Informatics, Electrical and Electronics Engineering (Infotron)
Volume 6, Number 1, (2026) pp. 14-27
ISSN 2798-0197
Sinta 4 Accreditation – No SK. 177/ E/ KPT/ 2024
http://riset.unisma.ac.id/index.php/infotron/article/view/24722
DOI: 10.33474/infotron.v6i1.24722

Texture-Based Classification of Cocoa Pod Diseases and Pests Using GLCM and LDA on Field-Acquired Dataset

Redhitya Wempi Ansori a[1], Abd. Charis Fauza b[2], Agus Yulianto a[3]*
[a]*Department of Indonesian Language Education, Universitas Nahdlatul Ulama Blitar, Kota Blitar, Indonesia*
[b]*Department of Computer Science, Universitas Nahdlatul Ulama Blitar, Kota Blitar, Indonesia*
email: [1]redhityawempi@unublitar.ac.id, [2]abdcharis@unublitar.ac.id, [3]agusyulianto@unublitar.ac.id*

**Corresponding Author*

Article History:

Received 16 December 2025
Reviewed 14 January 2026
Revised 11 May 2026
Accepted 23 May 2026

A B S T R A CT

Cocoa pod diseases and pests are major factors that reduce cocoa yield quality and productivity. This study proposes a texture-based image classification approach for identifying cocoa pod conditions, including diseases and pest damage, using the Gray Level Co-Occurrence Matrix (GLCM) for feature extraction and Linear Discriminant Analysis (LDA) for classification. The original dataset, consisting of 420 field-acquired images across four classes (blackpod, healthy, helopeltis, and tupai), was expanded through augmentation to 1,260 images to improve model robustness. Model performance was evaluated using two complementary evaluation schemes: hold-out testing with active data augmentation and stratified k-fold cross validation. Experimental results show that the hold-out evaluation achieved an accuracy of 82.11%, indicating good practical performance when trained on augmented data. Meanwhile, stratified 5-fold cross validation produced an average accuracy of 77.70% with a standard deviation of 4.49%, providing a more conservative estimation of model generalization under limited original data conditions. The results demonstrate that classification performance is strongly influenced by the evaluation strategy and data availability. Overall, the combination of GLCM and LDA proves effective for texture-based classification of cocoa pod diseases and pests and can be implemented in a web-based application to support practical usage. However, further improvements in robustness and generalization are expected through the inclusion of more diverse original images and additional feature representations.

Keywords: Cocoa Pests, Cocoa Pod Diseases GLCM, LDA, Texture-Based Classification

1. Introduction

Over the past three years, cocoa (Theobroma cacao L.) production has experienced a global decline [1]. In Indonesia, data from the Central Bureau of Statistics (Badan Pusat Statistik, BPS) indicate that national cocoa production decreased annually from 767.4 thousand tons in 2018 to 706.5 thousand tons in 2021. As a consequence, by March 2024, cocoa prices reached a historical peak of USD 9,000 per metric ton, equivalent to IDR 142.39 million at an exchange rate of IDR 15,821 per dollar. For comparison, at the beginning of 2024, cocoa was traded at less than USD 4,200 per ton [2]. This significant price surge has raised concerns regarding the sustainability of cocoa-related industries in Indonesia [3]. As one of the country's leading plantation commodities, cocoa plays a crucial role as a raw material for food, pharmaceutical, and cosmetic industries, and contributes significantly to the national economy as the second-largest export commodity after palm oil, followed by rubber and coffee [4].

One of the main factors contributing to the decline in cocoa production worldwide is the high incidence of diseases and pest attacks [2], [5], [6]. These biological threats adversely affect both the quality and quantity of cocoa yields. Among the most destructive diseases is black pod disease, caused

Informatics, Electrical and Electronics Engineering (Infotron)
Volume 6, Number 1, (2026) pp. 14-27
ISSN 2798-0197

by Phytophthora palmivora, which is characterized by blackish-brown lesions on the pod surface and internal pulp decay [7]. This disease alone can result in yield losses of up to 30–50% [8], [9]. In addition to diseases, pest attacks such as those caused by insects and animals can also damage cocoa pods and reduce marketable yield. Currently, the identification of diseased or pest-damaged cocoa pods is largely conducted through manual visual inspection by farmers. This approach is subjective, prone to inconsistency, and often fails to provide early and accurate detection [9]. Delayed identification may allow diseases to spread to the stalks and bark, potentially causing stem cankers and more severe production losses [8].

To overcome these limitations, computational image-based approaches have been proposed to support objective and automated classification of cocoa pod conditions. In this study, a texture-based classification framework is developed using the Gray Level Co-Occurrence Matrix (GLCM) for feature extraction and Linear Discriminant Analysis (LDA) for classification. GLCM is a statistical technique that characterizes image texture by analyzing the spatial relationship between pixel intensity values. The process involves converting RGB images into grayscale and constructing co-occurrence matrices based on predefined distances and orientations [10]. Texture features were chosen because cocoa pod diseases, such as blackpod and helopeltis damage, exhibit distinct surface roughness and patterns that are more discriminative than color alone under varying light conditions. The resulting texture features provide a quantitative description of surface patterns associated with diseases and pest damage on cocoa pods.

LDA is a supervised classification method that aims to maximize the separability between different classes while minimizing intra-class variance [11]. By integrating GLCM-based texture features with LDA classification, this study seeks to classify multiple cocoa pod conditions, including healthy pods, disease-infected pods, and pest-damaged pods, in a more consistent and objective manner. The proposed approach is expected to support early identification of cocoa pod diseases and pests, thereby contributing to improved crop management and the long-term sustainability of cocoa production.

