Institut Riset dan Publikasi Indonesia (IRPI)
MALCOM: Indonesian Journal of Machine Learning and Computer Science
Journal Homepage: https://journal.irpi.or.id/index.php/malcom
Vol. 6 Iss. 1 January 2026, pp: 94-105
ISSN(P): 2797-2313 | ISSN(E): 2775-8575

Performance of K-Nearest Neighbors and Advanced Metaheuristic Algorithms for Feature Selection in Classifying the Purity of Civet Coffee

Shinta Widyaningtyas[1*], Muhammad Arwani[2], Ririn Fatma Nanda[3]

[1]Departemen of Agricultural Techology, Politeknik Negeri Jember, Indonesia
[2]Departemen Agroindustrial Technology, Universitas Nahdlatul Ulama Indonesia, Indonesia
[3]Departemen Agroindustrial Technology, Universitas Jambi, Indonesia

E-Mail: [1]shinta_widya@polije.ac.id, [2]m.arwani@unusia.ac.id, [3]ririnfatma.nanda@unja.ac.id

Received Oct 28th 2025; Revised Dec 09th 2025; Accepted Dec 22th 2025; Available Online Jan 16th 2026
Corresponding Author: Shinta Widyaningtyas

Abstract

Various studies have shown that feature selection can improve classification accuracy, particularly in agriculture. However, most of these studies still use conventional metaheuristic algorithms, which have certain limitations, including a tendency to get stuck in local optima. Therefore, this study explores the potential of advanced metaheuristic algorithms for selecting colour and texture features to classify the purity of civet coffee. This study used k-Nearest Neighbour (K-NN) model optimized with several advanced metaheuristic algorithms, i.e. Bare Bones Particle Swarm Optimisation (BBPSO), Modified Generalised Flower Pollination Algorithm (MGFPA), Enhanced Salp Swarm Algorithm (ESSA), Improved Salp Swarm Algorithm (ISSA), and Two-Stage Modified Grey Wolf Optimizer (TMGWO). The results show that feature selection can improve model accuracy. The best model was obtained from a combination of K-NN and TMGWO with an accuracy of 0.981, precision of 0.982, recall of 0.981, F1-Score of 0.981, and Area Under Curve (AUC) close to 1 with three selected features, i.e. blue correlation, s_hsl_correlation, and s_hsv_correlation. Furthermore, the results of this study indicate that the development of advanced metaheuristic algorithms can overcome the weaknesses of conventional algorithms, as demonstrated by improvements in classification model accuracy and the number of selected features.

Keywords: Advanced Metaheuristic Algorithms, Civet Coffee, Feature Selection, K-Nearest Neighbors

1. INTRODUCTION

Ensuring the purity of civet coffee is very important given the increasing number of cases of civet coffee counterfeiting. Typically, green civet coffee beans are counterfeited with regular green coffee beans to meet market demand amid production constraints. Therefore, the development of advanced technology for testing the purity of civet coffee is needed. To date, there is no fast, inexpensive, real-time, and non-destructive tool for classifying the purity of civet coffee. Conventional purity testing of civet coffee is destructive because it requires a series of chemical tests and experts. One of the sensors that can be used in the development of non-destructive and real-time tools for classifying the purity of civet coffee is computer vision.

Computer vision has been used in various studies to detect the quality of materials, especially agricultural products, such as grading tea commodities [1], evaluating the quality of Congou black tea [2], early detection of Colletotrichum Kahawae disease in coffee cherries [3], and the roasting level of coffee [4]. Computer vision will capture the colour and texture features of civet coffee. However, not all colour and texture features are relevant in classifying the purity of civet coffee. Using all colour and texture features can reduce model performance due to noise, outliers, and data redundancy [5]. Feature selection consistently improves classification accuracy across various machine learning models [6], [7]. The results [8] show an increase in accuracy across several machine learning models, including k-Nearest Neighbours, Decision Trees, and Multilayer Perceptron. In addition, study [9] shows an increase in accuracy of 9.7% in liver disease prediction. A study [10] also shows that feature selection can reduce validation MSE by 30.174% and 64.602% using Artificial Neural Network (ANN) and Random Forest (RF) modelling, respectively. Therefore, this study presents a feature selection of colour and texture that can classify the purity of civet coffee.

