ICo SCIENCE: Journal Of Intelligent Computing And Advanced Data Science ***Original Paper***
VOL 1 Issue 2 (2026)

Banana Leaf Disease Classification Using HSV and LBP Feature Extraction with Support Vector Machine

Novriza Rahayu[1*], Farhan Muhammad[2], Sylvia Indri Yani[3], Agung Fadillah[4], Akbar Idaman[5]

[1, 2, 3, 4, 5] Department of Informatics, Universitas Satya Terra Bhinneka, Medan, Indonesia

ABSTRACT

Banana leaf diseases are one of the primary factors contributing to the decline in the quality and productivity of banana plants, as manual identification remains time-consuming, subjective, and prone to misclassification due to the similarity of symptoms among different diseases. This raises the question of whether a combination of Hue, Saturation, Value (HSV) and Local Binary Pattern (LBP) feature extraction can provide an effective alternative for banana leaf disease classification using a Support Vector Machine (SVM). While previous studies have relied on deep learning methods or combined HSV with Histogram of Oriented Gradients (HOG) features for this task, the combination of HSV and LBP for banana leaf disease classification remains largely unexplored. Using the Banana Leaf Spot Diseases (BananaLSD) dataset, comprising four classes (Cordana, Healthy, Pestalotiopsis, and Sigatoka), the original images were first divided into training and testing sets using an 80:20 ratio to prevent data leakage, after which data augmentation was applied exclusively to the training set. All images were center cropped before HSV and LBP features were extracted, combined, and classified using an SVM with a Radial Basis Function (RBF) kernel. On an independent test set of original, non-augmented images, the proposed model achieved an accuracy of 87.30%, precision of 89.06%, recall of 87.30%, and F1-score of 87.80%, with consistent results confirmed through Stratified Group 5-Fold Cross-Validation and an ablation study showing that the HSV and LBP combination outperformed either feature type alone. These findings indicate that combining HSV and LBP features offers a reliable, feature based alternative to deep learning for automated banana leaf disease identification.

ARTICLE INFO

Keywords:
Banana Leaf Disease Classification, Feature Extraction, HSV, Local Binary Pattern, Support Vector Machine

Histori:
Received: 04 July 2026
Accepted: 27 July 2026
Published: 29 July 2026

* Corresponding Author at Department of Informatics, Universitas Satya Terra Bhinneka, Medan, Indonesia.
E-mail address: novrizarahayu18@gmail.com (Novriza Rahayu)

1. INTRODUCTION

Bananas are one of the most economically valuable tropical horticultural commodities and play a significant role in Indonesia's food sector and agribusiness industry. They are widely cultivated due to their high consumption rate, adaptability to diverse environmental conditions, and rich nutritional content. In addition to serving as a staple food source, bananas are a major agricultural commodity that contributes substantially to both regional and national economies. According to a study on the development potential of banana cultivation, bananas rank as the fourth most important food crop in the world after wheat, rice, and maize (Sirappa, 2021). Furthermore, data from the Indonesian Central Bureau of Statistics (BPS) indicate that bananas recorded the highest fruit production in Indonesia in 2025, reaching 98,174,214.89 quintals (Badan Pusat Statistik, 2026). This high level of production highlights the importance of maintaining sustainable banana cultivation to ensure stable production and high-quality yields.

Despite their considerable economic potential, banana productivity remains constrained by various factors, one of the most significant being leaf diseases. Banana leaf diseases are commonly caused by fungal pathogens such as Sigatoka (Mycosphaerella spp.), Cordana (Cordana musae), and Pestalotiopsis spp. These pathogens can induce necrotic leaf spots, discoloration, morphological deformation, and premature leaf senescence. Such damage disrupts the photosynthetic process, thereby reducing both the quality and quantity of banana yields. Moreover, if these diseases are not detected and treated promptly, the severity of infection may increase, leading to substantial economic losses for farmers (Prasetyo & Utami, 2024). Therefore, the accurate detection and classification of banana leaf diseases are essential for supporting more effective disease management strategies.

In practice, banana leaf disease identification is still predominantly carried out using conventional visual inspection by farmers or agricultural experts (Muflich & Kusrini, 2026). However, this approach has several limitations, including high subjectivity, dependence on the observer's expertise, and limited efficiency when applied to large-scale plantations. In addition, manual identification is time-consuming and susceptible to misclassification due to the similarity of symptoms among different diseases (Bhuiyan et al., 2023). These challenges have encouraged the adoption of computer vision and machine learning techniques as alternative solutions for image-based plant disease identification. With the rapid advancement of digital imaging technology and computational methods, automated plant disease diagnosis systems have gained increasing attention because they offer improved efficiency and classification accuracy (Surnanto et al., 2026).

Based on these challenges, this study aims to apply a combination of Hue, Saturation, Value (HSV) and Local Binary Pattern (LBP) feature extraction methods with a Support Vector Machine (SVM) classifier to classify banana leaf diseases. The proposed approach is expected to provide an accurate and efficient feature-based alternative for banana leaf disease classification.

2. LITERATURE REVIEW

2.2. Banana Leaf Disease Classification

Numerous studies on banana leaf disease classification have employed deep learning techniques, particularly Convolutional Neural Networks (CNNs). One study utilized a MobileNetV2-based CNN model to classify four categories of banana leaf diseases and achieved an accuracy of 81% (Pratiwi & Prastika Wahyudi, 2026). Another study developed a banana leaf disease classification system using a CNN model and reported an accuracy of 82%, with a precision of 85%, recall of 83%, and an F1-score of 82% (Pratama et al., 2024). These findings indicate that CNN-based methods demonstrate promising performance in recognizing disease patterns from digital images of banana leaves.

More advanced deep learning approaches have also been implemented through transfer learning and fine-tuning techniques. A comparative study between ResNet50 and EfficientNetV2M revealed that EfficientNetV2M achieved an accuracy of 94.92%, outperforming ResNet50, which obtained an accuracy of 91.88% (Gimmas Manggara & Kalifia, 2025). In addition, another study combined CNN-based feature extraction using MobileNetV2 with a Support Vector Machine (SVM) classifier and achieved an accuracy of 97.8%, with precision, recall, and F1-score values ranging from 97% to 98% (Lase et al., 2025). Although deep learning methods generally provide high classification performance, they typically require substantial computational resources, longer training times, and large amounts of training data to achieve optimal performance. These requirements present significant challenges for deploying such models on devices with limited computational capabilities.

