



Efficiency and Comparative Performance of LBP-Based Random Forest and SVM for Toraja Buffalo Classification

Abdul Rachman Manga^{*1}, Anik Nur Handayani², Heru Wahyu Herwanto³, Rosa Andrie Asmara⁴, Syamsul Bahri⁵

^{1,2,3}Department of Electrical Engineering, Universitas Negeri Malang, Malang, Indonesia

¹Department of Computer Science, Universitas Muslim Indonesia, Makassar, Indonesia

⁴Information Technology Department, Politeknik Negeri Malang, Malang, Indonesia

⁵Graduate School of Environmental Engineering, University of Kitakyushu, Fukuoka, Japan

*abdul.rachman.2205349@students.um.ac.id

Abstract

Toraja buffalo holds significant cultural and economic value, yet automated classification remains challenging due to subtle visual differences between types. This study evaluates the efficiency of Local Binary Pattern (LBP) combined with Random Forest (RF) and Support Vector Machine (SVM) for classifying six Toraja buffalo types: Balian, Lotong Boko, Pudu, Saleko, Todi, and Ulu. Unlike complex deep learning approaches, this research focuses on a computationally efficient framework by extracting texture features from multiple body parts: head, eyes, horns, body, and tail to capture distinctive patterns. The methodology involves multi-part feature fusion and a comparative analysis of ensemble versus kernel-based learners. Experimental results demonstrate that the LBP-Random Forest model significantly outperforms SVM, achieving a superior accuracy of 92.08% compared to 64.17%. The findings highlight that the proposed LBP-RF integration provides a robust and resource-efficient alternative for livestock image classification, balancing high diagnostic accuracy with lower computational requirements.

Keywords: buffalo toraja; feature extraction; image classification; LBP; random forest; SVM

How to Cite: A. R. Manga^{*}, A. N. Handayani, H. W. Herwanto, R. A. Asmara, and S. Bahri, "Efficiency and Comparative Performance of LBP-Based Random Forest and SVM for Toraja Buffalo Classification", *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 10, no. 4, pp. 1045 - 1054, Aug. 2026.

Permalink/DOI: <https://doi.org/10.29207/resti.v10i4.7536>

Received: January 29, 2026

Accepted: June 21, 2026

Available Online: August 10, 2026

*This is an open-access article under the [CC BY 4.0 License](#)
Published by [Ikatan Ahli Informatika Indonesia](#)*

1. Introduction

Species identification and classification are fundamental elements in biodiversity and conservation studies, as they enable researchers to understand biodiversity at various levels, from molecular and morphological to ecosystem levels [1], [2]. Accurate identification procedures enable the deeper revelation of evolutionary links between species, bolster taxonomic databases and provide vital information for identifying uncommon, endangered, or environmentally vulnerable species. In order to build more methodical and evidence-based conservation methods, such as evaluating extinction threats, setting conservation priorities, and allocating resources as efficiently as possible in attempts to preserve and restore natural ecosystems, accurate classification is also essential [3].

Furthermore, species identification and classification are not only important in the context of ecological conservation, but also highly relevant in sustainable agriculture and livestock farming practices [4], especially those related to domestic species that have high economic and cultural value. One example is the Toraja Buffalo, which is a unique breed of livestock from the Toraja region of South Sulawesi, Indonesia. This buffalo is now an essential component of the Toraja people's social and cultural identity rather than just a useful farm animal. In various traditional ceremonies such as *Rambu Solo'* and *Rambu Tuka'*, buffaloes play a central role as symbols of social status, spiritual offerings, and representations of respect for ancestors [5].

From an economic perspective, Toraja buffalo are also valuable assets in agrarian households, reflecting the

level of family welfare and serving as a source of long-term investment [6]. However, the existence of Toraja buffaloes today faces major challenges, ranging from climate change and environmental degradation to modernization and shifting cultural values, which have led to a decline in the breeding of native buffalo breeds [7]. Therefore, efforts to preserve Toraja buffalo are not only important from a biological perspective, but also urgent in terms of preserving the cultural heritage and local identity of the Toraja people [8].

In the context of preservation and documentation, the process of automatically classifying Toraja buffalo types based on visual characteristics is a highly promising approach [7]. Each type of buffalo, such as *Balian*, *Lotong Boko*, *Pudu*, *Saleko*, *Todi* and *Ulu*, has unique phenotypic characteristics, particularly in terms of skin colour, horn shape, and body pattern, which can be detected through image processing. Therefore, the development of a computer vision-based classification system is important to support a more efficient, objective, and sustainable identification process [9], [10].

With the advancement of technology, digital image processing approaches are increasingly being used for visual classification tasks, including livestock classification [11], [12]. One of the important steps in this process is feature extraction. It attempts to transform visual data from pictures into useful numerical representations so that classification systems can process them [13], [14]. Previous studies extracted edge features from buffalo pictures using the Canny Edge Detection method, which were then further classified using machine learning techniques [15]. Additionally, Support Vector Machine (SVM) has been utilized in earlier research to classify buffalo photos; on the Nili-Ravi and Khundi buffalo datasets, accuracy rates of over 84% have been recorded [16].

However, most previous studies have focused on general buffalo classification datasets and have not specifically investigated the classification of Toraja buffalo breeds. Furthermore, limited attention has been given to exploiting texture information from multiple buffalo body regions, even though different anatomical parts may contain complementary visual characteristics that are useful for distinguishing visually similar breeds. These limitations highlight the need for a more effective and computationally efficient approach tailored to Toraja buffalo classification.

Recent studies have demonstrated that lightweight texture-based approaches remain highly competitive for image classification tasks, particularly in situations where computational resources are limited. Local Binary Pattern (LBP) has been widely applied in animal recognition, medical image analysis, and texture classification because of its low computational complexity, robustness to illumination variations, and effectiveness in capturing local texture information. LBP-based techniques require fewer training samples and less computing power than deep learning

techniques, which makes them appropriate for real-world implementation in resource-limited settings.

Inspired by these benefits, the main feature extraction technique used in this study to represent the texture features of Toraja buffalo photos is Local Binary Pattern (LBP). Two popular machine learning methods, Support Vector Machine (SVM) and Random Forest (RF), are then used to classify the retrieved LBP features. Due to their robustness against noise, computational efficiency during training and prediction, and capacity to handle high-dimensional feature spaces, both classifiers have shown excellent performance in image classification tasks.