DOI: https://doi.org/10.57152/malcom.v6i1.2345

Feature selection requires algorithm that can find the best solutions, such as metaheuristic algorithms. These algorithm are inspired by natural principles or phenomena, such as animal behaviour, plants, and other biological processes [11]. The main idea is to translate these natural principles or phenomena into mathematical models to find optimal solutions in complex search spaces. However, in their implementation, some metaheuristic algorithms have several weaknesses, such as being easily trapped in local optima and slow convergence. Therefore, advanced metaheuristic algorithms have been developed, such as Bare Bones Particle Swarm Optimization (BBPSO), which was developed from the Particle Swarm Optimization algorithm; Modified Generalized Flower Pollination Algorithm (MGFPA), which was developed from the Flower Pollination Algorithm (FPA); Enhanced Salp Swarm Algorithm (ESSA) and Improved Salp Swarm Algorithm (ISSA) developed from the Salp Swarm Algorithm (SSA); Two-Stage Modified Grey Wolf Optimizer developed from the Grey Wolf Optimizer (GWO). Several conventional algorithm and bio-inspired developments have been used in feature selection in various studies, such as PSO for corn seed classification [12], BBPSO for classification in 16 datasets [13], FPA for spam email detection [14], MGFPA for classification in 18 UCI datasets [15], SSA for intrusion detection dataset classification [16], ESSA for solving global optimization and complex engineering processes [17], ISSA for classification on 23 UCI datasets [18], TMGWO can reduce features by nearly 97.5% and improve classification accuracy across various datasets [19].

In carrying out feature selection, the use of advanced bio-inspired algorithms needs to be supported by machine learning classification models such as k-Nearest Neighbours (K-NN). These model have been widely used to solve classification problems. The best model criteria are based on the highest accuracy, precision, recall, and F1-score values from the confusion matrix. In addition, the best model criteria also consider the number of colour and texture features selected for the design of automatic detection tools. The novelty of this research is that no previous studies have reported the performance of advanced metaheuristic algorithms in classifying the purity of civet coffee using computer vision data. Previous studies focused on exploring colour features, texture features, and combinations of both using conventional metaheuristic algorithms and machine learning [20], [21].

Based on this background, the present study aims to develop a feature selection model using the K-NN model and an advanced metaheuristic approach for classifying the purity of civet coffee. Specifically, the study evaluates the model in terms of improvements in classification accuracy, the number of selected features, and compares the performance of the advanced and conventional metaheuristic algorithms. The results of this study are expected to serve as a reference for the development of methods for detecting the purity of civet coffee.

2. MATERIALS AND METHOD

Several previous studies have employed conventional metaheuristic algorithms such as Particle Swarm Optimisation (PSO), Flower Pollination Algorithm (FPA), Salp Swarm Algorithm (SSA), and Grey Wolf Optimisation (GWO) for feature selection across various datasets. However, a major drawback of some of these algorithms is their tendency to become trapped in local optima. As a result, conventional metaheuristic algorithms have undergone further development. These developments include Bare Bones Particle Swarm Optimisation (BBPSO), which eliminates the velocity parameter; Modified Generalised Flower Pollination Algorithm (MGFPA), which incorporates a Lévy flight strategy; Enhanced Salp Swarm Algorithm (ESSA) and Improved Salp Swarm Algorithm (ISSA), which increase search diversity; and Two-Stage Modified Grey Wolf Optimizer (TMGWO), which applies a two-stage strategy to improve exploration and exploitation. These advanced metaheuristic algorithms have been shown to yield higher classification accuracy in several previous studies, making them suitable for application in assessing the purity of civet coffee. The research design is illustrated in Figure 1.

2.1. Materials

This study used green civet coffee beans and regular coffee from PTPN XII Banyuwangi Indonesia. The coffee was mixed at four levels of purity: very low (0-25%), low (25-50%), medium (50-75%), and high (75-100%). Feature extraction was performed using Visual Basic 6.0, while feature selection was performed using Python 3.14.0 and Google Collaboration. The entire feature selection process was performed using a computer with the following specifications: Intel Core i7-1165G7 2.80 GHz, 8 GB RAM.

2.2. Methods

2.2.1. Image Acquisition

Coffee mixed in various percentages was placed on a mini studio tray. Then, a Nikon Coolpix camera was placed 10 cm above the sample. Image acquisition was performed 33 times, resulting in 528 images in bitmap format (.bmp). Next, the images were cropped to clean up the images from noise and standardized the image size. The image data was ready for feature extraction.

Figure 1. Research Design

2.2.2. Feature Extraction

Feature extraction is conducted to obtain texture and colour feature values. The results of feature extraction are a co-occurrence matrix of colour features, including RGB, HSV, HSL, L*a*b; while texture features include entropy, contrast, homogeneity, sum mean, variance, correlation, maximum probability, inverse difference moment and cluster tendency. A summary of the feature extraction results is provided in a Microsoft Excel file. Then, the extracted feature data is uploaded to Google Drive for classification using Python and Visual Studio Code.