2.3. HSV Feature Extraction

The HSV (Hue, Saturation, Value) color space is a color model that represents images using three primary components. Hue represents the type of color, Saturation indicates the intensity or purity of the color, and Value reflects the brightness level. Compared with the RGB color space, HSV provides a clearer separation of color, intensity, and illumination information, making it more suitable for digital image analysis under varying lighting conditions, including images of plant leaves (Muchtar & Muchtar, 2024).

The effectiveness of the HSV color space has been demonstrated in various studies. A combination of HSV and a Convolutional Neural Network (CNN) for herbal leaf classification achieved an accuracy, precision, recall, and F1-score of 98% under well-lit conditions (Alfitriana Riska et al., 2023). Furthermore, HSV feature extraction combined with the K-Nearest Neighbors (KNN) algorithm for classifying the maturity levels of chayote fruit achieved an accuracy of up to 95% (Dzulhijjah et al., 2021). Another study also reported that the HSV color space produced superior image enhancement results for MRI images, outperforming both RGB and YCbCr, with an enhancement rate of 66%, compared to 31% for RGB and 27% for YCbCr (Hamzah & Wahyusari, 2023). Based on these findings, this study employs the HSV color space to represent the color characteristics of banana leaves prior to the classification process using SVM.

2.4. Local Binary Pattern

Local Binary Pattern (LBP) is a texture feature extraction method introduced by Ojala et al. It operates by comparing the intensity value of a center pixel with those of its neighboring pixels to generate a binary pattern that represents the texture characteristics of an image (Fatimah & Agustin, 2025). LBP offers several advantages, including low computational complexity, robustness to grayscale intensity variations, and the ability to effectively capture local texture patterns. These characteristics have made LBP one of the most widely used feature extraction methods in image classification research (Neneng et al., 2020).

The effectiveness of LBP has been demonstrated in numerous studies. The application of LBP features for oil palm leaf disease classification achieved an accuracy of up to 97% (Simanjuntak & Udjulawa, 2022). In addition, a study on grape leaf disease classification that combined LBP features with HSV histograms showed that the two feature extraction methods effectively represented both the texture and color characteristics of leaves, thereby improving the disease classification process (Safitri et al., 2024). Based on these advantages, this study employs LBP to extract the texture characteristics of banana leaves and combines them with HSV features as input for the Support Vector Machine (SVM) classifier.

2.5. Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning algorithm designed to solve classification problems by constructing an optimal hyperplane that separates different classes within the feature space (Maulana Alfaruq et al., 2023). The primary objective of SVM is to identify the decision boundary that provides the maximum margin between the closest data points, known as support vectors, from different classes. By maximizing this margin, SVM is able to improve the model's generalization capability, allowing it to achieve better performance when classifying previously unseen data. For non-linearly separable data, SVM employs kernel functions to transform the original data into a higher-dimensional feature space, enabling more effective class separation while preserving the underlying relationships among the data points (Permata Aulia et al., 2021).

SVM demonstrates strong performance in classifying high-dimensional data while maintaining high classification accuracy even when the amount of training data is relatively limited (Permata Aulia et al., 2021). This advantage makes SVM particularly suitable for image classification tasks, where each image is represented by a large number of extracted features. Furthermore, the algorithm is capable of handling complex feature distributions by identifying the optimal decision boundary that minimizes classification errors. These characteristics have made SVM a widely adopted classification algorithm in feature extraction–based studies, especially those involving image analysis and pattern recognition. Therefore, this study employs SVM as the classification method to process the extracted HSV and LBP features for banana leaf disease identification, as the combination of these features provides complementary color and texture information that can support accurate disease classification.

2.6. Kombinasi HSV dan LBP

The combination of color and texture features has been widely adopted to improve image classification performance because these two types of features provide complementary information about the visual characteristics of an object. Color features describe variations in appearance that are often associated with changes caused by disease, whereas texture features capture differences in surface patterns that may not be distinguishable through color information alone. By integrating both types of features, the resulting feature representation becomes more informative and capable of describing the condition of plant leaves more comprehensively. HSV effectively represents color characteristics, while LBP is capable of extracting local texture patterns, allowing the two methods to complement each other in capturing the visual characteristics of an object. Compared with feature extraction methods such as Histogram of Oriented Gradients (HOG), which primarily focus on shape and edge orientation, LBP is considered more effective in capturing the texture patterns of leaf disease lesions (Surnanto et al., 2026).

Several studies have demonstrated that combining HSV and LBP features enhances the representation of image characteristics and contributes to improved classification performance across various agricultural applications. In tomato leaf disease classification, the integration of color and texture features achieved an accuracy of up to 100% (Cahyani Putri et al., 2025). Furthermore, the combination of HSV and LBP has produced promising classification results for determining the roasting level of coffee beans (Fahmi Wibawa et al., 2021) and has also been shown to effectively represent both color and texture characteristics in grape leaf disease classification (Safitri et al., 2024). These findings indicate that the complementary strengths of HSV and LBP can be applied successfully in different image classification tasks involving variations in color and surface texture. Based on these findings, this study combines HSV and LBP features to obtain a more comprehensive representation of the color and texture characteristics of banana leaves before performing classification using SVM, with the expectation of improving the model's ability to distinguish among different banana leaf disease categories.

2.7. Research Gap

Based on previous studies, banana leaf disease classification has been predominantly conducted using Convolutional Neural Networks (CNNs) and other deep learning approaches. Although these methods generally achieve high classification accuracy, they require substantial computational resources, making them less suitable for deployment on devices with limited hardware capabilities (Pratiwi & Prastika Wahyudi, 2026) (Pratama et al., 2024) (Gimmas Manggara & Kalifia, 2025). In contrast, studies employing conventional feature extraction techniques for banana leaf disease classification remain limited, with existing research primarily focusing on the combination of HSV and Histogram of Oriented Gradients (HOG) features (Surnanto et al., 2026). To the best of our knowledge, the application of a combined HSV and Local Binary Pattern (LBP) feature extraction approach for banana leaf disease classification has received little attention. HSV effectively represents color characteristics by separating hue, saturation, and brightness components, whereas LBP efficiently captures local texture patterns that serve as distinctive indicators of leaf diseases (Hayati, 2023). Therefore, integrating these two feature extraction methods has the potential to provide a more comprehensive image representation. However, a direct computational cost comparison was not conducted in this study, and this potential efficiency advantage remains to be empirically verified in future work.

