim of this study was to find an efficient method to address the problem of disease similarity caused by two types of diseases occurring on the same leaf as well as the influence of external light. First, this study obtained a cucumber leaf disease dataset in a naturally complex greenhouse background, which included not only powdery mildew, downy mildew, and healthy leaves, but also a combination of powdery mildew and downy mildew. Second, this study used the latest method, EfficientNet, to construct a classification model for these four types of diseases, achieving a model accuracy of 97%, and proving that EfficientNet-B4 is the most suitable method for this study. Then, this study built a two-class classification model for similar cucumber diseases using EfficientNet-B4 improved with the latest optimizer, Ranger, and achieved an unexpected accuracy of 96%. The experimental results showed that the improved method has a significant effect on the classification of similar cucumber diseases.

Furthermore, other research conducted plant disease identification using deep learning. In this experiment, researchers examined CNN, VGG-16, VGG-19, and ResNet-50 models on the

PlantVillage image dataset containing 10,000 images to detect plant infections, and obtained accuracy rates of 98.60%, 92.39%, 96.15%, and 98.98% for CNN, VGG-16, VGG-19, and ResNet-50, respectively. This study indicates that ResNet-50 outperforms the other models with an accuracy of 98.98% [16] .

Further research on disease classification using traditional image classification methods with Machine Learning (ML), such as the study by [17] on disease classification in corn leaf images, utilized Local Binary Pattern (LBP) feature extraction combined with the K-Nearest Neighbors (k-NN) algorithm for classification purposes. The performance of the k-NN method in this study was quite high, with an accuracy of around 81.1%, an Area Under the Curve (AUC) value of 94.1%, an F1 Score of 80.9%, Precision of 81.8%, and Recall also of 81.1%. These scores indicate the effectiveness of the method used in classifying diseases in corn leaves.

Then there was research on rice leaf disease classification using the Random Forest Classifier. In this study, the classification process was conducted using augmented images, applying the Color Histogram feature extraction method, and classifying using the Random Forest algorithm. Additionally, this study compared different feature extraction methods and algorithms to obtain the best results, achieving the highest accuracy of 99.65% with the proposed method [18].

Recent research on developing new methods for plant leaf identification by combining shape features and CNN features has also been conducted. This method used a shape descriptor called IMTD and CNN feature analysis. This method has been tested on various plant leaf datasets, demonstrating adequate recognition results both in automatic leaf image acquisition without manual intervention and in supervised and parameter-adjusted leaf image classification by an operator. Overall, this method outperformed other plant identification methodologies previously developed [19].

From the above research, it is concluded that although deep learning (DL) features produce excellent performance in plant disease classification, some researchers also focus on artificial features based on machine learning (ML) for plant disease classification on datasets with a small number of samples, and the results are satisfactory with competitive accuracy levels. The use of complex architectures in DL, when applied to small datasets, is prone to overfitting. Therefore, the authors propose a study to develop a new model that can classify cucumber plant diseases in the Cucumber Disease Recognition Dataset. This study implements a combination of image texture features derived from texture feature extraction using Local Binary Pattern (LBP) and convolutional features from VGG-16, followed by a classification process using the Random Forest (RF) algorithm to reduce overfitting. This research aims to improve the reliability and accuracy of the model in classifying types of cucumber plant diseases in the images within the dataset.

2. Materials and Methods

2.1. Dataset

This research uses the Cucumber Disease Recognition Dataset, which is publicly available on the Mendeley website and was published in 2023. The dataset was collected by Nusrat Sultana and colleagues, containing 1280 original cucumber disease images and 6400 augmented cucumber disease images. The dataset comprises 8 classes of cucumber disease images: Anthracnose, Bacterial Wilt, Belly Rot, Downy Mildew [20], [21], Fresh Cucumber, Fresh Leaf, Gummy Stem Blight, and Pythium Fruit Rot, sample data can be seen in Fig. 1 [3]. This study uses the augmented dataset, consisting of 6400 cucumber disease images. Augmentation was performed using scaling, shifting, shearing, cropping, random rotation, and brightness adjustment [22] as shown in Fig. 2. The scaling process was performed using bicubic interpolation, bilinear interpolation, and nearest neighbour interpolation. Brightness was adjusted using histogram equalisation. The `fill_mode()` function from the Keras package with Python was used to estimate parameters. The parameters we used in the augmentation process were: rotation at angles of 45°, 60°, and 90°, width shift range, and height shift range of 0.2. From each class consisting of 160 original images, 800 additional images were obtained. The number of images used in testing can be seen in Table 1.