2.2.3. Model Design

1. Bare Bones Particle Swarm Optimization (BBPSO)
   BBPSO is a simplified variant of standard Particle Swarm Optimization (PSO) that eliminates the velocity term and replaces it with Gaussian probabilistics, making this parameter-free and easier to implement [22]. In conventional PSO, particles move by adjusting their speed based on personal best (pbest) and global best (gbest). Meanwhile, in BBPSO, the position of new particles is generated stochastically through a Gaussian distribution whose mean is between the pbest and gbest [23]. A Gaussian distribution is used to control the behavior of the particles [24]. The main advantages of BBPSO over conventional PSO are that it is simpler, more efficient, and has fast convergence. The BBPSO workflow is as follows:
   a. Initialization: Randomly generate a population of particles in the search space. Each particle has a position (a potential solution).
   b. Evaluation: Calculate the fitness value (quality) for each particle based on the objective function.
   c. Update Personal and Global Best: For each particle, update:
      1) pBest: The best position ever achieved by the particle itself.
      2) gBest: The best position ever achieved by any particle in the entire population.

   d. Update Position: For each dimension d of particle i, the new position is calculated as: New_Position[i][d] = N(μ, σ)

      Where:
      1) μ = (pBest[i][d] + gBest[d]) / 2 (the mean between pBest and gBest)
      2) σ = |pBest[i][d] - gBest[d]| (the absolute difference between them)
      3) N() is a function that generates a random number from a Gaussian distribution.

   e. Repeat steps 2-4 until a stopping criterion is met (e.g., maximum number of iterations is reached or a sufficiently good solution is found).

2. Modified Generalized Flower Pollination Algorithm (MGFPA)
   MGFPA is a development of the Flower Pollination Algorithm (FPA), which is inspired by the process of flower pollination. This algorithm modifies the strategy for selecting solutions, searching for solutions, and increasing flexibility in exploring and exploiting the search space. MGFPA is more efficient at solving complex optimization problems, although it requires more computation and time to reach convergence [25]. The MGFPA workflow is as follows:
   a. Initialization: Generate a population of flowers (solutions) randomly.
   b. Evaluation: Calculate the fitness value of each flower.
   c. Pollination: In each iteration, for each flower i:
      1) Global Pollination (Cross-pollination): With a probability p, perform:
         New_Flower[i] = Current_Flower[i] + L * (gBest - Current_Flower[i]), where L is a step size drawn from a Lévy flight distribution (for long-range exploration). The Modification (M) is often applied to parameter p or the calculation of L.
      2) Local Pollination (Self-pollination): With a probability (1-p), perform:
         New_Flower[i] = Current_Flower[i] + ε * (Flower[j] - Flower[k]) where Flower[j] and Flower[k] are two randomly selected flowers from the same population, and ε is a random number between [0,1]. This is local exploitation.

   d. Selection: Compare the new flower with the old one. If the new flower is better, replace the old flower.
   e. Elitism Strategy: Modifications often include an elitism mechanism to ensure the best flower (gBest) is not lost.
   f. Repeat steps 2-4 until the stopping criterion is met.

3. Enhanced Salp Swarm Algorithm (ESSA)
   ESSA is an extension of the Salp Swarm Algorithm (SSA). The SSA algorithm is inspired by the behaviour of salp groups in the ocean [26], [27]. The salp chain is divided into two groups: leaders and followers. Leaders update their position towards the food source as the best solution, while followers move in a chain following the leaders. However, the main drawback of SSA is that if the leader gets stuck in a local optimum, the entire salp chain will move in the direction of that local solution [28], [29]. In this context, reduced exploration of the search space. Therefore, ESSA adds an adaptive parameter that decreases with each iteration and random perturbations to the position of the

followers so that they do not always follow the leader, thereby increasing the likelihood of escaping the local optimum. The ESSA workflow is as follows:

a. Initialization: Initialize the salp population. Each salp is a potential solution. The population is divided into two groups: Leader (the first salp in the chain) and Followers (the rest).
b. Evaluation: Calculate the fitness of each salp. Identify the salp with the best fitness as the food source (F).
c. Update Leader Position: The leader's position is updated with an equation that directs it towards the food source:
   Leader_Position = F + c1 * ((ub - lb) * c2 + lb), where c1 is a coefficient that decreases over iterations (balancing exploration/exploitation), and c2, c3 are random numbers. If c3 < 0.5, the leader moves toward the food; otherwise, it moves away (for exploration).
d. Update Follower Positions: The followers positions are updated based on the movement of the salp in front of them (not directly towards the food), mimicking a chain movement.
   Follower_Position[i] = (Follower_Position[i] + Follower_Position[i-1]) / 2
e. Elitism Strategy: Before or after updating positions, an elitism strategy is applied. For example, a number E of the best salps (elite) from the old population can be directly carried over to the new population, or used to guide the update of other salps, ensuring the best information is preserved.
f. Repeat steps 2-5 until the stopping criterion is met.

