Journal of Intelligent Decision Support System (IDSS) 8 (3) (2025), pp. 154-161

Published by: Institute of Computer Science (IOCS)

Journal of Intelligent Decision Support System (IDSS)

Journal homepage: www.idss.iocspublisher.org

Naïve bayes classification for oil palm leaf disease based on color and texture features

Alvin Yoga Kesuma[1], Natasya Ate Malem Bangun[2], Meida Cahyo Untoro[3]

[1,2,3]*Teknik Informatika, Institut Teknologi Sumatera, Bandar Lampung, Indonesia*

Article Info

Article history:

Received Aug 26, 2025
Revised Sep 2, 2025
Accepted Sep 16, 2025

Keywords:

Feature Extraction;
Image Processing;
Machine Learning;
Naïve Bayes;
Oil Palm Disease.

ABSTRACT

This study presents a comparison between standard Naïve Bayes classifier and its Genetic Algorithm-optimized variant for automated classification of oil palm leaf diseases. The system incorporates RGB color features alongside texture features extracted using the Gray Level Co-Occurrence Matrix. A dataset of of 225 JPG images of oil palm leaves, divided into training and testing sets in an 80:20 split is used. The methodology consisted of preprocessing, feature extraction, and classification. In the preprocessing phase, images were manually cropped, resized to 256 × 256 pixels, and background elements were removed. Feature extraction was then performed to obtain RGB color values and GLCM-based texture values, including contrast, correlation, energy, and homogeneity. Classification was conducted using two variants of the Naïve Bayes algorithm: one with default parameters and another optimized via GA for the Laplace smoothing hyperparameter. Model performance was assessed using a confusion matrix, with accuracy, precision, and recall serving as the primary evaluation metrics. Experimental results showed that both models achieved identical performance, with an accuracy of 51%, a precision of 52%, and a recall of 51%. These findings suggest that the Naïve Bayes classifier, even in its baseline form, demonstrates low discriminative performance for oil palm leaf disease detection, and when enhanced through GA-based optimization, it still provides only limited effectiveness. Therefore, this research highlights the need to pursue alternative methodologies, such as deep learning techniques or the adoption of more discriminative feature representations, aimed at improving both the accuracy and robustness of image-based disease detection in agriculture.

This is an open access article under the CC BY-NC license.

Corresponding Author:

Alvin Yoga Kesuma,
Program Studi Teknik Informatika,
Institut Teknologi Sumatera,
Jl. Terusan Ryacudu, Way Huwi, Kec. Jati Agung, Kab. Lampung Selatan, Lampung, 35365, Indonesia
Email: alvinyoga33@gmail.com

1. Introduction

Oil palm (Elaeis guineensis Jacq.) constitutes a cornerstone of Indonesia's plantation industry, exerting a pivotal influence on the national economy. As a major producer of vegetable oil used across various industries, the oil palm industry has expanded significantly, now spanning 22 provinces and exploiting numerous agro-ecologically suitable regions across the archipelago (Nurkholis & Sitanggang, 2020)(Puspitasari et al., 2022). This widespread cultivation has positioned Indonesia as the world's largest producer and exporter of palm oil, with a substantial portion of production also meeting domestic demand (Yuliani et al., 2019)(Abdul et al., 2022).

Despite its economic importance, oil palm cultivation faces significant challenges, particularly from leaf disorders that can severely impact plant health and productivity. These disorders arise from fungal infections, pests, genetic factors, or environmental stressors. One common disease is leaf spot, caused by the fungus Curvularia, which manifests as yellow spots that evolve into large (7–8 cm), oval, brown lesions with yellow margins and occasionally oily centers. Twisted leaves typically occur in seedlings due to improper planting techniques, leading to abnormal leaf curling above ground level. Wrinkled leaves result from genetic abnormalities that hinder the development of the central lamina, though this condition is generally mild. More concerning is leaf yellowing, primarily caused by Ganoderma boninense, a pathogenic fungus that can lead to plant death if not detected and managed early. Another notable condition is rolled leaves, characterized by curling along the central vein, which may stem from genetic traits, aphid infestations, or herbicide exposure (Afriliya, 2019). Collectively, these disorders underscore the need for timely and accurate diagnosis to ensure effective disease management and sustainable yield.