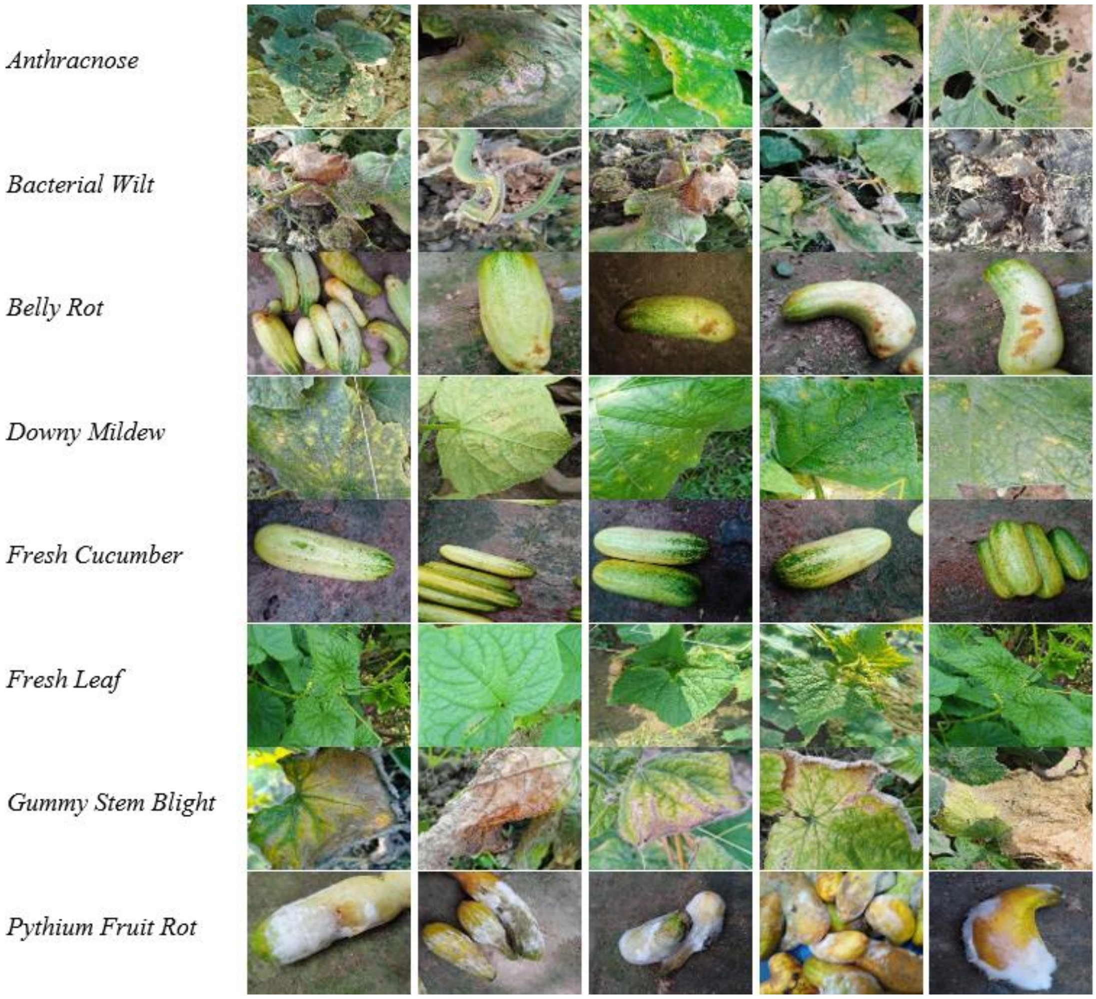

Fig. 1. Sample data from the cucumber disease recognition dataset

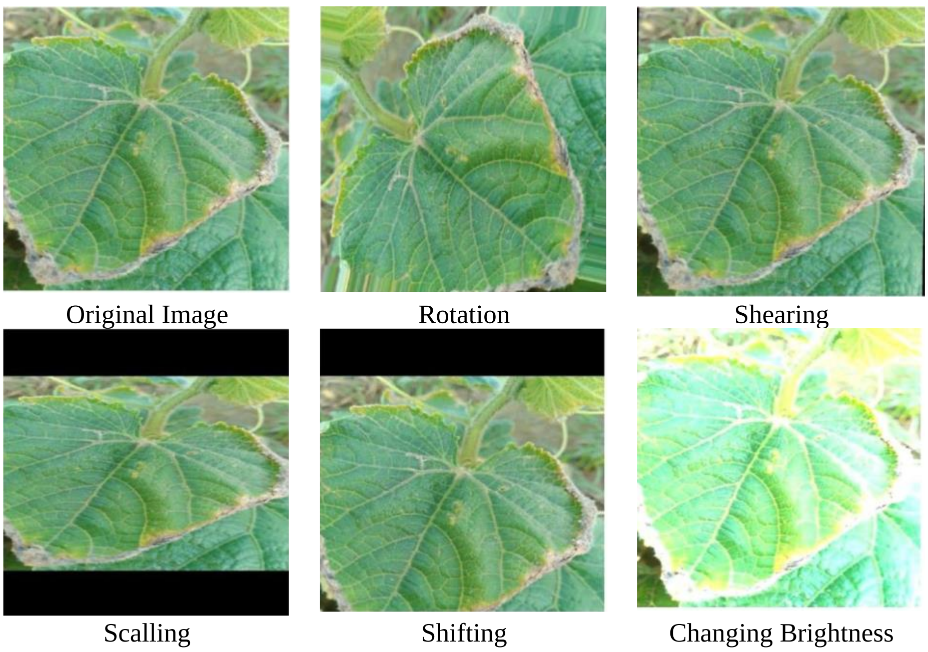

Fig. 2. Image augmentation on cucumber disease dataset

Table 1. Cucumber disease recognition dataset

Disease Class	Original Images	Augmented Images
Anthracnose	160	800
Bacterial Wilt	160	800
Belly Rot	160	800
Downy Mildew	160	800
Fresh Cucumber	160	800
Fresh Leaf	160	800
Gummy Stem Blight	160	800
Pythium Fruit Rot	160	800
Total	1280	6400

2.2. Image Classification

Image classification is a method used to group pixels within an image into classes that share similar types or categories [23], [24]. In a general context, there are two common approaches to image classification: supervised and unsupervised [25]. Supervised image classification methods are effective techniques, especially for grouping images containing similar objects. In this approach, input images with unique characteristics of each object are used as the basis for identifying and classifying images into specific categories or classes. These unique characteristics can include various aspects, such as shape, colour, texture, or specific patterns present in the images [26].

2.3. Local Binary Pattern (LBP)

Local Binary Patterns (LBP) is a statistical approach used in image processing that allows us to obtain effective features from images. LBP was first introduced in 1994 as a way to describe specific visual characteristics and was specifically designed for two-dimensional texture patterns [27], [28]. LBP works by comparing each pixel in an image with the surrounding pixels and then converting these comparisons into a binary pattern [29], [30]. This binary pattern represents the local texture around the pixel. The LBP code is calculated from a 3×3 window representation on an image. The intensity of each pixel in the original image is represented in a matrix pattern. By shifting the 3x3 filter from the top left corner to the bottom right corner for each pixel individually to find the intensity of each weighted point, we finally obtain the LBP code for each point in the original image. Then, we represent the image histogram based on the LBP code. LBP features can be extracted from the histogram and each LBP cell histogram can be normalized using L2 normalization. An illustration of the LBP process with a 3×3 matrix can be seen in Fig. 3 along with the calculation below [31], [32]:

$$A = \begin{pmatrix} g1 & g2 & g3 \\ g8 & g0 & g4 \\ g7 & g6 & g5 \end{pmatrix} \tag{1}$$

his matrix shows a 3×3 grayscale pixel block, with the center located at $g0$. To obtain the threshold value, LBP subtracts the coordinates of each neighbour as follows:

$$LBP1 = \begin{pmatrix} (g1 - g0) & (g2 - g0) & (g3 - g0) \\ (g8 - g0) & g0 & (g4 - g0) \\ (g7 - g0) & (g6 - g0) & (g5 - g0) \end{pmatrix} \tag{2}$$

Then, convert the values into binary values of 0 and 1

$$LBP1 = \begin{pmatrix} s(g1 - g0) & s(g2 - g0) & s(g3 - g0) \\ s(g8 - g0) & g0 & s(g4 - g0) \\ s(g7 - g0) & s(g6 - g0) & s(g5 - g0) \end{pmatrix} \tag{3}$$

$$S(x) = \{\begin{matrix} 1 & if\ x > 0 \\ 0 & if\ x < 0 \end{matrix} \tag{4}$$

Vol. 4, No. 3, 2024, pp. 1429-1445

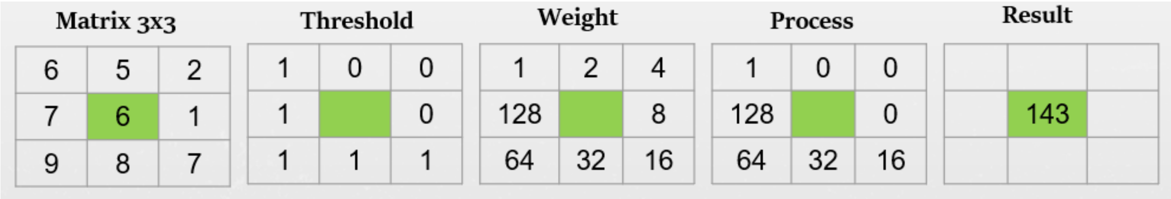

Fig. 3. Illustration of the LBP feature extraction process with a 3×3 matrix

This method is very effective in identifying texture patterns because it can capture small details and intensity variations in images. LBP is used in various applications such as facial recognition, texture analysis, and object detection due to its advantages in extracting robust texture information that is resistant to changes in lighting and rotation [33], [34].

2.4. Transfer Learning Visual Geometry Group (VGG-16)

The Visual Geometry Group network, better known as VGG-16, is recognized as one of the most accurate feature extractors in the field of computer vision. Its architecture, detailed in Fig. 4, processes input images with fixed dimensions of 3×224×224 pixels [35], [36]. The images are then processed through a series of diverse convolutional layers, each with different receptive fields, allowing the VGG-16 model to capture features at various scales. The VGG-16 network structure is characterized by consistency in the stride level, namely 3×3 with a stride of one for convolutional layers, and 2×2 with a stride of two for pooling layers, ensuring that every detail in the image can be explored efficiently [37].

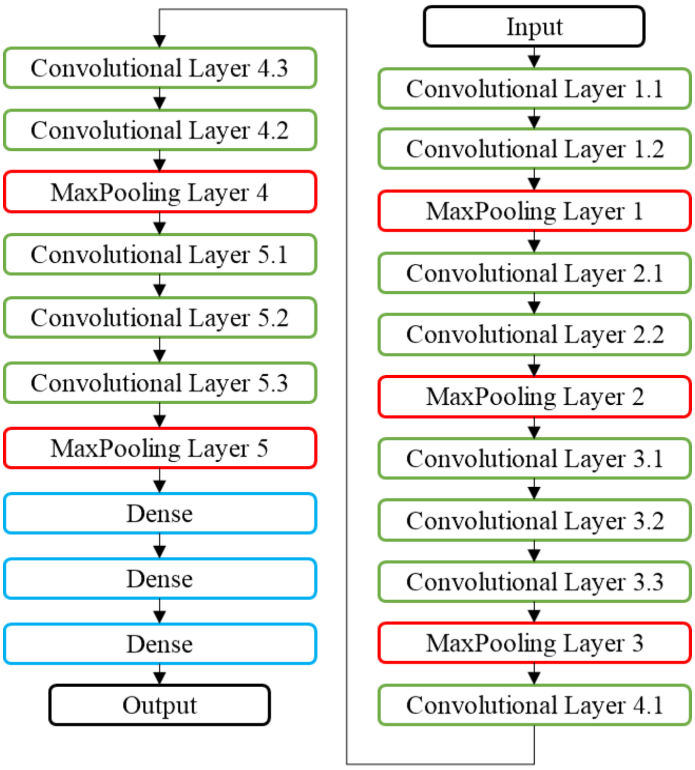

Fig. 4. Transfer learning architecture of visual geometry group VGG-16

In its architectural design, the first two convolutional layers are set with 64 and 128 filters, respectively, as shown in Fig. 4 with green colour blocks, aiming to capture the basic features of the image [38]. The progression in convolutional layers uses an increasing number of filters, namely 256, 512, and 512, for the subsequent layers. This combination allows the network to extract more complex and refined information from the input images. Additionally, to maintain the feature map size consistent with the input image, the edge pixels of the image are padded before performing convolution operations. This ensures that no information is lost during the convolution process [39].

