Benchmarking the LGBM, Random Forest, and XGBoost Models Based on Accuracy in Classifying Melon Leaf Disease

Chaerur Rozikin[1], Agus Buono[2], Sri Wahjuni[3], Chusnul Arif[4], Widodo[5]
Dept. Computer Science, IPB University, Bogor, Indonesia[1, 2, 3]
Dept. Informatics, University of Singaperbangsa, Karawang, Indonesia[1]
Dept. Civil and Environmental, IPB University, Bogor, Indonesia[4]
Dept. Plant Protection, IPB University, Bogor, Indonesia[5]

***Abstract*—Leaf diseases in melon plants cause losses for melon farmers. However, melon plants become less productive or even die. Downy mildew is a foliar disease that spreads rapidly in melon plants. Determining the level of downy mildew in melon leaves is important. Determining the level of downy mildew disease, farmers can carry out preventive treatment according to the severity level of downy mildew disease. This study aimed to create a classification model for the level of downy mildew disease on melon leaves using combined features and to compare the classification models, namely the LGBM, Random Forest, and XGBoost models. The combined features consist of colour, texture, Shannon entropy, and Canny edge features. The combined features are used as input for a classification model to predict the level of downy mildew leaf disease in melon plants. Model evaluation was carried out with three scenarios of data sharing: the first scenario, 90% training data and 10% test data; the second scenario, 80% training data and 20% test data; and the third scenario, 70% training data and 30% test data. The results of the evaluation of the model with the confusion matrix show that for the first and second scenarios, the highest accuracy was achieved by the Random Forest algorithm, with 72% and 73% accuracy, respectively. For the third scenario, the highest accuracy was obtained using the XGBoost algorithm.**

Keywords—Classification; Downy mildew; LGBM; disease level; melon leaves

I. INTRODUCTION

Melon is a fruit commodity with a high selling price, and many farmers cultivate it. Cultivating melons is difficult because there are many diseases associated with melon plants. Based on their causes, melon plant diseases are divided into three types: viruses, bacteria, and fungi [1]. One type of disease that infects the leaves is downy mildew. Downy mildew disease spreads very quickly, and if it is not controlled correctly, it can cause melon plants to die [2]. Determining the level of downy mildew disease on melon leaves is essential; this is done to determine the development of downy mildew disease that infects the leaves. In addition, by determining the level of downy mildew disease, farmers can carry out preventive treatments according to the level of development of downy mildew disease.

To overcome this problem, image processing (IP) and machine learning (ML) approaches can be used to classify the levels of downy mildew disease on melon leaves to help farmers treat downy mildew disease on melon leaves. IP and ML have been widely used to detect, identify, and classify leaf diseases [3]. This has been proven by many related scientific publications, including the classification of tomato leaf disease with public datasets using multiple feature extraction techniques, namely colour histograms, Hu Moments, Haralick and Local Binary Pattern features and classification models using Random Forest and decision tree classification; model evaluation results in decision tree classification with 90% accuracy and 94% Random Forest model [4]. Classification of banana leaf disease into four disease classes: healthy leaves, Sigatoka-infected leaves, Pestalotiopsis infected leaves, and Cordana-infected leaves using DenseNet and Inception. The result is that the model using the DenseNet method with an oversampling scheme is superior, with an accuracy of 84.73% [5]. Classification of grape leaf disease into two classes, namely healthy leaves and leaf spot (Cercospora), using Deep Forest, the evaluation results showed an accuracy of 96.25% [6]. Segmentation of cucumber leaf disease to detect cucumber leaf disease points using an improved saliency method and deep feature selection with an accuracy of 97.23% [7]. To detect and classify leaf diseases, feature extraction is required, and the feature extraction results are used as a dataset for classification [8]. In general, the feature extraction used by researchers is colour, texture, shape, and edge features [9]. The feature extraction results are then classified using a classification algorithm. Currently, many leaf disease classification algorithms have been developed, including SVM, Naive Bayes, Decision Tree, KNN, Random Forest, AdaBoost, Neural Network, Rule Base Classifier, Fuzzy Classifier [10] [11].