Early identification of leaf diseases is crucial for maintaining the health and productivity of oil palm plantations (Widians & Rizkyani, 2020). Recent technological advances have enabled the development of automated disease classification systems using machine learning models. These systems leverage digital image processing to analyze visual characteristics particularly texture and color from leaf images to detect and classify diseases (Komalasari et al., 2020). In this process, input images are transformed into meaningful output data through feature extraction and classification. Objects are grouped based on distinctive attributes, allowing for the differentiation between healthy and diseased leaves (Sholihin et al., 2017). In this study, we apply digital image processing techniques to distinguish between healthy and diseased oil palm leaves by analyzing their texture and color structures. Texture is evaluated using the Gray Level Co-Occurrence Matrix (GLCM) method, which captures spatial relationships between pixel intensities to reveal structural differences in leaf surfaces. Color analysis, on the other hand, is performed by examining RGB color values, enabling the detection of disease-related changes in pigmentation and overall leaf condition (Amatullah et al., 2021).

Among various classification methods, the Naïve Bayes algorithm has demonstrated strong performance in plant disease identification tasks involving color and texture features (Fansyuri & Hariansyah, 2020). Its straightforward design and strong performance in object classification, especially when dealing with objects that have independent features, make it well suited for this application. (Afriansyah et al., 2024). For instance, Ardi et al. employed Naïve Bayes to identify diseases in tomato plants using GLCM and Naïve Bayes, achieving an accuracy of 80% with a test dataset of 15 tomato plant image disease images (Nainggolan et al., 2022). Similarly, Felicia et al. compared Naïve Bayes and K-Nearest Neighbors (KNN) for apple classification, using Local Binary Pattern (LBP) for texture and HSV color space for color features. With 100 images across five apple types, Naïve Bayes outperformed KNN, reaching 97% accuracy versus 82% (Febriana et al., 2021). In another study, Gracia et al. applied Naïve Bayes and K-Nearest Neighbors (KNN) to classify edible and poisonous mushrooms using a large dataset of 8,124 images. The model achieved an accuracy of 92%, demonstrating its effectiveness in differentiating between edible and poisonous mushroom (Batubara et al., 2023).

Building on these findings, this study adopts the Naïve Bayes classifier for identifying oil palm leaf diseases based on texture and color features extracted from digital images. Unlike deep learning approaches such as Convolutional Neural Networks (CNNs), which require large annotated datasets and high computational power for training, Naïve Bayes is particularly suitable for studies with limited sample sizes and constrained resources, which are conditions often encountered in agricultural research, including oil palm disease detection. Moreover, while ensemble methods like Random Forest can achieve high accuracy, they are more prone to overfitting if not carefully tuned and offer less interpretability compared to probabilistic models. In contrast, Naïve Bayes provides a simple, fast, and statistically sound classification process that performs well with structured feature inputs such as color (RGB) and texture (GLCM), making it an efficient choice for this application.

Model performance will be rigorously evaluated using a Confusion Matrix, which provides a detailed breakdown of classification outcomes by identifying True Positives, True Negatives, False Positives, and False Negatives. This enables the calculation of key performance metrics such as accuracy, precision, and recall, offering a comprehensive assessment of the model's predictive capability (Normawati & Prayogi, 2021). To further enhance classification accuracy, we integrate the Genetic Algorithm (GA) as an optimization technique. GA will be used to fine-tune the smoothing hyperparameter (also known as Laplace smoothing) in the Naïve Bayes algorithm, which is a critical factor influencing model performance, especially in the presence of sparse data. By simulating evolutionary processes such as selection, crossover, and mutation, GA efficiently searches the parameter space to identify optimal values. This optimization improves the model's adaptability to variations in leaf appearance and enhances its generalization ability, ultimately leading to more accurate and reliable disease classification (Tahir & Sugianto, 2024).

While previous studies have applied machine learning to plant disease detection, most focus on crops other than oil palm or use basic classification models without optimization. Research on oil palm leaf diseases often relies on manual diagnosis or standard classifiers with fixed parameters, limiting accuracy and adaptability. This study aims to advance automated detection by implementing and comparing two variants of the Naïve Bayes classifier: a baseline model with default parameters, and a version optimized using a Genetic Algorithm to improve model accuracy. This work contributes to more accurate, automated detection of oil palm leaf disorders, supporting timely disease management and improved plantation productivity.

2. Method

This study employs a systematic approach to develop and optimize a classification model for identifying diseases in oil palm leaves, leveraging the Naïve Bayes algorithm enhanced with optimization techniques. The research process is structured into five key phases, as outlined in Figure 1: Research Flow.