2. Method

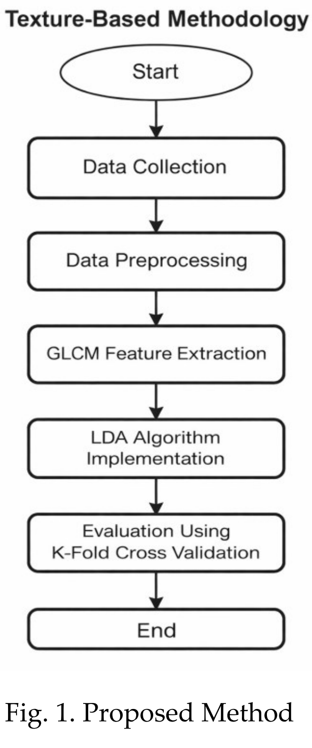

Fig. 1. Proposed Method

DOI: 10.33474/infotron.v6i1.24722

Informatics, Electrical and Electronics Engineering (Infotron)
Volume 6, Number 1, (2026) pp. 14-27
ISSN 2798-0197

As illustrated in Figure 1, the proposed research methodology follows a sequential process for classifying cocoa pod diseases and pest damage based on texture analysis. The process begins with data collection, where cocoa pod images representing different conditions are gathered. This step is followed by data preprocessing, which aims to standardize the images and improve their quality prior to analysis. Next, texture feature extraction is performed using the Gray Level Co-Occurrence Matrix (GLCM) to capture spatial relationships between pixel intensities that characterize surface patterns of cocoa pods. The extracted texture features are then used as input for the Linear Discriminant Analysis (LDA) algorithm, which performs classification by maximizing the separability between different classes.

To assess the performance of the proposed classification model, k-fold cross validation is employed as the primary evaluation method, allowing each data subset to be used alternately for training and testing. In addition, an alternative evaluation scheme using hold-out testing with active data augmentation is applied as a comparative approach. This comparison aims to analyze the effect of different evaluation strategies on model performance, particularly under conditions of limited original data.

2.1 Data Collection and Augmentation

The dataset used for training and testing in this study was obtained through direct image acquisition of cocoa pods at a cocoa plantation owned by PT. Kampung Coklat Blitar, East Java, Indonesia. Image collection was conducted under natural field conditions to capture real variations in cocoa pod appearance. The original dataset consisted of 420 images, with a distribution of approximately 105 images per class. The acquired images represent four classes: blackpod, healthy, helopeltis, and tupai.

To address the limitation of the small original dataset and to prevent overfitting, a data augmentation process was implemented. This process expanded the dataset from the initial 420 images to a total of 1,260 images, where 840 additional images were generated through rotation and flipping operations. Augmentation was strategically applied to increase the diversity of texture patterns while preserving the semantic characteristics of each cocoa pod class. From this augmented pool, the samples were distributed for model training and evaluation. Specifically, 100 images per class (totaling 400 images) were allocated for the training phase to maintain a balanced model performance, while the remaining images, including the original non-augmented samples, were utilized for testing and validation. This approach ensures that the model is trained on a sufficiently diverse dataset while being evaluated against authentic field conditions.

2.2 Preprocessing Stage

The preprocessing stage is the initial step in image processing and aims to prepare cocoa pod images for further analysis. In this study, preprocessing consists of image cropping to focus on the cocoa pod region, image resizing to standardize image dimensions, and grayscale conversion to transform RGB images into intensity-based images [12]. These steps ensure uniform image quality and enable effective texture feature extraction using GLCM.

2.3 GLCM Feature Extraction

GLCM feature extraction is performed to obtain texture features that are used for model training. In this study, several GLCM texture features are extracted, including correlation, contrast, energy, homogeneity, entropy, mean, and variance [13]. These features represent the spatial relationships between pixel intensities and characterize the texture patterns of cocoa pod surfaces. The workflow of the GLCM feature extraction process is illustrated in Figure 2.

DOI: 10.33474/infotron.v6i1.24722

Informatics, Electrical and Electronics Engineering (Infotron)
Volume 6, Number 1, (2026) pp. 14-27
ISSN 2798-0197

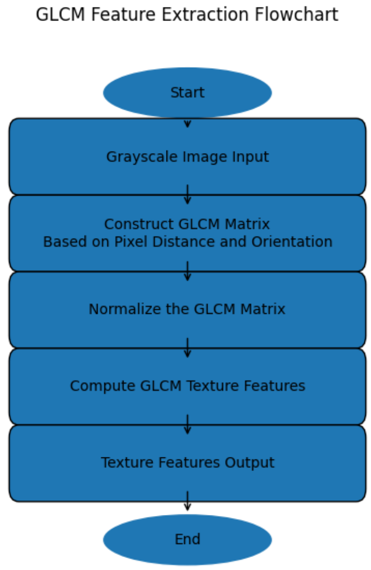

Fig. 2. Flowchart of GLCM

a. GLCM Matrix Construction. The GLCM matrix is constructed by calculating the frequency of occurrence of pixel pairs with specific intensity values based on a predefined pixel distance (d) and orientation angle (θ).
b. GLCM Matrix Normalization. Each element of the GLCM matrix is divided by the total number of pixel pairs in the matrix. This normalization process converts the GLCM into a probability distribution matrix.
c. Computation of GLCM. Several GLCM-based texture features are extracted in this study, including energy, contrast, correlation, homogeneity, mean, variance, and entropy. These features are defined as follows:
    a. Energy is used to measure the uniformity of pixel intensity distribution in an image and is calculated using Equation (1):

$$\text{Energy } = \sum\sum(GLCM(i,j))^2 \tag{1}$$

    b. Contrast measures the intensity difference between neighboring pixels in an image and is computed using Equation (2):

$$\text{Contrast } = \sum\sum|i-j|^2 GLCM(i,j) \tag{2}$$

    c. Correlation measures the linear relationship between the gray levels of pixel pairs and is defined by Equation (3):

$$\text{Correlation } = \sum\sum\frac{(i-\mu_i)(j-\mu_j)(GLCM(i,j))}{\sqrt{\sigma_i\sigma_j}} \tag{3}$$