In the final stage of this architecture, VGG-16 is equipped with three fully connected layers, with the first two layers containing 4096 neurons each, serving to integrate the features extracted by the previous convolutional layers. The final fully connected layer compresses the complex feature array into a 1000-dimensional feature vector, which can then be used for tasks such as classification. The strength of the VGG-16 structure lies in its exceptional ability to extract highly detailed features from images, making it an invaluable tool in various image processing and computer vision applications [40]-[42].

2.5. Random Forest

Random Forest (RF) is an ensemble method based on decision trees. RF reduces the level of overfitting by combining several overfit evaluators, namely decision trees, to form an ensemble learning algorithm [30], [43]-[45]. Each decision tree can obtain an appropriate classification result. By using the voting results from each decision tree in the forest or the set of decision trees, the category of the sample to be tested is determined based on the principle of the minority following the majority, and the category with the higher votes across all decision trees is determined as the final result [46]-[50] as shown in Fig. 5.

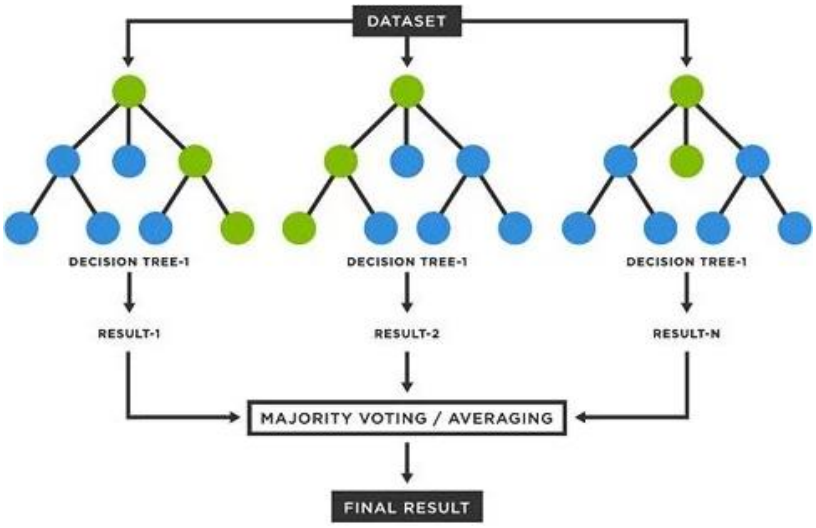

Fig. 5. Illustration of the random forest (RF) process

2.6. Proposed Method

In this study, a method is proposed for the classification of cucumber plant disease images by combining image features, namely LBP (Local Binary Pattern) texture features and features generated from VGG-16, as shown in Fig. 6.

2.6.1. Image Preprocessing

The initial step in this research is to prepare the cucumber disease image dataset downloaded from the website www.mendeley.com, titled Cucumber Disease Recognition Dataset. Next, image preprocessing is performed to make the images easier to process in subsequent steps. Several data preprocessing techniques are applied, including:

1. Resizing: Changing the image size from 2296×1724 to 227×227 to speed up the computation process and serve as input for the VGG-16 model.
2. Normalization: Converting the image into a tensor so that each pixel in each image falls within the range of 0–255, then normalizing each tensor with a mean and standard deviation of 0.5.
3. Splitting data: The preprocessed data is then divided into three datasets: training data, validation data, and test data, with respective proportions of 70%, 10%, and 20%.
4. Converting images to grayscale: This step is done to facilitate the extraction of LBP (Local Binary Pattern) features.

2.6.2. Image Feature Extraction

Image feature extraction is a crucial process in this research. The feature extraction process is applied to the dataset we previously divided into training data and test data. The feature extraction techniques used in this system design include Local Binary Pattern (LBP) texture feature extraction and visual feature extraction using the VGG-16 model [51]. The results of the feature extraction process yield four image features: LBP and VGG-16 features for test data, and LBP and VGG-16 features for training data, as shown in Fig. 6.

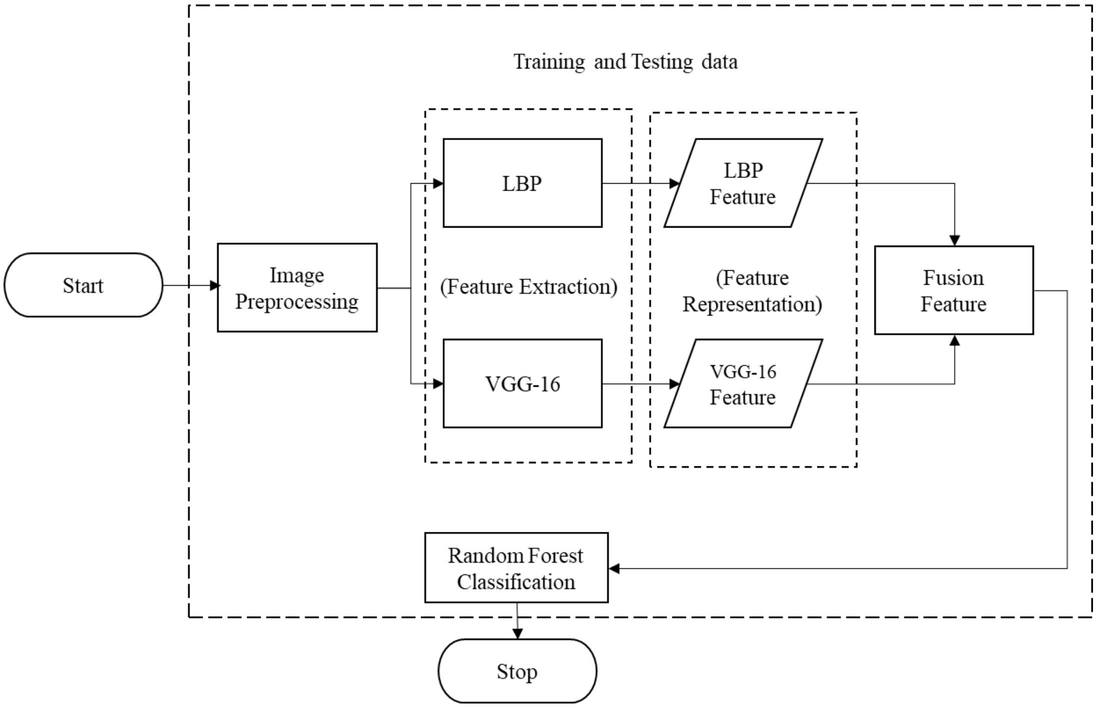

Fig. 6. Design of the proposed method

2.6.3. Image Classification Process

Before the image classification process, the four features generated in the previous step are combined. The combined features result in LBP + VGG-16 features for training data and LBP + VGG-16 features for testing data, as shown in Fig. 6. The next step in this research is training the model using the Random Forest Classifier (RF) on the combined LBP + VGG-16 features of the training data for image classification. Then, the trained model is tested using the combined LBP + VGG-16 features of the testing data. The final stage of the system design and implementation involves evaluating the model testing results and analyzing the outcomes of this evaluation.