The purpose of this study is to assess how well Toraja buffalo breeds can be classified using LBP feature extraction in conjunction with SVM and RF classifiers. To determine the best strategy for this application, the study also contrasts the two classifiers' performances. It is anticipated that the results will aid in the creation of precise, effective, and useful picture categorization systems that support conservation efforts, cultural heritage preservation, and livestock identification.

There are three primary areas where this study is innovative. It first presents a multi-part feature fusion approach that uses Local Binary Pattern descriptors to merge texture data taken from five buffalo body regions: the head, eyes, horns, body, and tail. Second, using the same fused feature representation, it offers a thorough comparison of Random Forest and Support Vector Machine classifiers. Third, the suggested framework is specifically made to be a computationally effective and lightweight solution for Toraja buffalo classification, which makes it appropriate for deployment in settings with constrained computational resources while preserving competitive classification performance.

2. Methods

This study began with the collection of a video dataset, which was then converted into an image dataset as the main material for image processing. The images underwent a pre-processing stage involving resizing to 150x150 pixels and normalization to ensure data consistency and readiness for processing. Next, feature extraction was performed using LBP, which extracts local texture features from the images. To provide a more complete feature representation, the extracted features are then concatenated according to the buffalo's physical parts, including the body, eye, head, horn, and tail. The SVM and Random Forest algorithms are then used to move this feature data to the classification stage once it has been divided into 80% training data and 20% testing data. To better understand the model's efficacy in classifying buffalo kinds, performance metrics like accuracy, precision, recall, and F1-score are measured, as seen in Figure 1.

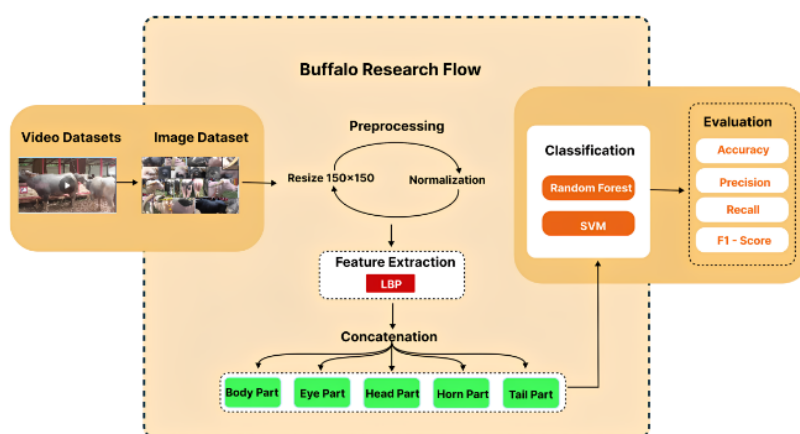


Figure 1. Research Workflow

2.1 Dataset

The dataset used in this research was collected from videos of six Toraja buffalo breeds, namely Balian, Lotong Boko, Pudu, Saleko, Todi, and Ulu. Video recordings were captured directly under outdoor farm conditions with natural illumination and varying viewing angles. The recorded videos were subsequently converted into image frames through a video frame sampling process, resulting in a dataset of buffalo images with diverse visual representations [17], [18].

For each buffalo breed, the extracted images were manually organized into five body-part categories, namely body, eyes, head, horns, and tail. Each body-part category contains 1,200 images distributed across the six buffalo breeds, resulting in a total of 6,000 images across all body-part inputs. Consequently, each breed contributes approximately 200 images to each body-part category. Each body-part input had 960 training images and 240 testing images after the dataset was split in an 80:20 ratio. In order to avoid class bias during model training and evaluation and to guarantee balanced learning, the training data were divided almost equally among the six classes.

The variety of image viewpoints is crucial because it makes the model more resilient in identifying Toraja buffaloes in spite of differences in viewpoint, orientation, and surroundings. For instance, distinguishing features like head form and horn structure that might not be apparent from other angles might be seen in photos taken from the front or side. As a result, models trained with this dataset have a greater capacity to generalize to data that has never been seen before and was acquired under various circumstances [19].

Additionally, the model can acquire distinguishing characteristics linked to particular buffalo breeds by using several inputs that represent various body parts. Characteristics such as horn shape and size, eye appearance, body texture, and tail structure can provide important visual cues for breed identification. By integrating information from multiple body regions, the model can capture complementary features that may not

be sufficiently represented by a single input, thereby improving classification performance [20].

Furthermore, the variation in image content and acquisition conditions provides a broader context for the model to learn representative patterns from the data [21]. This variation enables the model to adapt to differences in position, lighting conditions, and individual appearances among buffaloes [22]. As a result, models trained on such diverse datasets are expected to generate more accurate and reliable predictions when applied to new data.

This study intends to create a more robust and generalizable classification model that may successfully serve applications related to livestock management, breed identification, and Toraja buffalo conservation using a dataset structure that incorporates various body-part inputs and visual viewpoints.

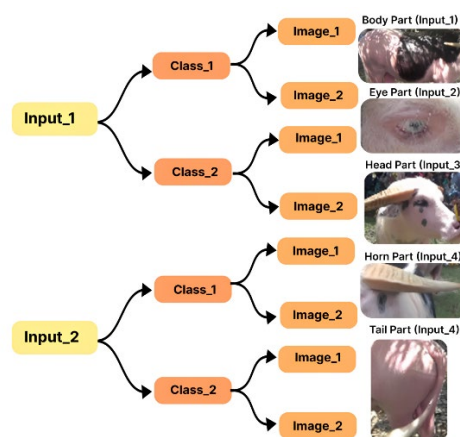


Figure 2. Directory structure of multi-input multi-class dataset

Figure 2 Shows the kind of dataset with a variety of inputs and class categories, where the classes belong to various buffalo species and the inputs pertain to various Toraja buffalo sections. Images of the Toraja Buffalo's head, horns, tail, eyes, and general body structure are included in this collection, which aims to document the buffalo's physical changes from a variety of angles. The utilization of photographs captured from various perspectives offers a more thorough depiction of every component of the buffalo, facilitating the identification

of distinctive characteristics linked to each species or subspecies. The detailed distribution of the dataset, which comprises six types of Toraja buffalo and five distinct body part inputs (body, eyes, head, horns, and tail), including the total number of images and the training-testing split, is shown in Table 1.

Table 1. Dataset Information

Input	Number of Class	Total Image	Train	Test
Body Part	6	1200	960	240
Eye Part	6	1200	960	240
Head Part	6	1200	960	240
Horn Part	6	1200	960	240
Tail Part	6	1200	960	240

2.2 Preprocessing

An essential step in this research is data preprocessing, which prepares the data in the format needed to be compatible with machine learning models [23]. Preprocessing guarantees that the model's training data is in ideal condition and prepared for additional processing [24]. In order to make data easier for machine learning algorithms to read and process, this stage entails a number of actions to clean, transform, and organize the data [25].

