Institut Riset dan Publikasi Indonesia (IRPI)
**MALCOM: Indonesian Journal of Machine Learning and Computer Science**
Journal Homepage: https://journal.irpi.or.id/index.php/malcom
Vol. 6 Iss. 1 January 2026, pp: 332-340
ISSN(P): 2797-2313 | ISSN(E): 2775-8575

# *Optimization of Machine Learning Model for Sugarcane Leaf Detection Using Ensemble Methods*

**Sekar Dana Chiatra[1*], Sri Lestari[2]**

[1,2]Department of Informatics Engineering,
Institute of Informatics and Business Darmajaya, Indonesia

E-Mail: [1]sekardanachiatra@gmail.com, [2]srilestari@darmajaya.ac.id

*Received Dec 04th 2025; Revised Jan 14th 2026; Accepted Jan 20th 2026; Available Online Jan 31th 2026*
*Corresponding Author: Sekar Dana Chiatra*


**Abstract**

*Sugarcane leaf identification plays an important role in supporting agricultural monitoring and crop management. However, variations in leaf appearance and background conditions often reduce the performance of single classification models. This study aims to analyse and compare the performance of Support Vector Machine (SVM), AdaBoost, and ensemble models that combine SVM and AdaBoost for sugarcane leaf image classification. The data used in this study was obtained from a public dataset available on Kaggle and modified through image selection with a total of 1,391 data points grouped into two classes, namely sugarcane (587 image data points) and non-sugarcane (804 image data points), as well as pre-processing steps. Feature extraction was performed to represent leaf characteristics prior to classification. The models were evaluated using accuracy, precision, recall, and F1-score metrics. The experimental results showed that the SVM-AdaBoost ensemble model achieved the best performance among all models tested, with an accuracy value of 93.55% and an F1-score of 93.47%, demonstrating its effectiveness in improving classification reliability. These findings indicate that ensemble learning can improve classification performance for sugarcane leaf images and can be considered as an alternative approach for agricultural image analysis applications.*

*Keywords: AdaBoost, Ensemble Learning, Image Classification, Sugarcane Leaf, Support Vector Machine*

## 1. INTRODUCTION

Plantation is an activity that utilizes science, technology, capital, and management to grow certain types of plants on soil or growing media that are suitable for the ecosystem environment. This activity also includes the processing and marketing of these crops with the aim of improving the welfare of business actors and the community (in Indonesia, this is regulated by Law No. 18 of 2004, Article 1, paragraph 1). One type of plant included in the plantation sector is sugarcane. Sugarcane (Saccharum officinarum) is a plantation crop that has high commercial value because it can increase the country's foreign exchange earnings. This plant belongs to the Poaceae family, a group of grass-like plants, and can grow in lowlands in tropical regions and in some subtropical areas. The greatest benefit of sugarcane is as a raw material for producing granulated sugar [1].

Recent developments in digital image processing and machine learning have played an important role in the development of automated systems capable of recognizing plants based on image data. Various methods have been studied to recognize leaves, including Support Vector Machine (SVM) as done in the study "Detection of Sugarcane Leaf Objects Using Classification Methods in Machine Learning" which showed an accuracy of 96% [2], K -Nearest Neighbor (K-NN) as conducted in the study "Detection of Plant Types Based on Leaf Shape Using K-NN" with an accuracy of 94% [3], Random Forest in the study "Implementation of Machine Learning Algorithms for Sugarcane Leaf Disease Detection: Performance Comparison Analysis" with an accuracy of 64.8% [4], and a boosting-based technique, namely Adaptive Boosting, which occurred in the study "Detection of Diseases in Rice Plants Using Optimized AdaBoost Classifier" with an accuracy of 91.37% [5]. Among these methods, SVM is well-known for its ability to effectively handle high-dimensional data and maintain consistent results even when the amount of training data is limited.

However, relying on only one classification model often limits the system's ability to process complex image data. Differences in lighting conditions, background noise, and visual similarities between leaves of different plant species can affect the accuracy of the classification process. To address these challenges, an ensemble learning approach has been introduced as an effective strategy that combines several classification models to produce more accurate and reliable predictions. One commonly used ensemble method is Adaptive Boosting (AdaBoost), which repeatedly emphasizes difficult-to-classify samples during training [6].

In addition to choosing the right classification model, the use of feature extraction techniques also greatly affects classification performance. Histogram of Oriented Gradients (HOG) is often used because it is able to accurately describe edge structures, object shapes, and texture patterns found in an image. HOG has demonstrated good robustness to changes in lighting and geometric transformations, making it suitable for object recognition tasks, such as plant leaf classification [7].