4. Improved Salp Swarm Algorithm (ISSA)
   In general, ISSA was developed to address the weaknesses of SSA [30]. ISSA was developed by updating the leader's position with a non-linear adaptive function that increases exploration at the beginning and exploitation at the end, and the presence of random follower perturbations to avoid local optima traps. The main advantage of ISSA is its effectiveness in solving high-dimensional problems. The ISSA workflow is as follows:
   a. Initialization: Instead of pure random generation, use a Chaos Map to initialize the salp positions. This helps improve population diversity and initial quality.
   b. Evaluation: Calculate fitness values and identify the food source (F).
   c. Update Position: Update the leader and follower positions as in the standard SSA.
   d. Opposition-Based Learning: After the position update, apply OBL to a portion of the population (e.g., the worst salps). For each selected salp, calculate its opposite position (X_opp = ub + lb - X). Evaluate the fitness of this opposite position. If the opposite position is better, replace the original salp with it. This helps explore the opposite region of the search space, which may be more promising.
   e. Repeat steps 2-4 until the stopping criterion is met.

5. Two-Stage Modified Grey Wolf Optimizer (TMGWO)
   In general, TMGWO is a variant of Grey Wolf Optimizer inspired by the leadership hierarchy and hunting mechanisms of grey wolves to find global solutions [31]. TMGWO is designed to improve GWO performance by modifying its strategy through two search stages. These stages aim to improve convergence speed and avoid local optima. The ISSA workflow is as follows:
   a. Initialization Phase:
      1) Initialize Parameters: Set the population size, maximum iterations (Max_iter), and the switch iteration (switch_iter) that divides Stage 1 and Stage 2.
      2) Initialize Population: Generate the initial population of grey wolves randomly within the search boundaries.

   b. Global Exploration Stage (Iterations 1 to switch_iter) : the objective of this stage is encourage widespread exploration
   c. Local Exploitation Stage (Iterations switch_iter+1 to Max_iter) : the objective of this stage are rapid and precise convergence to the global optimum
   d. Evaluation & Update: After the position update in either stage, evaluate the fitness of all new positions and update the α, β, and δ wolves.
   e. Repeat the process (the active stage) until the maximum iteration Max_iter is reached. The position of the Alpha wolf is returned as the best-found solution.

2.2.4. Classification Model

K-Nearest Neighbors (K-NN) is a distance-based classification model. This model works by finding k nearest neighbours based on the euclidean distance from the input data, then determining the class or target value based on the majority of neighbours. K-NN is classified as a lazy learning algorithm because there is no complex training process. The major advantage of K-NN is easy to implement and does not require complex computation. This study uses the K-NN k-value parameter. The k-value in K-NN refers to the number of nearest neighbors used in determining the class, while leaf size is the key parameter that tunes the trade-off between accuracy and speed [32], [33]. The k-value in K-NN includes 3, 5, 7, 9, 11, 13 and 15. According to [34], selecting the k-value is important in building a classification model using K-NN. In this context, if the k-value is too small, there will be a lot of noise, which reduces classification accuracy; if it is too large, it can also cause errors due to limited values and indirectly affect accuracy.

2.2.5. Evaluation Model

Evaluate the classification model using accuracy, precision, recall, and F1-score based on the confusion matrix. In addition, the model will be evaluated using the Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC). The equations for accuracy, precision, recall, and F1-score are as equations 1-4.

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

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

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

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

3. RESULTS AND DISCUSSION

This study selected features in the colour and texture data of civet coffee using several advanced bio-inspired algorithms and machine learning models such as k-Nearest Neighbours (K-NN). These machine learning models were used because can generally solve classification problems in non-linear data. Table 1 shows the performance of K-NN and advanced metaheuristic algorithms in classifying the purity of civet coffee.

Table 1. Conventional K-NN

k-Value	Accuracy	Precision	Recall	F1-Score
3	0.877	0.878	0.877	0.877
5	0.889	0.889	0.889	0.889
7	0.871	0.872	0.871	0.871
9	0.889	0.889	0.889	0.889
11	0.871	0.872	0.871	0.872
13	0.868	0.867	0.866	0.866
15	0.849	0.849	0.849	0.848

Table 1 shows that conventional K-NN (K-NN without a metaheuristic optimization algorithm) can classify data on the colour and texture features of civet coffee purity with the highest accuracy of 0.889, followed by precision of 0.889, recall of 0.889, and F1-score of 0.889. In conventional K-NN, all colour and texture features are used as classification inputs. This makes it difficult to create an automatic detection tool for classifying the purity of civet coffee. This is because the more features that are extracted, the more it will affect the speed of the tool in detecting the purity of civet coffee.