The images in the Toraja Buffalo picture dataset range in resolution from good to low. The effectiveness and consistency of model training may be impacted by these variations in resolution. Therefore, every image in the dataset is reduced to 150x150 pixels to guarantee that the model is trained on the data more uniformly and effectively. This procedure lessens changes in image size, which might make training more difficult [26]. Consistent image sizes allow models to process and learn patterns from each image more quickly, cutting down on computation time and speeding up training [27]. Figure 3 shows an example of image resizing in the Toraja buffalo dataset.

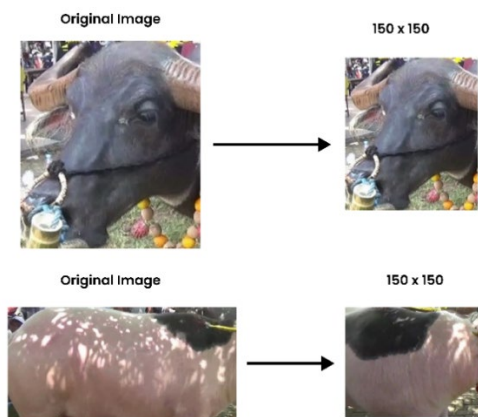


Figure 3. Resize Image

Normalization is an important step in data preprocessing, especially in image processing and the application of machine learning algorithms [25], [26]. In this study, `rgb2gray` is used to convert the input image into grayscale after resizing them to 150×150 pixels. This conversion produces floating-point grayscale intensities that are already scaled in the

$[0, 1]$ range, so explicit min-max scaling on pixel values is not required in the implementation. Instead, normalization is applied at the feature level to ensure that the extracted texture descriptors are comparable across images. Specifically, after computing Local Binary Pattern (LBP) codes, the resulting LBP histogram is normalized into a probability-like distribution by dividing each bin count by the total histogram sum [27-29]. The histogram normalization is defined in Equation 1.

$$h_{norm}(i) = \frac{h(i)}{\sum_j h(j) + \varepsilon} \quad (1)$$

$h(i)$ is the frequency of the i -th LBP histogram bin, $\sum_j h(j)$ is the total number of counted samples, and $\varepsilon = 10^{-7}$ is a small constant to prevent division by zero. For example, if histogram counts are $[50, 30, 20]$ with total 100, the normalized histogram becomes $[0.50, 0.30, 0.20]$. This feature-level normalization reduces scale variability and produces a stable feature vector for the subsequent classification stage.

2.3 Feature Extraction using LBP

One prominent technique for extracting features from images is the Local Binary Pattern (LBP) method, which is especially useful for capturing local texture properties [30], [31]. The basic principle of LBP is to compare the intensity of a center pixel g_c with its neighboring pixels g_p within a certain radius [32]. A neighbor is given a binary value of 1 if its intensity is higher than or equal to the center pixel; otherwise, it is given a value of 0 [33]. In this study, LBP is computed using $P = 8$ neighbors and radius $R = 1$ with the uniform pattern setting (method="uniform"). The LBP code is computed using Equation 2.

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p \quad (2)$$

The thresholding function $s(\cdot)$ is defined as Equation 3.

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (3)$$

To illustrate the computation, consider a 3×3 grayscale patch with the center pixel $g_c = 35$:

$$\begin{bmatrix} 34 & 36 & 39 \\ 32 & 35 & 40 \\ 30 & 33 & 37 \end{bmatrix}$$

Comparing each neighbor with g_c (clockwise from the top-left) yields the binary pattern 01111000. Using Equation 2, the decimal LBP value is obtained by summing the weighted bits, resulting in $LBP = 2 + 4 + 8 + 16 = 30$. This process is applied to all valid pixels to form an LBP-coded image. Since the uniform setting is used, LBP patterns with at most two bitwise transitions are assigned dedicated labels, while non-uniform patterns are grouped into a single label. Consequently, for $P = 8$, the output label space becomes $P + 2 = 10$ bins, which produces a compact feature vector.

Next, an LBP histogram is constructed as the final texture feature representation [34–36]. Following the implementation, the histogram is computed using $P + 2$ bins over the label range $[0, P + 1]$. Finally, the histogram is normalized using Equation 1 so that the sum of all bins equals 1, making the feature representation robust to differences in pixel counts and image variations. Figure 4 illustrates the feature extraction workflow from the original RGB image, grayscale conversion, LBP-coded output, and the normalized histogram. The resulting normalized LBP histogram vector is then used as input for the SVM and Random Forest classification models in the next stage [37–39].

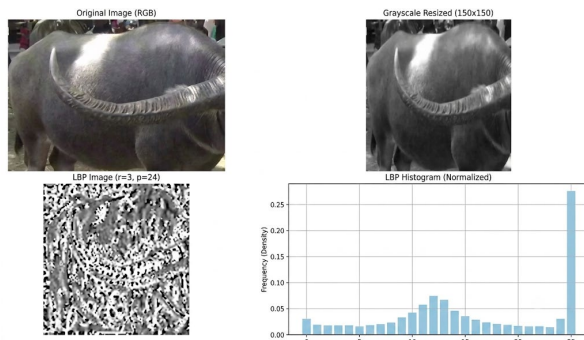


Figure 4. Feature Extraction LBP

2.4 Classification

The SVM algorithm was used in this study to classify features that were extracted using the LBP approach. Because of its ease of use and capacity to handle linearly separable data, the SVM model was built using a linear kernel [28–30]. The regulatory parameter is set at $C = 1.0$. It regulates the ratio of margin maximization to classification error minimization. The goal of SVM with a linear kernel is to identify the best hyperplane to divide two classes [31],[32] which is formulated in Equation 4.

$$f(x) = w^T x + b \quad (4)$$

x is a feature vector generated from the extraction process using the LBP method, w is a weight vector that indicates the direction and orientation of the hyperplane, and b is the bias or offset value of the hyperplane relative to the origin point [33],[34]. The function $f(x)$ is used to determine the class of data based on the sign of the calculation result. The data is categorized as belonging to the positive class if the value of $f(x)$ is more than or equal to zero, and as belonging to the negative class if it is less than zero [35], [36]. This approach is effective in producing optimal class separation, especially when the data can be separated linearly in feature space. Figure 5 shows the concept of margin and hyperplane.