To address this research gap, the present study proposes a banana leaf disease classification approach that combines HSV and Local Binary Pattern (LBP) feature extraction with a Support Vector Machine (SVM) classifier. SVM was selected because of its ability to handle high-dimensional data, deliver strong classification performance, and provide computational efficiency that is well suited for feature extraction based classification systems (Nasution et al., 2024). Through this approach, the study aims to develop an accurate, feature based banana leaf disease classification system as an alternative to deep learning-based methods.

3. METHOD, DATA, AND ANALYSIS

3.1. Research Workflow

The research was conducted through a systematic sequence of stages to ensure that the entire analysis process was carried out in accordance with the objectives of the study. The overall research workflow is illustrated in Figure 1.

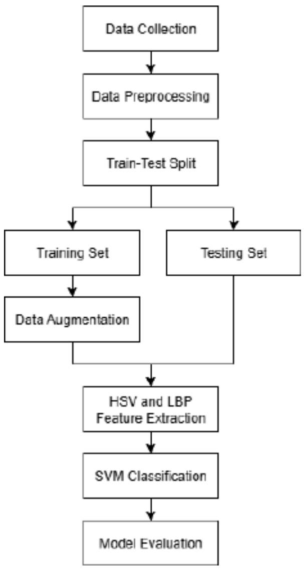

Figure 1. Research Workflow

3.2. Data Collection

The initial stage of this study involved collecting a banana leaf disease dataset that served as the primary data source for the image classification process. This stage is essential because the quality and diversity of the dataset directly influence the effectiveness of feature extraction and the overall performance of the classification model. Therefore, selecting an appropriate dataset is a crucial step in ensuring that the research objectives can be achieved reliably.

The dataset used in this study was the Banana Leaf Spot Diseases (BananaLSD) dataset, originally collected in Bangladesh, which is publicly available through the Mendeley Data repository (Arman et al., 2023) and was accessed via the Kaggle platform (https://www.kaggle.com/datasets/shifatearman/bananalsd). The dataset was selected because it contains images of both healthy banana leaves and leaves affected by different diseases, making it suitable for developing and evaluating automated disease classification methods. Furthermore, the BananaLSD dataset has been widely adopted in previous studies, allowing the proposed approach to be evaluated using a publicly accessible benchmark dataset.

Although the BananaLSD dataset provides both OriginalSet and AugmentedSet folders, only the OriginalSet was used in this study. The OriginalSet contains 937 original banana leaf images distributed across four classes, namely Cordana with 162 images, Healthy with 129 images, Pestalotiopsis with 173 images, and Sigatoka with 473 images.. To prevent data leakage during model evaluation, the original images were first divided into training and testing sets using an 80:20 split. To ensure reproducibility, the train test split, the random selection of rotation angles during augmentation, and the model training process were all conducted using a fixed random seed (random_state = 42). Data augmentation was subsequently performed only on the training set, while the testing set consisted exclusively of original, non-augmented images. This experimental design ensured that model performance was evaluated using completely independent test data.

3.3. Data Preprocessing

In this study, image preprocessing was performed using the center crop technique before the feature extraction stage. Image preprocessing plays an important role in improving the quality and consistency of the input data by minimizing unnecessary visual information that may interfere with the classification process. By applying the same preprocessing procedure to all images, the extracted features become more representative of the actual characteristics of the banana leaves.

The center crop technique was employed to focus each image on the primary region of interest, namely the banana leaf, while removing less relevant areas around the image boundaries, such as the background, shadows, and surrounding objects that do not contribute to disease identification (Asykurani Abdillah et al., 2023). These irrelevant regions may introduce unwanted variations in color and texture, potentially affecting the feature extraction process and reducing classification performance. Consequently, removing these regions enables the model to concentrate on visual characteristics that are more closely associated with banana leaf disease symptoms.

In the proposed preprocessing procedure, 85% of the central region of each image was retained, while the outer 15% was discarded. This crop ratio was selected to preserve the main leaf area while reducing the influence of irrelevant background information. The same center crop operation was applied to both the training and testing images before feature extraction to ensure a consistent preprocessing procedure throughout the experiment. As a result, the extracted color and texture features are expected to more accurately represent the characteristics of the banana leaves rather than the surrounding environment.

3.4. Data Augmentation

Data augmentation was performed to increase the quantity and diversity of the training images, enabling the model to learn a broader range of banana leaf characteristics while reducing the risk of overfitting. To prevent data leakage, the original dataset was first divided into training and testing sets using an 80:20 split. Data augmentation was then applied exclusively to the training set, whereas the testing set consisted entirely of original, non-augmented images. This experimental design ensured that model evaluation was conducted using completely independent test data.

The augmentation process consisted of three geometric transformation techniques: horizontal flipping, rotation, and shear transformation. Horizontal flipping generated mirrored versions of the images to introduce positional variations. Rotation was performed by randomly selecting one of four rotation angles ($-20°$, $-10°$, $10°$, and $20°$) to improve the model's ability to recognize banana leaves captured from different orientations. In addition, shear transformation was applied using a random shear factor within the range of -0.2 to 0.2, enabling the model to better accommodate geometric distortions that may occur during image acquisition. These augmentation techniques preserved the disease-related visual characteristics of the banana leaves while increasing the diversity of the training samples (Sanjaya & Ayub, 2020).

Following augmentation, additional images were generated until each class contained 400 training images, resulting in a balanced training dataset of 1,600 images across the four classes. Meanwhile, the testing dataset remained unchanged and consisted only of the original images obtained after the initial train–test split. This augmentation strategy improves the diversity of the training data while maintaining an unbiased evaluation using independent, non-augmented test images.