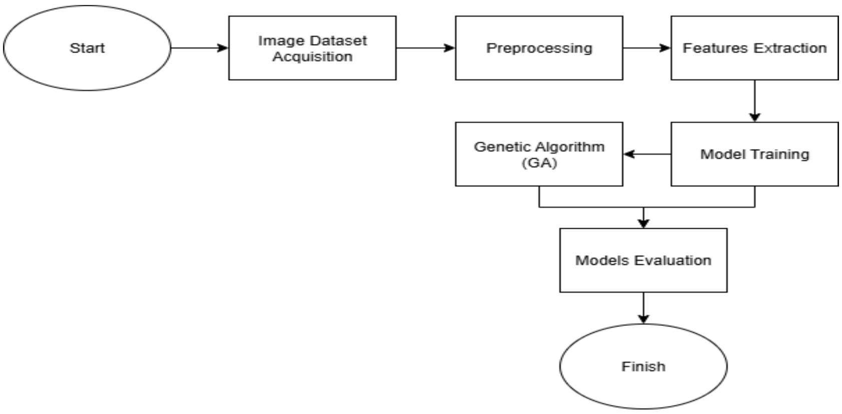


Figure 1 Research flow

Image Dataset Acquisition

This research begins with the data collection stage, which involves acquiring images of oil palm leaves affected by five types of diseases: leaf spot, yellowing leaves, curling leaves, twisting leaves, and wrinkled leaves. The images are captured using a smartphone to obtain samples of leaves infected with various diseases. The total dataset consists of 225 images in JPG format with an original resolution of 960 x 1280 pixels. The dataset is then divided into 80% training data and 20% testing data (Suryani et al., 2021).

Table 1 Oil palm leaves images dataset

Disease	Dataset	Disease	Dataset	Disease	Dataset
Leaf Spot		Twisting Leaves		Curling Leaves	

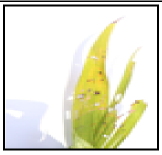
Disease	Dataset	Disease	Dataset	Disease	Dataset
Yellowing Leaves		Wrinkled Leaves			

Image Dataset Preprocessing

After data collection, the preprocessing stage is carried out to ensure the quality and consistency of the images used for classification. The process begins with manual cropping of the original images to focus exclusively on the area of interests which are the plants leaves (Wardhani et al., 2021). All images are then resized to 256 × 256 pixels to standardize the dataset size (Kurniatin et al., 2023). The images were resized to 256×256 pixels to strike a balance between computational efficiency and visual fidelity which offers significantly faster processing times while still retaining sufficient detail for accurate disease classification, as fine-grained texture or microscopic features are not required for distinguishing macroscopic symptoms like leaf spots, curling, or yellowing. The final preprocessing step is image segmentation, performed using the Python rembg module to remove the background from the leaves. This results in a standardized, noise-free dataset optimized for the subsequent classification process.

Features Extraction

The next stage is feature extraction, where both color and texture characteristics of each image are extracted to form a meaningful representation for subsequent processing in the model (Pahlevi et al., 2021). For color features, the average intensity of the red, green, and blue channels in each image is calculated. These averages are then normalized by dividing each color's average by the sum of all three channels. The resulting normalized values always add up to one and indicate the relative proportion of red, green, and blue in the image, which ensures that the color signature remains consistent under different lighting conditions (Afriansyah et al., 2023). For texture features, the Gray Level Co-occurrence Matrix (GLCM) is used to capture spatial relationships between pixel intensities. From this matrix, four measures are derived: contrast, which reflects how much neighboring pixels differ in intensity; correlation, which measures the degree of linear relationship between pixel values; energy, which indicates the uniformity or repetitiveness of texture patterns; and homogeneity, which shows how similar neighboring pixels are to each other (Yasmin et al., 2024).

Model Training

Following the feature extraction process, the subsequent stage involved model training using the Naïve Bayes classifier. The Naïve Bayes algorithm, which is a probabilistic classification approach grounded in Bayes' theorem introduced by Thomas Bayes, estimates the probability of a given class based on a set of observed features (Roring et al., 2022)(Muchtar et al., 2024). It assumes that all features are conditionally independent given the class label, allowing for straightforward and efficient computation. This method is particularly effective for classification tasks with numerical data and performs well even with limited training samples. It leverages prior probabilities and likelihood estimates to compute the posterior probability of each class, enabling accurate predictions.