One of the classification techniques utilized in this study to assess performance against texture features derived from the LBP extraction procedure was the Random Forest algorithm.

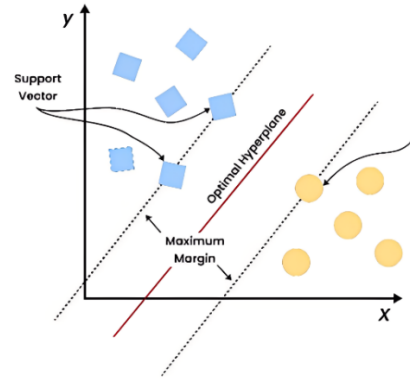


Figure 5. Architecture SVM

To guarantee repeatability of results, the Random Forest model was constructed using 100 decision trees, $n_estimators$ and a $random_state$. Using the bagging technique, each tree in the ensemble is trained on a random subset of the training data. The majority vote from all trees is then used to make the prediction [37–39]. Mathematically, the final prediction in Random Forest for classification cases can be expressed as Equation 5.

$$\hat{y} = \text{mode}(h_1(x), h_2(x), \dots, h_T(x)) \quad (5)$$

$h_T(x)$ is the prediction result of the t^{th} tree, and T is the total number of trees in the forest. Figure 6 shows the architecture of the Random Forest model.

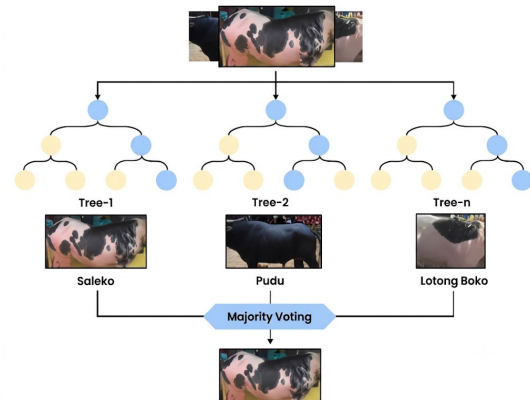


Figure 6. Architecture Random Forest

2.5 Evaluation

Accuracy, precision, recall, and F1-score were the four main metrics utilized to assess the classification model's performance. These metrics were used to assess how well the model classified data, especially in datasets with multiple classes [40], [41].

Accuracy: Calculates the percentage of accurate predictions among all the data. Equation 6 provides the mathematical formula for determining accuracy.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

Precision: Shows the proportion of accurate positive forecasts. Equation 7 provides the mathematical formula for determining precision.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (7)$$

Recall: Measures how much positive data is correctly classified. The mathematical formula for calculating recall is given in Equation 8.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (8)$$

F1-Score: is the harmonic mean of precision and recall, which balances both. Equation 9 provides the mathematical formula for determining the F1-Score.

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (9)$$

3. Results and Discussions

3.1 Research Settings

To support image processing and machine learning tasks, this investigation was carried out on a computer with a 12th generation Intel® Core™ i5 processor and 16 GB RAM. Python was chosen for the implementation because of its versatility and effectiveness in machine learning and data processing operations. In addition, Google Colab was utilized as the development environment to facilitate coding, experimentation, and model evaluation.

Prior to the machine learning stage, the dataset was organized into two main directories, namely training and testing, where each directory contained six subfolders corresponding to the Toraja buffalo classes: Balian, Lotong Boko, Pudu, Saleko, Todi, and Ulu. As described in Section 2.1, each body-part input (body, eyes, head, horns, and tail) consisted of 1,200 images that were divided using an 80:20 ratio, resulting in 960 training images and 240 testing images. The training data were distributed approximately equally among the six classes to ensure balanced learning during model development.

The machine learning models may be trained and assessed on separate data subsets thanks to this dataset organization, which clearly separates training and testing data. Model performance was assessed using accuracy, precision, recall, and F1-score metrics to provide a comprehensive evaluation of classification efficacy. When there are slight differences in class support in the testing data, using a variety of assessment metrics helps guarantee that model performance is evaluated equitably.

3.2 Experiment Result

A number of metrics, including accuracy, precision, recall, and F1-Score, are employed in the model's analysis and performance. In relation to the total amount of data, accuracy quantifies how well the model predicts positive and negative classes. While recall displays the proportion of pertinent predictions that were correctly detected from all real positive classes, precision displays the percentage of pertinent predictions from the anticipated positive class. The false positive rate is the frequency with which the model

incorrectly predicts a positive class. The F1 score measures the precision and recall harmonic mean, a graph using validation data that shows accuracy and loss throughout the training process. All assessment metrics for the SVM and random forest models with LBP feature extraction are displayed in Table 2.

Table 2. Comparison of Random Forest and SVM Model Performance with LBP Feature for Toraja Buffalo Classification

Model	Buffalo Class	Precision	Recall	F1-Score
RF	Balian	95.00 %	97.44 %	96.20 %
	Lotong Boko	91.84 %	93.75 %	92.78 %
	Pudu	93.94 %	88.57 %	91.18 %
	Saleko	90.00 %	94.74 %	92.31 %
	Todi	89.58 %	95.56 %	92.47 %
	Ulu	93.33 %	80.00 %	86.15 %
SVM	Balian	50.00 %	46.15 %	48.00 %
	Lotong Boko	75.00 %	50.00 %	60.00 %
	Pudu	53.66 %	62.86 %	57.89 %
	Todi	75.51 %	97.37 %	85.06 %
	Saleko	80.56 %	64.44 %	71.60 %
	Ulu	51.17 %	68.57 %	59.26 %

To clarify how the values in Table 1 are obtained, one example of metric computation is presented for the Random Forest model based on the classification report output. The total testing set consists of 240 images, with minor variations in class support across categories. As an example, for the Ulu class (support = 35), the reported recall of 80.00% follows $\text{Recall} = \frac{TP}{TP+FN}$, yielding $TP = 0.80 \times 35 = 28$ and $FN = 35 - 28 = 7$. Next, the precision of 93.33% is computed using $\text{Precision} = \frac{TP}{TP+FP}$; thus, $0.9333 = \frac{28}{28+FP}$, which gives $FP \approx 2$ (since $\frac{28}{28+2} = 0.9333$). The F1-score is then calculated as $F1 = \frac{2PR}{P+R}$, resulting in $F1 = \frac{2(0.9333)(0.8000)}{0.9333+0.8000} \approx 0.8615$ or 86.15%, consistent with Table 2. Therefore, the precision, recall, and F1-score values reported in Table 2 are directly derived from the confusion-matrix components (TP, FP, FN), and the same procedure is applied to the other classes.