Classification of the severity of tea leaf blight using deep learning methods such as VGG16 deep networks is an interesting application in the fields of agriculture and pest management. In this case, the tea leaf blight was divided into two levels: mild and severe. The test results of the proposed model had an average accuracy of 84.5%, which was considered a good result [12]. To classify the severity of leaf diseases in tomato plants, deep learning using the ResNet architecture was used. This classification differentiates tomato leaves into three categories: healthy leaves, leaves affected by mild diseases, and leaves affected by severe diseases. The test results of this model showed an average accuracy of 88.2% [13]. To classify the severity of bacterial leaf streak leaf disease in rice plants, deep learning with BLSNet architecture was used. This classification divides the severity of the disease into five levels: level 0, leaves with no lesions; level 1, lesions less than 10%; level 2: Lesions 11-25%, level 3: Lesions 26-45%, level 4: Lesions 46-65%, and level 5, lesions > 65%. The BLSNet model test results showed an average accuracy of 98.2% [14].

A melon leaf image dataset was collected from a farmer's garden under natural conditions. Furthermore, the dataset is extracted for colour features by calculating the average colour value, standard deviation, skewness, texture, Shannon entropy, Canny edge, and colour histogram. Extraction of GLCM texture features with distances of 1, 3, and 5 and angles of 0^0, 45^0, 90^0, and 135^0 to obtain homogeneity, entropy, energy, contrast, and correlation values. Shannon entropy feature extraction is a feature extraction method used to obtain the value of the information acquisition between classes. Edge feature extraction values were obtained using Canny Edge.

The results of feature extraction were then combined into combined features. The combined features are used as inputs for the LGBM, Random Forest and XGBoost classification models. The model was evaluated using a confusion matrix and a scenario of dividing the training data and test data into three scenarios: 90% training data and 10% test data, 80% training data and 20% test data, and 70% training data and 30% test data. A confusion matrix was used to determine the best model for predicting the level of downy mildew disease in melon leaves.

The purpose of this study was to create a classification model for the levels of downy mildew disease on melon leaves, namely healthy leaves (DS), downy mildew level 1 (DM1), downy mildew level 2 (DM2), and downy mildew level 3 (DM3), using a combination of texture, colour, entropy Shannon, and edge features. The second was to compare the classification models, namely, LGBM, Random Forest, and XGBoost.

II. Research Method

A melon leaf image dataset was collected from a farmer's garden under natural conditions. Furthermore, the dataset is extracted for colour. This study has several stages; namely, data collection, pre-processing, feature extraction, classification, and evaluation of the model created using the confusion matrix (see Fig. 1).

A. Data Acquistion

Melon was planted from 17 October 2022 to 20 December 2022. The planting site was in Sukatani village, Sukatani sub-district, Purwakarta district, West Java. Melon plants were planted with as many as 30 plants, and for each melon plant, a sample of five leaves was selected randomly and then marked to differentiate. After 26 days after planting (HST), pictures of melon leaves were taken every two days from 11 November 2022 to 15 December 2022. Using a smartphone camera, smartphone specifications are shown in Table I.

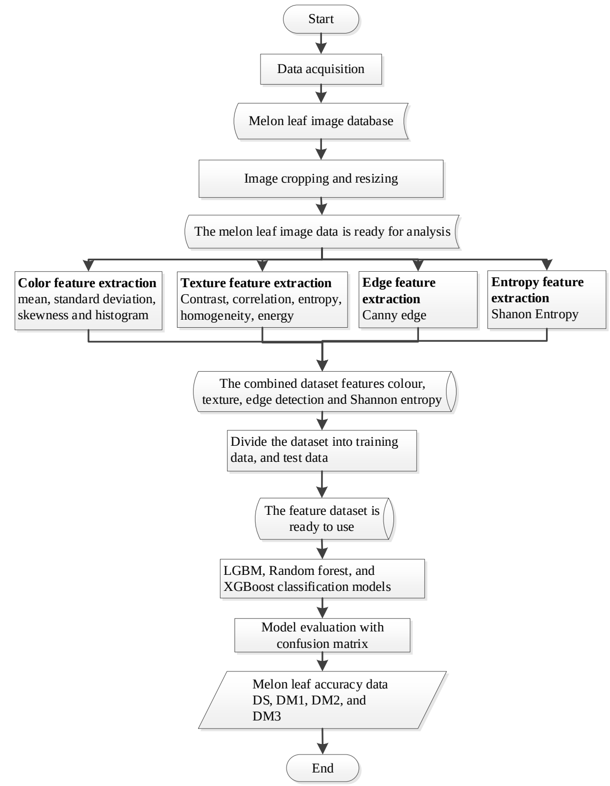

Fig. 1. Research stages.

TABLE I. Smartphone Specifications

No	Name	Description
1	Smartphone	Infinix Note 11 NFC
2	Camera resolution	50 MP, f/1.6, (wide), PDAF, 2 MP, f/2.4, (depth)
3	Operating system	Android 11

When photographing melon leaves, it is necessary to consider technical factors. The technical factors considered are listed in Table II.