In this study, Naïve Bayes was implemented for numerical datasets by first calculating the average (mean) value of each feature for every class, which represents the typical value of that feature. Next, the spread of the data around this mean was measured using the standard deviation. With these two values, the probability of observing a given feature in a class was estimated using a Gaussian distribution, which assumes the data follows a bell-shaped curve. Finally, the algorithm combined these probabilities with the prior probability of each class to determine the posterior probability. The class with the highest posterior probability was then selected as the prediction result.

Model Optimization using Genetic Algorithm (GA)

To improve the performance of the Naïve Bayes classifier, the Laplace smoothing (α) hyperparameter is optimized using the Genetic Algorithm (GA) [16]. Laplace smoothing works by adding a small constant value, α, to the probability calculation so that no feature is assigned a zero probability when it does not appear in the training data. In this study, instead of relying on a fixed α, GA is employed to automatically search for the most effective value. The optimization is carried out by creating multiple instances of the Naïve Bayes model, each with a different α value, and evaluating their accuracy as the fitness score. Through iterative processes of selection, crossover, and mutation over several generations, GA converges on the optimal α. The purpose of this process is to obtain the best Laplace smoothing hyperparameter value, which is then used to train the final optimized Naïve Bayes model.

Model Evaluation

The performance evaluation of the Naïve Bayes classification model in identifying oil palm leaf diseases is carried out using a Confusion Matrix, which provides a detailed breakdown of classification outcomes through the components: True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN) (Khasanah et al., 2021). Two models are evaluated and compared. The first is the original Naïve Bayes model that uses a default smoothing parameter. The second is an optimized Naïve Bayes model in which the Laplace smoothing parameter (α) is tuned using the Genetic Algorithm (GA). This comparison aims to assess the improvement in classification performance achieved through hyperparameter optimization. Each model was evaluated using key performance metrics, including accuracy, precision, and recall, which were derived from the confusion matrix. The results are analyzed to determine whether the GA-optimized model enhances the classification accuracy and reliability in detecting oil palm leaf diseases compared to the standard Naïve Bayes model.

Table 2 Confusion matrix for naïve bayes classification

		Real Values	
		True	False
Predicted Values	True	TP *(True Positive)*	FP *(False Positive)*
	False	FN *(False Negative)*	TN *(True Negative)*

3. Results and Discussions

This research produced several model evaluations for important findings in the classification of oil palm leaf diseases using the optimized Naïve Bayes method. The results and discussion are divided into several sub-sections, including:

Confusion Matrix Analysis

The classification results of oil palm leaf diseases are analyzed in detail using the confusion matrix presented in Table 3. This matrix provides a comprehensive view of the model's performance by comparing actual and predicted disease classes.

Table 2 Confusion matrix of classification results

Actuall vs Pred	Leaf Spot	Wrinkled Leaves	Twisting Leaves	Curling Leaves	Yellowing Leaves
Leaf Spot	6	0	1	0	2
Wrinkled Leaves	1	2	0	6	0
Twisting Leaves	0	0	4	4	1
Curling Leaves	0	3	2	4	0
Yellowing Leaves	2	0	0	0	7

From this confusion matrix, several patterns of classification errors can be observed. The Leaf Spot and Yellowing Leaves categories show relatively high classification accuracy, while Wrinkled Leaves and Curling Leaves exhibit significant mutual confusion, which suggest that their color and texture features may be visually or statistically similar, making them harder to distinguish using the current feature set. The Twisting Leaves class also shows notable misclassification into Curling Leaves, reinforcing the likelihood of overlapping visual characteristics among deformation-based symptoms.

Notably, the confusion matrix presented above applies to both the original Naïve Bayes classifier and the improved version that uses Genetic Algorithm (GA) to optimize the Laplace smoothing parameter (α). Despite the GA-based optimization process successfully identifying an optimal α value, the resulting classification predictions were identical to those of the standard Naïve Bayes model with default smoothing. This suggests that Laplace smoothing may not be the limiting factor in this classification task, possibly because the dataset size or feature distributions already mitigate zero-probability issues, or because the optimal α found by GA was numerically close to the default value.

Performance Matrix Analysis

Based on the confusion matrix, key performance metrics such as precision, recall, and F1-score were computed for each disease category to evaluate the model's effectiveness in detail as listed in Table 4. These metrics offer deeper insight beyond overall accuracy, highlighting where the classifier excels or struggles in balancing false positives and false negatives.