Several previous studies have examined the application of machine learning and ensemble methods in classification processes, including the study “Automatic detection of individual oil palm trees from UAV images using HOG features and SVM classifiers.” This study showed that the HOG-SVM algorithm successfully grouped plant types based on leaf shape [8]. Furthermore, in the study “Application Of The Haar-Like Feature Method And Adaboost Algorithm In Determining the Classification of Coffee Plant Pests,” the Adaptive Boosting algorithm successfully performed classification with 80% accuracy [9]. Based on the findings of the study "Application Of The Support Vector Machine, Light Gradient Boosting Machine, Adaptive Boosting, And Hybrid Adaboost-Svm Model On Customers Churn Data" it can be concluded that the SVM-AdaBoost ensemble algorithm is a boosting of the SVM algorithm, which successfully displayed better performance than SVM with an accuracy of 97.25%, while SVM only achieved 89.75% [10].

Based on these considerations, this study suggests an ensemble learning approach that combines Support Vector Machine (SVM) and AdaBoost methods to detect sugarcane leaves using digital image data and Histogram of Oriented Gradients (HOG) features. Unlike deep learning approaches that are highly dependent on the availability of large amounts of data, the proposed method emphasizes model efficiency and robustness when data availability is limited. The main contributions of this study consist of three main aspects, namely: (1) using the HOG-based feature extraction method to effectively represent the characteristics of sugarcane leaf images, (2) designing an ensemble-based classification model that combines SVM and AdaBoost to improve detection capabilities through a boosting mechanism, and (3) conducting a comprehensive evaluation of model performance using various metrics such as accuracy, precision, recall, and F1-score in order to test and ensure the effectiveness of the developed approach.

## 2. MATERIALS AND METHOD

This study employed an experimental approach by applying several machine learning models to detect sugarcane leaves based on digital image data. The research methodology consisted of dataset collection, image preprocessing, feature extraction, classification model training using Support Vector Machine (SVM), AdaBoost, and the ensemble SVM-AdaBoost model, followed by performance evaluation using multiple evaluation metrics. The Research Flow is shown in Figure 1.

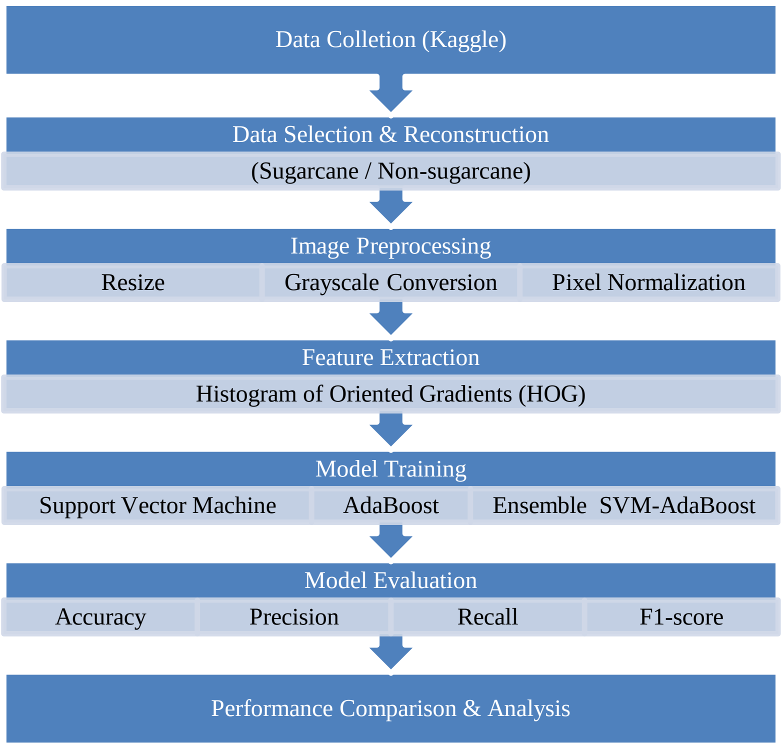


**Figure 1.** Research Flow of Sugarcane Leaf Detection Using Ensemble SVM-AdaBoost

### 2.1. Research Flow

This study follows a structured research workflow to ensure systematic development and evaluation of the proposed sugarcane leaf detection model. The overall research flow is illustrated in Figure 1 and consists of several sequential stages, namely dataset preparation, image preprocessing, feature extraction, model training, and performance evaluation.

The research begins with dataset collection and preparation. The image dataset was obtained from a publicly available Kaggle dataset. The collected images were not used directly; instead, a selection and reconstruction process was conducted to ensure relevance to the research objective. The images were reorganized into two main classes, namely sugarcane and non-sugarcane, resulting in a modified public dataset suitable for binary classification tasks.

Next, image preprocessing was performed to enhance image quality and ensure consistency across the dataset. All images were resized to a uniform resolution to maintain dimensional consistency. The images were then converted into grayscale format to reduce computational complexity while preserving essential structural information. Pixel intensity normalization was also applied to standardize the range of pixel values and improve the robustness of feature extraction.