TABLE II. Photographing Technique

No	Parameter (variabel konfirmasi)	Description
1	Shooting frequency	Once every two days
2	The distance between the leaf object and the camera	20 cm and 30 cm centred on the leaf object, and the leaf object does not exceed the camera frame
3	Angle position between camera and object (leaf)	Centered on the leaf object
4	Position between camera frame and object (leaf)	The leaf object is centred and does not exceed the smartphone camera frame.

B. Melon Leaf Image Dataset

The melon leaf data collected for a total of 1861 images were then labelled with the grading of melon leaf downy mildew disease. The labelling process involves a plant protection lab to determine the grading of downy mildew. The labelling process included healthy leaf labels (DS), downy mildew level 1 (DM1), downy mildew level 2 (DM2), and downy mildew level 3 (DM3). Table III shows the number of images from the DS, DM1, DM2, and DM3.

TABLE III. DETAILS THE AMOUNT OF DATA

No	Level of disease severity	Amount of data
1	Healthy Leaves (DS)	665 images
2	Downy Mildew level 1 (DM1)	449 images
3	Downy Mildew level 2 (DM2)	253 images
4	Downy Mildew level 3 (DM3)	494 images
Total amount of data		1861 images

Labelled downy mildew level 1 begins to show signs of disease until it spreads 20% on leaves, labelled downy mildew level 2, downy mildew disease begins to spread 20% - 30% on leaves, and downy mildew level 3 on melon leaves is more than 30% level of leaf disease. Fig. 2 shows (a) DS, (b) DM1, (c) DM2, and (d) DM3.

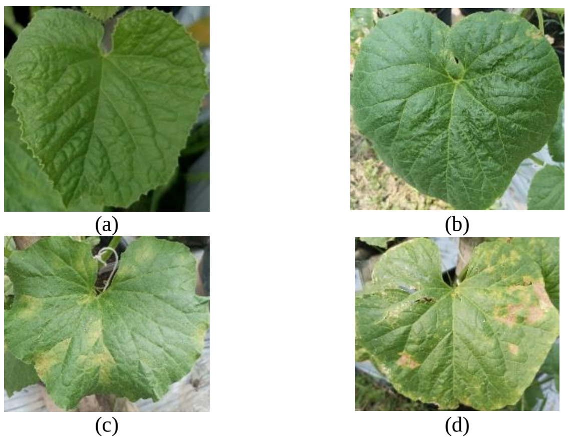

Fig. 2. Examples of melon leaves (a) DS, (b) DM1, (c) DM2, and (d) DM3.

C. Preprocess

Melon leaf image data were then preprocessed. The preprocessing involved cutting the image data and changing the size of the image data. Cutting the melon leaf image data aims to remove unwanted objects so that only melon leaf objects are produced. After cutting, the size was changed from the initial size of 2087 pixels × 2087 pixels to 128 pixels × 128 pixels. Changing the image data size aims to accelerate the computational process when performing feature extraction. Subsequent preprocessing changes the colour from RGB to grayscale.

D. Colour, Texture, and Edge Feature Extraction Melon Leaf Image Database

After preprocessing, the feature extraction process was performed. Feature extraction is performed to obtain the value from the image. In this study, the feature extraction included colour, texture, and shape features. Feature extraction obtains the average colour value in Eq. (1), standard deviation in Eq. (2), and skewness in Eq. (3) [15].

$$\mathrm{Mean} = \frac{1}{\mathrm{M\ X\ N}} \sum_{\mathrm{X}=1}^{\mathrm{M}} \sum_{\mathrm{y}=1}^{\mathrm{n}} \mathrm{M}_{\mathrm{xy}} \quad (1)$$

$$SD = \sqrt{\frac{1}{M\ X\ N} \sum_{x=1}^{M} \sum_{y=1}^{N} \left(M_{xy} - m\right)^2} \quad (2)$$

$$Skewness = \frac{\sum_{x=1}^{M} \sum_{y=1}^{N} \left(M_{xy} - m\right)^3}{(M\ x\ N)\ x\ SD^3} \quad (3)$$

The extracted colour feature values are blue mean, green average, red average, blue standard deviation, green standard deviation, red standard deviation, blue kurtosis, green kurtosis, red kurtosis, blue skewness, green skewness, and red skewness.

The subsequent colour feature extraction is a histogram obtained by extracting the histogram values using Eq. (4) [16].

$$h(r_k) = n_k \quad (4)$$

where is n_k the number of pixels with intensity level.