Table 3 Performance metrics per disease category

Disease Category	Precision	Recall	F1-Score
Leaf Spot	0,67	0,67	0,67
Wrinkled Leaves	0,40	0,22	0,29
Twisting Leaves	0,57	0,44	0,50
Curling Leaves	0,29	0,44	0,35
Yellowing Leaves	0,70	0,78	0,74

The Yellowing Leaves category achieves the strongest performance, with the highest F1-score (0.74), indicating a good balance between correctly identifying true cases and minimizing misclassification of other classes as Yellowing. Leaf Spot follows with a moderate F1-score of 0.67. In contrast, the Wrinkled Leaves category shows the weakest results (F1-score = 0.29), driven by both low precision and low recall which suggest that the model frequently mislabels other classes as Wrinkled Leaves, while also failing to recognize most actual Wrinkled samples. This aligns with the confusion matrix, which revealed heavy misclassification between Wrinkled and Curling Leaves.

Twisting and Curling Leaves show intermediate but still suboptimal F1-scores (0.50 and 0.35, respectively), further confirming the model's difficulty in distinguishing between morphologically similar leaf deformations. Overall, the model achieves 51% accuracy, with average precision and recall both at 51%, indicating no strong bias toward over- or under-prediction across classes, but also confirming limited discriminative power overall.

This relatively low accuracy contrasts with the strong performance of Naïve Bayes reported in other plant and agricultural classification studies, such as 80% in tomato disease detection (Nainggolan et al., 2022), 97% in apple type classification (Febriana et al., 2021), or 92% in mushroom edibility recognition (Batubara et al., 2023). Those tasks typically involved diseases or objects with visually distinct color patterns, sharp texture boundaries, or categorical differences that RGB and GLCM features can effectively capture. In contrast, oil palm leaf deformations, particularly curling, twisting, and wrinkling, all exhibit highly overlapping visual characteristics that are poorly differentiated by basic color and statistical texture descriptors. This suggests that the limitation lies not in the classifier itself, but in the mismatch between the complexity of the symptoms and the simplicity of the features used which is a challenge less pronounced in prior studies where symptom categories were more visually separable.

Optimization Method Analysis

The performance of the original Naïve Bayes classifier is compared with its enhanced version that incorporates Laplace smoothing parameter optimization using the Genetic Algorithm (GA). As shown in Table 5, both models achieve identical results across all evaluation metrics.

Table 4 Performance comparison between original naïve bayes and ga-optimized naïve bayes

Metrik	Naïve Bayes	Naïve Bayes with GA
Total Accuracy	51%	51%
Macro Precision	52%	52%
Macro Recall	51%	51%
F1-Score	51%	51%

The complete agreement in performance indicates that the optimal α value identified by GA did not yield any improvement over the default smoothing setting. This suggests that, for this dataset and feature set, the baseline Naïve Bayes model is already operating near its performance ceiling with respect to Laplace smoothing. It is likely that the probability estimates are relatively stable across a range of α values, possibly due to the balanced class distribution in the training set and the limited dimensionality of the feature space, which reduce the sensitivity of the model to smoothing adjustments.

More importantly, this outcome implies that the primary constraints on classification performance lie elsewhere: in the Naïve Bayes assumption of feature independence, the simplicity of the probabilistic model, or most critically, the limited discriminative power of the input features (RGB color and GLCM texture). These limitations are further underscored when comparing our 51% accuracy to the 80–97% reported in prior Naïve Bayes studies involving more separable visual classes or richer feature representations.

4. Conclusions

This study investigated the application of the Genetic Algorithm (GA) to optimize the Laplace smoothing parameter in the Naïve Bayes classifier for the task of oil palm leaf disease classification. The results show that, despite the automated search for an optimal α value, the GA-optimized model achieves performance identical to that of the original Naïve Bayes classifier. Both models attain 51% accuracy and match in precision, recall, and F1-scores across all classes. This outcome suggests that the default smoothing setting is already effective for this dataset, and further tuning offers no advantage. It indicates that performance bottlenecks lie beyond hyperparameter selection.

The primary limitations appear to be related to the feature representation and the inherent assumptions of the Naïve Bayes algorithm. The use of only color (RGB) and texture (GLCM) features proves insufficient to distinguish between diseases with similar visual symptoms, such as Wrinkled and Curling Leaves. This pattern is clearly reflected in both the confusion matrix and the low F1-scores for these categories. These findings confirm that the model’s inability to resolve fine-grained visual differences stems not from suboptimal smoothing, but from the limited discriminative power of the input features and the classifier’s assumption of feature independence.