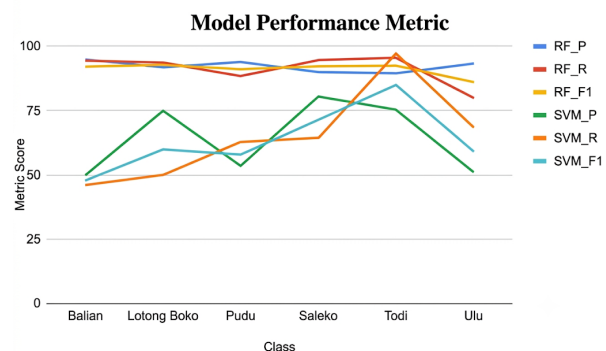


Figure 7. Model Performance Metric

Figure 7 illustrates the per-class performance comparison between Random Forest and SVM based on the precision, recall, and F1-score metrics. As shown, the Random Forest model consistently yields higher values across most classes, indicating better generalization when using LBP-based texture features.

3.3 Comparison of SVM and Random Forest Architecture Accuracy on LBP Feature Extraction

Testing was conducted at this stage to find the best architecture for classifying Toraja buffalo using LBP feature extraction. The accuracy obtained for the Random Forest model was 92.08%, while the SVM model achieved an accuracy of 64.17%. The results of this test are shown in Table 3.

Table 3. Average Accuracy Result of Random Forest and SVM Obtained from Classification Performance in Table 2

Model	Accuracy Result
Random Forest	92.08 %
SVM	64.17 %

Based on the data in Table 3, the performance of the Random Forest model with LBP feature extraction outperforms the SVM model with LBP feature extraction. Random Forest with LBP features provides better results than SVM due to Random Forest's ability to handle non-linear and complex data such as LBP extraction results. This model also performs automatic feature selection and is not overly sensitive to parameters, making it more stable and effective in processing diverse texture features. In contrast, SVM requires careful parameter optimization and is less flexible in handling features with non-linear relationships, resulting in lower performance on LBP data. The confusion matrix results for the Random Forest and SVM models are shown in Figure 8 and Figure 9.

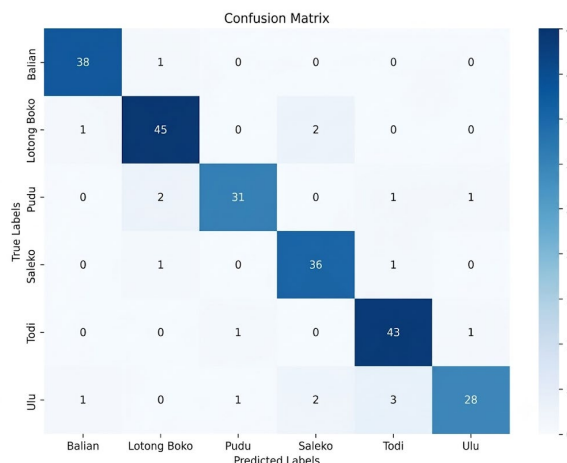


Figure 8. Confusion Matrix of Random Forest Model

The confusion matrix in Figure 8 represents that the model recognizes majority classes such as *Lotong Boko*, *Todi*, and *Balian* well. However, errors occur in classes such as Pudu and Ulu, which are often misclassified as other classes. This is due to visual similarities between classes. The model still has difficulty distinguishing classes with similar visual characteristics when using only texture features from LBP.

The confusion matrix in Figure 9 indicates that the SVM model was only able to recognize some of the data from certain classes well, particularly in the *Saleko* (37) and *Todi* (29) classes. However, the model failed to

accurately recognize data from several other classes. For example, the *Balian* class was frequently misclassified as Ulu (17 times), and the *Lotong Boko* class was incorrectly identified as Balian (11 times) and Pudu (9 times). These errors indicate that the model still struggles to distinguish between classes with high visual similarity. Additionally, the performance of SVM, which relies on optimal margin, can be disrupted if data between classes overlap in the feature space, as may occur with LBP features. This leads to a decline in classification performance, particularly for minority classes or those with unclear feature distributions.

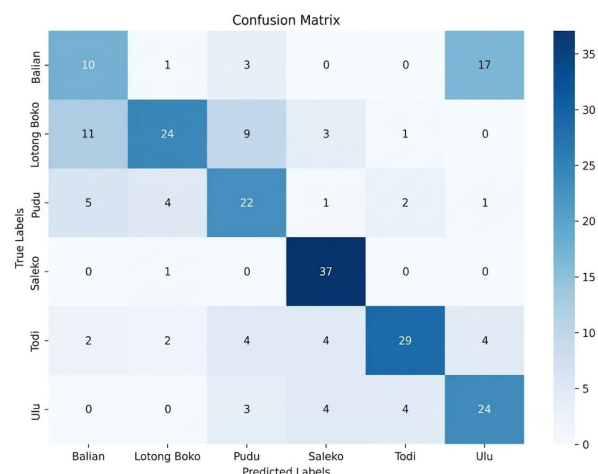


Figure 9. Confusion Matrix of SVM Model

3.4 Comparison of Previous Research

Two kinds of models were created based on the study's findings. Among these models are the SVM model with LBP feature extraction and the Random Forest model. The random forest model with LBP feature extraction yielded an accuracy of 92.08% out of these two models. A comparison was then made with previous studies that used the same dataset, namely the Toraja buffalo dataset. This comparison is listed in Table 4.

Table 4. Comparison Result with Previous Research

References	Feature Extraction	Method	Accuracy
[14]	Canny	KNN	73.33 %
[14]	Canny	Naïve Bayes	63.75 %
[14]	Canny	Decision Tree	71.25 %
[14]	Canny	SVM	60.42 %
[15]	-	SVM	84.07 %
[42]	-	VGG16	93.33 %
Proposed Method	LBP	RF	92.08 %
Proposed Method	LBP	SVM	64.17 %