After preprocessing, feature extraction was carried out using the Histogram of Oriented Gradients (HOG) method. HOG was employed to capture important visual characteristics of sugarcane leaves, particularly edge information, shape, and texture patterns. The extracted HOG features were represented as numerical feature vectors, which served as input data for the classification models.

The subsequent stage involved model training and classification. Three different classification models were developed and evaluated in this study: Support Vector Machine (SVM), AdaBoost an ensemble model combining SVM, and AdaBoost. In the ensemble approach, an SVM was used as the base learner within the AdaBoost framework to improve classification performance by adaptively emphasizing misclassified samples during training.

Finally, model evaluation and performance analysis were conducted to assess the effectiveness of each classification model. The performance of the SVM, AdaBoost, and ensemble SVM-AdaBoost models was evaluated using multiple evaluation metrics, including Accuracy, Precision, Recall, and F1-score. The obtained results were then compared to identify the model that achieved the best overall performance in detecting sugarcane leaf images.

### 2.2. Dataset

The data used in this study was obtained from the Kaggle platform, which provides a collection of images that are openly available for research purposes. The original dataset consisted of 829 image data divided into 30 plant classes, namely sugarcane, almond, banana, cardamom, cherry, chili, clove, coconut, coffee, cotton, cucumber, peanuts, gram, jowar, jute, lemon, maize, mustard-oil, olive-trees, papaya, pearl millet, pineapple, rice, soybean, sunflower, tea, tobacco plant, tomato, vigna radiata, and wheat. From these 30 classes, the data was reprocessed into only two classes, sugarcane (class 1) and non-sugarcane (class 0), due to research requirements. However, after recalculating, it turned out that the number of sugarcane and non-sugarcane image data was very unbalanced. There were 804 non-sugarcane images, while there were only 25 sugarcane images. Therefore, the researcher added data by searching for images of healthy sugarcane leaves on the Kaggle platform, obtaining 562 images, bringing the total number of sugarcane images to 587 and non-sugarcane images to 804 [11], [12].

All of these image data (a total of 1,391) are in 3 formats (.jpg, .jpeg, and .png) and are stored in a folder named data_tesis in .zip format, to facilitate image import into Google Colab for image preprocessing.

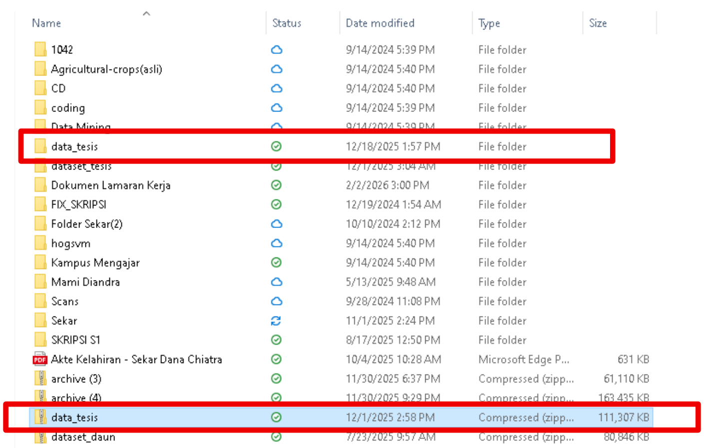


**Figure 2.** Dataset Folder

### 2.3. Image Preprocessing

The first stage in this research is to preprocess the images. This stage aims to improve image quality and ensure data consistency before entering the feature extraction stage. The first task at this stage is to import the library and upload the dataset. All libraries and datasets required for this research must be ready in Google Colab.

All images in the dataset are resized to the same size (img = resize(img, (64, 128, 3))) so that the input dimensions are standardized and computational complexity during model training is reduced. This resizing process also helps reduce variations caused by different image capture conditions. In addition to resizing the images, the RGB images were converted to grayscale (img_gray = rgb2gray(img)). The conversion to grayscale was done to reduce the image data dimensions while retaining important structural and textural information from the sugarcane leaves. This step is often used in image-based classification tasks to simplify data representation without significantly reducing classification performance. By performing these pre-processing steps, the resulting images are ready for the next stage, which is feature extraction[13].

### 2.4. Feature Texture

The second stage in this study is feature extraction. Feature extraction is an important stage in converting original image data into numerical form that can be used by machine learning algorithms. In this study, the chosen feature extraction method was the Histogram of Oriented Gradients (HOG) because it captures structural information, such as edges, object contours, and texture patterns. HOG does not only rely on pixel intensity, but also analyzes the distribution of gradient directions, making it very effective in object recognition tasks such as plant leaf classification [7], [14].

Based on these parameters, each leaf image is converted into a numerical form in the form of HOG feature vectors that show the distribution of light waves in each part of the image in a regular and adjusted manner. This extraction yields a set of features that better describe the shape, edges, and texture of leaves than using only the original pixel intensities. These features are then used as input in the machine learning stage. Thus, the feature extraction process acts as a bridge between raw image data and classification algorithms, so that in the next stage, Support Vector Machine (SVM), AdaBoost, and ensemble methods can be trained to effectively classify sugarcane and non-sugarcane leaves [7].

