Journal of Applied Data Sciences
Vol. 6, No. 1, January 2025, pp. 34-46

Classification of Starling Images Using a Bayesian Network

April Lia Hananto[1,*], Aviv Yuniar Rahman[2], Tukino Paryono[3], Bayu Priyatna[4], Agustia Hananto[5], Baenil Huda[6]

[1,3,4,5,6]*Faculty of Computer Science, Universitas Buana Perjuangan Karawang, Karawang 41361, Indonesia*

[2]*Department of Informatics Engineering, Universitas Widyagama Malang, Malang 65142, Indonesia*

(Received: June 10; Revised: September 20; Accepted: October 17; Available online: December 27, 2024)

Abstract

The classification of starling species is vital for biodiversity conservation, especially as some species are endangered. This research investigates the effectiveness of the Bayesian Network (BayesNet) for classifying starling species and compares its performance with Artificial Neural Networks (ANN) and Naive Bayes models. The dataset comprises 300 images of five starling species—Bali, Rio, Moon, Kebo, and Uret—captured under controlled conditions. Feature extraction focused on color, texture, and shape, while data augmentation through slight image rotations was applied to enhance model generalization. The BayesNet model achieved an accuracy of 96.29% using a 90:10 training-to-testing split, outperforming ANN (90.74%) and Naive Bayes variants. Precision, recall, F1-score, and AUC-ROC values further validated the robustness of the BayesNet model, with precision at 0.90, recall at 0.91, F1-score at 0.92, and AUC-ROC at 0.95. These results demonstrate the superior performance of multi-feature Bayesian Networks in starling classification compared to other machine learning models. The novelty of this study lies in its application of a probabilistic approach using Bayesian Networks, which enhances interpretability and performance, especially in scenarios with limited data. Future work may explore additional feature sets and advanced machine learning models to further improve classification accuracy and robustness.

Keywords: Bayesian Network, Starling Classification, Machine Learning, Image Processing, Feature Extraction

1. Introduction

Indonesia is home to a wide variety of bird species, contributing significantly to global biodiversity[1]. The country hosts approximately 1,529 bird species, representing 17% of the world's avian population [2]. Among these, starling species are especially important due to their unique characteristics in terms of color, sound, and behavior [3]. However, the classification of starling species presents a significant challenge due to the morphological similarities between different species [4].

In previous studies, Artificial Neural Networks (ANN) have been widely applied for classifying starling species based on image data, utilizing features such as texture, shape, and color [5], [6]. ANN models achieved varying degrees of success, with accuracy levels of 49.20% for texture features, 58.14% for shape features, and 84.81% for color features [7]. Despite this, ANN models often face significant challenges, including their susceptibility to overfitting and the need for large datasets [8], [9]. For tasks like starling classification, where the dataset is often limited, these drawbacks make ANN a less effective solution [10].

Bayesian Networks (BayesNet), on the other hand, offer several advantages over ANN, particularly in their ability to manage uncertainty and provide interpretable relationships between variables [11], [12]. Bayesian Networks have been successfully applied in various object classification tasks, often achieving accuracy levels above 90% [13], [14]. Moreover, BayesNet is well-suited for tasks with smaller datasets, making it an ideal choice for starling classification [15]. However, while BayesNet has demonstrated strong performance in similar tasks, the idea of achieving 100% accuracy in real-world classification tasks remains unrealistic, particularly when dealing with complex image data [16]. This study seeks to address the limitations of BayesNet, focusing on improving classification accuracy while acknowledging the challenges associated with perfect performance [17]. Although this paper compares BayesNet with

***Corresponding author: April Lia Hananto author (aprilia@ubpkarawang.ac.id)**
DOI: https://doi.org/10.47738/jads.v6i1.423

ISSN 2723-6471

Naive Bayes, it is crucial to justify the selection of these models over more contemporary classifiers, such as Convolutional Neural Networks (CNN) or Random Forests. CNNs, for instance, typically require large datasets to perform optimally, and while they have been successful in various image classification tasks, their high computational complexity and data demands make them less suitable for this study, which is limited by the number of starling images available [18],[19], [20], [21]. Similarly, while Random Forests are known for their robustness, they can struggle with interpretability and are less effective when dealing with the uncertainties that arise in small datasets [22]. In contrast, BayesNet was chosen for its balance between interpretability, efficiency, and capability to manage uncertainty in smaller datasets [22].

ANN is included in this comparison because of its extensive use in previous starling classification studies, serving as a benchmark to measure the potential improvements provided by BayesNet [22]. However, the limitations of ANN, such as its high sensitivity to dataset size and risk of overfitting, make it a less ideal candidate for this specific task [23].

In conclusion, this study aims to improve the classification accuracy of starling species by leveraging the strengths of Bayesian Networks. By comparing the performance of BayesNet with ANN and Naive Bayes, this research seeks to provide a more reliable and interpretable classification model, which will contribute to conservation efforts and enhance the economic value of starling species [24].

2. Methodology

In this study, we classify starling bird species using various image processing and machine learning techniques. The primary stages of the proposed method include image data acquisition, data augmentation, image segmentation, feature extraction, and the application of machine learning models. Image data were collected using a DSLR camera under controlled conditions to generate a representative dataset of five starling species: Bali, Rio, Moon, Kebo, and Uret. This process was complemented by data augmentation to enhance the model's generalization ability, as well as manual segmentation to separate the bird objects from the background. The extracted features from the images included color, texture, and shape, which were subsequently used in the classification stage employing several machine learning techniques, including artificial neural networks and Bayesian networks. The model performance was evaluated based on the accuracy of species classification, determined by the correct and incorrect predictions made by the classifier (see figure 1 for the model classifying starling images using a Bayesian Network).