3.5 Analysis Result

The results of the study indicate that the Random Forest model with LBP feature extraction achieved superior classification performance compared to SVM. This is evidenced by the Random Forest accuracy of 92.08%, which substantially exceeds the SVM accuracy of 64.17%. Furthermore, precision, recall, and F1-score values obtained by Random Forest are consistently

higher across most classes. These results suggest that Random Forest is more effective in handling the high-dimensional texture features generated by LBP, which exhibit complex and non-linear distributions. The ensemble-based learning mechanism of Random Forest also contributes to better generalization performance and demonstrates better generalization performance than SVM on the evaluated dataset. The confusion matrix further supports these findings, showing that the Random Forest model successfully recognizes majority classes such as *Lotong Boko*, *Todi*, and *Balian* with high accuracy. Although misclassifications are still observed in the *Pudu* and *Ulu* classes, the number of errors remains relatively small. The lower recognition performance observed for these classes can be attributed to their visual similarity with other buffalo breeds. Since LBP primarily captures local texture information, subtle differences in body shape, horn geometry, and overall appearance may not be fully represented in the extracted features. Consequently, samples exhibiting similar texture distributions tend to overlap in the feature space, increasing the likelihood of misclassification. This phenomenon is particularly evident among buffalo breeds that share comparable skin patterns and body textures. In contrast, the SVM model exhibits more substantial and uneven classification errors, particularly in the *Balian* and *Lotong Boko* classes, which are frequently misclassified as other categories. This highlights the limitations of SVM when dealing with overlapping texture-based feature distributions derived from LBP descriptors.

Compared with previous studies employing conventional classification methods such as KNN, Naïve Bayes, and Decision Trees using Canny-based features, the Random Forest model proposed in this study demonstrates a considerable improvement in classification accuracy. Although its accuracy is slightly lower than that reported by the VGG16-based approach [39] (93.33%), the proposed LBP-RF framework achieves a competitive accuracy of 92.08% while requiring substantially lower computational resources. Unlike deep learning architectures such as VGG16, which demand extensive training data, longer training times, and higher hardware requirements, the LBP-RF approach provides a more lightweight and practical solution for real-time deployment in resource-constrained environments. Therefore, the combination of LBP feature extraction and Random Forest classification can be considered an effective and efficient approach for the automatic classification of Toraja buffalo breeds while supporting conservation, livestock management, and cultural heritage preservation efforts.

3.6 Research Limitation

Despite the high accuracy achieved by the LBP-Random Forest model, this study has several limitations that should be considered. First, although the dataset was collected under natural outdoor illumination, the

captured conditions did not fully represent extreme environmental variations such as heavy rain, dense shadows, or very low-light situations. Therefore, the model's performance under such challenging conditions remains to be investigated. Second, the current classification framework relies on the manual cropping of buffalo body parts, which limits its potential for fully automated, end-to-end processing. Third, the proposed approach relies primarily on texture-based descriptors. While LBP is highly effective in capturing local texture patterns and offers computational efficiency, it is less sensitive to global structural information, such as horn curvature, body proportions, and spatial relationships among buffalo body parts. Consequently, important discriminative shape characteristics may not be fully represented in the extracted feature set. Moreover, LBP may not capture deep semantic information as comprehensively as modern deep learning architectures, particularly when dealing with visually similar buffalo breeds or images containing more complex and cluttered backgrounds. These limitations may contribute to feature overlap among classes with similar visual appearances and can affect classification performance in challenging scenarios. Therefore, future studies could explore the integration of texture descriptors with shape-based features or deep feature representations to improve classification robustness, feature discrimination capability, and overall model generalization.

From a practical perspective, the proposed LBP-Random Forest framework offers significant advantages for deployment in real-world livestock management systems. Because the method relies on lightweight feature extraction and conventional machine learning algorithms, it can operate on standard computers or low-cost edge devices without requiring dedicated GPU hardware. This capability makes the approach suitable for field-based buffalo identification and can potentially support digital livestock registration, breeding management, and livestock monitoring activities in rural areas where computational resources may be limited. In addition, the ability to automatically distinguish Toraja buffalo breeds may contribute to breed documentation, conservation programs, and the preservation of cultural heritage associated with indigenous buffalo populations.

4. Conclusions

The effectiveness of Random Forest (RF) and Support Vector Machine (SVM) in categorizing Toraja buffalo photos using Local Binary Pattern (LBP) feature extraction was examined in this work. According to the findings, the RF model outperformed SVM, which only attained 64.17%, with an accuracy of 92.08%. The RF model also showed more steady and consistent performance when evaluated using precision, recall, and F1-score.

When compared to previous research that employed SVM on Nili-Ravi and Khundi buffalo datasets with an

accuracy of 84% [15], the findings of this study indicate a significant performance improvement through the application of LBP with RF. These results confirm that RF is more effective in handling complex data distributions and non-linear texture features generated by LBP.

For future research, this approach can be further developed by combining LBP with other feature extraction methods or hybrid architectures, as well as expanding the dataset with more diverse image variations. In addition, the development of real-time classification systems could be a potential direction to support cultural conservation, livestock management, and the digitalization of local species.

Acknowledgements

This study was supported by the Doctoral Dissertation Research Grant Program administered by the *Lembaga Penelitian dan Pengabdian kepada Masyarakat* (LPPM), Universitas Negeri Malang. The authors would like to express their sincere gratitude to all individuals and institutions whose guidance, assistance, and support contributed to the successful completion of this research. Their valuable contributions were essential throughout the research process. The authors also hope that the findings of this study will advance the development of data classification methods and provide meaningful support for the conservation of the Toraja Buffalo.