### 2.5. Support Vector Machine (SVM)

After the images are extracted using HOG, the next stage, or the third stage in this study, is classification. In this study, classification is performed using three methods, namely Support Vector Machine (SVM), Adaptive Boosting (AdaBoost), and Ensemble SVM-AdaBoost. This section discusses the classification of sugarcane leaves using the Support Vector Machine method.

There is a process that must be carried out before classification, namely data splitting. Data splitting is the process of dividing a dataset into two sets, a training set and a test set. This process is carried out so that the model created not only remembers the data it has seen, but can also apply this knowledge to new data it has never seen before. The dataset is split into training and test sets at 80:20, using stratified random sampling to maintain balanced class distributions across both sets. The random_state setting ensures the experiment results can be reproduced.

However, there is one more process that needs to be done before entering the classification stage, namely feature normalization. Features extracted using HOG are normalized using StandardScaler, so that each feature has a mean of zero and a standard deviation of one. The standardization process is carried out by calculating the statistical values only from the training data, then using them to standardize the test data, so that data leakage does not occur (a condition where information from the test data is unintentionally targeted during the model learning process) [15].

### 2.6. Adaptive Boosting (AdaBoost)

At the same time, classification does not rely solely on SVM; it also uses the AdaBoost method for comparison. This section discusses the Adaptive Boosting or AdaBoost method in this study. Boosting is a technique in ensemble learning that combines several simple models (weak models) to create a better model (strong model). The main principle of boosting is to train models sequentially, with each subsequent model focusing on data previously misclassified by earlier models. Initially, all data has the same weight. After the first model is trained, incorrectly classified data is given a higher weight, so that the next model pays more attention to that data. This process is repeated continuously until the specified number of estimators is reached, and the final decision is obtained by combining the results of all models in a weighted manner. One of the well-known boosting algorithms is AdaBoost, short for Adaptive Boosting, which was first introduced by Yoav Freund and Robert Schapire in 1996. It is called adaptive because this method automatically adjusts the weight of the data based on previous errors. In its application, AdaBoost usually uses Decision Stump, which is a Decision Tree with a depth of one, as the base learner [6], [18].

In this study, the AdaBoost method was chosen as a comparison model with Support Vector Machine (SVM) because both use fundamentally different theoretical approaches. SVM is a method that uses the margin principle: it finds the best separating hyperplane with the maximum margin between classes through mathematical optimization. Meanwhile, AdaBoost is a method that uses an adaptive weighting scheme, gradually combining several weak models to create a strong model. These differences in characteristics allow for a more in-depth analysis of the ability of HOG features to classify sugarcane leaf images, as well as providing a good basis for developing ensemble methods that can improve the model's ability to adapt to various types of data [10], [19].

This model uses a basic learner, a Decision Tree with max_depth=1, resulting in a decision stump as a weak classifier. Selecting a depth of 1 creates a simple model with limited data grouping capabilities, but its performance can be improved through boosting. Furthermore, the algorithm is configured with n_estimators=50, meaning 50 weak models are trained slowly, one by one. In each round, the training data weights are updated so that samples misclassified in the previous round receive greater attention in the next. The learning_rate=1.0 parameter controls how much influence each weak learner has on the final model, while random_state=42 helps ensure the experiment results can be reproduced. The training process is carried out using feature data that has been standardized to the same scale (X_train_scaled) to maintain stable learning and avoid features that are too large from affecting the results due to scale differences. With this approach, the AdaBoost model is expected to improve the accuracy rate in classifying data by combining the prediction results from various simple models that complement each other, so that errors in classification can be minimized [5], [9].

## 2.7. Ensemble SVM-AdaBoost

In addition to using the AdaBoost method, this study also used the SVM-AdaBoost ensemble as a comparison method for classification. This section will discuss the SVM-AdaBoost ensemble. In this study, the ensemble learning approach was used with the VotingClassifier method to combine the best Support Vector Machine model (best_svm) and the AdaBoost model (adaboost_model). Both models were registered in the estimators parameter as a pair of names and model objects, namely (‘svm’, best_svm) and (‘ada’, adaboost_model), so that the system could recognize and manage each classifier separately in the ensemble structure. The parameter voting=‘soft’ is used to apply the soft voting method, where each model provides a prediction probability for each class, then these probabilities are summed and divided to determine the class with the highest value as the final result [20].

This approach was chosen because it accounts for each model's confidence when making predictions, leading to more consistent decisions than the hard voting method, which relies solely on the most frequent label. The training process was carried out by calling the fit(X_train_scaled, y_train) function, so that both models could learn using normalized feature data. This aims to maintain uniform scale across features and to improve consistency and stability in the optimization process. With this configuration, the ensemble model is expected to leverage the strengths of each model: SVM, which focuses on optimal margins, and AdaBoost, which can adapt to classification errors, thereby improving the model's ability to adapt to new data and improving overall classification results [21], [22].