2.7. Scenario

The test scenarios in this study are as follows. First, we conducted performance tests for image classification using the VGG-16 model alone. Second, we conducted performance tests for image classification using a combination of the LBP model and Random Forest. Third, we conducted performance tests for image classification using a combination of the VGG-16 model and Random Forest. Lastly, we conducted performance tests for image classification using a combination of features (LBP and VGG-16) and Random Forest. All tests used the same dataset and dataset composition.

3. Results and Discussion

3.1. Cucumber Disease Classification with VGG-16

The VGG-16 model on the Cucumber Disease Recognition dataset used 4,480 images for training data, 1,280 for validation data, and 640 for test data, totalling 6,400 images. The VGG-16 model training was designed with a batch size parameter of 128, a learning rate of 0.001, and the number of epochs determined using the early stopping function from the Python library, where epochs stop if there is no further increase in accuracy. In this case, the number of epochs reached 16, as shown in the accuracy and loss graph in Fig. 7. Based on the graph, the average validation accuracy was 0.85, and the average validation loss was 1.85.

After the model was trained, the VGG-16 model was tested using the prepared test data. The evaluation metrics used in this research were precision, recall, accuracy, and F1-score. The performance results of the VGG-16 model on the Cucumber Disease Recognition dataset are presented in Table 2, which shows an accuracy of 23.5%, indicating that the model experienced overfitting.

From the table, it is evident that the model displayed significant variation in its results based on the type of disease recognized. For example, the model was very accurate for "Bacterial Wilt" with a precision of 0.97. In contrast, the accuracy for "Anthracnose" was very low, at only 0.06. Furthermore, the model completely failed to identify diseases like "Belly Rot" and "Pythium Fruit Rot," as evidenced by zero scores across all measurements. One notable result was the model's ability to always recognize "Gummy Stem Blight" perfectly with a recall of 1.00, despite its low precision of 0.16.

A confusion matrix of the VGG-16 model on the Cucumber Disease Recognition dataset is provided in Fig. 7. According to the confusion matrix, 489 data points were misclassified, and 151 data points were correctly classified. We can also see the representation of recognized or unrecognized images in the confusion matrix.

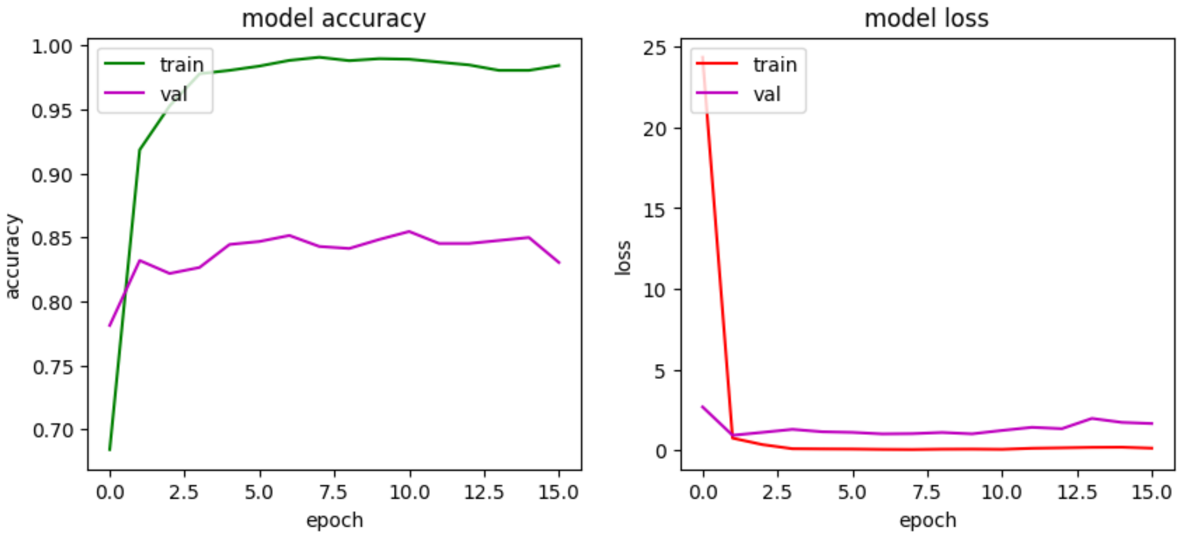

Fig. 7. Accuracy and loss graph of the VGG-16 model on the cucumber disease dataset training recognition

3.2. Cucumber Disease Classification with LBP and Random Forest

Table 3 presents the results of cucumber disease classification using Random Forest (RF) with Local Binary Pattern (LBP) feature extraction. The Random Forest model was designed with n-estimator = 500, meaning that the model builds 500 decision trees randomly on the dataset. The experimental results for precision, recall, and F1-score for each class in the Cucumber Disease Recognition dataset show that the Bacterial Wilt class stands out as the best-detected disease, with high precision (0.59), very high recall (0.79), and an F1-score of (0.67). This indicates that the classification model using Random Forest (RF) with Local Binary Pattern (LBP) feature extraction is effective in identifying Bacterial Wilt disease cases with high accuracy.

The model also has the lowest precision and recall values (0.28 and 0.21) for the Anthracnose disease class. This shows that the model is less effective in detecting Anthracnose cases. The confusion metrics for the classification model using Random Forest (RF) with Local Binary Pattern

Vol. 4, No. 3, 2024, pp. 1429-1445

(LBP) feature extraction are presented in Fig. 8. Based on Fig. 9, the overall accuracy of the model reaches 0.4803, meaning 48.03% of the disease classifications are predicted correctly. The misclassification rate is 0.5197, indicating that approximately 51.97% of the predictions are incorrect.

Table 2. Performance results of the VGG-16 model on the cucumber disease recognition dataset

Label	Precision	Recall	F1-Score	Acc
Anthracnose	0.06	0.03	0.03	0.235
Belly Rot	0	0	0	
Fresh Cucumber	0.61	0.34	0.44	
Gummy Stem Blight	0.16	1.00	0.28	
Bacterial Wilt	0.97	0.36	0.53	
Downy Mildew	0.70	0.09	0.16	
Fresh Leaf	0.24	0.07	0.11	
Pythium Fruit Rot	0	0	0	

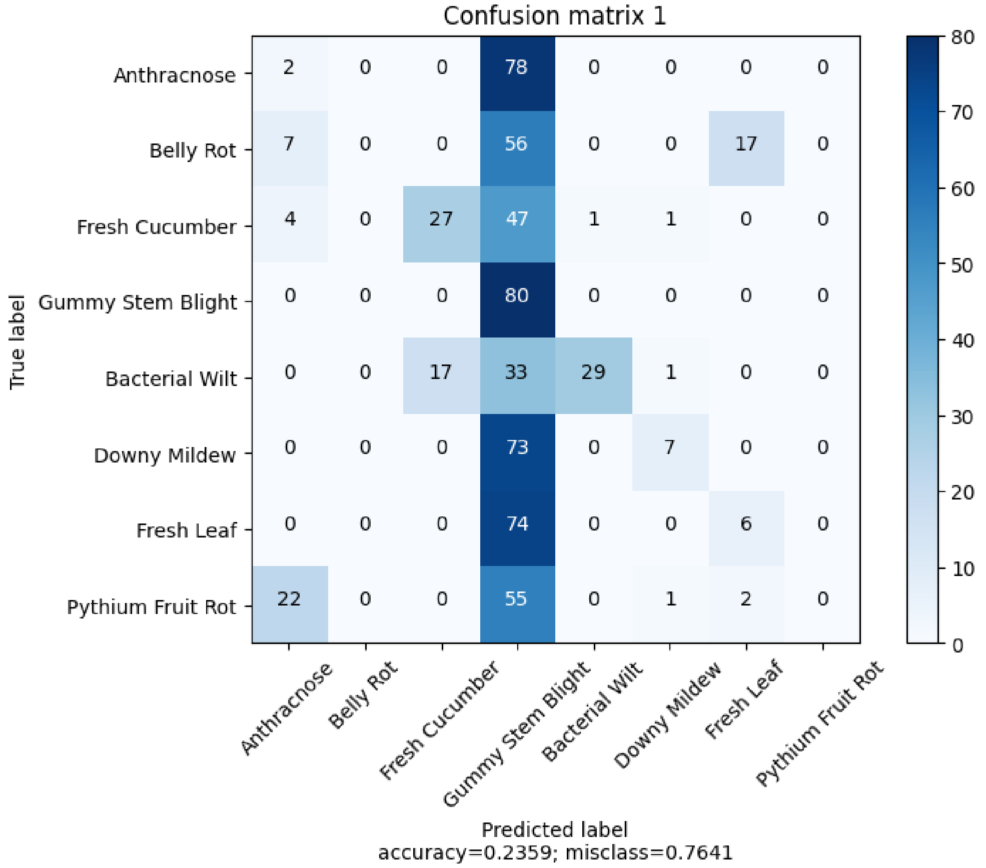

Fig. 8. Confusion matrix of the VGG-16 model on the cucumber disease recognition testing dataset