3.5. HSV and LBP Feature Extraction

The feature extraction stage was performed to obtain representative color and texture information from the banana leaf images for the classification process. This stage plays an important role because the extracted features serve as the input to the Support Vector Machine (SVM) classifier, enabling it to distinguish healthy leaves from diseased ones based on their visual characteristics. In this study, feature extraction was carried out by combining the HSV (Hue, Saturation, Value) color space and the Local Binary Pattern (LBP) texture descriptor to provide complementary information describing banana leaf images.

HSV feature extraction was employed to represent the color characteristics of banana leaf images, particularly the color variations associated with different disease symptoms. The Hue component represents the type of color, Saturation indicates the color intensity, and Value describes the brightness level of the image (Rahayu et al., 2026). Each preprocessed image was converted from the BGR color space to HSV using the OpenCV library. Color information was then represented using normalized histograms for each HSV channel. The Hue channel was computed over the range of 0–180 using 50 histogram bins, whereas the Saturation and Value channels were computed over the range of 0–256 using 60 histogram bins each. The histogram values of each channel were normalized by dividing them by the total number of pixels to reduce the influence of image size. Finally, the three normalized histograms were concatenated to produce a 170 dimensional HSV feature vector.

Texture characteristics were extracted using the Local Binary Pattern (LBP) method. LBP operates by comparing the intensity value of a center pixel with those of its neighboring pixels, producing binary patterns that effectively represent the surface texture of the leaves (Fatimah & Agustin, 2025). Before LBP computation, each image was converted from the BGR color space to grayscale using the OpenCV grayscale conversion function. LBP was then computed using a radius of 3, 24 neighboring points (8 × radius), and the uniform mapping method. The resulting LBP image was represented as a normalized histogram consisting of 26 histogram bins, producing a 26 dimensional texture feature vector. This descriptor effectively captures local texture patterns that are commonly associated with banana leaf disease symptoms.

The final feature representation was obtained by concatenating the HSV and LBP feature vectors, resulting in a 196 dimensional feature vector for each image. By combining color and texture information, the extracted features provide a more comprehensive representation of banana leaf characteristics and serve as the input for the subsequent SVM classification process.

3.6. SVM Classification

The classification stage in this study was performed using the Support Vector Machine (SVM) algorithm. SVM was selected because of its strong capability to classify high-dimensional data while maintaining excellent classification performance (Nasution et al., 2024) (Eldo et al., 2024). Furthermore, SVM has been widely applied in image classification tasks because it constructs an optimal decision boundary that maximizes the margin between different classes, making it well suited for distinguishing banana leaf diseases based on extracted color and texture features.

The combined 196 dimensional HSV–LBP feature vectors were used as the input to the SVM classifier. Before model training, feature standardization was performed using StandardScaler to ensure that all features were represented on a comparable numerical scale. To prevent data leakage, the StandardScaler was incorporated into a machine learning pipeline and fitted only on the training data. The same fitted scaler was subsequently applied to transform the testing data before classification.

The SVM classifier was implemented using the Support Vector Classification (SVC) model with a Radial Basis Function (RBF) kernel. The classifier was configured with C = 1.0, gamma = scale, and the one-versus-rest (OvR) decision strategy for multiclass classification. During training, the classifier learned the discriminative characteristics of the four banana leaf classes from the standardized HSV and LBP feature vectors. The trained model was then evaluated using an independent testing set consisting exclusively of original, non-augmented images to assess its ability to classify previously unseen samples.

Through this process, the proposed model classified banana leaf images into four categories, namely Cordana, Healthy, Pestalotiopsis, and Sigatoka. By combining complementary color information extracted from the HSV color space with texture information obtained using the LBP descriptor, the SVM classifier is expected to effectively distinguish among different banana leaf disease categories.

3.7. Model Evaluation

Model evaluation was conducted to assess the performance and reliability of the proposed HSV–LBP–SVM approach for banana leaf disease classification. The evaluation process consisted of two stages. First, Stratified Group 5-Fold Cross-Validation was performed on the training set to assess the consistency and robustness of the model while preventing augmented images originating from the same original image from appearing in different folds. This grouping strategy reduces the risk of information leakage during cross-validation and provides a more reliable estimate of model performance on the training data.

Following model training, the final evaluation was conducted using an independent testing set consisting exclusively of original, non-augmented images that were not used during the training or augmentation processes. This evaluation strategy provides an unbiased assessment of the model's ability to classify previously unseen banana leaf images.

Model performance was evaluated using a confusion matrix, classification report, and four quantitative performance metrics: accuracy, weighted precision, weighted recall, and weighted F1-score. The confusion matrix summarizes the classification results for each class, while the classification report presents the precision, recall, F1-score, and support of each class. These evaluation measures provide a comprehensive assessment of the effectiveness of the proposed classification approach.

4. RESULT AND DISCUSSION

4.1. Data Collection

The data collection stage resulted in a dataset of banana leaf images obtained from the OriginalSet of the Banana Leaf Spot Diseases (BananaLSD) dataset, which was originally collected in Bangladesh. (Arman et al., 2023). The dataset consists of 937 original banana leaf images in JPG format with a resolution of 224 × 224 pixels using the RGB color model. The images are categorized into four classes, namely Cordana (162 images), Healthy (129 images), Pestalotiopsis (173 images), and Sigatoka (473 images). These classes represent healthy banana leaves and leaves affected by different disease types, providing variations in color, texture, lighting conditions, and leaf orientation that are useful for training and evaluating the classification model.

To prevent data leakage, the original dataset was first divided into training (80%) and testing (20%) subsets before any augmentation process was performed. The resulting split produced 748 original images for training and 189 original images for testing. The class wise distribution of the training and testing datasets is presented in Table 1.

The testing dataset consisted entirely of original, non-augmented images to provide an unbiased evaluation of the proposed classification model. Meanwhile, data augmentation was applied only to the training set to increase the diversity of training samples without affecting the independence of the testing data.

4.2. Data Preprocessing

The preprocessing stage was performed by applying the center crop technique to all images before feature extraction. In this study, 85% of the central region of each image was retained, while the outer regions were removed to reduce the influence of irrelevant background information. This preprocessing procedure was applied consistently to both the training and testing datasets to ensure that all images underwent the same preparation process prior to feature extraction.