References

- [1] L. Nanni, D. Cuza, and S. Brahmam, "AI-Powered Biodiversity Assessment: Species Classification via DNA Barcoding and Deep Learning," *Technologies*, vol. 12, p. 240, Nov. 2024, doi: [10.3390/technologies12120240](https://doi.org/10.3390/technologies12120240).
- [2] T. Miller, G. Michoński, I. Durlak, P. Kozłowska, and P. Biczak, "Artificial Intelligence in Aquatic Biodiversity Research: A PRISMA-Based Systematic Review," *Biology (Basel)*, vol. 14, no. 5, pp. 1–41, 2025, doi: [10.3390/biology14050520](https://doi.org/10.3390/biology14050520).
- [3] S. Nematollahi, S. Fakheran, A. Jafari, S. Pourmanafi, and F. Kienast, "Applying a systematic conservation planning tool and ecological risk index for spatial prioritization and optimization of protected area networks in Iran," *J. Nat. Conserv.*, vol. 66, p. 126144, 2022, doi: <https://doi.org/10.1016/j.jnc.2022.126144>.
- [4] V. Kremsa, "Sustainable management of agricultural resources (agricultural crops and animals)," 2021, pp. 99–145. doi: [10.1016/B978-0-12-824342-8.00010-9](https://doi.org/10.1016/B978-0-12-824342-8.00010-9).
- [5] A. Rachman Manga, H. Herawati, and P. Purnawansyah, "Utilization of Deep Learning YOLO V9 for Identification and Classification of Toraja Buffalo Breeds," *Ilk. J. Ilm.*, vol. 17, pp. 12–19, Apr. 2025, doi: [10.33096/ilkom.v17i1.2349.12-19](https://doi.org/10.33096/ilkom.v17i1.2349.12-19).
- [6] A. Baan, M. Deli, G. Allo, and A. Anto, "Heliyon The cultural attitudes of a funeral ritual discourse in the indigenous Torajan, Indonesia," *Heliyon*, vol. 8, no. February, p. e08925, 2022, doi: [10.1016/j.heliyon.2022.e08925](https://doi.org/10.1016/j.heliyon.2022.e08925).
- [7] P. W. Prihandini et al., "Biodiversity of Indonesian indigenous buffalo: First review of the status, challenges, and development opportunities," *Vet. World*, vol. 16, no. 11, pp. 2217–2229, 2023, doi: [10.14202/vetworld.2023.2217-2229](https://doi.org/10.14202/vetworld.2023.2217-2229).
- [8] T. Maulana, H. Iskandar, S. Said, and A. Gunawan, "The Current Status and Potential Development of Genetic Resources of Indigenous Toraja Spotted Buffalo in Indonesia: A Systematic Review," *World's Vet. J.*, vol. 13, pp. 617–625, Dec. 2023, doi: [10.54203/scil.2023.wvj66](https://doi.org/10.54203/scil.2023.wvj66).
- [9] P. Chand and S. Lal, "Vision-Based Detection and Classification of Used Electronic Parts," *Sensors*, vol. 22, no. 23, 2022, doi: [10.3390/s22239079](https://doi.org/10.3390/s22239079).
- [10] Z. Wu, Y. Chen, B. Zhao, X. Kang, and Y. Ding, "Review of Weed Detection Methods Based on Computer Vision," *Sensors*, vol. 21, p. 3647, May 2021, doi: [10.3390/s21113647](https://doi.org/10.3390/s21113647).
- [11] D. A. Borges Oliveira, L. G. Ribeiro Pereira, T. Bresolin, R. E. Pontes Ferreira, and J. R. Rebouças Dorea, "A review of deep learning algorithms for computer vision systems in livestock," *Livest. Sci.*, vol. 253, p. 104700, 2021, doi: <https://doi.org/10.1016/j.livsci.2021.104700>.
- [12] G. Li et al., "Practices and applications of convolutional neural network-based computer vision systems in animal farming: A review," *Sensors*, vol. 21, no. 4, pp. 1–42, 2021, doi: [10.3390/s21041492](https://doi.org/10.3390/s21041492).
- [13] V. Patil, A. Vaidya, and M. Patil, "Automated Brain Tumor Classification on Brain MRI Images uses CNN Feature Extraction and Machine Learning Techniques," *Basrah J. Sci.*, vol. 43, pp. 708–736, Dec. 2025, doi: [10.29072/basjs.20250309](https://doi.org/10.29072/basjs.20250309).
- [14] V. Patil, J. Hrishikesh, and M. Patil, "Infant's MRI Brain Tissue Segmentation using Integrated CNN Feature Extractor and Random Forest," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 11, pp. 71–79, Jan. 2023, doi: [10.17762/ijritcc.v11i1s.6002](https://doi.org/10.17762/ijritcc.v11i1s.6002).
- [15] A. Rachman Manga, A. Nanda, A. Handayani, H. Wahyu, R. Andrie, and D. Lantara, "Comparison Between Single-Input and Multi-Input Classification with the Application of Canny Feature Extraction and Classification Algorithms on the Toraja Buffalo Dataset," 2025. doi: [10.1109/IMCOM64595.2025.10857554](https://doi.org/10.1109/IMCOM64595.2025.10857554).
- [16] Y. Pan, H. Jin, J. Gao, and H. T. Rauf, "Identification of Buffalo Breeds Using Self-Activated-Based Improved Convolutional Neural Networks," *Agriculture*, vol. 12, p. 1386, Sep. 2022, doi: [10.3390/agriculture12091386](https://doi.org/10.3390/agriculture12091386).
- [17] A. R. Manga, "Data Buffalo Toraja," *Mendeley Data*, V2, doi: <https://dx.doi.org/10.17632/kbft73pdkw.1>.
- [18] A. R. Manga, M. A. Attalah, R. Puspitasari, S. Paembonan, D. Lantara, and M. A. Asis, "Multilabel Classification Buffalo Datasets Using Transfer Learning Models," in *2025 9th International Conference On Electrical, Electronics And Information Engineering (ICEEIE)*, 2025, pp. 1–5. doi: [10.1109/ICEEIE66203.2025.11253914](https://doi.org/10.1109/ICEEIE66203.2025.11253914).
- [19] B. Yang et al., "Identification of Species by Combining Molecular and Morphological Data Using Convolutional Neural Networks," *Syst. Biol.*, vol. 71, Sep. 2021, doi: [10.1093/sysbio/syab076](https://doi.org/10.1093/sysbio/syab076).
- [20] M. Alweshah, L. Rababa, M. Ryalat, A. Almomani, and M. Ababneh, "African Buffalo Algorithm: Training the Probabilistic Neural Network to Solve Classification Problems," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, Jul. 2020, doi: [10.1016/j.jksuci.2020.07.004](https://doi.org/10.1016/j.jksuci.2020.07.004).
- [21] R. Singh, S. Chun, and V. Atluri, *Developing Machine Learning Models to Automate News Classification*. 2020. doi: [10.1145/3396956.3397001](https://doi.org/10.1145/3396956.3397001).
- [22] J. Pathmarasa, U. Wijenayake, R. Wijesinghe, and B. Silva, *An Exploratory Study of Diverse Models and Datasets for Transfer Learning based Image Classification on Sparse Data*. 2024. doi: [10.1109/MERCon63886.2024.10688577](https://doi.org/10.1109/MERCon63886.2024.10688577).
- [23] A. Parashar, A. Parashar, W. Ding, M. Shabaz, and I. Rida, "Data preprocessing and feature selection techniques in gait recognition: A comparative study of machine learning and deep learning approaches," *Pattern Recognit. Lett.*, vol. 172, pp. 65–73, 2023, doi: <https://doi.org/10.1016/j.patrec.2023.05.021>.
- [24] I. Chahid, A. Kerkour Elmiad, and B. Mohammed, *Data Preprocessing For Machine Learning Applications in Healthcare: A Review*. 2023. doi: [10.1109/SITA60746.2023.10373591](https://doi.org/10.1109/SITA60746.2023.10373591).
- [25] Herman, M. D. A. Putra, H. Nasir, H. Darwis, S. H. Mansyur, and M. N. M. Noor, "Measuring the Performance of VGG-16, VGG-19, and a Concatenated Model Architecture in Toraja Carving Classification," in *2025 19th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, 2025, pp. 1–7. doi: [10.1109/IMCOM64595.2025.10857556](https://doi.org/10.1109/IMCOM64595.2025.10857556).