    Where:
    $\mu_i = \sum(i)(GLCM(i,j))$
    $\mu_j = \sum(j).(GLCM(i,j))$
    $\sigma_i = \sum(i-\mu_i)^2.(GLCM(i,j))$
    $\sigma_j = \sum(j-\mu_j)^2.(GLCM(i,j))$
    d. Homogeneity measures the local uniformity of gray levels in an image and is calculated using Equation (4):

$$\text{Homogeneity } = \frac{\sum\sum(GLCM(i,j)}{1+(i-j)^2} \tag{4}$$

DOI: 10.33474/infotron.v6i1.24722

Informatics, Electrical and Electronics Engineering (Infotron)
Volume 6, Number 1, (2026) pp. 14-27
ISSN 2798-0197

e. Mean represents the average gray-level value of the image and is computed using Equation (5):

$$Mean = \frac{\mu_i + \mu_j}{2} \tag{5}$$

Where:
$\mu_i = \sum(i).(GLCM(i,j))$
$\mu_j = \sum(j).(GLCM(i,j))$

f. Variance indicates the degree of gray-level variation in an image and is calculated using Equation (6):

$$Variance = \frac{\sigma_i + \sigma_j}{2} \tag{6}$$

$\sigma_i = \sum(i - \mu_i)^2.(GLCM(i,j))$
$\sigma_j = \sum(j - \mu_j)^2.(GLCM(i,j))$

g. Entropy measures the randomness or irregularity of gray-level distribution in an image, reflecting its texture complexity. It is calculated using Equation (7):

$$\text{Entropy} = -\sum\sum GLCM\,(i,j).log(GLCM(i,j)) \tag{7}$$

2.4 Texture-Based Classification using LDA

The GLCM texture features extracted from cocoa pod images are used to train a machine learning model based on Linear Discriminant Analysis (LDA). As illustrated in Figure 3, the LDA classification process begins with the input of training and testing data, followed by class-wise data grouping. The algorithm then computes class means, the global mean, and covariance matrices to construct discriminant functions [14][15][16]. Finally, classification results are obtained by assigning each sample to the class with the highest discriminant function value.

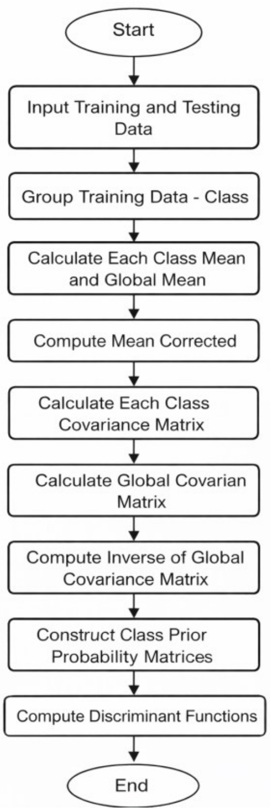

Fig. 3. Flowchart of LDA

Informatics, Electrical and Electronics Engineering (Infotron)
Volume 6, Number 1, (2026) pp. 14-27
ISSN 2798-0197

The computation steps of Linear Discriminant Analysis (LDA), as illustrated in **Figure 3**, are described as follows:

1. Input training data and testing data.
2. Group the training data into class-specific matrices denoted as X_i, where i represents the class index and the total number of classes.
3. Compute the mean matrix of each class (μ_i), and calculate the global mean of all data (μ).
4. Compute the mean-corrected data (X_i^0) by subtracting the global mean (μ) from each value in X_i
5. Compute the covariance matrix (C_i) from each class of (X_i) using the following equation:

$$C_i = \frac{(X_i^0)^T X_i^0}{n_i} \tag{8}$$

   Where:
   n_i = is the number of samples in class i
   $(X_i^0)^T$= the transpose of the mean-corrected data.

6. Compute the pooled within-group covariance matrix or global covariance matrix (C) using Equation (9):

$$\mathrm{C(a,b)} = \frac{1}{n}\sum_{i=1}^{g} n_i \,.\, C_i\,(a,b) \tag{9}$$

   Where:
   n = the total number of samples from all classes,
   n_i = the number of samples in class i,
   g = the number of classes,
   $C_i\,(a,b)$= represents the covariance value of the-i at row a dan column b

7. Compute the inverse of the global covariance matrix C (C^i).
8. Compute the prior probability of each class (P_i), using Equation (10)

$$P_i = \frac{n_i}{N} \tag{10}$$

   Where:
   n_i= The number of samples in class i,
   N = The total number of samples across all classes.

9. Compute the discriminant function (f_i), for each class using Equation (11). The classification result is determined by assigning the object to the class with the highest discriminant function value:

$$f_i = \mu_i C^{-1} X_k^T - \frac{1}{2}\mu_i C^{-1}\mu_i^T + ln(p_i) \tag{11}$$

   Where:
   f_i = discriminant function of class i
   μ_i = mean vector of class i
   μ_i^T = transpose of the mean vector of class i
   C^{-1} = inverse of covariance matrix,
   p_i = prior probability of class i,
   X_k^T = transpose of the input data.