The center crop operation successfully focused each image on the banana leaf, which represents the primary region of interest, while minimizing the presence of background objects, shadows, and other irrelevant regions. As a result, the processed images exhibited a more consistent spatial composition, allowing the subsequent feature extraction process to focus on disease related visual characteristics rather than background variations.

The preprocessing results are illustrated using representative images from each disease class before and after the center crop operation. Figure 2 presents the original images prior to preprocessing, whereas Figure 3 shows the corresponding images after preprocessing. As shown in Figure 3, the center crop technique effectively preserves the main leaf region while reducing unnecessary background areas. Consequently, the preprocessed images provide more representative inputs for the subsequent feature extraction and classification stages, improving the consistency of the extracted color and texture features across the dataset.

Figure 2. Original Images Before Preprocessing

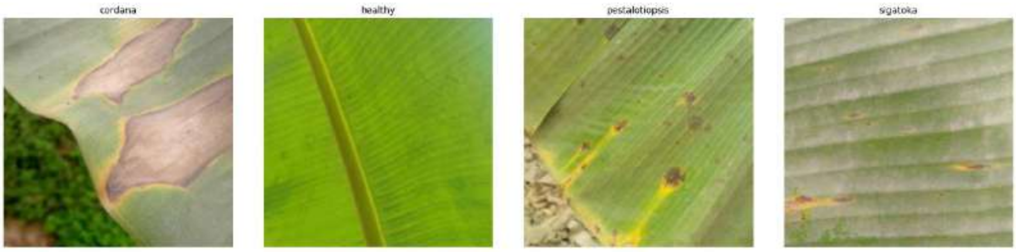

Figure 3. Images After Preprocessing

4.3. Data Augmentation

Data augmentation was performed exclusively on the training dataset after the original images had been divided into training and testing subsets. This strategy was adopted to prevent data leakage and ensure that the testing dataset remained completely independent throughout the evaluation process. The testing dataset therefore consisted only of original, non-augmented images.

The augmentation process employed three geometric transformation techniques: horizontal flipping, rotation, and shear transformation. Horizontal flipping generated mirrored versions of the banana leaf images, while rotation was randomly applied using angles of $-20°$, $-10°$, $10°$, and $20°$ to simulate different leaf orientations. In addition, shear transformation was performed using a random shear factor within the range of -0.2 to 0.2 to produce moderate geometric variations. These transformations increased the diversity of the training data while preserving the disease-related visual characteristics of the banana leaves.

Following the augmentation process, each class contained 400 training images, resulting in a balanced training dataset of 1,600 images. Representative examples of the augmented images are presented in Figure 4. The generated image variations demonstrate that the applied augmentation techniques successfully increased the diversity of the training data while maintaining the distinctive visual characteristics of each disease class. Consequently, the augmented training dataset provides a wider range of image variations, enabling the SVM classifier to learn more representative color and texture patterns without compromising the independence of the testing dataset.

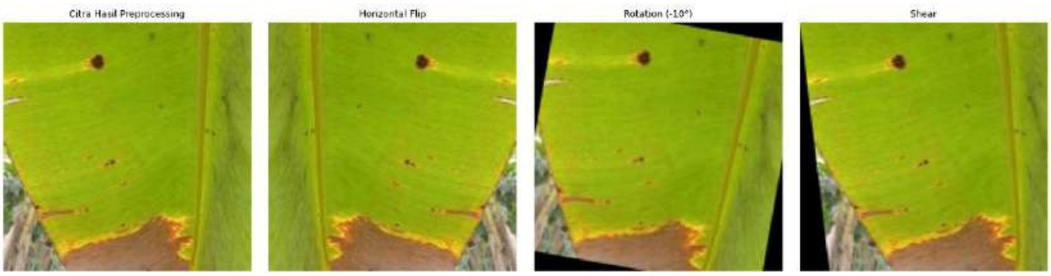

Figure 4. Images After Applying Horizontal Flip, Rotation, and Shear Transformation

4.4. HSV and LBP Feature Extraction

After preprocessing, feature extraction was performed to obtain representative color and texture information from the banana leaf images. In this study, color features were extracted using the HSV color space, whereas texture features were extracted using the Local Binary Pattern (LBP) descriptor. The extracted features were subsequently combined to form the input feature vectors for the Support Vector Machine (SVM) classifier.

The HSV feature extraction process generated normalized histograms for the Hue, Saturation, and Value channels, representing the color distribution of each banana leaf image. The extracted HSV features consisted of 170 dimensions, comprising 50 histogram bins for the Hue channel and 60 histogram bins each for the Saturation and Value channels. Representative results of HSV feature extraction, including the original banana leaf image and its corresponding HSV histogram, are presented in Figure 5. The normalized histograms capture the distribution of color information associated with healthy leaves and different disease symptoms, providing discriminative color features for the classification process.

Texture information was extracted using the Local Binary Pattern (LBP) method after converting each image to grayscale. The resulting LBP image was represented by a normalized histogram consisting of 26 feature dimensions. Representative LBP extraction results are shown in Figure 6, which includes the grayscale image, the corresponding LBP map, and the LBP histogram. The LBP map highlights local texture patterns on the leaf surface, while the histogram summarizes the distribution of these patterns into a compact numerical representation suitable for classification.

The final feature vector for each image was obtained by concatenating the 170 dimensional HSV feature vector with the 26 dimensional LBP feature vector, resulting in a 196 dimensional combined feature vector. By integrating complementary color and texture information, the resulting feature representation provides a more comprehensive description of banana leaf characteristics, enabling the SVM classifier to distinguish among the four disease classes more effectively.