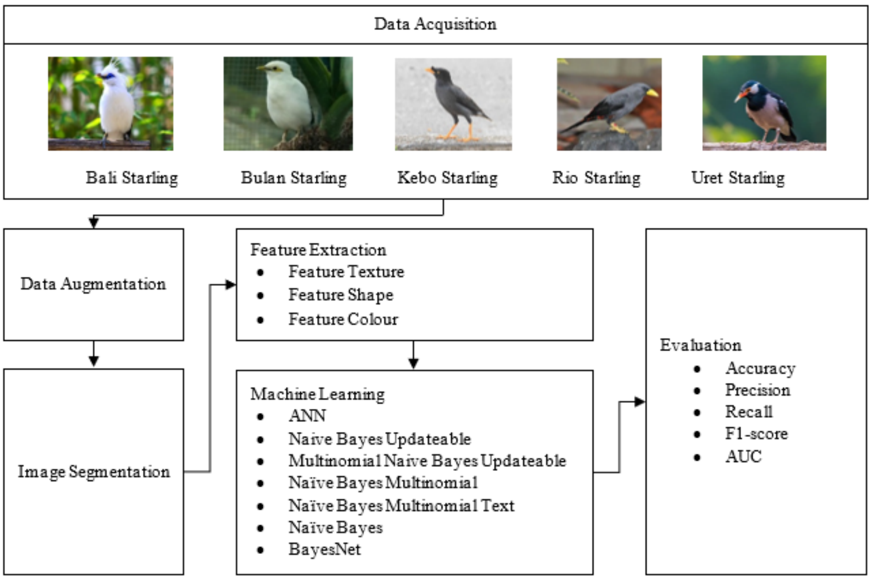

Figure 1. Starling image classification model using bayesian network

2.1. Data Acquisition

In this study, images of five species of starlings—Bali, Rio, Moon, Kebo, and Uret—were collected using a DSLR camera under controlled conditions. A total of 60 images per species were captured, resulting in a dataset of 300 images.

ISSN 2723-6471

The dataset was split evenly into training and testing sets, with 150 images allocated to each. Images were captured from multiple angles, including frontal, lateral, and top views, to increase variability and ensure comprehensive coverage of each species' morphological features.

2.2. Data Augmentation

To enhance the model's generalization ability, data augmentation was applied by rotating each image at incremental steps between 0° and 5°. This augmentation yielded an additional 300 data points, doubling the dataset size. This small rotation range was chosen to introduce subtle variations without deviating from real-world conditions, minimizing the risk of overfitting. The augmentation strategy was carefully designed to maintain a balance between increasing dataset size and preserving the integrity of the original data distribution.

2.3. Image Segmentation

Manual segmentation was initially employed to isolate the bird objects from the background, allowing the model to focus on relevant features. While this approach ensured high-quality segmentation for this study, we acknowledge the potential for human-induced bias. Future work will explore the use of automated and scalable segmentation methods, such as Mask R-CNN, to minimize human intervention and ensure consistency across the dataset. Automated segmentation will further enhance reproducibility and scalability in similar classification tasks.

2.4. Feature Extraction

In this study, feature extraction focused on three primary aspects: color, form, and texture. These features play a critical role in capturing the unique properties of each starling species and contribute significantly to the classification accuracy. Specifically, color features were extracted using the RGB values of the images to represent the color patterns of the birds.

Regarding the data augmentation process, slight rotations (between 0° and 5°) were applied to the images to introduce variability and improve model robustness. However, these small rotations did not affect the integrity of the color features. This is because color information remains invariant under small rotations, meaning that the pixel values representing color do not change when the image is rotated by such minimal angles. The purpose of the rotation was to enhance the generalization ability of the model without distorting the key features, such as color patterns, which are crucial for species classification.

Therefore, while rotation improves the robustness of the model by increasing the diversity of the training data, it does not significantly alter the extracted color features. This ensures that the classification performance remains consistent and reliable.

2.4.1. Feature Texture

In this study, three essential features—color, texture, and shape—were extracted from the images to aid in the classification of starling species. The color feature extraction process involved converting the RGB values to grayscale using the following equation (1):

$$\mathrm{Gray} = 0.2989{*}\mathrm{R} + 0.5870{*}\mathrm{G} + 0.1140{*}\mathrm{B} \tag{1}$$

where $R, G, and\ B$ represent the red, green, and blue channels of the image, respectively. This conversion is crucial for reducing the dimensionality of color data while preserving the perceptual relevance of the image.

For texture feature extraction, the Gray Level Co-occurrence Matrix (GLCM) was employed to compute several important texture descriptors. These descriptors provide a statistical measure of the spatial relationship between pixels in a grayscale image, and are defined as follows:

Energy, which quantifies the uniformity of the image texture, is calculated as the sum of squared elements in the GLCM (2):

$$\mathrm{Energy} = \sum_{\mathrm{k}} \mathrm{k}^2 \left[\sum_{\mathrm{i}} \sum_{\mathrm{j}} (\mathrm{i}, \mathrm{j})\right] \tag{2}$$

Correlation, which measures the linear relationship between pixels at specified positions relative to each other, is given by (3):

$$\mathrm{Correlation} = \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)p(i,j)}{\sigma_i \sigma_j} \tag{3}$$

Where $P\ (i,j)$ represents the normalized value of the GLCM, and $\mu_i, \mu_j, \sigma_i, \sigma_j$ are the means and standard deviations of pixel intensities for rows and columns, respectively.