Texture feature extraction was carried out to obtain the distance and angle values by taking energy in Eq. (5), correlation values in Eq. (6), contrast in Eq. (7), entropy in Eq. (8), and homogeneity in Eq. (9) [4].

$$Energy = \sum_{i,j} (p(i,j))^2 \quad (5)$$

$$Correlation = \frac{\sum_i \sum_j (ij) P\ (i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (6)$$

$$Cont = \sum_{n=0}^{N_g - 1} N^2 \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i,j) \right\}, |i - j| = n \quad (2)$$

$$Entropy = -\sum_i \sum_j P(i,j) log\big(P(i,j)\big) \quad (8)$$

$$Homogeneity = \sum_{i,j} \frac{p(i,j)}{1 + |i - j|} \quad (9)$$

The extracted GLCM feature values with variations of distance 1, 3, 5 and angles 0^0, 45^0, 90^0, 135^0.

The extracted edge feature value is Canny. The canny edge feature removes noise by using a Gaussian filter with the following Eq. (10) [17].

$$G(x.y) = \frac{1}{2\pi\sigma^2} exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (10)$$

where y is the distance from the origin on the vertical axis, x is the distance from the origin on the horizontal axis, and σ is the standard deviation of the Gaussian distribution.

The next step is calculating the image gradient by calculating the gradient magnitude (G) and angle gradient (θ) with the Eq. (11) and Eq. (12).

$$G = \sqrt{\left(G_x^2 + G_y^2\right)} \quad (11)$$

$$\theta = tan\left(\frac{G_y}{G_x}\right) \quad (12)$$

G_x represents the horizontal and G_y vertical gradients, respectively.

Shannon stated that the measure of the amount of information $H(p)$ contained in a series of events $p_1 \ldots p_n$ must meet three conditions, namely, H must be continuous in p_i, secondly if all p_i have the same probability, so $p_i = \frac{1}{N}$, then H should be a monotonic rising function of N, and H must be additive [18]. Eq. (13) extracts Shannon entropy features.

$$H(p) = -k \sum_{i=1}^{N} p_i ln p_i \quad (13)$$

E. Combined Feature Dataset

The results of colour, texture, edge, and entropy feature extraction were in the form of colour, texture, edge, and entropy feature datasets. DFColour denotes the colour feature dataset, DFTexture denotes the texture feature dataset, DFEdge denotes the edge feature dataset, and DFEntropy denotes the entropy feature dataset. Then, the combined feature dataset can be formulated using Eq. (14).

$$DFCombined = DFColour \cup DFTekstur \cup DFEdge \cup DFEntropy \quad (3)$$

where DFCombined is the combined feature dataset. The total DFCombined has as many as 838 features.

F. Train Data, and Test Data

The DFCombined feature divides the data, namely, training data and test data. The scenarios for dividing the training and test data are shown in Table IV.

TABLE IV. DATA DIVIDING SCENARIO

Scenario	Training Data %	Test Data %
1	90	10
2	80	20
3	70	30

G. LGBM Random Forest, XGBoost

Furthermore, the combined DF dataset will be classified for grading downy mildew using three classification models, namely LGBM, Random Forest and XGBoost, according to the scenario of dividing the training data and test data. The results will be compared based on the accuracy values of the three models.

H. Evaluation of the Confusion Matrix Model

It is necessary to develop an evaluation model to measure the performance of the LGBM algorithm. The confusion matrix measures the performance of a classification algorithm by creating a detailed table of the amount of data that is classified correctly or incorrectly. The confusion matrix measures the accuracy, precision, recall, and F1 score [19]. The accuracy value is obtained from the amount of positive data predicted to be positive and the amount of harmful data predicted to be negative divided by the total amount of data, as shown in Eq. (15). The precision value is obtained from the number of opportunities for positive predictive data and the reality of the positive data, as shown in Eq. (16). The recall value is obtained from the number of positive data opportunities, and the prediction results are positive, as shown in Eq. (17). The F1 score was obtained from the recall and precision between the predicted and actual data as shown in Eq. (18).

$$Accuray = \frac{TP+TN}{TP+FP+FN+TN} \quad (15)$$

$$Precision = \frac{TP}{TP+FP} \quad (16)$$

$$Recall = \frac{TP}{TP+FN} \quad (17)$$

$$F1 = \frac{2\ x\ Recall\ x\ Precision}{Recall+Precision} \quad (18)$$

Where TP = True Positive, FP = False Negative, TN = True Negative, FN = False Negative F1 = F-Measure.