Theoretically, this research contributes to the plant disease classification literature by demonstrating that, in low-dimensional feature spaces with visually ambiguous classes, hyperparameter optimization, even via advanced methods like GA, may yield no measurable gain if the underlying model and feature set lack sufficient discriminative capacity. This contrasts with prior Naïve Bayes studies reporting high accuracy (80–97%) on diseases with more visually distinct symptoms, where basic color and texture features were more effective. This insight challenges the common assumption that performance gaps can be resolved through parameter tuning alone. Instead, it redirects attention to representation learning and model architecture as the primary levers for improvement, a principle that holds across similar domains in agricultural image analysis.

Therefore, future work should focus on enriching the feature set with morphological descriptors, such as leaf shape, margin deformation, or vein patterns, that can better capture subtle differences between disease types. Additionally, transitioning to more powerful classification frameworks, such as deep learning models (e.g., CNNs) or ensemble methods (e.g., Random Forest, XGBoost), may yield better performance by capturing complex, non-linear relationships in the data. Expanding the dataset in size and diversity, including variations in lighting, background, and growth stage, will also enhance model generalization and robustness.

In summary, while GA-based optimization is a valuable tool for automating parameter selection, its effectiveness depends on the context. For this particular classification task, improvements are more likely to come from advances in feature engineering and model complexity rather than hyperparameter tuning alone. This insight underscores the importance of aligning optimization strategies with the specific characteristics and limitations of the dataset and problem domain. It redirects future efforts toward more impactful areas of development: richer representations, stronger models, and more diverse data.

References

Abdul, I., Wulan Sari, D., Haryanto, T., & Win, T. (2022). Analysis of factors affecting the technical inefficiency on Indonesian palm oil plantation. *Scientific Reports*, *12*(1), 3381.

Afriansyah, M., Saputra, J., Ardhana, V. Y. P., & Sa'adati, Y. (2024). Algoritma Naive Bayes Yang Efisien Untuk Klasifikasi Buah Pisang Raja Berdasarkan Fitur Warna. *Journal of Information Systems Management and Digital Business*, *1*(2), 236–248.

Afriansyah, M., Saputra, J., Sa'adati, Y., & Ardhana, V. Y. P. (2023). Optimasi Algoritma Nai? ve Bayes Untuk Klasifikasi Buah Apel Berdasarkan Fitur Warna RGB. *Bulletin of Computer Science Research*, *3*(3), 242–249.

Afriliya, F. (2019). Keanekaragaman Jenis-Jenis Penyakit dan Cara Pengendaliannya di Pembibitan Kelapa Sawit (Elaeis Guinensis Jacq) PT. Perkebunan Nusantara I Langsa. *Biologica Samudra*, *1*(1), 34–40.

Amatullah, L., Ein, I., & Santoni, M. M. (2021). Identifikasi Penyakit Daun Kentang Berdasarkan Fitur Tekstur dan Warna Dengan Menggunakan Metode K-Nearest Neighbor. *Prosiding Seminar Nasional Mahasiswa Bidang Ilmu Komputer Dan Aplikasinya*, *2*(1), 783–791.

Batubara, G. M. C., Desiani, A., & Amran, A. (2023). Klasifikasi Jamur Beracun Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbors. *Jurnal Ilmu Komputer Dan Informatika*, *3*(1), 33–42.

Fansyuri, M., & Hariansyah, O. (2020). Pengenalan Objek Bunga dengan Ekstraksi Fitur Warna dan Bentuk Menggunakan Metode Morfologi dan NaÃ ve Bayes. *Jurnal Sistem Dan Informatika (JSI)*, *15*(1), 70–80.

Febriana, F., Riva, L. S., Piero, R. S. S., Ikramsyah, M. A., & Santoni, M. M. (2021). Perbandingan Klasifikasi Naive-Bayes dan KNN untuk Mengidentifikasi Jenis Buah Apel dengan Ekstraksi Ciri LBP dan HSV. *Prosiding Seminar Nasional Mahasiswa Bidang Ilmu Komputer Dan Aplikasinya*, *2*(2), 191–201.

Khasanah, N., Komarudin, R., Afni, N., Maulana, Y. I., & Salim, A. (2021). Skin cancer classification using random forest algorithm. *Sisfotenika*, *11*(2), 137–147.