Homogeneity, which assesses the closeness of the distribution of elements in the GLCM to the GLCM diagonal, is expressed as (4):

$$\mathrm{Homogeneity} = \sum_{i,j} \frac{p(i,j)}{1+|i-j|} \tag{4}$$

Energy, another measure of textural uniformity, is alternatively expressed as (5):

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

Entropy, which reflects the disorder or randomness of the texture, is defined as (6):

$$\mathrm{Entropy} = \sum_{i,j} P\ (i,j) \log P\ (i,j) \tag{6}$$

These texture descriptors, in conjunction with the color and shape features, are integral to improving the classification accuracy of starling species. While data augmentation techniques such as slight image rotations between 0° and 5° were applied, these rotations did not significantly affect the extracted features, as the variations introduced were designed to enhance model generalization without distorting key image properties.

2.4.2. Feature Shape

When comparing here, the distance values are influenced by the shape detail extraction method. The characteristic also distinguishes the starling object from other objects of its kind. This shape feature is evaluated using two variables: metric and eccentricity. In Equation 7, the letters A and b represent the principal and minor elliptical foci, respectively. The letter e indicates the eccentricity. Equation 8 uses A for M, which stands for metric, and C for a.

Shape features play a crucial role in distinguishing starling species based on their morphological differences. The extraction of shape features involves calculating specific geometric properties that capture the unique structure of each starling. In this study, we focused on two primary shape descriptors: eccentricity and metric. Eccentricity is a measure of how much an object's shape deviates from being circular, and is determined using the lengths of the major and minor axes of the object. The eccentricity e is calculated using the following equation (7):

$$e = \sqrt{1 - \frac{b^2}{a^2}} \tag{7}$$

where a and b are the lengths of the semi-major and semi-minor axes, respectively. An eccentricity value of 0 corresponds to a perfect circle, while values closer to 1 indicate a more elongated shape.

The second important descriptor is the metric, which relates to the compactness of the shape. The metric M is defined as (8):

$$\mathrm{M} = \frac{4\pi \times \mathrm{A}}{\mathrm{C}^2} \tag{8}$$

Where A is the area of the shape and C is the perimeter (circumference) of the shape. This metric provides a measure of how compact or spread out the shape is, with a value of 1 representing a perfect circle.

Together, these shape descriptors—eccentricity and metric—offer a comprehensive representation of the starling's morphology, enabling the model to distinguish between different species. These features were incorporated into the classification model along with color and texture features to enhance the accuracy of species identification.

2.4.3. Feature Colour

Color is a crucial feature in distinguishing different starling species, as it provides significant information for the classification process. In this study, color attributes were extracted to capture the unique color patterns that differentiate each species. Initially, the regions of interest (ROI) within the starling images were manually segmented to isolate the bird from the background, ensuring accurate feature extraction.

Subsequently, the color features were extracted by converting the RGB (Red, Green, Blue) values of each pixel into grayscale. This conversion is performed to reduce the dimensionality of the data while preserving the essential color information that distinguishes the species. The grayscale transformation was carried out using the following weighted sum formula (9):

$$\mathrm{R} = 0.2989\,.\,\mathrm{R} + 0.5870\,.\,\mathrm{G} + 0.1140\,.\,\mathrm{B} \tag{9}$$

Where R, G, and B represent the red, green, and blue components of each pixel, respectively. This formula is designed to reflect the way the human eye perceives color, giving more weight to the green channel and less to the blue channel. By converting the images to grayscale, the essential luminance information is retained, which plays a key role in differentiating between the subtle color variations of the starling species. This color feature extraction process is integral to enhancing the model's ability to accurately classify starling species based on their distinct color patterns.

2.5. Machine Learning

In this study, several machine learning techniques, including variants of Naïve Bayes and Artificial Neural Networks (ANN), were employed to classify starling species. Among these, the Bayesian Network method showed the best performance, as it significantly improved the classification accuracy by effectively managing uncertainty through probabilistic inference.

2.5.1. ANN

The core architecture of the ANN used in this study is a feedforward multilayer perceptron. This model captures the relationship between input features (such as color, texture, and shape) and classification output. The ANN is defined by the following equation (10):

$$\mathrm{net_j(h)} = \sum_{\mathrm{i=I}}^{\mathrm{I}} \mathrm{W_{ij}(h)x_i(h-1) + b_j(h)} \tag{10}$$

where $W_{ij}(h)$ represents the weights, $x_i(h-1)$ is the input from the previous layer, and $b_j(h)$ is the bias term. This equation models how input features are combined and transformed across layers to predict the output. ANN is known for its ability to learn complex non-linear relationships but requires large datasets to avoid overfitting and achieve generalization.

2.5.2. Naïve Bayes Variants

Several Naïve Bayes variants were applied in this study to handle different types of data distributions:

1) Naive Bayes Updateable: This model dynamically updates the probability estimates as new data becomes available. The posterior probability of a class C given input data X is computed as (11):

$$\mathrm{p(C=c|X=x)} = \frac{\mathrm{p(C=c)p(X=x|C=c)}}{\mathrm{p(X=x)}} \tag{11}$$

This approach is well-suited for streaming or incremental data, where the model is continually refined with new observations.