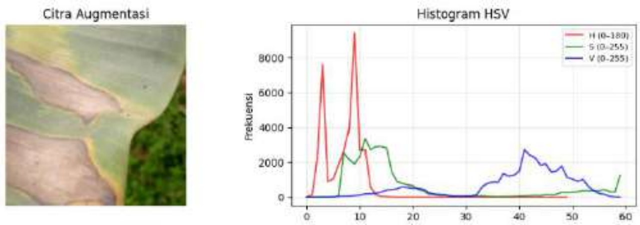

Figure 5. HSV Feature Extraction

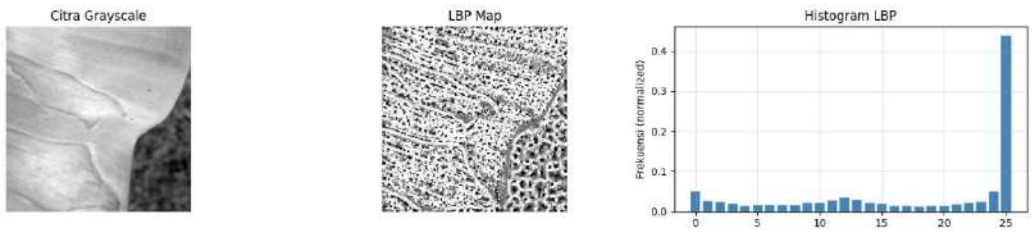

Figure 6. LBP Feature Extraction

4.5. SVM Classification

The classification process was performed using the Support Vector Machine (SVM) algorithm implemented through the Support Vector Classification (SVC) model with a Radial Basis Function (RBF) kernel. The classifier was trained using the combined 196-dimensional HSV–LBP feature vectors extracted from the augmented training dataset. Prior to model training, feature standardization was performed using StandardScaler within a machine learning pipeline to ensure that all features were represented on a comparable numerical scale while preventing data leakage during model development.

The SVM classifier was configured with C = 1.0, gamma = scale, and the one-versus-rest (OvR) decision strategy for multiclass classification. The RBF kernel was selected because of its ability to handle non-linear data distributions, enabling the model to achieve more effective classification performance by mapping the input features into a higher-dimensional space where class boundaries can be separated more efficiently (Andriyani et al., 2025).

After the training process was completed, the final model was evaluated using an independent testing dataset consisting exclusively of original, non-augmented banana leaf images. The detailed classification performance on the testing dataset is presented in the following subsection.

4.6. Model Evaluation

The trained model was evaluated using an independent testing dataset consisting exclusively of original, non-augmented banana leaf images. Prior to this evaluation, Stratified Group 5-Fold Cross-Validation was performed on the training set to assess model consistency. The cross-validation results showed a mean accuracy of 85.38% (standard deviation 2.50%), a mean weighted precision of 86.42% (standard deviation 2.31%), a mean weighted recall of 85.38% (standard deviation 2.50%), and a mean weighted F1-score of 85.52% (standard deviation 2.52%). These values were slightly lower than the accuracy obtained on the test set (87.30%), which is reasonable given that the training set contained augmented images with greater visual variability, whereas the test set consisted of original images with more consistent visual characteristics. The consistency between the cross-validation and test set results further indicates that the model did not suffer from overfitting caused by data leakage.

In addition, an ablation study was conducted to compare the performance of HSV features, LBP features, and the combined HSV–LBP features under the same experimental settings. The comparison results are presented in Table 2.

Table 2. Model Evaluation Results

Feature Extraction Method	Accuracy	Precision	Recall	F1-Score
HSV + SVM	86.24%	88.86%	86.24%	87.06%
LBP + SVM	64.02%	66.85%	64.02%	64.66%
HSV + LBP + SVM	87.30%	89.06%	87.30%	87.80%

As shown in Table 2, the proposed HSV and LBP feature combination achieved the highest overall performance among the three feature extraction strategies. Using HSV features alone produced competitive results, while LBP alone yielded substantially lower performance. These findings indicate that color information plays a dominant role in banana leaf disease classification, whereas the addition of texture information extracted by LBP provides complementary information that further improves the classification performance.

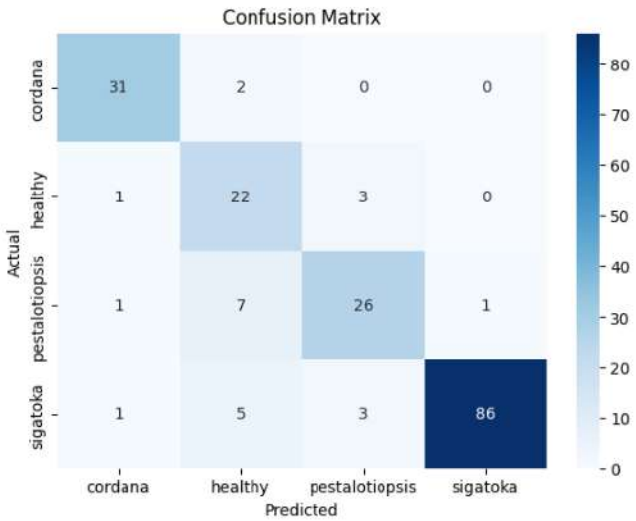

Figure 7. Confusion Matrix

The confusion matrix of the proposed model is presented in Figure 7. Overall, the classifier correctly classified 165 of the 189 testing images, corresponding to an overall accuracy of 87.30%. The Cordana and Sigatoka classes achieved the highest classification performance, whereas most misclassifications occurred between the Healthy and Pestalotiopsis classes. This observation suggests that these classes share similar color distributions and texture patterns, making them more difficult to distinguish using handcrafted HSV and LBP features. Nevertheless, the proposed model demonstrated satisfactory performance in classifying unseen original banana leaf images.

Table 3. Classification Report

Class	Precision	Recall	F1-Score	Support
Cordana	91%	94%	93%	33
Healthy	61%	85%	71%	26
Pestalotiopsis	81%	74%	78%	35
Sigatoka	99%	91%	95%	95

According to Table 3, the Sigatoka class achieved the highest classification performance, whereas the Healthy class obtained the lowest precision among the four classes. The relatively lower precision indicates that several images predicted as Healthy actually belonged to other disease classes, reflecting the similarity of visual characteristics between healthy leaves and leaves showing early disease symptoms.