III. RESULTS AND DISCUSSIONS

The melon leaf disease image dataset consisted of healthy leaves (DS), downy mildew grade 1 (DM1), downy mildew grade 2 (DM2), and downy mildew grade 3 (DM3), which were divided into four classes: DS, DM1, DM2, and DM3.

A. Preprocess

Melon leaf image data were preprocessed. Pre-processing involved cutting the image data and changing the size of the image data. Cutting the melon leaf image data aims to remove unwanted objects such that only melon leaf objects are produced. The original and cropped images are shown in Fig. 3(a) original image and Fig. 3(b).

(a) Original image (b) Cropping image

Fig. 3. (a) Original image and (b) Cropping image.

After cutting, the size was changed from the initial size of 2087 pixels × 2087 pixels to 128 pixels × 128 pixels. Changing the image data size aims to speed up the computational process when performing feature extraction, and the subsequent preprocessing changes the colour from RGB to grayscale, as shown in Fig. 4(a) and Fig. 4(b).

(a) RGB image (b) Grayscale image

Fig. 4. Image conversion (a) RGB color (b) grayscale.

B. Feature Extraction

Colour feature extraction was performed to obtain the colour feature values. The extracted colour feature values are the average red, green, and blue colour values. Column Sample leaf number is sample of leaf, column meanR is the average value of red, Column meanG is the average value of green, and Column meanB is the average value of blue see Table V.

TABLE V. THE AVERAGE VALUE OF THE COLOURS R, G, AND B

Sample leaf number	meanR	meanG	meanB
1	130.8272	131.0041	130.8621
2	130.806	130.9863	130.8507
3	130.8316	131.0139	130.8828

Variants of colour features were extracted from red, green, and blue variances. Column Sample leaf number is sample of leaf, column VarianceR is the red variation value, column VarianceG is the green colour variation value, and column VarianceB is the green colour variation value (see Table VI).

TABLE VI. VARIANCE VALUES R, G, AND B

Sample leaf number	varianceR	varianceG	varianceB
1	3373.02	3491.309	3611.722
2	3371.538	3489.867	3610.48
3	3371.666	3489.709	3610.031

Skewness values were extracted to obtain the red, green, and blue skewness values. Column Sample leaf number is sample of leaf, column SkewnessR is the red skewness value, column SkewnessG is the green skewness value, and column SkewnessB is the blue skewness value (see Table VII).

TABLE VII. SKEWNESS VALUES R, G, AND B

Sample leaf number	skewnessR	skewnessG	skewnessB
1	-0.06524	-0.05881	-0.04562
2	-0.06546	-0.05898	-0.04581
3	-0.06603	-0.05949	-0.04637

The results of the colour feature extraction obtained the feature values of the nine features.

Subsequent colour feature extraction uses a histogram. A histogram was used to determine the distribution of colours in the image. The histogram feature value was obtained by calculating the histogram value of each pixel in the image. The histogram feature extraction resulted in 512 histogram features.

The subsequent feature extraction is a texture feature using GLCM. The values extracted from the GLCM texture features were the energy, correlation, dissimilarity, homogeneity, and contrast. Variation in GLCM values at distances of 1, 3, 5 and angle variations, namely0^0, 45^0, 90^0, 135^0. The following iteration 1 is used:

Iteration 1

```
distances = [1, 3, 5]
angles = [0, np.pi/4, np.pi/2, 3*np.pi/4]
    for d in distances:
        for a in angles:
            GLCM = graycomatrix(img, [d], [a])
            GLCM_Energy = graycoprops(GLCM,
            'energy')[0]
            df[f'Energy_d{d}_a{a}'] = GLCM_Energy
            GLCM_corr = graycoprops(GLCM,
            'correlation')[0]
            df[f'Corr_d{d}_a{a}'] = GLCM_corr
            GLCM_diss = graycoprops(GLCM,
            'dissimilarity')[0]
            df[f'Diss_sim_d{d}_a{a}'] = GLCM_diss
            GLCM_hom = graycoprops(GLCM,
            'homogeneity')[0]
            df[f'Homogen_d{d}_a{a}'] = GLCM_hom
            GLCM_contr = graycoprops(GLCM, contrast')[0]
            df[f'Contrast_d{d}_a{a}'] = GLCM_contr
```

Thus, the distance and angle texture features formed 60 texture features. Table VIII shows an example of the feature extraction results for a distance of 1 and an angle of 0^0. Column Sample leaf number is sample of leaf, column Energy_d1_0^0 is the energy value, column Corr_d1_0^0 is the coorelation value, column Diss_sim_d1_0^0 is the dissimiliarity value, column Homogen_d1_0^0 is the homogeinity value, column Contrast_d1_0^0 is the contrast value.