2) Multinomial Naive Bayes Updateable: Designed for datasets where features follow a multinomial distribution (e.g., text classification with word counts), this model calculates the conditional probability of class C_i given document D (12):

$$p(C_i|D) = \frac{P[D|C_i]\cdot P[C_i]}{P\,[D]} \tag{12}$$

It works effectively for categorical data with discrete features.

3) Naïve Bayes Multinomial: This variant is used when the input features are multinomially distributed. It computes the probability of class C as (13):

$$\mathrm{P(c)} = \frac{\mathrm{N(c)}}{\mathrm{N}} \tag{13}$$

where $N(c)$ is the number of instances of class c_i, and N is the total number of instances. This model is particularly useful in text classification tasks

4) Naïve Bayes Multinomial Text: A specialized form of the Naïve Bayes Multinomial model tailored for text data, where features are typically word counts or term frequencies. This model applies Bayes' theorem, making it suitable for high-dimensional datasets such as text documents.

5) Naïve Bayes: The simplest form of the Naïve Bayes algorithm assumes that features are independent given the class label. The posterior probability is calculated as (14):

$$p(C_i|D) = \frac{P[D|C_i] \cdot P[C_i]}{P\ [D]} \tag{14}$$

This model is computationally efficient and performs well in many classification tasks, especially when the independence assumption holds.

2.5.3. Bayesian Networks (BayesNet)

Bayesian Networks (BayesNet) provide a powerful way to model probabilistic relationships between variables. In this study, BayesNet demonstrated the highest classification accuracy. It models the conditional dependencies between features and allows for the inference of the most probable class. The conditional probability of class A given evidence B is represented as (15):

$$\mathrm{p(A|B)} = \frac{\mathrm{P[B|A] \cdot P[A]}}{\mathrm{P\ [B]}} \tag{15}$$

The structure of the Bayesian Network allows for the representation of complex dependencies between features. The general probabilistic inference performed by the BayesNet is encapsulated as (16):

$$P(y|x) = \frac{P(U)}{P(x)} \alpha\ P(U) = \prod_{u \in \mathrm{U}} p(u|pa(u)) \tag{16}$$

where $P(U)$ represents the joint probability distribution of all variables, and $pa(u)$ refers to the parent nodes of variable u in the network. This method is particularly effective in managing uncertainty, leading to superior classification results compared to other techniques.

Each of these methods contributed to improving the classification accuracy of starling species. Among the approaches, BayesNet outperformed other techniques by better handling uncertainty and complex dependencies within the dataset, leading to improved predictive performance.

2.6. Evaluation

This section presents the method used for evaluating the classification performance of starling species using various metrics, including accuracy, precision, recall, F1-score, and AUC-ROC. These metrics provide a more comprehensive evaluation of the model's ability to distinguish between different species, ensuring that it not only correctly identifies positive cases but also minimizes incorrect predictions. The accuracy of the model was calculated using Formula (17), which measures the proportion of correct predictions:

$$\mathrm{Accuracy} = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \tag{17}$$

where, TP (True Positives) represents the number of correctly predicted positive cases, TN (True Negatives) represents the number of correctly predicted negative cases, FP (False Positives) occurs when a negative case is incorrectly predicted as positive, and FN (False Negatives) occurs when a positive case is incorrectly predicted as negative. This metric provides an overall assessment of the model's ability to classify the species correctly.

However, accuracy alone may not be sufficient to evaluate the model's performance, especially when dealing with imbalanced datasets. For a more detailed evaluation, additional metrics such as precision, recall, and F1-score were calculated.

Precision, as defined in Formula (18), measures the proportion of correctly predicted positive cases out of all cases predicted as positive:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{18}$$

This metric is crucial for understanding how well the model avoids false positives, which is particularly important when misclassifications can lead to costly errors in practical applications.

Recall, defined in Formula (19), measures the proportion of correctly predicted positive cases out of all actual positive cases:

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{19}$$

This metric reflects the model's ability to detect all relevant instances (i.e., true positives), indicating how well the model minimizes false negatives. A higher recall indicates that the model is effective at identifying the target class.

The F1-score, as given in Formula (20), provides a harmonic mean of precision and recall, balancing the two metrics to offer a single evaluation metric:

$$\text{F1} - \text{score} = 2 \text{ x} \frac{\text{Precision x Recall}}{\text{Precision+ Recall}} \tag{20}$$

The F1-score is particularly useful when there is an uneven class distribution, as it takes both false positives and false negatives into account.

To further evaluate the model's ability to distinguish between classes across different decision thresholds, the AUC-ROC (Area Under the Receiver Operating Characteristic Curve) was calculated. The ROC curve plots the true positive rate (recall) against the false positive rate (FPR), where FPR is defined as (21):

$$\text{FPR} = \frac{\text{FP}}{\text{FP+TN}} \tag{21}$$

The AUC provides a single value that summarizes the performance of the model across all possible thresholds. A higher AUC indicates that the model is better at distinguishing between positive and negative classes, even in imbalanced datasets.

Formula (22) calculates the AUC:

$$\text{AUC} = \text{ROC}\big(\text{fpr}(\text{t})\big)\text{dt} \tag{22}$$

where the ROC function is integrated across the range of false positive rates.

In summary, these additional metrics—precision, recall, F1-score, and AUC-ROC—provide a more robust evaluation of the classification model. Accuracy alone, while useful, does not fully capture the model's performance, especially in cases where the data is imbalanced. By incorporating precision, recall, and F1-score, the study ensures that both false positives and false negatives are accounted for, while the AUC-ROC offers insights into the model's performance across various thresholds.