- [26] M. Shahnaz and A. Mollah, "On the Performance of Convolutional Neural Networks with Resizing and Padding," 2022, pp. 51–62. doi: [10.1007/978-981-19-0105-8_6](https://doi.org/10.1007/978-981-19-0105-8_6).
- [27] T. D. Rahmani, A. S. Safitri, I. Maulidin, I. Wijayanto, and S. Hadiyoso, "Employing Up-Sampling and Augmentation for Imbalanced Datasets in Skin Cancer Classification Using CNN Architecture," in *2024 8th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, 2024, pp. 453–457. doi: [10.1109/ICITISEE63424.2024.10729972](https://doi.org/10.1109/ICITISEE63424.2024.10729972).
- [28] H. Darwis, R. Puspitasari, Purnawansyah, W. Astuti, D. Atmajaya, and M. Hasnawi, "A Deep Learning Approach for Improving Waste Classification Accuracy with ResNet50 Feature Extraction," in *2025 19th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, 2025, pp. 1–8. doi: [10.1109/IMCOM64595.2025.10857536](https://doi.org/10.1109/IMCOM64595.2025.10857536).
- [29] A. Manga, A. Utami, H. Azis, Y. Salim, and A. Faradibah, "Optimizing classification models for medical image diagnosis: a comparative analysis on multi-class datasets," *Comput. Sci. Inf. Technol.*, vol. 5, pp. 205–214, Nov. 2024, doi: [10.11591/csit.v5i3.p205-214](https://doi.org/10.11591/csit.v5i3.p205-214).
- [30] A. R. Manga', A. N. Handayani, H. W. Herwanto, R. A. Asmara, Y. I. Sulistya, and K. Kasmira, "Analysis of the Ensemble Method Classifier's Performance on Handwritten Arabic Characters Dataset," *Ilk. J. Ilm.*, vol. 15, no. 1, pp. 186–192, 2023, doi: [10.33096/ilkom.v15i1.1357.186-192](https://doi.org/10.33096/ilkom.v15i1.1357.186-192).
- [31] Y. Salim, A. F. Rakasyah, H. Darwis, Harlinda, Irawati, and A. R. Manga, "Comparative Analysis of Machine Learning Algorithms and Ensemble Techniques for Diverse Image Classification Tasks," in *2025 19th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, 2025, pp. 1–7. doi: [10.1109/IMCOM64595.2025.10857517](https://doi.org/10.1109/IMCOM64595.2025.10857517).
- [32] Purnawansyah et al., "Congestion Predictive Modelling on Network Dataset Using Ensemble Deep Learning," *J. Appl. Data Sci.*, vol. 5, no. 4, pp. 1597–1613, 2024, doi: [10.47738/jads.v5i4.333](https://doi.org/10.47738/jads.v5i4.333).
- [33] A. Roy and S. Chakraborty, "Support vector machine in structural reliability analysis: A review," *Reliab. Eng. Syst. Saf.*, vol. 233, p. 109126, 2023, doi: <https://doi.org/10.1016/j.res.2023.109126>.
- [34] P. Chhajaj, M. Shah, and A. Kshirsagar, "The applications of artificial neural networks, support vector machines, and long-short term memory for stock market prediction," *Decis. Anal. J.*, vol. 2, p. 100015, 2022, doi: <https://doi.org/10.1016/j.dajour.2021.100015>.
- [35] M. Bansal, A. Goyal, and A. Choudhary, "A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning," *Decis. Anal. J.*, vol. 3, p. 100071, 2022, doi: <https://doi.org/10.1016/j.dajour.2022.100071>.
- [36] X. Ding, J. Liu, F. Yang, and J. Cao, "Random radial basis function kernel-based support vector machine," *J. Franklin Inst.*, vol. 358, no. 18, pp. 10121–10140, 2021, doi: <https://doi.org/10.1016/j.jfranklin.2021.10.005>.
- [37] Herman, H. Darwis, Nurfauziyah, R. Puspitasari, D. Widyawati, and A. Faradibah, "Comparative Analysis of Anxiety Disorder Classification Using Algorithm Naïve Bayes, Decision Tree and K-NN," in *2025 19th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, 2025, pp. 1–6. doi: [10.1109/IMCOM64595.2025.10857485](https://doi.org/10.1109/IMCOM64595.2025.10857485).
- [38] R. F. Syam, "Performance Comparison Analysis of Classifiers on Binary Classification Dataset," *Indones. J. Data Sci.*, 2023, doi: <https://doi.org/10.56705/ijodas.v4i2.77>.
- [39] R. J. Suhatri, R. D. Syah, M. Hermita, B. Gunawan, and W. Silfianti, "Evaluation of Machine Learning Models for Predicting Cardiovascular Disease Based on Framingham Heart Study Data," *Ilk. J. Ilm.*, vol. 16, no. 1, pp. 68–75, 2024, doi: [10.33096/ilkom.v16i1.1952.68-75](https://doi.org/10.33096/ilkom.v16i1.1952.68-75).
- [40] H. Azis, L. Syafie, F. Fattah, and ..., "Unveiling Algorithm Classification Excellence: Exploring Calendula and Coreopsis Flower Datasets with Varied Segmentation Techniques," *2024 18th Int.*, 2024, doi: [10.1109/IMCOM60618.2024.10418246](https://doi.org/10.1109/IMCOM60618.2024.10418246).
- [41] Harlinda, A. Rendi, H. Azis, D. Indra, L. N. Hayati, and N. Kurniati, "Classification of Cia-Cia Letters Using MobileNetV2 and CNN Methods," in *2025 19th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, 2025, pp. 1–6. doi: [10.1109/IMCOM64595.2025.10857478](https://doi.org/10.1109/IMCOM64595.2025.10857478).
- [42] A. Rachman and Y. Salim, "Evaluation of VGG16 Performance in Multi-Input and Multi-Class Classification of Toraja Buffalo Breeds," vol. 8, no. 1, pp. 1–10, 2024, doi: <https://doi.org/10.29099/ijair.v8i1.1.1203>.