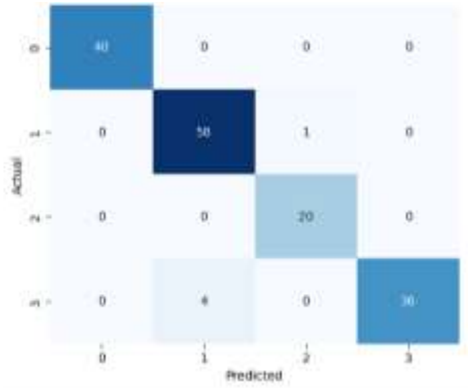
Table 2 shows the classification results using K-NN and advanced metaheuristic algorithms. Compared to Table 1, the classification model shows improved performance in terms of accuracy, precision, recall, and F1-Score. This indicates that not all colour and texture features are relevant in classifying the purity of civet coffee. The use of all colour and texture features can actually reduce the accuracy of the classification results. Feature selection improves classification accuracy by reducing dimensionality and removing irrelevant features [35]. Table 2 also shows that the best model in each advanced metaheuristic algorithm, BBPSO, achieved its best performance at a k-value of 11 with an accuracy of 0.968; precision of 0.970; recall of 0.969; and F1-Score of 0.968 with 16 selected features. The best MGFPA algorithm was achieved with a k-value of 5, with an accuracy of 0.981; precision of 0.982; recall of 0.981; and F1-Score of 0.981 with 17 selected features. The best classification models for the ESSA and ISSA algorithms were obtained at a k-value of 9 with an accuracy of more than 0.9 with 19 and 16 selected features, respectively. The last metaheuristic algorithm is TMGWO, which achieved the best accuracy at a k-value of 15 with an accuracy of 0.981; precision of 0.982; recall of 0.981; and an F1-Score of 0.981 with 3 selected features. The results of the study show that TMGWO is the most reliable metaheuristic algorithm in feature selection.

Compared to MGFPA, TMGWO was chosen as the best classification model because it considers the number of selected features. The advantages of TMGWO include shorter computation time compared to other metaheuristic algorithms, better classification accuracy even when using fewer features, and less susceptibility to local optima due to its two-stage search process [19]. This advantage also enabled this study to achieve the highest accuracy with only three selected features compared to other metaheuristic algorithms. The three selected features are blue correlation, s_hsl_correlation, and s_hsv_correlation. Correlation in texture feature measures correlation between two pixels in the pixel pairs. The correlation is expected to be high if the gray-levels of the pixel pairs are highly correlated [21]. The blue colour, HSL saturation and HSV saturation in the correlation show texture variations in various colour spaces. The combination of these three features is quite representative in distinguishing the purity levels of civet coffee, supported by classification accuracy results of > 0.9.

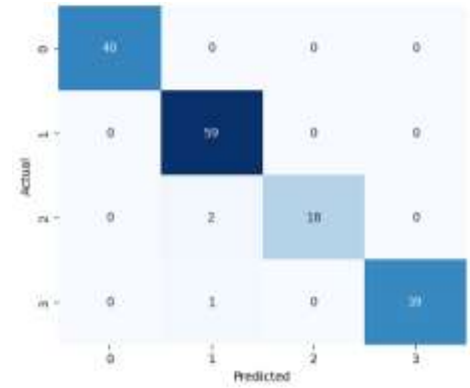
Table 2. K-NN and Advanced Metaheuristic Algorithms Performance

Advanced Metaheuristic Algorithms	k-Value	Accuracy	Precision	Recall	F1-Score	Number of Selected Feature
Bare Bones Particle Swarm Optimization (BBPSO)	3	0.937	0.939	0.937	0.937	14
	5	0.937	0.938	0.937	0.937	21
	7	0.918	0.919	0.918	0.918	12
	9	0.925	0.926	0.925	0.924	19
	11	0.968	0.970	0.969	0.968	16
	13	0.925	0.925	0.925	0.925	14
	15	0.918	0.920	0.918	0.918	8
Modified Generalized Flower Pollination Algorithm (MGFPA)	3	0.937	0.939	0.937	0.937	18
	5	0.981	0.982	0.981	0.981	17
	7	0.925	0.927	0.925	0.925	16
	9	0.931	0.931	0.931	0.931	30
	11	0.931	0.931	0.931	0.931	17
	13	0.925	0.925	0.925	0.925	23
	15	0.918	0.919	0.918	0.918	18
Enhanced Salp Swarm Algorithm (ESSA)	3	0.951	0.952	0.951	0.951	10
	5	0.945	0.946	0.945	0.955	4
	7	0.945	0.945	0.955	0.945	12
	9	0.962	0.962	0.962	0.962	10
	11	0.957	0.958	0.957	0.957	9
	13	0.957	0.958	0.957	0.957	10
	15	0.941	0.941	0.941	0.941	3
Improved Salp Swarm Algorithm (ISSA)	3	0.943	0.945	0.943	0.943	21
	5	0.943	0.944	0.943	0.944	24
	7	0.931	0.931	0.931	0.931	17
	9	0.975	0.976	0.975	0.975	16
	11	0.937	0.937	0.937	0.937	18
	13	0.967	0.967	0.967	0.967	23
	15	0.931	0.931	0.931	0.931	19
Two-Stage Modified Grey Wolf Optimizer (TMGWO)	3	0.962	0.9630	0.9620	0.9620	3
	5	0.969	0.971	0.969	0.968	4
	7	0.951	0.952	0.951	0.951	7
	9	0.969	0.969	0.969	0.968	4
	11	0.937	0.937	0.937	0.937	2
	13	0.950	0.950	0.950	0.950	6
	15	0.981	0.982	0.981	0.981	3