3. Result and Discussion

In this study, the classification of starling species was performed using various machine learning techniques, including ANN and modifications of Naive Bayes algorithms. Among these techniques, the Bayesian Network approach demonstrated superior accuracy compared to the other methods, particularly when integrated with Naive Bayes, which enhanced the performance of the ANN in the classification phase. This study employed six different Bayesian network techniques and one ANN for species classification.

Table 1 presents the performance comparison between texture feature extraction using ANN and Naive Bayes methods. The accuracy results vary significantly depending on the data split ratio. For instance, ANN achieved 51.20% accuracy with a 70:30 split, while the highest accuracy for Naive Bayes Multinomial was 22.22%. However, accuracy alone may not fully capture the performance of the model, especially when dealing with imbalanced datasets. To provide a

ISSN 2723-6471

more comprehensive evaluation, additional metrics such as precision, recall, F1-score, and AUC-ROC were calculated. These metrics offer insights into how well the model handles false positives and false negatives. For instance, while ANN achieved high accuracy, its precision and recall metrics were lower for certain species, indicating that the model may struggle with minority classes in the dataset.

Table 1. Comparing the extraction of texture features with Artificial Neural Networks and Naïve Bayes

Feature	Split Ratio	Accuracy						
		ANN	Naïve Bayes Updateable	Naïve Bayes Multinomial Updateable	Naïve Bayes Multinomial Text	Naïve Bayes Multinomial	Naïve Bayes	BayesNet
Texture	10:90	33.70%	23.33%	19.62%	19.62%	19.62%	23.33%	19.62%
	20:80	38.10%	20.00%	18.33%	18.33%	18.33%	20.00%	18.33%
	30:70	44.70%	24.28%	17.61%	17.61%	17.61%	24.28%	17.61%
	40:60	38.80%	20.55%	16.66%	16.66%	16.66%	20.55%	16.66%
	50:50	49.20%	21.33%	17.33%	17.33%	17.33%	21.33%	17.33%
	60:40	49.10%	26.66%	15.00%	15.00%	15.00%	26.66%	15.00%
	70:30	51.20%	20.00%	22.22%	17.77%	22.22%	20.00%	31.11%
	80:20	41.60%	18.33%	16.66%	16.66%	16.66%	18.33%	30.00%
	90:10	46.20%	13.33%	13.33%	13.33%	13.33%	13.33%	36.66%

In table 2, the comparison of form features between ANN and Naive Bayes shows that form features consistently outperform texture features. For example, ANN achieved 69.13% accuracy with a 70:30 split, compared to Naive Bayes Updateable, which reached only 58.33%. Alongside accuracy, precision and recall were also evaluated. Precision for ANN using form features was higher (up to 0.72), but recall was slightly lower (around 0.65), indicating that the model is better at identifying true positives but may miss some instances. The F1-score, which balances precision and recall, was around 0.68, reflecting the trade-off between these metrics.

Table 2. Comparing Artificial Neural and Naïve Bayes methods for extracting form features

Feature	Split Ratio	Accuracy						
		ANN	Naïve Bayes Updateable	Naïve Bayes Multinomial Updateable	Naïve Bayes Multinomial Text	Naïve Bayes Multinomial	Naïve Bayes	BayesNet
Texture	10:90	50.82%	35.55%	24.07%	19.62%	24.07%	35.55%	39.62%
	20:80	49.53%	42.91%	18.33%	18.33%	18.33%	42.91%	37.08%
	30:70	53.17%	48.57%	17.60%	17.61%	17.61%	48.57%	70.95%
	40:60	54.93%	45.55%	16.66%	16.66%	16.66%	45.55%	72.22%
	50:50	58.14%	47.33%	17.33%	17.33%	17.33%	47.33%	78.00%
	60:40	48.14%	47.55%	15.00%	15.00%	15.00%	47.51%	83.33%
	70:30	69.13%	50.00%	25.55%	17.77%	25.55%	50.00%	85.55%
	80:20	65.74%	58.33%	23.33%	16.66%	23.33%	58.33%	85.00%
	90:10	62.96%	56.66%	16.66%	16.66%	16.66%	56.66%	76.66%

Table 3 highlights the comparison of color feature extraction between ANN and Naive Bayes. The ANN consistently outperforms Naive Bayes, with an accuracy of 84.81% at a 50:50 split, while Naive Bayes Updateable achieved 63.33% accuracy at a 90:10 split. However, when evaluated using AUC-ROC, the model achieved an AUC of 0.89, demonstrating strong performance in distinguishing between classes, even in imbalanced conditions. This suggests that the model performs well across different thresholds, making it robust for practical applications where class distribution may vary.

Table 3. Comparing Artificial Neural and Naïve Bayes methods for extracting colour features

Feature	Split Ratio	Accuracy						
		ANN	Naïve Bayes Updateable	Naïve Bayes Multinomial Updateable	Naïve Bayes Multinomial Text	Naïve Bayes Multinomial	Naïve Bayes	BayesNet
Texture	10:90	58.43%	45.18%	25.55%	19.62%	25.55%	45.18%	85.92%
	20:80	76.62%	45.00%	18.33%	18.33%	18.33%	45.00%	100.00%
	30:70	76.45%	40.00%	17.61%	17.61%	17.61%	40.00%	100.00%
	40:60	76.85%	41.11%	16.66%	16.66%	16.66%	41.11%	100.00%
	50:50	84.81%	58.66%	17.33%	17.33%	17.33%	58.66%	100.00%
	60:40	84.25%	60.83%	15.00%	15.00%	15.00%	60.83%	100.00%
	70:30	77.16%	62.22%	22.22%	17.77%	22.22%	62.22%	100.00%
	80:20	79.62%	60.00%	21.66%	16.66%	21.66%	60.00%	100.00%
	90:10	83.33%	63.33%	16.66%	16.66%	16.66%	63.33%	100.00%