## 2.8. Evaluation Metrics

After completing the training and testing process for the classification model, the next step is to evaluate the performance of each algorithm to determine their ability to classify sugarcane leaf images. This evaluation aims to provide a quantitative basis for comparing the performance of the SVM, the AdaBoost, and the SVM-AdaBoost ensemble models. To obtain a comprehensive analysis, several assessment metrics are used, derived from the confusion matrix and discriminant curve analysis.

Model performance is assessed based on four main parts of the confusion matrix, namely true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values. Based on these components, the evaluation metrics are formulated as follows.

Accuracy measures the level of correctness of predictions across the entire test data sample and is formulated as equation 1.

$$\text{Accuracy} = \frac{(\text{TP}+\text{TN})}{(\text{TP}+\text{TN}+\text{FP}+\text{FN})} \tag{1}$$

The accuracy value represents the model's overall performance in classifying sugarcane and non-sugarcane images. However, this metric is sensitive to class distribution imbalances [23]. Precision measures how accurately the model predicts positive classes and is formulated as equation 2.

$$\text{Precision} = \frac{\text{TP}}{(\text{TP}+\text{FP})} \tag{2}$$

A high precision value indicates that most predictions categorized as positive classes are actually positive classes, thereby minimizing the number of false positive predictions [23]. Recall measures the model's ability to detect all samples that are actually positive classes and is formulated as equation 3.

$$\text{Recall} = \frac{\text{TP}}{(\text{TP+FN})} \tag{3}$$

A high recall value indicates that the model is able to recognize most of the data that belongs to the positive class, thereby reducing false negative predictions [23]. The F1-score is the harmonic mean of precision and recall, formulated as equation 1.

$$\text{F1} - \text{score} = 2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \tag{4}$$

The use of the harmonic mean imposes a heavier penalty on low precision or recall values, resulting in a more balanced performance measure. In addition to using metrics based on confusion matrices, this study also analyzed ROC curves to assess the model's ability to distinguish between classes at different decision thresholds. ROC curves illustrate the relationship between true positive rate (TPR) and false positive rate (FPR) [24].

$$\text{TPR} = \frac{\text{TP}}{(\text{TP+FN})} \tag{5}$$

$$\text{FPR} = \frac{\text{FP}}{(\text{FP+TN})} \tag{6}$$

The ROC curve is obtained by plotting TPR against FPR for various probability threshold levels. AUC is the area under the ROC curve, and is generally defined as equation 7.

$$\text{AUC} = \int_0^1 \text{TPR (FPR) d(FPR)} \tag{7}$$

AUC values range from 0 to 1. A value of 0.5 means that the results are the same as random guessing, while a value close to 1 means that classes can be distinguished very well [24]. With these combined metrics, the assessment not only emphasizes overall prediction accuracy but also considers the balance of classification error types and the model's ability to distinguish between classes. This approach provides a strong and objective basis for comparing the performance of the three algorithms used in the study.

## 3. RESULTS AND DISCUSSION

### 3.1. Performance Evaluation Based on Classification Metrics

This section presents the results of experiments evaluating the three classification models: Support Vector Machine (SVM), AdaBoost, and the proposed SVM-AdaBoost ensemble model. The performance of each model was assessed using Accuracy, Precision, Recall, and F1 Score to comprehensively evaluate its effectiveness in the classification process.

**Table 1.** Performance Comparison of Classification Models

| Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| SVM | 0.928315 | 0.928567 | 0.928315 | 0.928393 |
| AdaBoost | 0.917563 | 0.921181 | 0.917563 | 0.917940 |
| Ensemble (SVM + AdaBoost) | 0.931900 | 0.932333 | 0.931900 | 0.932007 |

Based on the performance evaluation results shown in Table 1, the SVM–AdaBoost ensemble model achieved the best performance among the single models. The Support Vector Machine model as a baseline obtained an accuracy of 92.83%, precision of 0.9286, recall of 0.9283, and F1-score of 0.9284, indicating stable and balanced performance between false positive and false negative errors. Meanwhile, the AdaBoost model achieved an accuracy of 91.75% with precision of 0.9212, recall of 0.9176, and F1-score of 0.9179, showing competitive performance but slightly below SVM. The combination of the two models through an ensemble approach resulted in improved performance with an accuracy of 93.19%, precision of 0.9323, recall of 0.9319, and F1-score of 0.9320. Although the improvement is relatively small numerically, the results are consistent across all evaluation metrics, indicating that the integration of the two algorithms is able to optimally utilize the strengths of each model, thereby improving the overall accuracy and stability of the predictions [13].

### 3.2. ROC Curve and AUC Analysis

After evaluating the model's performance based on classification metrics such as accuracy, precision, recall, and F1-score, the next step is to test the model's ability to distinguish between existing classes more comprehensively using decision thresholds. For this purpose, the Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC) values are used.