TABLE VIII. TEXTURE VALUE EXTRACTION

Sample leaf Number	Energy_d1_0^0	Corr_d1_0^0	Diss_sim_d1_0^0	Homogen_d1_0^0	Contrast_d1_0^0
1	0.01236	0.85204	17.16683	0.08106	659.80450
2	0.01156	0.76719	22.35888	0.06753	1088.4191
3	0.01532	0.76395	15.14720	0.08373	504.07068

The entropy feature is used to make it easier to deal with uncertainty in classifying diseases into DS, DM1, DM2, and DM3 classes, so that the presence of entropy can increase the value of information between classes in the classification so that it can improve prediction results by measuring the highest information gain. The following is an example of the entropy shanon feature value. Column Sample leaf number is sample of leaf, column Entropy is the entropy value which is shown in Table IX.

TABLE IX. EXTRACTION OF ENTROPY VALUES

Sample leaf Number	Entropy
1	7.542966
2	7.573822
3	7.020211

Edge feature extraction is used to determine points that experience a drastic change in brightness, typically in a line or curve, known as an edge. The edge feature values can be extracted using the Canny edge method [20]. The Canny edge feature extraction results are in the form of 256 features.

C. The Combined Features

The extraction results of colour, texture, entropy, and Canny features yielded 521 colour features, 60 texture features, one entropy feature, and 256 Canny edge features. Then, these features are combined so that the total number of features extracted from feature extraction is 838.

D. The Scenario of Dataset Division

After the combined features were obtained, they were used as datasets. The combined feature dataset was divided into training and test data for modelling. The scenario of dividing the dataset into training and test data was performed using three comparison scenarios for further details (see Table X).

TABLE X. THREE SCENARIOS FOR DIVIDING DATA

No	Level of disease severity	Total Data	Scenario 1		Scenario 2		Scenario 3	
			90%	10%	80%	20%	70%	30%
1	DS	665	598	67	532	133	465	200
2	DM1	449	404	45	359	90	314	135
3	DM2	253	227	26	202	51	177	76
4	DM3	494	444	50	404	99	345	149

E. LGBM, Random Forest, and XGBoost Models

1) Scenario 1: The dataset was divided into training and test data at a ratio of 90% training data and 10% test data. The classification results for LGBM, Random Forest, and XGBoost are shown in Fig. 5.

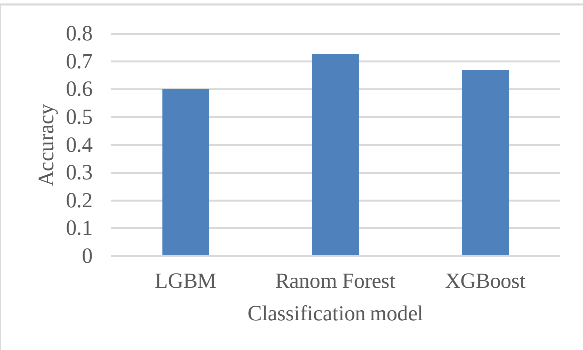

Fig. 5. Accuracy with 90% training data and 10% test data.

The results of the accuracy test for the classification of downy mildew with classes DS, DM1, DM2, and DM3, with a comparison of 90% training data and 10% test data, showed that the accuracy performance of the LGBM model was 60%, Random Forest was 72%, and XGBoost was 67%. The best accuracy performance of 72% was obtained by the Random Forest model.

2) Scenario 2: The dataset was divided into training and test data at a ratio of 80% training data and 20% test data. The classification results for LGBM, Random Forest, and XGBoost are shown in Fig. 6.

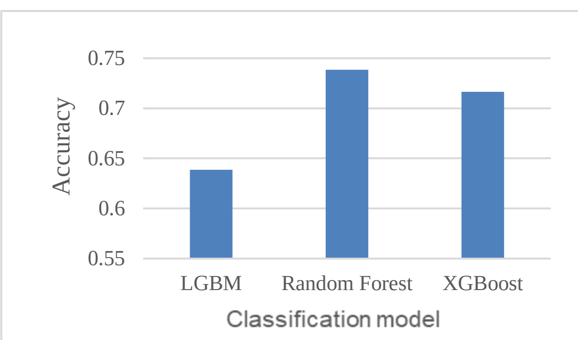

Fig. 6. Accuracy with 80% training data and 20% test data.