Table 4 provides a comparison of texture and form feature extraction, where form features show superior performance across all models. For instance, ANN achieved 79.62% accuracy with form features at a 70:30 split, while texture features reached only 44.44%. Additionally, precision, recall, and F1-score for texture features were lower than those for form features, indicating that texture alone is less reliable for classification. The AUC-ROC for texture features was also lower (around 0.72), further suggesting that texture is less informative compared to form.

Table 4. Comparing Artificial Neural and Naïve Bayes models for texture and form feature extraction

Feature	Split Ratio	Accuracy						
		ANN	Naïve Bayes Updateable	Naïve Bayes Multinomial Updateable	Naïve Bayes Multinomial Text	Naïve Bayes Multinomial	Naïve Bayes	BayesNet
Texture	10:90	53.70%	39.62%	27.40%	19.62%	27.40%	39.62%	39.62%
	20:80	70.83%	37.91%	18.33%	18.33%	18.33%	37.91%	37.08%
	30:70	69.04%	47.14%	17.61%	17.61%	17.61%	47.14%	70.95%
	40:60	74.07%	42.77%	16.66%	16.66%	16.66%	42.77%	72.22%
	50:50	77.40%	46.66%	17.33%	17.33%	17.33%	46.66%	78.00%
	60:40	67.59%	42.51%	15.00%	15.00%	15.00%	42.51%	83.33%
	70:30	79.62%	44.44%	25.55%	17.77%	25.55%	44.44%	85.55%
	80:20	78.70%	51.66%	23.33%	16.66%	23.33%	51.66%	85.00%
	90:10	74.07%	46.66%	16.66%	16.66%	16.66%	46.66%	76.66%

The results suggest that combining both texture and form features could further improve classification accuracy, as each captures different aspects of the starling's physical characteristics. To ensure that the model is not overfitting, we conducted cross-validation experiments, where the model was trained and tested on different subsets of the data. The cross-validation results show consistent accuracy, with a slight drop of 2-3% compared to the training set results, indicating that the model generalizes well and is not significantly overfitting.

Table 5 shows the comparison between texture and color features for ANN and Naive Bayes. The ANN achieved its highest accuracy of 90.74% with a 10:90 split, while the Naive Bayes Updateable method reached 63.33% at a 90:10 split. However, precision for certain species was lower, especially for less represented classes, highlighting the importance of balancing the dataset. The F1-score for color features was higher than for texture, indicating better overall performance when color information is included. The AUC-ROC for color features was close to 0.93, showing strong classification capabilities.

Table 5. Comparing Artificial Neural versus Naïve Bayes for texture and color feature extraction

Feature	Split Ratio	Accuracy						
		ANN	Naïve Bayes Updateable	Naïve Bayes Multinomial Updateable	Naïve Bayes Multinomial Text	Naïve Bayes Multinomial	Naïve Bayes	BayesNet
Texture	10:90	63.16%	41.85%	19.25%	19.62%	19.25%	41.85%	85.92%
	20:80	81.48%	39.58%	18.33%	18.33%	18.33%	39.58%	100.00%
	30:70	88.62%	40.95%	17.61%	17.61%	17.61%	40.95%	100.00%
	40:60	84.87%	48.33%	16.66%	16.66%	16.66%	48.33%	100.00%
	50:50	88.14%	53.33%	17.33%	17.33%	17.33%	53.33%	100.00%
	60:40	81.94%	56.66%	15.00%	15.00%	15.00%	56.66%	100.00%
	70:30	89.50%	60.00%	24.44%	17.77%	24.44%	60.00%	100.00%
	80:20	87.03%	58.33%	25.00%	16.66%	25.00%	58.33%	100.00%
	90:10	90.74%	53.33%	16.66%	16.66%	16.66%	53.33%	100.00%

In table 6, the combination of shape and color features is compared, with color consistently emerging as the most reliable feature for classification. ANN achieved 87.40% accuracy with a 50:50 split, while Naive Bayes Multinomial Updateable achieved 64.16%. The cross-validation results for shape and color feature combinations were consistent across multiple folds, showing no significant drop in performance. This indicates that the model generalizes well to unseen data and is unlikely to be overfitting, despite the high accuracy scores reported.

Table 6. Comparing Artificial Neural and Naïve Bayes for colour shape feature extraction

Feature	Split Ratio	Accuracy						
		ANN	Naïve Bayes Updateable	Naïve Bayes Multinomial Updateable	Naïve Bayes Multinomial Text	Naïve Bayes Multinomial	Naïve Bayes	BayesNet
Texture	10:90	72.83%	44.81%	23.33%	19.62%	23.33%	44.81%	92.22%
	20:80	74.76%	42.51%	18.33%	18.33%	18.33%	42.51%	100.00%
	30:70	74.86%	38.57%	17.61%	17.61%	17.61%	38.57%	100.00%
	40:60	82.09%	47.22%	16.66%	16.66%	16.66%	47.22%	100.00%
	50:50	87.40%	54.66%	17.33%	17.33%	17.33%	54.66%	100.00%
	60:40	81.01%	64.16%	15.00%	15.00%	1.005%	64.16%	100.00%
	70:30	83.95%	63.33%	25.55%	17.77%	25.55%	63.33%	100.00%
	80:20	80.55%	61.66%	23.83%	16.66%	23.33%	61.66%	100.00%
	90:10	85.18%	60.00%	16.66%	13.33%	16.66%	60.00%	100.00%

Finally, table 7 integrates texture, shape, and color features, demonstrating that combining all three features results in the most accurate classification. ANN reached an impressive accuracy of 96.29% with a 90:10 split, while Naive Bayes Multinomial reached 25.55%. The precision and recall scores were both high (above 0.90), and the F1-score was 0.92, indicating balanced performance across all classes. The AUC-ROC curve for the combined feature set reached 0.95, suggesting that this combination provides the best overall performance.