2.5 Evaluation Stage

The final stage of this study is system evaluation, which aims to assess the performance of the proposed texture-based classification method. Two evaluation schemes are employed to provide a comprehensive analysis of model performance under different testing conditions. The first evaluation scheme uses hold-out testing with active data augmentation. In this approach, the dataset is divided into training and testing subsets. Data augmentation techniques are applied only to the training data to increase sample variability, while the testing data consist of original images. This scheme is used to evaluate the practical performance of the model when trained with enriched data.

DOI: 10.33474/infotron.v6i1.24722

Informatics, Electrical and Electronics Engineering (Infotron)
Volume 6, Number 1, (2026) pp. 14-27
ISSN 2798-
0197

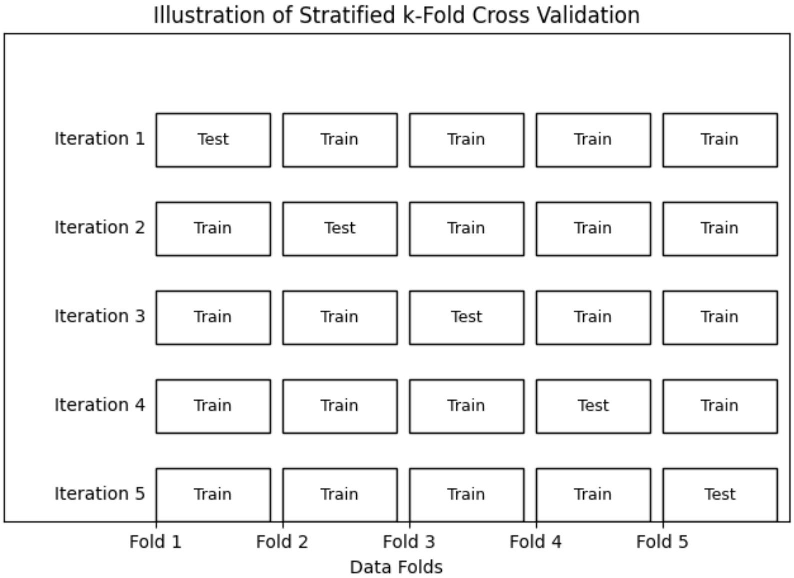

Fig. 4. Ilustration of k-Fold Cross Validation

The second evaluation scheme applies stratified k-fold cross validation, where the dataset is partitioned into several folds with balanced class distributions. In each iteration, one fold is used as testing data and the remaining folds are used for training. The evaluation results are obtained by averaging the performance across all folds, providing an estimate of model stability and robustness. By employing both evaluation schemes, this study analyzes the impact of different evaluation strategies on classification performance, particularly in the context of limited original data, and provides a more reliable assessment of the proposed GLCM–LDA classification approach.

3. Results and Discussion

3.1 Preprocessing Result

Data The cocoa pod images obtained from the field were processed to generate input data suitable for texture-based classification. This stage, referred to as preprocessing, aims to improve image quality, standardize image characteristics, and increase data variability prior to feature extraction. In this study, preprocessing consists of cropping, resizing, sharpening, data augmentation, and grayscale conversion, all of which were implemented using Python in the Google Colab environment.

Initially, each cocoa pod image was processed individually through cropping to isolate the region of interest, namely the cocoa pod surface, while removing irrelevant background elements such as leaves and surrounding objects. This step ensures that the extracted texture features accurately represent the cocoa pod surface. The cropped images were then subjected to resizing to standardize image dimensions and reduce computational complexity during subsequent processing stages. In addition, a sharpening operation was applied using image filtering techniques to enhance edge details and surface texture patterns, allowing disease and pest-related characteristics to be more clearly distinguished. Examples of the cropping, resizing, and sharpening processes are shown in Figure 5.

Informatics, Electrical and Electronics Engineering (Infotron)
Volume 6, Number 1, (2026) pp. 14-27
ISSN 2798-0197

Fig. 5. Examples of cropping, resizing, dan sharpening processes

To overcome the limitation of a small number of original images, a data augmentation stage was performed to generate additional training samples. Augmentation was implemented programmatically in Python by applying rotation and flipping operations to the original images. Rotation was performed by applying small-angle transformations to introduce orientation variability, while flipping was applied vertically to simulate different viewing perspectives. The results of the augmentation process are presented in Figure 6.

Fig. 6. The results of the augmentation process

Through the augmentation process, the dataset size increased from an initial 420 images to 1,260 images, where 840 additional images were generated from rotation and flipping operations. This augmentation strategy increases the diversity of texture patterns in the training data while preserving the semantic meaning of each cocoa pod class. Since the extraction of GLCM texture features requires grayscale images, the final step of preprocessing involved grayscale conversion and image resizing to a uniform dimension of 128x128 pixels. This process converts RGB images, which contain three color channels (Red, Green, and Blue), into single-channel grayscale images. Grayscale conversion simplifies pixel representation while retaining intensity information that is essential for texture analysis. The grayscale pixel matrix representation of cocoa pod images is illustrated in Figure 7.

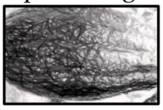

Fig. 7. The grayscale pixel matrix representation of cocoa pod images

Overall, the preprocessing stage produces standardized, enhanced, and sufficiently varied image data, which form a reliable foundation for subsequent GLCM feature extraction and LDA-based classification.

3.2 Results of GLCM Texture Feature Extraction

Table 1 presents the mean values of GLCM texture features extracted from cocoa pod images for each class (blackpod, healthy, helopeltis, and tupai). The results show distinct texture characteristics among the four classes, indicating that GLCM features are capable of capturing surface texture variations associated with healthy conditions, diseases, and pest damage.