3.3. Cucumber Disease Classification with VGG-16 and Random Forest

Table 4 presents the results of cucumber disease classification using Random Forest (RF) with VGG-16 feature extraction. The Random Forest model in this instance was designed with n-estimator = 500 and random state = 42. Based on the data in Table 4. The cucumber disease 'Belly Rot' shows the best performance with a precision of 0.92 and a perfect recall of 1.00, resulting in an F1-Score of 0.96. The disease 'Pythium Fruit Rot' also provides excellent results with a precision of 0.91 and a recall of 0.99, yielding an F1-Score of 0.95. On the other hand, the disease 'Gummy Stem Blight' has the lowest performance with a precision of 0.63 and a recall of 0.76, resulting in an F1-Score of 0.69. Overall, the model achieves an accuracy of 82%, indicating that this method is quite effective in identifying cucumber diseases based on the analysed data. The confusion matrix is presented in Fig. 10 to facilitate the representation of classification results for each cucumber disease class.

Table 3. Performance of classification using LBP and RF on the cucumber disease recognition dataset

Label	Precision	Recall	F1-Score	Acc
Anthracnose	0.28	0.21	0.24	0.496
Belly Rot	0.59	0.76	0.66	
Fresh Cucumber	0.49	0.51	0.50	
Gummy Stem Blight	0.64	0.47	0.55	
Bacterial Wilt	0.59	0.79	0.67	
Downy Mildew	0.55	0.49	0.53	
Fresh Leaf	0.32	0.39	0.35	
Pythium Fruit Rot	0.47	0.35	0.40	

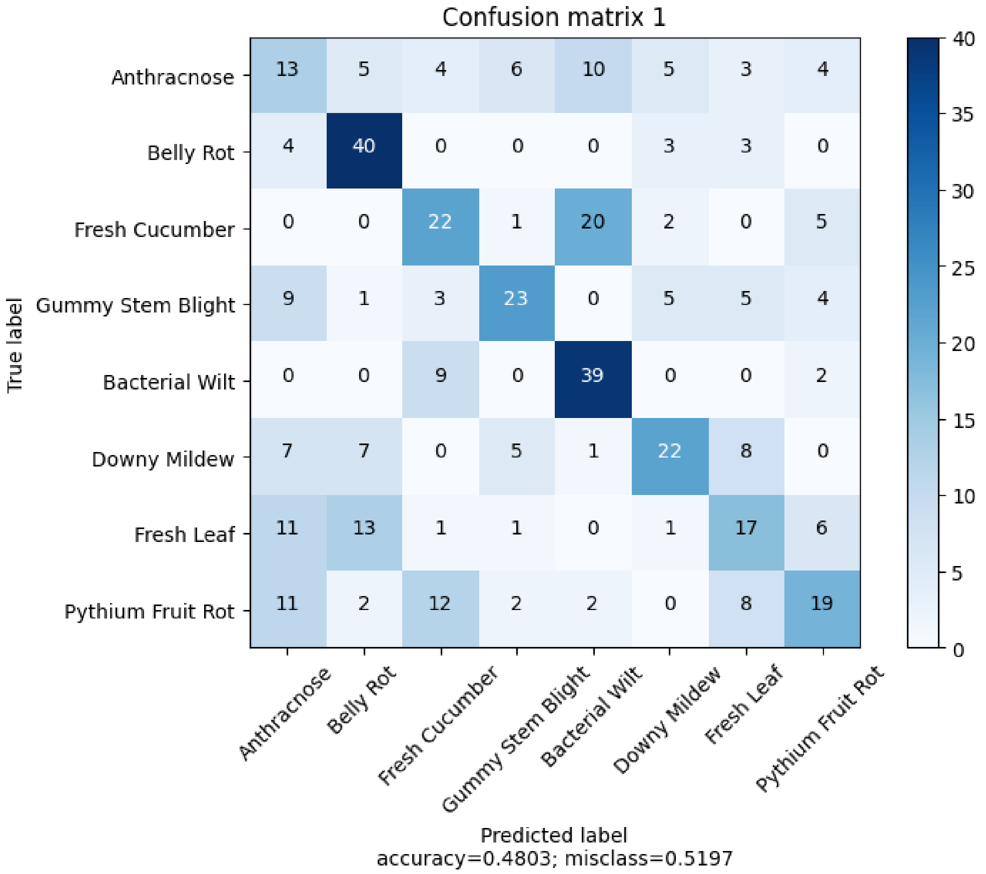

Fig. 9. Confusion matrix of the RF classification model with LBP features on the cucumber disease recognition testing dataset

Table 4. Performance results of classification using VGG-16 and RF on the cucumber disease recognition dataset

Label	Precision	Recall	F1-Score	Acc
Anthracnose	0.85	0.55	0.67	0.82
Belly Rot	0.92	1.00	0.96	
Fresh Cucumber	0.96	0.80	0.87	
Gummy Stem Blight	0.63	0.76	0.69	
Bacterial Wilt	0.88	0.97	0.92	
Downy Mildew	0.75	0.81	0.78	
Fresh Leaf	0.73	0.68	0.70	
Pythium Fruit Rot	0.91	0.99	0.95	

3.4. Cucumber Disease Classification with the Proposed Method

Table 5 presents the performance results of classifying various diseases in cucumber plants using combined features and the Random Forest (RF) method on the Cucumber Disease Recognition dataset. From Table 5 the disease 'Belly Rot' shows the best classification performance, with a precision of 0.92 and a perfect recall of 1.00, resulting in an F1-Score of 0.96. 'Pythium Fruit Rot' also shows excellent results with a precision of 0.92 and a recall of 0.99, giving an F1-Score of 0.95. 'Bacterial Wilt' has a precision of 0.90 and a recall of 0.99, with an F1-Score of 0.94, indicating very effective detection. On the other hand, 'Gummy Stem Blight' shows lower performance with a precision of 0.67 and a recall of 0.78, giving an F1-Score of 0.72.