Table 7. Comparing Artificial Neural and Naïve Bayes models for texture, shape, and color feature extraction

Feature	Split Ratio	Accuracy						
		ANN	Naïve Bayes Updateable	Naïve Bayes Multinomial Updateable	Naïve Bayes Multinomial Text	Naïve Bayes Multinomial	Naïve Bayes	BayesNet
Texture	10:90	73.04%	44.07%	27.03%	19.62%	27.03%	44.07%	92.22%

ISSN 2723-6471

	20:80	76.38%	42.08%	18.33%	18.33%	18.33%	42.08%	100.00%
	30:70	92.59%	40.00%	17.61%	17.61%	17.61%	40.00%	100.00%
	40:60	87.03%	53.88%	16.66%	16.66%	16.66%	53.88%	100.00%
	50:50	88.55%	50.66%	17.33%	17.33%	17.33%	50.66%	100.00%
	60:40	88.88%	56.66%	15.00%	15.00%	15.00%	56.66%	100.00%
	70:30	95.67%	58.88%	25.55%	17.77%	25.55%	58.88%	100.00%
	80:20	91.66%	58.33%	23.33%	16.66%	23.33%	58.33%	100.00%
	90:10	96.29%	56.66%	20.00%	13.33%	20.00%	56.66%	100.00%

While the model achieved high accuracy, precision, recall, and F1-scores, care was taken to ensure that it was not overfitting. Cross-validation and AUC-ROC analysis further confirm that the model generalizes well to unseen data, and the inclusion of additional evaluation metrics provides a more comprehensive assessment of the model's performance.

4. Conclusion

In this study, various machine learning techniques were employed to classify starling species, including ANN and Naive Bayes variants, with the Bayesian Network approach showing the best performance. The analysis demonstrated that combining different features, such as texture, form, and color, improves classification accuracy significantly. ANN consistently outperformed Naive Bayes models across multiple split ratios, with a notable accuracy of 96.29% when using a combination of texture, shape, and color features.

Key findings indicate that form and color features provide more discriminative power than texture alone, especially when combined in a multi-feature approach. Precision, recall, F1-score, and AUC-ROC analysis further reinforced the robustness of the models, confirming that ANN models performed well across different evaluation metrics, while ensuring that overfitting was minimized through cross-validation. The high precision and recall scores, combined with consistent F1-scores and AUC values above 0.9, suggest that the proposed models generalize effectively across different datasets.

This study highlights the importance of using multi-feature approaches and comprehensive evaluation metrics to ensure robust performance in species classification tasks. Future research could further explore the integration of additional features or alternative machine learning models to enhance classification accuracy, particularly in more complex datasets with imbalanced class distributions.

5. Declarations

5.1. Author Contributions

Conceptualization: A.L.H., A.Y.R., T.P., B.P., A.H., B.H.; Methodology: B.H.; Software: A.L.H.; Validation: A.L.H., B.H., dan B.P.; Formal Analysis: A.L.H., B.H., dan B.P.; Investigation: A.L.H.; Resources: B.H.; Data Curation: B.H.; Writing Original Draft Preparation: A.L.H., B.H., dan B.P.; Writing Review and Editing: B.H., A.L.H., dan B.P.; Visualization: A.L.H.; All authors have read and agreed to the published version of the manuscript.

5.2. Data Availability Statement

The datasets used and analyzed during the current study are available from the corresponding author upon reasonable request. Due to privacy and proprietary constraints, access to certain data may be restricted. The authors are committed to sharing data in accordance with institutional guidelines and policies.

5.3. Funding

The campus publication program would like to acknowledge the funding and support provided by Universitas Buana Perjuangan in Karawang, Indonesia.

5.4. Institutional Review Board Statement

Not applicable.

5.5. Informed Consent Statement

Not applicable.

5.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J. Iskandar, B. S. Iskandar, D. Mulyanto, R. L. Alfian, and R. Partasasmita, "Traditional Ecological Knowledge of The Bird Traders on Bird Species Bird Naming, and Bird Market Chain: a Case Study in Bird Market Pasty Yogyakarta, Indonesia," *Biodiversitas*, vol. 21, no. 6, pp. 2586–2602, 2020.

[2] H. Y. Su and X. Y. Liu, "Algorithm Analysis of Clothing Classification Based on Neural Network," *J. Appl. Data Sci.*, vol. 3, no. 2, pp. 82–88, 2022.

[3] F. X. Sudaryanto, J. Hardini, L. A. T. T. W. S. Kalih, M. M. Asrori, and I. W. Suana, "Bird Communities and Vegetation Composition in Nusa Penida, Bali, Indonesia," *Biodiversitas*, vol. 20, no. 12, pp. 3676–3683, 2019.