Table 1. Result of GLCM

Feature / Class	Blackpod	Healthy	Helopeltis	Tupai
Contrast (0°)	353.98	197.86	338.82	435.30
Contrast (45°)	569.09	404.23	569.56	749.44
Contrast (90°)	531.30	374.82	491.88	705.61
Contrast (135°)	567.94	402.04	568.27	748.52
Dissimilarity (0°)	10.47	8.57	11.58	11.13
Dissimilarity (45°)	12.87	11.04	14.53	14.49
Dissimilarity (90°)	12.27	10.46	13.40	13.72
Dissimilarity (135°)	12.84	11.03	14.52	14.47
Homogeneity (0°)	0.279	0.274	0.221	0.258
Homogeneity (45°)	0.265	0.257	0.208	0.239
Energy (90°)	0.187	0.167	0.141	0.146

DOI: 10.33474/infotron.v6i1.24722

Informatics, Electrical and Electronics Engineering (Infotron)
Volume 6, Number 1, (2026) pp. 14-27
ISSN 2798-0197

Energy (135°)	0.187	0.167	0.141	0.146
Correlation (0°)	0.931	0.972	0.944	0.939
Correlation (45°)	0.890	0.950	0.912	0.895
Correlation (90°)	0.897	0.955	0.926	0.903
Correlation (135°)	0.890	0.951	0.912	0.895
ASM (0°)	0.0746	0.0621	0.0446	0.0438
ASM (45°)	0.0726	0.0602	0.0429	0.0420
ASM (90°)	0.0726	0.0602	0.0429	0.0421
ASM (135°)	0.0726	0.0602	0.0429	0.0420

Healthy cocoa pods consistently exhibit the lowest contrast and dissimilarity values across all orientations (e.g., contrast at 0° = 197.86 and dissimilarity at 0° = 8.57), reflecting smoother and more uniform surface textures. In contrast, the blackpod, helopeltis, and tupai classes show substantially higher contrast and dissimilarity values, with the tupai class presenting the highest contrast values across most orientations (e.g., contrast at 45° = 749.44). These results indicate more irregular surface patterns caused by disease infection and physical damage.

In terms of homogeneity and energy, healthy cocoa pods tend to have relatively higher values compared to pest-damaged classes, suggesting greater uniformity in pixel intensity distribution. Conversely, helopeltis and tupai classes show lower homogeneity and energy values, indicating more heterogeneous textures. This pattern is further supported by Angular Second Moment (ASM) values, where pest-damaged classes (helopeltis and tupai) consistently exhibit lower ASM values, reflecting increased texture randomness.

Correlation values across all classes remain relatively high, exceeding 0.89 in most orientations, which indicates strong spatial relationships between pixel intensities. Nevertheless, healthy cocoa pods achieve the highest correlation values (up to 0.97 at 0°), while disease- and pest-affected classes show slightly lower values. These subtle differences contribute to class separability when combined with other texture features.

Overall, the quantitative differences observed in contrast, dissimilarity, homogeneity, energy, ASM, and correlation values demonstrate that GLCM texture features provide discriminative information for distinguishing healthy cocoa pods from those affected by diseases and pests. These findings support the suitability of GLCM features as input for classification using Linear Discriminant Analysis (LDA).

3.2 Results of Texture-Based Classification using LDA

This subsection presents the results of texture-based classification of cocoa pod images using the Linear Discriminant Analysis (LDA) classifier, which was trained on texture features extracted through the Gray Level Co-Occurrence Matrix (GLCM). The classification results are demonstrated through the implementation of a web-based application, as illustrated in Figure 9 and Figure 10. Based on the extracted GLCM features, such as contrast, dissimilarity, homogeneity, energy, correlation, and ASM, the LDA model is capable of distinguishing cocoa pod surface conditions into four predefined classes, namely blackpod, healthy, helopeltis, and tupai. These texture features provide discriminative information that enables the LDA classifier to assign an input image to the most appropriate class based on its surface texture characteristics.

Figure 9 shows the initial interface of the developed web-based application, where users can upload a cocoa pod image for classification. The interface is designed to be simple and user-friendly, allowing practical usage by non-technical users. After an image is uploaded, the system automatically performs preprocessing and GLCM feature extraction before applying the trained LDA model.

Informatics, Electrical and Electronics Engineering (Infotron)
Volume 6, Number 1, (2026) pp. 14-27
ISSN 2798-0197

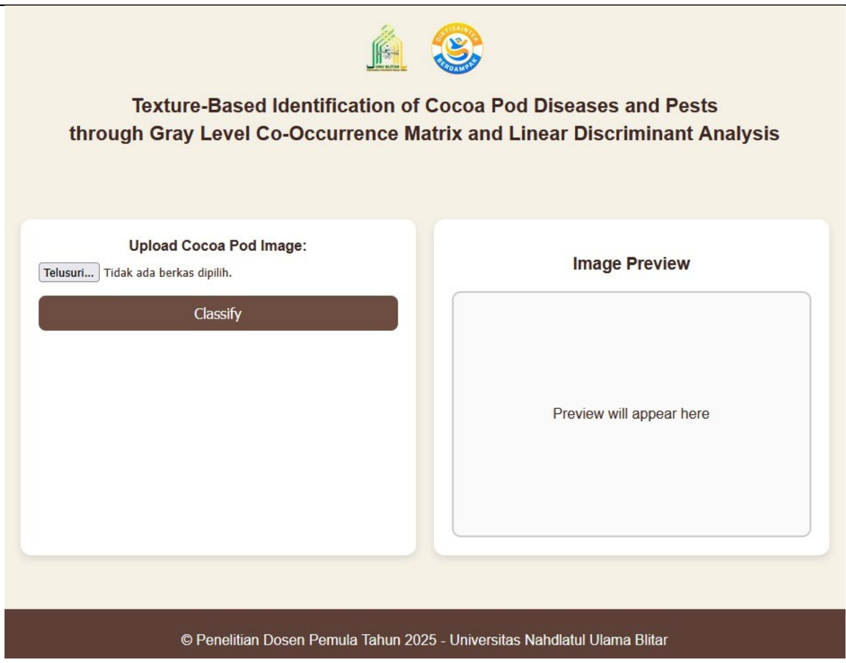

Fig. 8. Interface of the developed web-based application

Figure 10 presents an example of the classification result generated by the application. In this case, the uploaded cocoa pod image is successfully identified as helopeltis cacao, demonstrating the system's ability to recognize pest-related damage patterns. The classified result is displayed alongside the image preview, providing clear visual confirmation of the detected class.