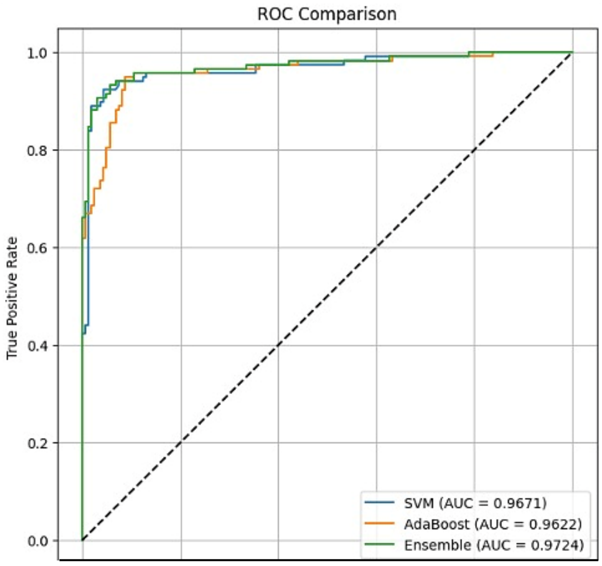


**Figure 3.** ROC Curve

Based on the ROC curve in Figure 3, all three models exhibit excellent discriminatory ability, as all curves lie well above the diagonal line representing random classification (AUC = 0.5). Each model has an AUC value exceeding 0.96, indicating excellent class discrimination ability. The SVM model achieved an AUC of 0.9671, while the AdaBoost model achieved 0.9622. The SVM–AdaBoost ensemble model performed best with an AUC value of 0.9724, meaning that this model was able to effectively distinguish between sugarcane and non-sugarcane classes at various threshold levels.

Visually, the ensemble curve is closer to the upper left corner of the graph, indicating a combination of high True Positive rates and low False Positive rates. Although the difference in AUC values between models is not very large, the continuous improvement in the ensemble model reinforces the previous evaluation results based on accuracy, precision, recall, and F1-score. This proves that combining SVM and AdaBoost not only improves prediction accuracy at a specific decision point, but also improves the model's ability to distinguish between classes as a whole [25].

### 3.3. Confusion Matrix Analysis

After evaluating using the ROC curve and AUC, the analysis continued with a confusion matrix to examine classification performance in more detail for each class. Using TP, TN, FP, and FN components, it was possible to identify error patterns and assess the model's balance between detecting sugarcane plants and distinguishing them from non-sugarcane objects.

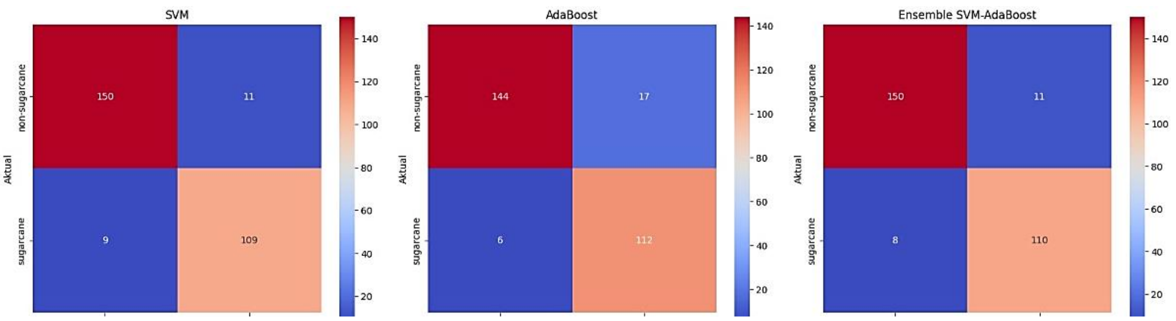


**Figure 4.** Confusion Matrix SVM, AdaBoost, and SVM-AdaBoost

Based on the confusion matrix in Figure 4, the Support Vector Machine model shows balanced performance with 150 true negatives (TN), 109 true positives (TP), 11 false positives (FP), and 9 false negatives (FN). The number of classification errors is quite low and evenly distributed across both classes, so that the precision and recall values are adequate. The AdaBoost model produces 144 true negatives, 17 false positives, 6 false negatives, and 112 true positives. This model has the best sugarcane detection capability because it has the smallest number of false negatives, even though it produces more false positives, thereby slightly reducing the precision rate. Meanwhile, the SVM–AdaBoost ensemble model obtained a TN = 150, FP = 11, FN = 8, and TP = 110, which shows a good balance between errors in the positive and negative classes. The number of false positives remains low as in SVM, but false negatives are lower than in single SVM, creating a more optimal combination of precision and recall. In general, the results of the confusion matrix show that the ensemble approach provides the most stable and consistent classification results in separating the sugarcane and non-sugarcane classes [17], [25].