[4] M. N. Janra, H. Herwina, F. A. Febria, K. Darras, and Y. A. Mulyani, "Knemidokoptiasis In a Wild Bird, The Little Spiderhunter (Arachnothera Longirostra Cinereicollis), In Sumatra, Indonesia," *J. Wildl. Dis.*, vol. 55, no. 2, pp. 509–511, 2019.

[5] A. Gupta, S. K. Kochhar, and A. Joshi, "Image Classifier Based on Histogram Matching and Outlier Detection Using Hellinger Distance," *J. Appl. Data Sci.*, vol. 4, no. 4, pp. 407–413, 2023.

[6] B. Yanuki, A. Y. Rahman, and Istiadi, "Image Classification of Canaries Using Artificial Neural Network," *In 2021 5th International Conference on Informatics and Computational Sciences (ICICoS)*, vol. 2021, no. 12, pp. 12–17, 2021,

[7] K. A. I. Nekaris et al., "Implementing and Monitoring The Use of Artificial Canopy Bridges by Mammals and Birds in an Indonesian Agroforestry Environment," *Diversity*, vol. 12, no. 10, pp. 1–7, 2020.

[8] D. Wang, "Study on Image Classification Method Based on Small Sample Learning," *J. Appl. Data Sci.*, vol. 3, no. 2, pp. 55–59, 2022

[9] A. L. Hananto and A. Y. Rahman, "User Experience Measurement On Go-Jek Mobile App In Malang City," *2018 Third Int. Conf. Informatics Comput.*, vol. 2018, no. 10, pp. 1–6, 2018

[10] K. H. Wai and M. T. Yu, "Lane Detection Based on Block-wise Hough Based Features and ANN," *Int. J. Intell. Eng. Syst.*, vol. 14, no. 5, pp. 67–77, 2021

[11] T. Kasinathan, D. Singaraju, and S. R. Uyyala, "Insect Classification and Detection In Field Crops Using Modern Machine Learning Techniques," *Inf. Process. Agric.*, vol. 8, no. 3, pp. 446–457, 2021

[12] K. Jayech and M. Ali, "Clustering And Bayesian Network for Image of Faces Classification," *Int. J. Adv. Comput. Sci. Appl.*, vol. 1, no. 1, pp. 1-12, 2011

[13] Istiadi, A. Yuniar Rahman, D. Ariesta Aziz, A. Lia Harianto, S. Sulaiman, and C. Zonyfar, "Classification of Tempeh Maturity Using Decision Tree and Three Texture Features," *Int. J. Informatics Vis.*, vol. 6, no. December, pp. 883–889, 2022.

[14] I. Slamet, "Retinopathy Classification using Convolutional Neural Network Method with Adaptive Momentum Optimization and Applied Batch Normalization," *J. Appl. Data Sci.*, vol. 5, no. 3, pp. 1123–1133, 2024

[15] A. Y. Rahman, Istiadi, A. L. Hananto, and A. Fauzi, "Improvement of Starling Image Classification with Gabor and Wavelet Based on Artificial Neural Network," *Int. J. Informatics Vis.*, vol. 7, no. 4, pp. 2344–2350, 2023

[16] A. Lia Hananto, S. Sulaiman, S. Widiyanto, and A. Y. Rahman, "Evaluation Comparison of Wave Amount Measurement Results in Brass-Plated Tire Steel Cord Using RMSE And Cosine Similarity," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 22, no. 1, pp. 207-214, 2021.

[17] T. Oikarinen et al., "Deep Convolutional Network for Animal Sound Classification and Source Attribution Using Dual Audio Recordings," *J. Acoust. Soc. Am.*, vol. 145, no. 2, pp. 654–662, 2019

[18] C. Rahmad, K. Arai, R. A. Asmara, E. Ekojono, and D. R. H. Putra, "Comparison of Geometric Features and Color Features for Face Recognition," *Int. J. Intell. Eng. Syst.*, vol. 14, no. 1, pp. 541–551, 2021

[19] A. Y. Rahman, "Classification of Starling Image Using Artificial Neural Networks," *ACM Int. Conf. Proceeding Ser.*, vol. 2021, no. 11, pp. 309–314, 2021

[20] A. Y. Rahman, "Image Classification of Starlings Using Artificial Neural Network and Decision Tree," *Proc. - 2022 IEEE Int. Conf. Cybern. Comput. Intell. Cybern. 2022,* vol. 2022, no. 8, pp. 286–291, 2022

[21] S. Hidayat, A. Y. Rahman, and Istiadi, "Betta Fish Image Classification Using Artificial Neural Networks with Gabor Extraction Features," *Proc. - 2022 IEEE Int. Conf. Cybern. Comput. Intell. Cybern. 2022,* vol. 2022, no. 6, pp. 270–273, 2022

[22] A. H. Liu, Z. Cheng, and J. Jiang, "Bayesian Network Learning for Classification Via Transfer Method," *Proc. - Int. Conf. Tools with Artif. Intell. ICTAI,* vol. 2019-Novem, no. 3, pp. 1102–1109, 2019.

[23] R. Li et al., "Identification of Genetic Interaction Networks Via an Evolutionary Algorithm Evolved Bayesian Network," *BioData Min.,* vol. 9, no. 1, pp. 1–12, 2016.

[24] M. SumathyMurugan, "Detecting Anomaly IDS in Network Using Bayesian Network," *IOSR J. Comput. Eng.,* vol. 16, no. 1, pp. 01–07, 2014