The results of the accuracy test for the classification of downy mildew with classes DS, DM1, DM2, and DM3, with a comparison of 80% training data and 20% test data, showed that the accuracy of the LGBM model was 63%, Random Forest 73%, and XGBoost 71%. The best accuracy performance of 73% was obtained by the Random Forest model.

3) Scenario 3: The dataset was divided into training and test data in a ratio of 70% training data and 30% test data. The classification results for LGBM, Random Forest, and XGBoost are shown in Fig. 7.

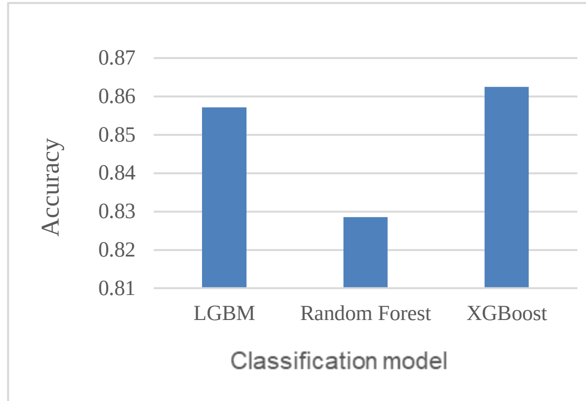

Fig. 7. Accuracy with 70% training data and 30% test data.

The results of the accuracy test for the classification of downy mildew disease with classes DS, DM1, DM2, and DM3, with a comparison of 70% training data and 30% test data, showed that the performance accuracy of the LGBM model was 85%, Random Forest was 82%, and XGBoost was 86%.

IV. DISCUSSION

The classification of melon leaf disease is divided into four classes, namely DS, DM1, DM2, and DM3, with feature extraction that uses colour, texture, Shannon entropy, and canny edge features. The results of feature extraction are in the form of combined features, with a total of 838 features. The combined features were then modelled using the LGBM, Random Forest, and XGBoost. Results of evaluating the confusion metrics with a comparison of training and test data scenarios, namely scenarios 1, 2, and 3. The results of testing the first scenario showed that the best accuracy performance model obtained by Random Forest was 72%. The results of testing scenario 2 with the best accuracy performance were obtained by Random Forest, that is, 73%. The results of testing scenario 3 showed that the best accuracy performance of 86% was obtained by the XGBoost model. The best average accuracy for testing the Random Forest, LGBM, and The use of more test data can be seen in scenarios 1, 2, and 3, and the accuracy performance of the Random Forest, LGBM and XGBoost models increased. The combined features of 838 features caused when the model was executed, the running process took a long time, and the accuracy of the model performance was at most 90%. While there were 838 combined features, there were still redundant and irrelevant features. The problem of redundant and irrelevant features requires further investigation.

V. Conclusion

Melon leaf disease, namely downy mildew, spreads very quickly; therefore, if it is not controlled properly, it can cause melon plants to die. Determining the level of downy mildew disease on melon leaves is important; this is done to determine the development of downy mildew disease that infects the leaves. One method to determine the level of downy mildew disease is to use a melon leaf disease classification model. The aim of this study was to create a classification model for downy mildew disease levels in melon leaves, namely DS, downy DM1, DM2, and DM3, using combined features, namely texture, colour, Shannon entropy, and edge features. The results of the extraction of colour, texture, entropy, and canny features resulted in 521 colour features, 60 texture features, 1 entropy feature, and 256 canny edge features; thus, the total number of combined features was 838 features. The model was evaluated using a confusion matrix, and a scenario was created by dividing the training data and test data into scenarios 1, 2, and 3. The results of accuracy testing for the classification of downy mildew disease with classes DS, DM1, DM2, and DM3 in scenario 1 showed that the accuracy performance of the LGBM model was 60%, Random Forest was 72%, and XGBoost was 67%. Accuracy comparison with scenario 2 and testing was carried out, with the accuracy performance results of the LGBM model being 63%, Random Forest 73%, and XGBoost 71%. Accuracy comparison with Secanrio 3 and model testing was carried out with the results of LGBM model accuracy performance testing of 85%, Random Forest 82%, and XGBoost 86%. Based on testing with scenarios 1, 2, and 3, the best average accuracy of testing the Random Forest, LGBM, and tested more data than in scenarios 1 and 2. The use of more test data can be observed in scenarios 1, 2, and 3, and the accuracy performance of the Random Forest, LGBM and XGBoost models increased.