### 3.4. Discussion

Based on the performance evaluation table, the SVM–AdaBoost ensemble model achieved the best results, with an accuracy of 0.9319, precision of 0.9323, recall of 0.9319, and an F1 score of 0.9320. Although the performance difference compared to the single model is relatively small, the continuous improvement across all metrics indicates increased classification stability. Theoretically, this can be explained by the basic characteristics of SVM, which focus on finding the optimal margin as a class separator, and by AdaBoost's ability to adjust weights for difficult-to-classify samples. The combination of both models in an ensemble approach helps reduce bias and variance errors more evenly than using only one model.

Analysis of the confusion matrix and ROC-AUC curve also reinforces these results, showing that the ensemble model has a better balance between false positives and false negatives and higher overall discriminatory power. This shows that the ensemble approach not only improves accuracy at a specific decision threshold but also maintains consistent performance across various classification thresholds. However, this relatively small numerical improvement needs to be critically reviewed, as in real-world applications its practical impact depends heavily on data size and the complexity of image variation in the field.

On the other hand, this study still has limitations, especially given the narrow scope of the dataset, so the model's ability to be applied across different environmental conditions cannot be fully ascertained. Variations in lighting, shooting angles, and plant growth phases can affect the model's performance when used in more complex situations. In addition, hyperparameter settings have not been thoroughly examined using a more comprehensive optimization approach. Therefore, future research should use larger, more diverse datasets, apply stricter cross-validation, and consider more advanced deep learning methods or ensemble strategies to improve the model's robustness and adaptability across diverse field situations.

## 4. CONCLUSION

This study evaluates and compares the performance of Support Vector Machine (SVM), AdaBoost, and the SVM–AdaBoost ensemble for classifying sugarcane and non-sugarcane leaves. The experimental results show that all models have ROC-AUC values of 0.9671 for SVM, 0.9622 for AdaBoost, and 0.9724 for ensemble. All values are above 0.96, indicating excellent ability to distinguish between classes. The ensemble model achieved the best results with an accuracy of 93.19%, precision of 93.23%, recall of 93.19%, and F1-score of 93.20%. These results are slightly better than those of models that only use one type. In addition, the ensemble model produced the fewest classification errors, namely 19 (FP + FN), compared to SVM with 20 errors and AdaBoost with 23 errors. This shows that the ensemble model is more effective in predicting data in this study.

These findings indicate that combining a margin-based approach with an adaptive boosting mechanism helps improve stability in the classification process. Although the improvement in results is not particularly large numerically, consistent results across various assessments suggest that the ensemble approach may be a stronger option for classifying agricultural images. Further research should examine the extent to which the model can be applied to larger, more diverse datasets and determine whether deep learning-based methods can be combined to make the model more robust to changes in environmental conditions.

## REFERENCES


[1] B. A. B. Ii and K. Pustaka, “Jurnal Tebu,” pp. 8–22, 2012.

[2] S. D. Chiatra, H. Sabita, J. Z. A. Pagar, A. No, G. Meneng, and K. Rajabasa, “Deteksi Objek Daun Tebu Dengan Menggunakan Metode Klasifikasi Pada Machine Learning,” *Semin. Nas. Has. Penelit. dan Pengabdi. Masy. 2025*, pp. 113–124, 2025, [Online]. Available: https://journal.darmajaya.ac.id/index.php/PSND/article/view/906

[3] S. Ananda, B. Nasution, D. Lestari, D. P. Azzahra, and D. Kiswanto, “Deteksi jenis tanaman berdasarkan bentuk daun menggunakan knn,” *J. Sist. Inf. dan Sist. Komput.*, vol. 10, no. 1, pp. 31–38, 2025, [Online]. Available: https://doi.org/10.51717/simkom.v10i1.646

[4] A. Wantoro *et al.*, “Implementasi Algoritma Machine Learning untuk Deteksi Penyakit Daun Tebu : Analisis Perbandingan Kinerja Abstrak Serangan penyakit daun seperti Penyakit Karat Daun ( Puccinia melanocephala ), Penyakit Bercak Daun ( Cercospora longipes ), dan Penyakit Moza,” *Semin. Nas.*

*Pembang. dan Pendidik. Vokasi Pertan.*, no. November, pp. 1683–1690, 2025, [Online]. Available: https://doi.org/10.51717/simkom.v10i1.646

[5] R. K. Dubey and D. K. Choubey, "Detection of Diseases in Rice Plant Using Optimized AdaBoost Classifier," *Int. J. Swarm Intell. Evol. Comput.*, vol. 14, no. 1000429, pp. 1–9, 2025, doi: 10.35248/2090-4908.25.14.429.

[6] Trivusi, "Algoritma AdaBoost: Pengertian, Cara Kerja, dan Kegunaannya," Trivusi. [Online]. Available: https://www.trivusi.web.id/2023/07/algoritma-adaboost.html

[7] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Jun. 2005, pp. 886–893 vol. 1. doi: 10.1109/CVPR.2005.177.