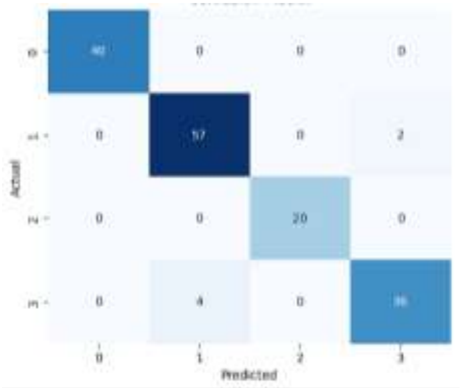
Compared to previous studies in the field of agricultural commodity quality classification, the performance achieved in the present study is notably higher. Studies related to coffee defect classification generally report overall accuracies ranging from 0.885 to 0.947 [36] and average accuracies between 0.9499 and 0.9566 [37], using colour and shape features with various classification methods such as Deep Neural Networks (DNN), Random Forest (RF), Support Vector Machine (SVM), and k-Nearest Neighbours (K-NN), without applying feature selection. Another study performed feature selection using correlation analysis and important feature ranking, which resulted in an increase in accuracy; however, the improvement was relatively small, less than 20%, in the classification of specialty coffees [38]. Additionally, the use of conventional metaheuristic algorithms such as FPA combined with K-NN produced a best accuracy of 0.918 using five selected features [39]. In contrast, the advanced metaheuristic algorithms employed in the present study—particularly the Two-Stage Modified Grey Wolf Optimizer (TMGWO)—achieved an accuracy of 0.981 with fewer selected features than those used in previous studies. These results indicate that advanced metaheuristic algorithms can more effectively explore the search space, producing a more compact and representative subset of features for distinguishing the purity level of civet coffee.



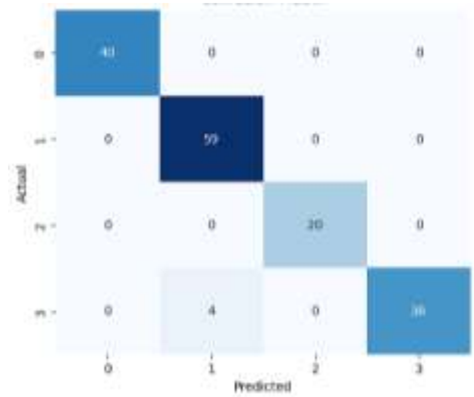
(a)



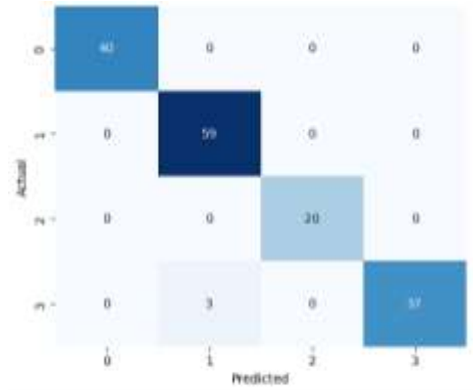
(b)



(c)



(d)



(e)

Figure 2. Confusion Matrix of the Best Model for Each Advanced Metaheuristic Algorithms (a) Bare Bones Particle Swarm Optimization (BBPSO); (b) Modified Generalized Flower Pollination Algorithm (MGFPA); (c) Enhanced Salp Swarm Algorithm (ESSA); (d) Improved Salp Swarm Algorithm (ISSA); (e) Two-Stage Modified Grey Wolf Optimizer (TMGWO)

Figure 2 shows the confusion matrix for the best model of each advanced metaheuristic algorithms. The confusion matrix is a table used to assess the performance of a classification model by comparing actual and predicted samples, showing how much the model confuses classes by mislabeling them [40]. Each confusion matrix shows the number of correct and incorrect predictions for each class. Based on the confusion matrix for each algorithm, all metaheuristic algorithms produce good classification performance with high true positive rates in all classes. The MGFPA and TMGWO algorithms show excellent classification performance as there are only 3 errors in the classification. This is consistent with the accuracy, precision, recall and F1-Score results in Table 2.