References

[1] L. Sinaga, N. Zahara, P. Tanaman, F. Pertanian, and U. Bengkulu, “Kajian Patogen Penyebab Penyakit Pada Tanaman Melon (Cucumis melo L.) di Bengkulu,” Konserv. Hayati, vol. 18, no. 1, pp. 22–25, 2022.

[2] A. Ozbahce et al., “Impact of different rootstocks and limited water on yield and fruit quality of melon grown in a field naturally infested with Fusarium wilt,” Sci. Hortic. (Amsterdam)., vol. 289, no. April, p. 110482, 2021.

[3] S. Kaur, S. Pandey, and S. Goel, “Plants Disease Identification and Classification Through Leaf Images: A Survey,” Arch. Comput. Methods Eng., vol. 26, no. 2, pp. 507–530, 2019.

[4] J. Basavaiah and A. Arlene Anthony, “Tomato Leaf Disease Classification using Multiple Feature Extraction Techniques,” Wirel. Pers. Commun., vol. 115, no. 1, pp. 633–651, 2020.

[5] Andreanov Ridhovan, Aries Suharso, and Chaerur Rozikin, “Disease Detection in Banana Leaf Plants using DenseNet and Inception Method,” J. RESTI (Rekayasa Sist. dan Teknol. Informasi), vol. 6, no. 5, pp. 710–718, Oct. 2022.

[6] J. Arora, U. Agrawal, and P. Sharma, “Classification of Maize leaf diseases from healthy leaves using Deep Forest,” J. Artif. Intell. Syst., vol. 2, no. 1, pp. 14–26, 2020.

[7] M. A. Khan, T. Akram, M. Sharif, K. Javed, M. Raza, and T. Saba, “An automated system for cucumber leaf diseased spot detection and classification using improved saliency method and deep features selection,” Multimed. Tools Appl., vol. 79, no. 25–26, pp. 18627–18656, 2020.

[8] L. C. Ngugi, M. Abelwahab, and M. Abo-Zahhad, “Recent advances in image processing techniques for automated leaf pest and disease recognition – A review,” Inf. Process. Agric., vol. 8, no. 1, pp. 27–51, 2021.

[9] K. K. Thyagharajan and I. Kiruba Raji, “A Review of Visual Descriptors and Classification Techniques Used in Leaf Species Identification,” Arch. Comput. Methods Eng., vol. 26, no. 4, pp. 933–960, 2019.

[10] G. Dhingra, V. Kumar, and H. D. Joshi, “Study of digital image processing techniques for leaf disease detection and classification,” Multimed. Tools Appl., vol. 77, no. 15, pp. 19951–20000, Aug. 2018.

[11] V. K. Vishnoi, K. Kumar, and B. Kumar, Plant disease detection using computational intelligence and image processing, no. 0123456789. Springer Berlin Heidelberg, 2020.

[12] G. Hu, H. Wang, Y. Zhang, and M. Wan, “Detection and severity analysis of tea leaf blight based on deep learning,” Comput. Electr. Eng., vol. 90, no. January 2020, p. 107023, 2021.

[13] P. Wspanialy and M. Moussa, “A detection and severity estimation system for generic diseases of tomato greenhouse plants,” Comput. Electron. Agric., vol. 178, no. January, p. 105701, 2020.

[14] S. Chen et al., “An approach for rice bacterial leaf streak disease segmentation and disease severity estimation,” Agric., vol. 11, no. 5, 2021.

[15] S. Sachar and A. Kumar, “Survey of feature extraction and classification techniques to identify plant through leaves,” Expert Syst. Appl., vol. 167, p. 114181, 2021.

[16] N. Salem, H. Malik, and A. Shams, “Medical image enhancement based on histogram algorithms,” Procedia Comput. Sci., vol. 163, pp. 300–311, 2019.

[17] E. A. Sekehravani, E. Babulak, and M. Masoodi, “Implementing canny edge detection algorithm for noisy image,” Bull. Electr. Eng. Informatics, vol. 9, no. 4, pp. 1404–1410, 2020.

[18] P. Bromiley, “Shannon entropy, Renyi entropy, and information,” Stat. Inf. Ser., no. 2004, pp. 1–8, 2004.

[19] J. Xu, Y. Zhang, and D. Miao, “Three-way confusion matrix for classification: A measure driven view,” Inf. Sci. (Ny)., vol. 507, pp. 772–794, 2020.

[20] A. Sharma, A. Mittal, S. Singh, and V. Awatramani, “Hand Gesture Recognition using Image Processing and Feature Extraction Techniques,” Procedia Comput. Sci., vol. 173, no. 2019, pp. 181–190, 2020.