5. CONCLUSION AND SUGGESTION

Based on the experimental results, the combination of HSV and Local Binary Pattern (LBP) feature extraction with a Support Vector Machine (SVM) classifier proved effective for banana leaf disease classification. Using an independent testing dataset consisting of original, non-augmented images, the proposed model achieved an accuracy of 87.30%, weighted precision of 89.06%, weighted recall of 87.30%, and weighted F1-score of 87.80%. The ablation study further demonstrated that combining HSV and LBP features produced better classification performance than using either HSV or LBP features individually, indicating that the integration of complementary color and texture information provides a more discriminative representation for banana leaf disease classification.

The confusion matrix analysis showed that most banana leaf images were correctly classified, although several misclassifications occurred, particularly between the Healthy and Pestalotiopsis classes. These errors suggest that certain disease symptoms exhibit similar color distributions and texture patterns, making them more difficult to distinguish using handcrafted feature descriptors. Nevertheless, the proposed approach demonstrated satisfactory performance in classifying unseen original banana leaf images.

This study has several limitations. First, the experiments were conducted using only the BananaLSD dataset, which was collected in Bangladesh. Therefore, the proposed model has not yet been validated using banana leaf images collected under different environmental conditions, including those in Indonesia. Second, this study evaluated only the SVM classifier, although the proposed feature combination may also be applicable to other machine learning algorithms. Future research should validate the proposed method using datasets collected from different geographical regions, investigate alternative machine learning and deep learning models, and compare the proposed approach with other feature extraction techniques to further improve classification performance and model generalizability.

Declaration of AI and AI-assisted technologies in the writing process

In preparing this manuscript, the authors utilized AI assisted tools, including ChatGPT and Google Translate, to support language refinement, translation, and improvement of overall clarity and readability. The authors remain fully responsible for the content of the manuscript and confirm that the use of these tools did not compromise the originality, validity, or integrity of the research.

REFERENCE

Alfitriana Riska, Purnawansyah, Darwis, H., & Astuti, W. (2023). Studi Perbandingan Kombinasi GMI, HSV, KNN, dan CNN pada Klasifikasi Daun Herbal. *The Indonesian Journal of Computer Science*, *12*(3). https://doi.org/10.33022/ijcs.v12i3.3210

Andriyani, W., Astuti, Y., Wisesa, B. A., & Hengki, H. (2025). Analisis Sentimen pada Ulasan Produk dengan SVM dan Word2Vec. *JIKO (Jurnal Informatika Dan Komputer)*, *9*(1), 173. https://doi.org/10.26798/jiko.v9i1.1498

Arman, S. E., Bhuiyan, Md. A. B., Abdullah, H. M., Islam, S., Chowdhury, T. T., & Hossain, Md. A. (2023). BananaLSD: A banana leaf images dataset for classification of banana leaf diseases using machine learning. *Data in Brief*, *50*, 109608. https://doi.org/10.1016/j.dib.2023.109608

Asykurani Abdillah, R., Wahyuni Program Studi Fisika, D., Fisika, J., & Universitas Tanjungpura Jalan Hadari Namawi, F. H. (2023). Identifikasi Penyakit COVID-19 dan Tuberkulosis Menggunakan Metode Convolutional Neural Network Arsitektur GoogLeNet Berdasarkan Citra Rontgen Thorax. *PRISMA FISIKA*, *11*(3), 65–70. www.kaggle.com

Badan Pusat Statistik. (2026, February 25). *Produksi Tanaman Buah-buahan dan Sayuran Tahunan Menurut Provinsi dan Jenis Tanaman, 2025*. Https://Www.Bps.Go.Id/Id/Statistics-Table/3/U0dKc1owczVSalJ5VFdOMWVETnlVRVJ6YlRJMFp6MDkjMw==/Produksi-Buah-Buahan-Menurut-Jenis-Tanaman-Menurut-Provinsi--2025.Html?Year=2025.

Bhuiyan, Md. A. B., Abdullah, H. M., Arman, S. E., Saminur Rahman, S., & Al Mahmud, K. (2023). BananaSqueezeNet: A very fast, lightweight convolutional neural network for the diagnosis of three prominent banana leaf diseases. *Smart Agricultural Technology*, *4*, 100214. https://doi.org/10.1016/j.atech.2023.100214

Cahyani Putri, E. D. S., Syah Putra, M. M., & Agustin, S. (2025). *Deteksi Penyakit Daun Tomat Berbasis Warna Dan Tekstur Dengan Algoritma K-Nearest Neighbor*. *03*(02). https://doi.org/10.33506/framework.v3i02.4644

Dzulhijjah, A. N., Anraeni, S., & Sugiarti, S. (2021). Klasifikasi Kematangan Citra Labu Siam Menggunakan Metode KNN (K-Nearest Neighbor) Dengan Ekstraksi Fitur HSV (Hue, Saturation, Value). *Buletin Sistem Informasi Dan Teknologi Islam*, *2*(2), 103–110. https://doi.org/10.33096/busiti.v2i2.808

Eldo, H., Ayuliana, A., Suryadi, D., Chrisnawati, G., & Judijanto, L. (2024). Penggunaan Algoritma Support Vector Machine (SVM) Untuk Deteksi Penipuan pada Transaksi Online. *Jurnal Minfo Polgan*, *13*(2), 1627–1632. https://doi.org/10.33395/jmp.v13i2.14186

Fahmi Wibawa, M., Rahman, M. A., & Widodo, A. W. (2021). *Penerapan Ruang Warna HSV dan Ekstraksi Fitur Tekstur Local Binary Pattern untuk Tingkat Kematangan Sangrai Biji Kopi* (Vol. 5, Number 7). http://j-ptiik.ub.ac.id

Fatimah, N. S., & Agustin, S. (2025). Klasifikasi Citra Batik Menggunakan Local Binary Pattern (LBP) dan Support Vector Machine (SVM). *Jurnal Algoritma*, *22*(1), 185–196. https://doi.org/10.33364/algoritma/v.22-1.2208

Gimmas Manggara, M., & Kalifia, A. D. (2025). Komparasi Model ResNet50 dan EfficientNetV2M dengan Penerapan Transfer Learning dan Fine Tuning pada Klasifikasi Penyakit Bercak Daun Tanaman Pisang. *Bulletin of Computer Science Research*, *6*(1), 293–302. https://doi.org/10.47065/bulletincsr.v6i1.855