Figure 3 shows the Receiver Operating Characteristic (ROC) curve for the best model of each advanced metaheuristic algorithms. The ROC curve is a graph that shows the ability of a classification model to distinguish between positive and negative classes at various probability thresholds. In general, the ROC curve shows good performance because the Area Under Curve (AUC) value is close to 1. This indicates that the classification performance of the advanced metaheuristic algorithms successfully classify the purity of civet coffee. The AUC value ranges from $0 \leq AUC \leq 1$. When the AUC value = 0.5 (diagonal line), it means that the classification model has no ability to distinguish between classes, while AUC < 0.5 indicates that the model unable distinguish between classes. Figure 2 shows that none of the AUC values are < 0.5, indicating that the model is reliable in distinguishing between classes.

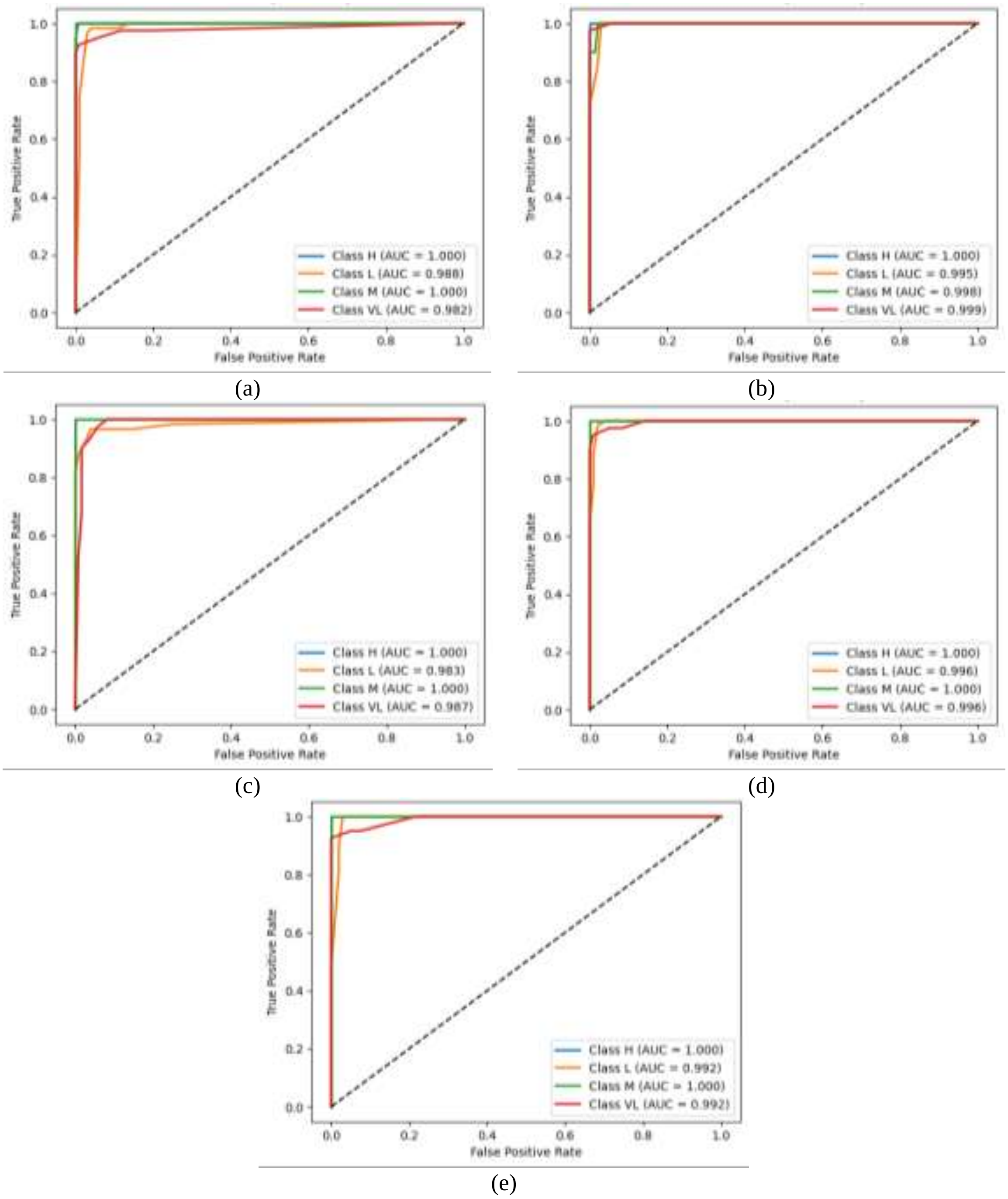


Figure 3. Receiver Operating Characteristic (ROC) of the Best Model for Each Advanced Metaheuristic Algorithms (a) Bare Bones Particle Swarm Optimization (BBPSO); (b) Modified Generalized Flower Pollination Algorithm (MGFPA); (c) Enhanced Salp Swarm Algorithm (ESSA); (d) Improved Salp Swarm Algorithm (ISSA); (e) Two-Stage Modified Grey Wolf Optimizer (TMGWO)