[8] Y. Wang, X. Zhu, and B. Wu, "Automatic detection of individual oil palm trees from UAV images using HOG features and an SVM classifier," *Int. J. Remote Sens.*, vol. 40, no. 19, pp. 7356–7370, 2019, doi: 10.1080/01431161.2018.1513669.

[9] S. Supiyandi, M. S. Hasibuan, and H. Harahap, "Penerapan Metode Haar-Like Feature Dan Algoritma Adaboost Dalam Penentuan Klasifikasi Hama Tanaman Kopi," *Smart Educ.*, vol. Vol 7, No, 2024, [Online]. Available: https://garuda.kemdiktisaintek.go.id/documents/detail/4627699

[10] F. Elena, R. Irawan, and B. Yong, "Application Of The Support Vector Machine , Light Gradient Boosting Machine , Adaptive Boosting , And Hybrid Adaboost-Svm Model On Customers Churn Data," *BAREKENG J. Math. Its Appl.*, vol. 19, no. 3, pp. 1957–1972, 2025, doi: https://doi.org/10.30598/barekengvol19iss3pp1957-1972.

[11] U. N. Fulari, R. K. Shastri, and A. N. Fulari, "Leaf Disease Detection Using Machine Learning," *J. Seybold Rep.*, vol. 15, no. 9, 2020.

[12] A. Syaputra, "Klasifikasi Penyakit Daun pada Tebu dengan Pendekatan Algoritma K-Nearest Neighbors , Multilayer Perceptron dan Support Vector Machine," *J. Ilm. Inform. Glob.*, vol. 15, no. 3, 2024.

[13] S. Choudhary and B. Saxena, "Image-Based Crop Disease Detection using Machine Learning Approaches : A Survey," *Int. J. Performability Eng.*, vol. 19, no. 2, pp. 122–132, 2023, doi: 10.23940/ijpe.23.02.p5.122132.

[14] A. Upadhyay, N. Singh, C. Krishna, and P. Singh, "Deep learning and Computer Vision in Plant Disease Detection : A Comprehensive Review of Techniques , Models , and Trends in Precision Agriculture," *ArtificialIntelligenceReview*, vol. 58, 2025.

[15] Y. Prabowo and K. N. Nasahara, "Detecting and Counting Coconut Trees in Pleiades Satellite Imagery Using Histogram of Oriented Gradients and Support Vector Machine," *Int. J. Remote Sens. Earth Sci.*, vol. 16, no. 1, p. 87, 2019, doi: 10.30536/j.ijreses.2019.v16.a3089.

[16] C. Cortes and V. Vapnik, "Support-Vector Networks," *MachineLearning*, vol. 20, no. 3, pp. 273–297, 1995.

[17] N. Khalid and N. A. Shahrol, "Evaluation the Accuracy of Oil Palm Tree Detection Using Deep Learning and Support Vector Machine Classifiers," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 1051, no. 1, 2022, doi: 10.1088/1755-1315/1051/1/012028.

[18] Y. Zheng *et al.*, "Development of a Phenology-Based Method for Identifying Sugarcane Plantation Areas in China Using High-Resolution Satellite Datasets," *Remote Sens.*, vol. 14, no. 5, 2022, doi: 10.3390/rs14051274.

[19] I. A. Elfitrianna and R. Prathivi, "Komparasi Metode SVM dan Adaboost untuk Klasifikasi Kanker Payudara," *J. Transform.*, vol. 22, no. 2, pp. 132–139, 2025.

[20] Z.-H. Zhou, *Ensemble methods : foundations and algorithms*. in Chapman & Hall/CRC machine learning & pattern recognition series. Boca Raton, FL: Taylor & Francis, 2012.

[21] F. I. Komputer and U. A. Yogyakarta, "Penerapan Metode Ensemble Untuk Meningkatkan Kinerja Algoritme Klasifikasi Pada Imbalanced Dataset," *TEKNOINFO*, vol. 13, no. 1, pp. 11–16, 2019.

[22] M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, "Ensemble deep learning : A review," vol. 3, 2022.

[23] D. M. . Powers, "Evaluation : From Precision , Recall And F-Measure To Roc , Informedness , Markedness & Correlation," *J. Mach. Learn. Technol.*, pp. 37–63, 2020.

[24] I. P. Carrascosa, "ROC AUC vs Precision-Recall for Imbalanced Data," Practical Machine Learning. Accessed: Sep. 09, 2025. [Online]. Available: https://machinelearningmastery.com/roc-auc-vs-precision-recall-for-imbalanced-data/

[25] W. Irmayani and K. Kunci, "Visualisasi Data Pada Data Mining Menggunakan Metode Klasifikasi Diterima : Diterbitkan :," *J. Khatulistiwa Inform.*, vol. IX, no. I, pp. 68–72, 2021.