Hamzah, F. A., & Wahyusari, R. (2023). Perbandingan Perbaikan Citra Magnetic Resonance Imaging (MRI) Menggunakan Ruang Warna RGB, HSV dan YCbCr Dengan Metode Histogram Equalization dan Contrast Streching. *JIIFKOM (Jurnal Ilmiah Informatika Dan Komputer)*, *2*(2), 22–26. https://doi.org/10.51901/jiifkom.v2i2.355

Hayati, N. (2023). Klasifikasi Jenis Bunga Mawar Menggunakan Algoritma K-Nearest Neighbour. *Jurnal Informatika Dan Riset*, *1*(1), 31–37. https://doi.org/10.36308/iris.v1i1.474

Lase, Y. Y., Purba, L. P., Prayudani, S., Lubis, A. R., & Adam, H. A. (2025). Analisis Deteksi Penyakit Daun Pisang Menggunakan Ekstraksi Fitur CNN (MobileNetV2) dan Klasifikasi SVM. *INSOLOGI: Jurnal Sains Dan Teknologi*, *4*(6), 1569–1578. https://doi.org/10.55123/insologi.v4i6.6590

Maulana Alfaruq, B., Erwanto, D., & Yanuartanti, I. (2023). Klasifikasi Kematangan Buah Tomat Dengan Metode Support Vector Machine. *Generation Journal*, *7*(3), 64–72. https://doi.org/10.29407/gj.v7i3.21092

Muchtar, M., & Muchtar, R. A. (2024). Perbandingan Metode Knn Dan Svm Dalam Klasifikasi Kematangan Buah Mangga Berdasarkan Citra Hsv Dan Fitur Statistik. *Jurnal Informatika Dan Teknik Elektro Terapan*, *12*(2). https://doi.org/10.23960/jitet.v12i2.4010

Muflich, A., & Kusrini, K. (2026). Identifikasi Penyakit Daun Pisang Berbasis Citra Warna dengan Ekstraksi ResNet50 dan Support Vector Machine. *Insect (Informatics and Security): Jurnal Teknik Informatika*, *12*(01), 106–118. https://doi.org/10.33506/insect.v12i01.5460

Nasution, A., Lubis, A. S., & Kiswanto, D. (2024). Identifikasi Variasi Paprika Berdasarkan Jenis Warna Paprika Berbasis Analisis Citra Digital Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Informasi, Sains Dan Teknologi*, *7*(2), 209–220. https://doi.org/10.55606/isaintek.v7i2.282

Neneng, Putri, N. U., & Susanto, R. E. (2020). Klasifikasi Jenis Kayu Menggunakan Support Vector Machine Berdasarkan Ciri Tekstur Local Binary Pattern. *CYBERNETICS*, *4*(02), 93–100.

Permata Aulia, T. M., Arifin, N., & Mayasari, R. (2021). Perbandingan Kernel Support Vector Machine (Svm) Dalam Penerapan Analisis Sentimen Vaksinisasi Covid-19. *SINTECH (Science and Information Technology) Journal*, *4*(2), 139–145. https://doi.org/10.31598/sintechjournal.v4i2.762

Prasetyo, A., & Utami, E. (2024). Detection and Classification of Banana Leaf Diseases: Systematic Literature Review. *Telematika*, *17*(2), 128–141. https://doi.org/10.35671/telematika.v17i2.2809

Pratama, M. D., Gustriansyah, R., & Purnamasari, E. (2024). Klasifikasi Penyakit Daun Pisang menggunakan Convolutional Neural Network (CNN). *Jurnal Teknologi Terpadu*, *10*(1), 1–6. https://doi.org/10.54914/jtt.v10i1.1167

Pratiwi, S. C., & Prastika Wahyudi, J. Z. (2026). Klasifikasi Penyakit Daun Pisang Menggunakan Metode CNN Berbasis Citra Digital. *Prosiding Seminar Nasional Teknologi Dan Sains*, *5*. https://doi.org/10.29407/z3f67269

Rahayu, N., Silitonga, M. T. A., Jordan, D., Nabila, S., & Pratama, A. (2026). Klasifikasi Jenis Kelamin Berdasarkan Citra Mata Menggunakan Fitur HSV dan HOG Dengan Algoritma SVM. *Jurnal Ilmiah ILKOMINFO - Ilmu Komputer & Informatika*, *9*(1), 25–34. https://doi.org/10.47324/ilkominfo.v9i1.423

Safitri, E., Heppy Ria Sibarani, R., SM Sidabutar, Y., & Kiswanto, D. (2024). Klasifikasi Penyakit Daun Anggur Berbasis Citra Menggunakan Metode K-Nearest Neighbors (KNN). *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(6), 12633–12642. https://doi.org/10.36040/jati.v8i6.12004

Sanjaya, J., & Ayub, M. (2020). Augmentasi Data Pengenalan Citra Mobil Menggunakan Pendekatan Random Crop, Rotate, dan Mixup. *Jurnal Teknik Informatika Dan Sistem Informasi*, *6*(2). https://doi.org/10.28932/jutisi.v6i2.2688

Simanjuntak, A. J. O., & Udjulawa, D. (2022). Klasifikasi Penyakit Daun Sawit Menggunakan Metode Jaringan Saraf Tiruan Dengan Fitur Local Binary Pattern. *Jurnal Algoritme*, *3*(1), 1–9. https://doi.org/10.35957/algoritme.v3i1.3158

Sirappa, M. P. (2021). Potensi Pengembangan Tanaman Pisang : Tinjauan Syarat Tumbuh dan Teknik Budidaya Pisang Dengan Metode Bit. *Jurnal Ilmiah Agrosaint*, *12*(2). https://doi.org/10.47178/6dyf7h16

Surnanto, Martanto, Pratama, D., Bahtiar, A., & Narasati, R. (2026). Klasifikasi Penyakit Daun Pisang Menggunakan Metode K-Nearest Neighbor (KNN) Berbasis Ekstraksi Fitur Warna Dan Tekstur. In *JANUARI 2026 JOURNAL OF COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE* (Number 01). JCSAI. https://ikmi.ac.id/page/18/?lang=de