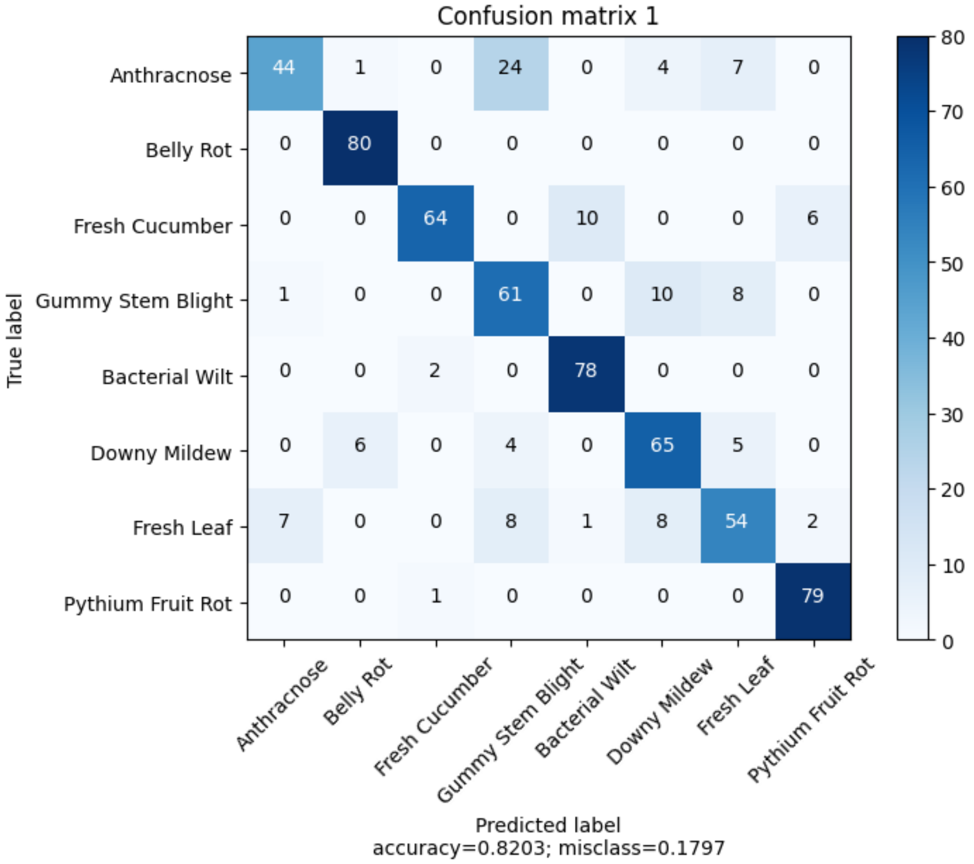

Fig. 10. Confusion matrix of the RF classification model with VGG-16 features on the cucumber disease recognition testing dataset

Overall, the model achieves an accuracy of 0.840, indicating a fairly high success rate in classifying cucumber diseases based on the combined features used. The confusion matrix is presented in Fig. 11 to facilitate the representation of classification results for each cucumber disease class.

Table 5. Classification performance results using combined features and RF on the cucumber disease recognition dataset

Label	Precision	Recall	F1-Score	Acc
Anthracnose	0.91	0.62	0.74	0.84
Belly Rot	0.92	1.00	0.96	
Fresh Cucumber	0.97	0.82	0.89	
Gummy Stem Blight	0.67	0.78	0.72	
Bacterial Wilt	0.90	0.99	0.94	
Downy Mildew	0.72	0.79	0.75	
Fresh Leaf	0.77	0.74	0.75	
Pythium Fruit Rot	0.92	0.99	0.95	

Table 6 presents the results of several tests for classifying cucumber plant diseases using different combinations of algorithms and techniques. In the first test, the VGG-16 model was used for cucumber plant disease classification. This model showed the lowest performance with a precision of 0.342, recall of 0.235, F1-Score of 0.193, and accuracy of 0.235. This indicates that using this model alone is not sufficiently effective for cucumber disease classification tasks. In the

Vol. 4, No. 3, 2024, pp. 1429-1445

second test, the combination of Local Binary Patterns (LBP) feature extraction with Random Forest (RF) classification showed a significant improvement, achieving a precision of 0.419, recall of 0.496, F1-Score of 0.487, and accuracy of 0.496. This shows that the combination of LBP textural features and RF classification is better than the first test for cucumber disease classification. The third test used a combination of VGG-16 visual feature extraction with RF classification. The results demonstrated a substantial improvement in all metrics, with a precision of 0.826, recall of 0.820, F1-Score of 0.816, and accuracy of 0.820. This shows the effectiveness of combining deep learning feature extraction using VGG-16 with conventional machine learning classification using RF. The fourth and final test used a combination of LBP textural feature extraction and VGG-16 visual feature extraction with RF classification. This technique showed the best results with a precision of 0.847, recall of 0.840, F1-Score of 0.838, and accuracy of 0.840. This provides evidence that the combined approach yields optimal classification performance for various cucumber plant diseases. Overall, these tests highlight the effectiveness of using a combination of feature extraction techniques and classification methods to improve the accuracy and reliability of cucumber disease classification.

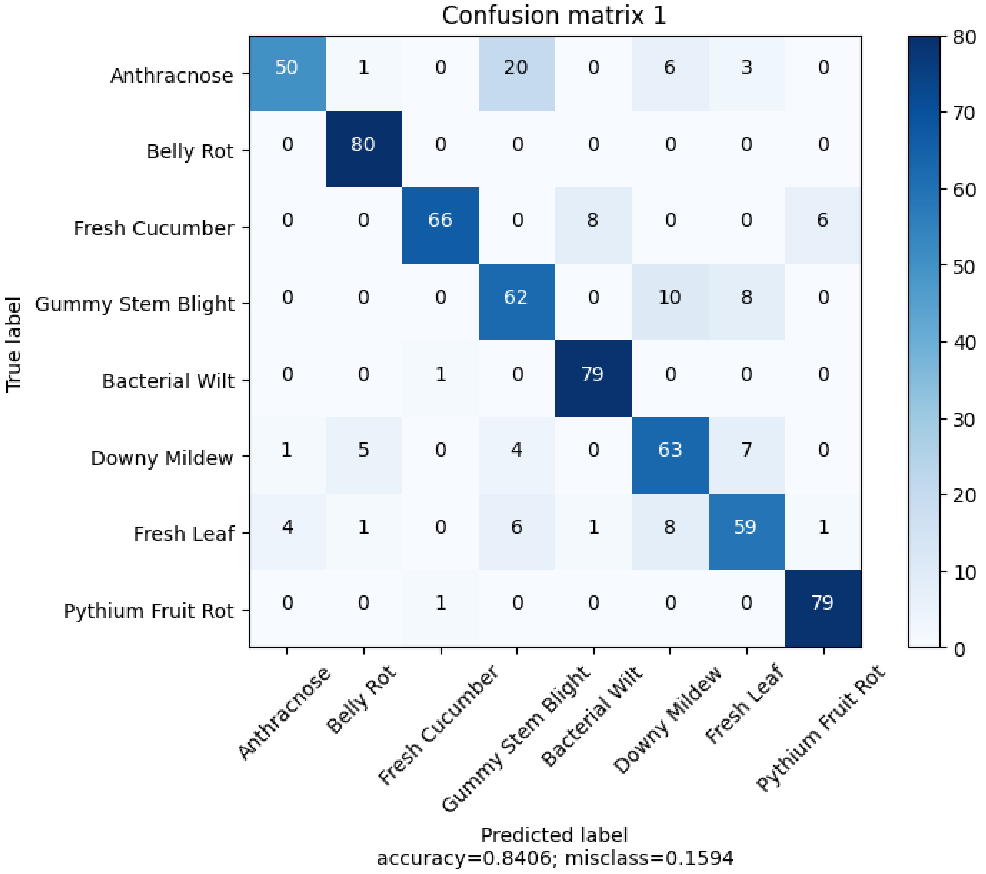

Fig. 11. Confusion matrix of the RF classification model with combined features on testing dataset cucumber disease recognition

Table 6. Comparison of model testing performance in cucumber plant disease classification

Scenario	Precision	Recall	F1-Score	Acc
VGG-16	0.91	0.62	0.74	0.235
LBP + RF	0.92	1.00	0.96	0.496
VGG-16 + RF	0.97	0.82	0.89	0.820
LBP+VGG-16+RF	0.67	0.78	0.72	0.840

4. Conclusion

The conclusion from the cucumber plant disease classification experiments shows that the proposed method, using a combination of Local Binary Patterns (LBP) texture features and VGG-16 visual features with the Random Forest method, provides the best performance. This method achieved a precision of 0.847, recall of 0.840, F1-Score of 0.838, and accuracy of 0.840, which are higher results compared to other methods or techniques tested with the same dataset, such as using VGG-16, the combination of LBP with Random Forest, and the combination of VGG-16 with Random Forest in cucumber disease classification. These results indicate that the combination of LBP's textural features with VGG-16's deep feature extraction capabilities can enhance the accuracy and effectiveness of cucumber plant disease classification. Therefore, this proposed method can be a more reliable solution for detecting and classifying cucumber plant diseases, thereby helping to improve plant production and reduce the negative impact of diseases on cucumber cultivation.