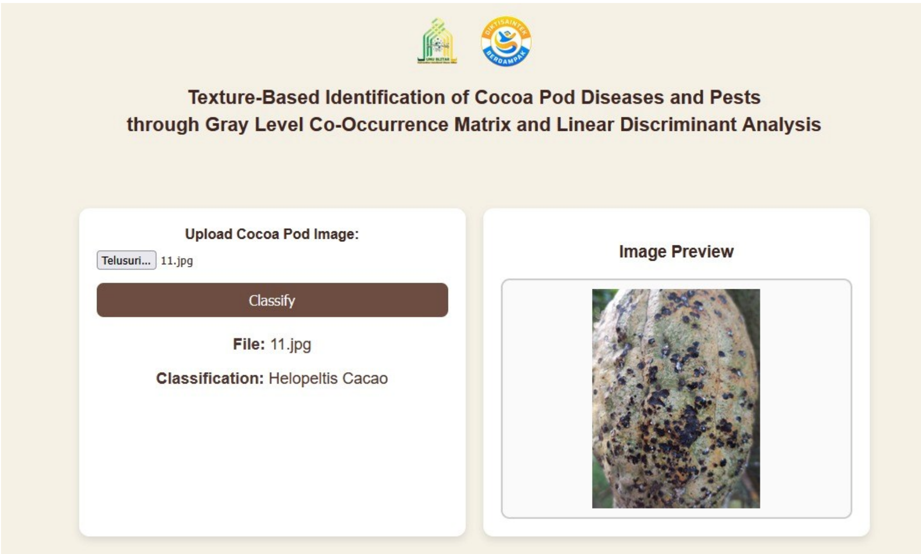

Fig. 9. Example of the classification result generated by the application

The results indicate that the proposed texture-based classification approach using GLCM and LDA can be effectively implemented in a web-based decision support system. The application is able to detect and classify cocoa pod conditions into four distinct classes based on texture information alone. This confirms the practical applicability of the proposed method for assisting cocoa pod disease and pest identification in real-world scenarios. Detailed performance evaluation and quantitative analysis of classification accuracy are discussed in the following subsection.

DOI: 10.33474/infotron.v6i1.24722

Informatics, Electrical and Electronics Engineering (Infotron)
Volume 6, Number 1, (2026) pp. 14-27
ISSN 2798-
0197

3.3 Performance Evaluation Using Hold-out with Active Augmentation

The performance of the proposed GLCM–LDA classification model was first evaluated using a hold-out testing scheme combined with active data augmentation. In this evaluation, the dataset was divided into training and testing subsets, where data augmentation was applied only to the training set to increase texture variability while avoiding data leakage. Based on the experimental results in Figure 10, the model achieved an overall classification accuracy of 82.11%, indicating good performance in distinguishing cocoa pod conditions.

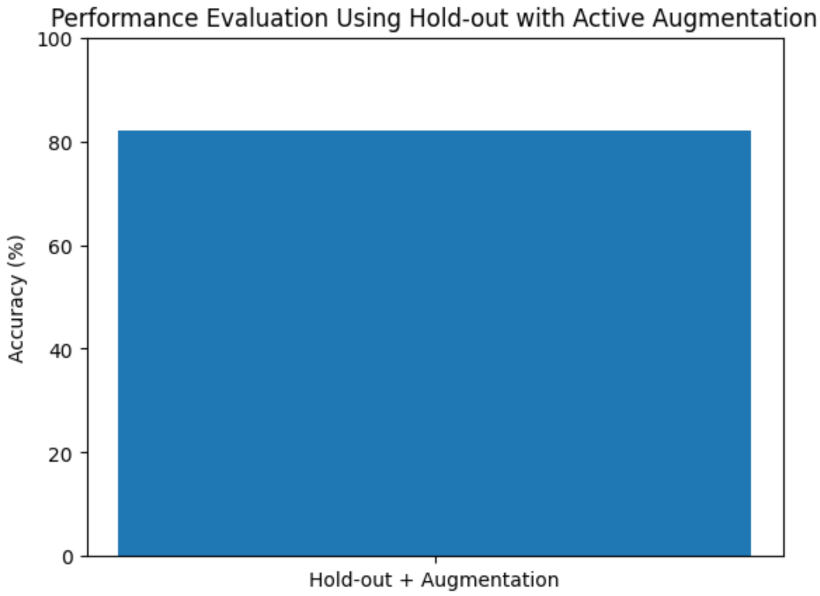

Fig. 10. Performance evaluation using Hold-out with Active Augmentation

Classification Report:

	precision	recall	f1-score	support
blackpod	0.79	0.74	0.77	31
healthy	0.68	0.83	0.75	30
helopeltis	0.89	0.77	0.83	31
tupai	0.97	0.94	0.95	31
accuracy			0.82	123
macro avg	0.83	0.82	0.82	123
weighted avg	0.83	0.82	0.82	123