Komalasari, D., Perlindungan, I., & Risnawati, R. (2020). *Pengenalan Tanaman Cabai dengan Teknik Klasifikasi Menggunakan Metode CNN*.

Kurniatin, L., Ariyanto, D. K. A., & Saputra, B. A. (2023). Analisis Perbandingan Hasil Pengolahan Citra Asli Dan Cropping Untuk Mengidentifikasi Karakteristik Tanaman Selada Menggunakan Metode Morfologi Dan Ekstrasi Ciri. *Jurnal Ilmiah SINUS*, *21*(1), 73–82.

Muchtar, M., Pasrun, Y. P., Rasyid, R., Miftachurohmah, N., & Mardiawati, M. (2024). Penerapan metode Naïve Bayes dalam klasifikasi kesegaran ikan berdasarkan warna pada citra area mata. *Jurnal Informatika Dan Teknik Elektro Terapan*, *12*(1).

Nainggolan, A., Rumapea, H., Silalahi, A. P., Sidauruk, L., & Sinambela, M. (2022). Identifikasi penyakit tanaman tomat berdasarkan citra penyakit menggunakan metode GLCM dan Naïve Bayes classifier. *J. Ilm. Tek. Inform*, *2*(1), 22–28.

Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes classifier dan confusion matrix pada analisis sentimen berbasis teks pada Twitter. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, *5*(2), 697–711.

Nurkholis, A., & Sitanggang, I. S. (2020). Optimalisasi model prediksi kesesuaian lahan kelapa sawit menggunakan algoritme pohon keputusan spasial. *Jurnal Teknologi Dan Sistem Komputer*, *8*(3), 192–200.

Pahlevi, T. R., Buaton, R., & Nurhayati, N. (2021). Identifikasi Jenis Bunga Menggunakan Ekstraksi Ciri Orde Satu Dan Algoritma Multi Support-Vector Machines (MULTISVM). *Jurnal Informatika Kaputama (JIK)*, *5*(1), 116–128.

Puspitasari, N., Rosmasari, R., Pratama, F. W., & Sulastri, H. (2022). Quality Classification of Palm Oil Varieties Using Naive Bayes Classifier. *Digital Zone: Jurnal Teknologi Informasi Dan Komunikasi*, *13*(1), 11–23.

Roring, C. B., Mulyana, D. I., Lubis, Y. T., & Zamzami, A. R. (2022). Klasifikasi Tingkat Kematangan Buah Jambu Bol Berdasarkan Warna Kulit Menggunakkan Metode Naïve Bayes. *Jurnal Pendidikan Tambusai*, *6*(1), 2938–2948.

Sholihin, M., Mujilahwati, S., & Wardhani, R. (2017). Classification of Batik Lamongan based on features of color, texture and shape. *Jurnal Ilmiah Kursor*, *9*(1).

Suryani, D., Yulianti, A., Maghfiroh, E. L., & Alber, J. (2021). Quality classification of palm oil products using Naïve Bayes method. *Sistemasi: Jurnal Sistem Informasi*, *11*(1).

Tahir, S. F., & Sugianto, C. A. (2024). Optimasi Naive Bayes Menggunakan Algoritma Genetika Pada Klasifikasi Komentar Cyberbullying Pada Media Sosial X. *Jurnal Informatika Dan Teknik Elektro Terapan*, *12*(3).

Wardhani, I. P., Putri, A. M., & Widayati, S. (2021). Algoritma Identifikasi Ciri Citra Pegunungan dengan Metode Copping. *Jurnal Ilmiah Komputasi*, *20*(2), 283–290.

Widians, J. A., & Rizkyani, F. N. (2020). Identifikasi hama kelapa sawit menggunakan metode certainty factor. *Ilk. J. Ilm*, *12*(1), 58–63.

Yasmin, N., Akbar, S. C. D., & Ramadhanu, A. (2024). Penerapan K-Means clustering untuk klasifikasi citra cabai keriting: Studi ekstraksi warna dan tekstur GLCM. *Indonesian Journal Computer Science*, *3*(2), 65–71.

Yuliani, A., Labellapansa, A., & Yulianti, A. (2019). Klasifikasi Citra Daun Kelapa Sawit Yang Terkena Dampak Hama Menggunakan Metode K-Nearest Neighbor. *Seminar Nasional Informatika Medis (SNIMed)*, 73–78.