Author Contribution: All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] A. Amzeri *et al.*, "Assessment of Genetic Parameters for Vitamin A, Vitamin C, and TSS Content Results in Melon Line Crosses at Five Maturity Stages," *International Journal of Agronomy*, vol. 2022, no. 1, pp. 1-10, 2022, https://doi.org/10.1155/2022/3661952.

[2] M. Thilagavathi, S. Abirami, "Application of Image Processing in Diagnosing Guava Leaf Diseases," *International Journal of Scientific Research and Management (IJSRM)*, vol. 5, no. 7, pp. 5927-5933, 2017, https://doi.org/10.18535/ijsrm/v5i7.19.

[3] J. Zhang, Y. Rao, C. Man, Z. Jiang, and S. Li, "Identification of cucumber leaf diseases using deep learning and small sample size for agricultural Internet of Things," *International Journal of Distributed Sensor Networks*, vol. 17, no. 4, 2021, https://doi.org/10.1177/15501477211007407.

[4] N. Sultana, S. B. Shorif, M. Akter, and M. S. Uddin, "A dataset for successful recognition of cucumber diseases," *Data in Brief*, vol. 49, p. 109320, 2023, https://doi.org/10.1016/j.dib.2023.109320.

[5] M. A. Khan *et al.*, "Cucumber Leaf Diseases Recognition Using Multi Level Deep Entropy-ELM Feature Selection," *Applied Sciences*, vol. 12, no. 2, p. 593, 2022, https://doi.org/10.3390/app12020593.

[6] R. Thangaraj, S. Anandamurugan, and V. K. Kaliappan, "Automated tomato leaf disease classification using transfer learning-based deep convolution neural network," *Journal of Plant Diseases and Protection*, vol. 128, pp. 73-86, 2021, https://doi.org/10.1007/s41348-020-00403-0.

[7] L. Deng and D. Yu, "Deep learning: Methods and applications," *Foundations and Trends® in Signal Processing*, vol. 7, no. 3-4, pp. 197-387, 2013, https://doi.org/10.1561/2000000039.

[8] C. Wang, P. Du, H. Wu, J. Li, C. Zhao, and H. Zhu, "A cucumber leaf disease severity classification method based on the fusion of DeepLabV3+ and U-Net," *Computers and Electronics in Agriculture*, vol. 189, p. 106373, 2021, https://doi.org/10.1016/j.compag.2021.106373.

[9] S. P. Jena, S. Chakravarty, S. P. Sahoo, S. Nayak, S. K. Pradhan, B. K. Paikaray, "Automatic Leaf Diseases Detection and Classification of Cucumber Leaves Using Internet of Things and Machine Learning Models," *International Journal of Web and Grid Services*, vol. 19, no. 3, pp. 350-388, 2023, https://doi.org/10.1504/IJWGS.2023.133506.

[10] H. Ulutas and M. Emin Sahin, "Classification of cucumber leaf diseases on images using innovative ensembles of deep neural networks," *Journal of Electronic Imaging*, vol. 32, no. 05, p. 053040, 2023, https://doi.org/10.1117/1.JEI.32.5.053040.

[11] S. Zhang, S. Zhang, C. Zhang, X. Wang, Y. Shi, "Cucumber leaf disease identification with global pooling dilated convolutional neural network," *Computers and Electronics in Agriculture*, vol. 162, pp. 422-430, 2019, https://doi.org/10.1016/j.compag.2019.03.012.

[12] G. Pradhan, R. T. Sharma, A. K. Shah and V. Kukreja, "Deep Learning for Cucumber Agriculture: A Hybrid CNN-SVM System for Disease Identification," *2024 International Conference on Automation*

and Computation (AUTOCOM), pp. 57-61, 2024, https://doi.org/10.1109/AUTOCOM60220.2024.10486173.

[13] M. M. Ozguven, "Deep Learning Algorithms for Automatic Detection and Classification of Mildew Disease in Cucumber," *Fresenius Environmental Bulletin*, vol. 29, no. 8, pp. 7081-7087, 2020, https://avesis.ankara.edu.tr/yayin/28dec5d1-bb44-4d35-948e-10d189f0284a/deep-learning-algorithms-for-automatic-detection-and-classification-of-mildew-disease-in-cucumber.

[14] S. M. Omer, K. Z. Ghafoor, and S. K. Askar, "An Intelligent System for Cucumber Leaf Disease Diagnosis Based on the Tuned Convolutional Neural Network Algorithm," *Mobile Information Systems*, vol. 2022, no. 1, pp. 1-16, 2022, https://doi.org/10.1155/2022/8909121.

[15] P. Zhang, L. Yang, and D. Li, "EfficientNet-B4-Ranger: A novel method for greenhouse cucumber disease recognition under natural complex environment," *Computers and Electronics in Agriculture*, vol. 176, p. 105652, 2020, https://doi.org/10.1016/j.compag.2020.105652.

[16] M. M. Islam *et al.*, "DeepCrop: Deep learning-based crop disease prediction with web application," *Journal of Agriculture and Food Research*, vol. 14, p. 100764, 2023, https://doi.org/10.1016/j.jafr.2023.100764.

[17] A. Rachmad, M. Syarief, S. Rifka, F. Sonata, W. Setiawan, and E. M. S. Rochman, "Corn Leaf Disease Classification Using Local Binary Patterns (LBP) Feature Extraction," *Journal of Physics: Conference Series*, vol. 2406, no. 1, p. 012020, 2022, https://doi.org/10.1088/1742-6596/2406/1/012020.

[18] H. Nhat-Duc, T. Van-Duc, "Comparison of histogram-based gradient boosting classification machine, random Forest, and deep convolutional neural network for pavement raveling severity classification," *Automation in Construction*, vol. 148, p. 104767, 2023, https://doi.org/10.1016/j.autcon.2023.104767.

[19] H. Wu, L. Fang, Q. Yu, J. Yuan, and C. Yang, "Plant leaf identification based on shape and convolutional features," *Expert Systems with Applications*, vol. 219, p. 119626, 2023, https://doi.org/10.1016/j.eswa.2023.119626.

[20] A. H. Dias *et al.*, "Anthracnose susceptibility for grapevines with resistance loci to downy and powdery mildew in Southern Brazil," *Vitis - Journal of Grapevine Research*, vol. 61, no. 3, pp. 93-100, 2022, https://doi.org/10.5073/vitis.2022.61.93-100.

[21] E. Ahn, L. K. Prom, and C. Magill, "Multi-Trait Genome-Wide Association Studies of Sorghum bicolor Regarding Resistance to Anthracnose, Downy Mildew, Grain Mold and Head Smut," *Pathogens*, vol. 12, no. 6, p. 779, 2023, https://doi.org/10.3390/pathogens12060779.

[22] N. Sultana, M. Jahan, and M. S. Uddin, "An extensive dataset for successful recognition of fresh and rotten fruits," *Data in Brief*, vol. 44, p. 108552, 2022, https://doi.org/10.1016/j.dib.2022.108552.

[23] G. Jia, H. K. Lam, and Y. Xu, "Classification of COVID-19 chest X-Ray and CT images using a type of dynamic CNN modification method," *Computers in Biology and Medicine*, vol. 134, p. 104425, 2021, https://doi.org/10.1016/j.compbiomed.2021.104425.

[24] Q. Zheng, M. Yang, X. Tian, N. Jiang, and D. Wang, "A full stage data augmentation method in deep convolutional neural network for natural image classification," *Discrete Dynamics Nature and Society*, vol. 2020, no. 1, pp. 1-11, 2020, https://doi.org/10.1155/2020/4706576.

[25] L. Wang, Z. Q. Lin, and A. Wong, "COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images," *Scientific Reports*, vol. 10, p. 19549, 2020, https://doi.org/10.1038/s41598-020-76550-z.

[26] M. A. Chandra and S. S. Bedi, "Survey on SVM and their application in image classification," *International Journal of Information Technology*, vol. 13, pp. 1-13, 2021, https://doi.org/10.1007/s41870-017-0080-1.