Fig. 11. Classification Report

Based on Figure 11, the confusion matrix and classification report reveal that the model performs consistently across all four classes. The tupai class shows the highest classification performance, achieving a precision of 0.97 and a recall of 0.94, which indicates that pest damage caused by tupai exhibits distinctive texture characteristics that are well captured by GLCM features. Similarly, the helopeltis class achieves strong precision (0.89) and a recall of 0.77, demonstrating effective identification of pest-related surface damage. The blackpod class achieves a recall of 0.74, with some misclassification occurring primarily with the healthy and helopeltis classes. This can be attributed to similarities in surface texture patterns, particularly in early stages of disease infection. The healthy class shows a high recall value (0.83), indicating that most healthy cocoa pods are correctly identified, although its precision is slightly lower (0.68), suggesting occasional confusion with disease or pest-damaged classes. Overall, the macro-average F1-score of 0.82 reflects balanced classification performance across all classes. These results demonstrate that the combination of GLCM texture

DOI: 10.33474/infotron.v6i1.24722

Informatics, Electrical and Electronics Engineering (Infotron)
Volume 6, Number 1, (2026) pp. 14-27
ISSN 2798-0197

features and LDA classification performs effectively when trained with augmented data, providing a realistic estimation of model performance in practical applications.

3.4 Performance Evaluation Using Stratified k-Fold Cross Validation

To further assess the robustness and generalization ability of the proposed model, stratified 5-fold cross validation was employed as a comparative evaluation scheme. This method ensures that each fold preserves the original class distribution, providing a more conservative and reliable estimation of model performance under limited data conditions. Based on Figure 13, the experimental results show that the classification accuracy varies across folds, ranging from 70.73% to 84.15%, with an average accuracy of 77.70% and a standard deviation of 4.49%. The observed variation indicates that model performance is sensitive to the specific composition of training and testing data in each fold, which is expected given the relatively small number of original images per class.

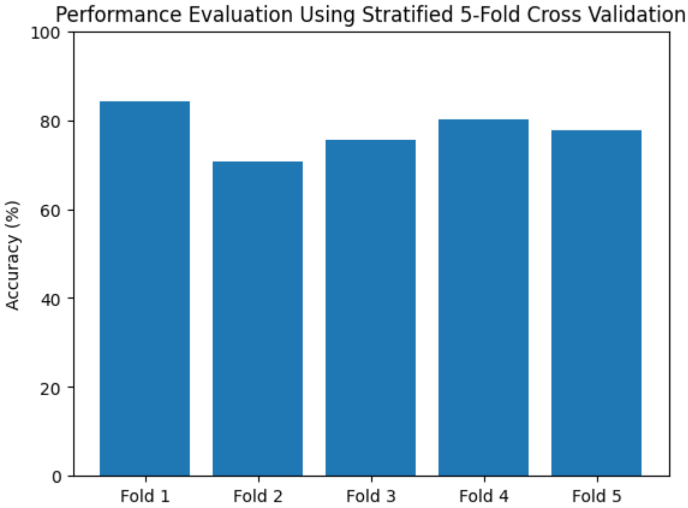

Fig. 12. Performance evaluation using Stratified 5-Fold Cross Validation

Compared to the hold-out evaluation with augmentation, the average accuracy obtained through stratified k-fold cross validation is lower. This difference reflects the absence of aggressive augmentation within each fold and highlights the impact of limited original data on model stability. Nevertheless, the cross-validation results provide valuable insight into the model's generalization capability and prevent overly optimistic performance estimation.

3.5 Comparative Discussion of Evaluation Schemes

The comparison between the two evaluation schemes demonstrates that the choice of evaluation strategy has a substantial influence on the reported performance of the proposed GLCM–LDA model. The hold-out evaluation with active augmentation achieved a higher accuracy of 82.11%, indicating that the increased variability introduced through augmentation enables the model to learn more discriminative texture patterns from limited original data. In contrast, stratified 5-fold cross validation resulted in a lower average accuracy of 77.70%, with a standard deviation of 4.49%, reflecting variations in performance across different folds. This evaluation scheme provides a more conservative estimation of model generalization by repeatedly testing the model on different data partitions, thereby exposing the sensitivity of the model to data scarcity and sample distribution.

These results suggest that performance obtained from a single hold-out split, even when combined with augmentation, may provide an optimistic estimate of classification capability. Conversely, stratified k-fold cross validation offers valuable insight into model robustness and stability under limited data conditions. Therefore, employing both evaluation strategies yields a more

DOI: 10.33474/infotron.v6i1.24722

Informatics, Electrical and Electronics Engineering (Infotron)
Volume 6, Number 1, (2026) pp. 14-27
ISSN 2798-0197

comprehensive and reliable understanding of model behavior, balancing practical performance with generalization assessment. However, it is important to acknowledge that the relatively small size of the original dataset remains a primary limitation, as it may increase the risk of overfitting despite the use of data augmentation. The discrepancy between hold-out and cross-validation results suggests that while the model learns well from augmented samples, its ability to generalize to completely unseen, non-augmented field data requires further validation with a larger collection of original images. Additionally, the current model relies exclusively on texture features; incorporating color and shape descriptors in future studies could potentially improve classification robustness against complex background noise and varying environmental conditions in cocoa plantations.

4. Conclusion

This study demonstrates that texture-based image analysis using the GLCM for feature extraction combined with LDA for classification can effectively identify cocoa pod diseases and pest damage. The evaluation results achieved an accuracy of 82.11% in hold-out testing and 77.70% in stratified 5-fold cross-validation, indicating that the proposed approach delivers reliable and consistent performance on field-acquired datasets. These findings highlight the potential of this method as an efficient alternative to manual inspection. However, this study has limitations, including a relatively small original dataset and reliance solely on texture features. Future work should involve extending the dataset, incorporating color and shape features, and experimenting with deep learning-based techniques to further improve adaptability under diverse field conditions.