[27] K. S. F. Azam, F. F. Riya, and S. T. Ahmed, "Leaf Detection Using Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), and Classifying with SVM Utilizing Claim Dataset," *Intelligent Data Communication Technologies and Internet of Things*, vol. 57, pp. 313-323, 2021, https://doi.org/10.1007/978-981-15-9509-7_27.

[28] I. Riaz, A. N. Ali, and H. Ibrahim, "Circular shift combination local binary pattern (CSC-LBP) method for dorsal finger crease classification," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 8, p. 101667, 2023, https://doi.org/10.1016/j.jksuci.2023.101667.

[29] D. Dominic, K. Balachandran, and C. Xavier, "A Mathematical Model to Explore the Details in an Image with Local Binary Pattern Distribution (LBP)," *Third Congress on Intelligent Systems*, pp. 439-451, 2023, https://doi.org/10.1007/978-981-19-9225-4_33.

[30] J. Y. R. Al-Awadi, H. K. Aljobouri, and A. M. Hasan, "MRI Brain Scans Classification Using Extreme Learning Machine on LBP and GLCM," *International journal of online and biomedical engineering*, vol. 19, no. 2, pp. 134-149, 2023, https://doi.org/10.3991/ijoe.v19i02.33987.

[31] S. H. Khaleefah, S. A. Mostafa, A. Mustapha, and M. F. Nasrudin, "Review of local binary pattern operators in image feature extraction," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 19, no. 1, pp. 23-31, 2020, http://doi.org/10.11591/ijeecs.v19.i1.pp23-31.

[32] V. D. Bharate, D. S. Chaudhari, and M. D. Chaudhari, "Emotion Recognition using Gamma Correction Technique Applied to HOG and LBP Features," *Advances in Signal and Data Processing*, vol. 307, pp 337-350, 2021, https://doi.org/10.1007/978-981-15-8391-9_25.

[33] P. Niyishaka and C. Bhagvati, "Image splicing detection technique based on Illumination-Reflectance model and LBP," *Multimedia Tools and Applications*, vol. 80, pp. 2161-2175, 2021, https://doi.org/10.1007/s11042-020-09707-7.

[34] N. Ahmed, N. Vij and Z. Kobti, "Comparative Study of LBP and HOG Feature Extraction Techniques for COVID-19 Pneumonia Classification," *2023 IEEE Symposium on Computers and Communications (ISCC)*, pp. 1-6, 2023, https://doi.org/10.1109/ISCC58397.2023.10217997.

[35] R. Yakkundimath, V. S. Jadhav, and B. S. Anami, "Automated classification of cervical cells using integrated VGG-16 CNN model," *International Journal of Medical Engineering and Informatics*, vol. 16, no. 2, pp. 185-197, 2024, https://doi.org/10.1504/IJMEI.2024.136966.

[36] T. A. T. K. Azaharan, A. K. Mahamad, S. Saon, Muladi, and S. W. Mudjanarko, "Investigation of VGG-16, ResNet-50 and AlexNet Performance for Brain Tumor Detection," *International journal of online and biomedical engineering*, vol. 19, no. 8, pp. 97-109, 2023, https://doi.org/10.3991/ijoe.v19i08.38619.

[37] J. A. Campos-Leal, A. Yee-Rendón and I. F. Vega-López, "Simplifying VGG-16 for Plant Species Identification," *IEEE Latin America Transactions*, vol. 20, no. 11, pp. 2330-2338, 2022, https://doi.org/10.1109/TLA.2022.9904757.

[38] J. Zhang, Q. Ma, X. Cui, H. Guo, K. Wang, D. Zhu, "High-throughput corn ear screening method based on two-pathway convolutional neural network," *Computers and Electronics in Agriculture*, vol. 175, p. 105525, 2020, https://doi.org/10.1016/j.compag.2020.105525.

[39] F. Liu, Y. Wang, F. -C. Wang, Y. -Z. Zhang and J. Lin, "Intelligent and Secure Content-Based Image Retrieval for Mobile Users," *IEEE Access*, vol. 7, pp. 119209-119222, 2019, https://doi.org/10.1109/ACCESS.2019.2935222.

[40] A. E. Minarno, B. Y. Sasongko, Y. Munarko, H. A. Nugroho, and Z. Ibrahim, "Convolutional Neural Network featuring VGG-16 Model for Glioma Classification," *International Journal on Informatics Visualization*, vol. 6, no. 3, pp. 660-666, 2022, https://dx.doi.org/10.30630/joiv.6.3.1230.

[41] S. Aulia and D. Rahmat, "Brain Tumor Identification Based on VGG-16 Architecture and CLAHE Method," *International Journal on Informatics Visualization*, vol. 6, no. 1, pp. 96-102, 2022, https://dx.doi.org/10.30630/joiv.6.1.864.

[42] R. K. Vamsee Kongara, V. Siva Charan Somasila, N. Revanth and R. D. Polagani, "Classification and Comparison Study of Rice Plant Diseases using Pre-Trained CNN Models," *2022 International Conference on Inventive Computation Technologies (ICICT)*, pp. 140-145, 2022, https://doi.org/10.1109/ICICT54344.2022.9850784.

[43] A. Tripathi, T. Goswami, S. K. Trivedi, and R. D. Sharma, "A multi class random forest (MCRF) model for classification of small plant peptides," *International Journal of Information Management Data Insights*, vol. 1, no. 2, p. 100029, 2021, https://doi.org/10.1016/j.jjimei.2021.100029.

[44] J. Tan, B. C. Sutanto, D. Marvelim, G. D. Pratama and S. Hasana, "Chili Plant Health Status Classification Using Random Forest," *2023 4th International Conference on Artificial Intelligence and Data Sciences (AiDAS)*, pp. 212-217, 2023, https://doi.org/10.1109/AiDAS60501.2023.10284701.

[45] S. K. Lakshmanaprabu, K. Shankar, M. Ilayaraja, A. W. Nasir, V. Vijayakumar, and N. Chilamkurti, "Random forest for big data classification in the internet of things using optimal features," *International Journal of Machine Learning and Cybernetics*, vol. 10, pp. 2609-2618, 2019, https://doi.org/10.1007/s13042-018-00916-z.

[46] G. F. Panggabean, A. Butar-butar, J. Siahaan, K. Sibarani and A. O. Silalahi, "Plants Classification Using Polarimetric SAR with the Random Forest Algorithm," *2023 29th International Conference on Telecommunications (ICT)*, pp. 1-6, 2023, https://doi.org/10.1109/ICT60153.2023.10374042.

[47] A. Pravin and C. Deepa, "Piper Plant Classification using Deep CNN Feature Extraction and Hyperparameter Tuned Random Forest Classification," *Transdisciplinary Journal of Engineering & Science*, vol. 13, pp. 233-258, 2022, https://doi.org/10.22545/2022/00202.

[48] K. A. Hemanthkumar and P. Shyamala Bharathi, "Improved Accuracy of Plant Leaf Classification using Random Forest Classifier over K-Nearest Neighbours," *2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, pp. 1-6, 2022, https://doi.org/10.1109/ICSES55317.2022.9914269.

[49] C. Martinez-Castillo, G. Astray, J. C. Mejuto, and J. Simal-Gandara, "Random Forest, Artificial Neural Network, and Support Vector Machine Models for Honey Classification," *eFood*, vol. 1, no. 1, pp. 69-76, 2020, https://doi.org/10.2991/efood.k.191004.001.

[50] J. Chen, Q. Li, H. Wang, and M. Deng, "A machine learning ensemble approach based on random forest and radial basis function neural network for risk evaluation of regional flood disaster: A case study of the yangtze river delta, China," *International Journal of Environmental Research and Public Health*, vol. 17, no. 1, p. 49, 2020, https://doi.org/10.3390/ijerph17010049.

[51] M. Bhagat, D. Kumar, and S. Kumar, "Bell pepper leaf disease classification with LBP and VGG-16 based fused features and RF classifier," *International Journal of Information Technology*, vol. 15, pp. 465-475, 2023, https://doi.org/10.1007/s41870-022-01136-z.