5. Acknowledgment

The authors would like to express their sincere gratitude to the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia, Universitas Nahdlatul Ulama Blitar, and PT Kampung Coklat Blitar for their support in the implementation of this research through the PDP scheme.

7. References

[1] R. Erina, “Gagal Panen Bikin Harga Biji Kakao Melonjak, Dunia Terancam Krisis Cokelat,” rmol.id. Accessed: Mar. 24, 2024. [Online]. Available: https://rmol.id/bisnis/read/2024/03/14/613015/gagal-panen-bikin-harga-biji-kakao-melonjak-dunia-terancam-krisis-cokelat

[2] A. R. Respati and S. R. D. Setiawan, “Pecah Rekor, Harga Kakao Naik Dua Kali Lipat sejak Awal Tahun,” kompas.com. Accessed: Mar. 24, 2024. [Online]. Available: https://money.kompas.com/read/2024/03/23/110300626/pecah-rekor-harga-kakao-naik-dua-kali-lipat-sejak-awal-tahun?page=all#google_vignette

[3] D. Andi, “Harga Kakao Melambung Tinggi, Bagaimana Dampaknya Bagi Indonesia?,” industri.kontan.co.id. Accessed: Mar. 24, 2024. [Online]. Available: https://industri.kontan.co.id/news/harga-kakao-melambung-tinggi-bagaimana-dampaknya-bagi-indonesia

[4] V. Zikria, “Analisis Wilayah dan Kontribusi Kakao Terhadap Pembangunan Daerah di Kabupaten Parigi Moutong,” *J. Agriuma*, vol. 4, no. 1, pp. 22–30, 2022, doi: 10.31289/agri.v4i1.7173.

[5] S. Alim, P. P. Lestari, and R. Rusliyawati, “Sistem Pakar Diagnosa Penyakit Tanaman Kakao Menggunakan Metode Certainty Factor Pada Kelompok Tani Pt Olam Indonesia (Cocoa) Cabang Lampung,” *J. Data Min. dan Sist. Inf.*, vol. 1, no. 1, p. 26, 2020, doi: 10.33365/jdmsi.v1i1.798.

[6] Minarni, E. Novira, E. Yulianti, and G. Yoga Swara, “Sistem Pakar Diagnosa Penyakit dan Hama Tanaman Kakao Dengan Metode Certainty Factor,” *SATIN - Sains dan Teknol. Inf.*, vol. 8, no. 2, pp. 79–88, 2022, doi: 10.33372/stn.v8i2.901.

Informatics, Electrical and Electronics Engineering (Infotron)
Volume 6, Number 1, (2026) pp. 14-27
ISSN 2798-0197

[7] B. Oduro, O. O. Apenteng, and H. Nkansah, "Assessing the effect of fungicide treatment on cocoa black pod disease in Ghana: Insight from mathematical modeling," *Stat. Optim. Inf. Comput.*, vol. 8, no. 2, pp. 374–385, 2020, doi: 10.19139/soic-2310-5070-686.

[8] J. Obeng, R. T. Awuah, A. W. Kena, and B. Armooh, "Identification of an aspergillus isolate with potential for biocontrol of phytophthora palmivora, causal agent of black pod disease of cocoa," *J. Crop Prot.*, vol. 10, no. 2, pp. 375–390, 2021.

[9] D. S. Tan *et al.*, "AuToDiDAC: Automated Tool for Disease Detection and Assessment for Cacao Black Pod Rot," *Crop Prot.*, vol. 103, no. January 2017, pp. 98–102, 2018, doi: 10.1016/j.cropro.2017.09.017.

[10] A. Anggrestianingsih, A. W. Widodo, and M. T. Furqon, "Implementasi Metode Linear Discriminant Analysis (LDA) Untuk Klasifikasi Pengambilan Mata Kuliah Pilihan," vol. 3, no. 10. pp. 10337–10343, 2019.

[11] K. Yupie, S. Ajib, and U. W. M. Ibnu, "KLASIFIKASI BATIK KUDUS BERDASARKAN POLA MENGGUNAKAN K-NN DAN," *Semin. Nas. Has. Penenlitian dan Pengabdi. pada Masy. V Tahun 2020*, pp. 509–514, 2020.

[12] Muhathir, S. M Hamdani, and D. A. Larasati, "Wayang Image Classification Using SVM Method and GLCM Feature Extraction," *JITE (Journal Informatics Telecommun. Eng.*, vol. 4, no. 2, pp. 373–382, 2021.

[13] Y. Adhitya, S. W. Prakosa, M. Köppen, and J. S. Leu, "Feature extraction for cocoa bean digital image classification prediction for smart farming application," *Agronomy*, vol. 10, no. 11, 2020, doi: 10.3390/agronomy10111642.

[14] I. P. Winda, W. W. Agus, and A. R. Muh, "Pemanfaatan Ciri Gray Level Co-occurrence Matrix (GLCM) Untuk Deteksi Melasma Pada Citra Wajah," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 11, pp. 10402–10409, 2019.

[15] D. Grace, F. Christin, and H. Buan, "Penerapan Metode Runge-Kutta Orde 4 pada Pemodelan Penularan Penyakit Cacar Monyet," vol. 5, no. 2622, pp. 24–26, 2022.

[16] Nurhalimah, P. S. W. I Gede, and B. Fitri, "Klasifikasi Kain Songket Lombok Bedasarkan Fitur GLCM dan Moment Invariant dengan Teknik Pengklasifikasian Linear Discriminant Analysis," vol. 2, no. 2, pp. 173–183, 2020.

DOI: 10.33474/infotron.v6i1.24722