Table 3. K-NN and Conventional Particle Swarm Optimization (PSO) Algorithm Performance

Metaheuristic Algorithms	Accuracy	Precision	Recall	F1-Score	Number of Selected Feature
Bare Bones Particle Swarm Optimization (BBPSO)	0.968	0.970	0.969	0.968	16
Particle Swarm Optimization (PSO)	0.925	0.925	0.925	0.925	28

Table 4. K-NN and Conventional Flower Pollination Algorithm (FPA) Algorithm Performance

Metaheuristic Algorithms	Accuracy	Precision	Recall	F1-Score	Number of Selected Feature
Modified Generalized Flower Pollination Algorithm (MGFPA)	0.981	0.982	0.981	0.981	17
Flower Pollination Algorithm (FPA)	0.937	0.938	0.937	0.937	45

Table 5. K-NN and Conventional Salp Swarm Algorithm (SSA) Algorithm Performance

Metaheuristic Algorithms	Accuracy	Precision	Recall	F1-Score	Number of Selected Feature
Enhanced Salp Swarm Algorithm (ESSA)	0.962	0.962	0.962	0.962	10
Improved Salp Swarm Algorithm (ISSA)	0.975	0.976	0.975	0.975	16
Salp Swarm Algorithm (SSA)	0.937	0.939	0.937	0.937	40

Table 6. K-NN and Conventional GWO Algorithm Performance

Metaheuristic Algorithms	Accuracy	Precision	Recall	F1-Score	Number of Selected Feature
Two Stage Modified Grey Wolf Optimizer (TMGWO)	0.981	0.982	0.981	0.981	3
Grey Wolf Optimizer (GWO)	0.956	0.956	0.956	0.956	5

This study also compares advanced metaheuristic algorithms with their conventional versions. Tables 3–6 generally show improvements in accuracy and efficiency in the number of selected features. Table 3 shows that BBPSO can increase accuracy by 4.44% compared to conventional PSO with a reduction in the number of selected features of 42.86%. Table 4 shows that MGFPA can increase accuracy by 4.49% compared to conventional FPA, while reducing the number of selected features by 62.22%. Table 5 shows that ESSA can increase accuracy by 2.59% from conventional SSA with a reduction in the number of selected features by 75%. Furthermore, Table 5 shows that ISSA can increase accuracy by 3.89% from conventional SSA with a reduction in the number of selected features of 60%. Table 6 shows that GWO can increase accuracy by 2.55% compared to conventional GWO, while reducing the number of selected features by 40%. The results of this study indicate that the development of advanced metaheuristic algorithms can improve the weaknesses of conventional metaheuristic algorithms. The results of study [21] show that GWO and Random Forest (RF) can achieve an accuracy of 0.981 with 5 selected colour and texture features. This study achieved better results than previous studies because the advanced metaheuristic algorithms successfully reduced the features to 3.

4. CONCLUSION

This study developed feature selection for classifying the purity of civet coffee using k-Nearest Neighbour (K-NN) optimized with advanced metaheuristic algorithms, namely Bare Bones Particle Swarm Optimization (BBPSO), Modified Generalized Flower Pollination Algorithm (MGFPA), Enhanced Salp Swarm Algorithm (ESSA), Improved Salp Swarm Algorithm (ISSA), and Two-Stage Modified Grey Wolf Optimizer (TMGWO). The results showed that the application of metaheuristic algorithms significantly improved the accuracy of the K-NN model while reducing the number of features used. Of all the models tested, K-NN and TMGWO showed the best performance with an accuracy of 0.981, precision of 0.982, recall of 0.981, F1-score of 0.981, AUC close to 1, and three selected features, i.e. blue correlation, s_hsl_correlation, and s_hsv_correlation. A comparison between conventional and advanced metaheuristic algorithms shows that modifications to the search mechanism and position updates in advanced metaheuristic algorithms consistently improve feature selection performance, as demonstrated by the accuracy of the classification model. This study still has several limitations, such as still using a K-NN machine learning model and limitations in the dataset. Therefore, further research could explore various types of civet coffee from several regions in Indonesia, using other machine learning models such as Random Forest (RF) and

Support Vector Machine (SVM), and comparing them with automatic feature extraction models such as deep learning.

REFERENCES

[1] M. T. Rahman, S. Ferdous, M. S. Jenin, T. R. Mim, M. Alam, and M. R. Al Mamun, “Characterization of tea (Camellia sinensis) granules for quality grading using computer vision system,” *J. Agric. Food Res.*, vol. 6, p. 100210, Dec. 2021, doi: 10.1016/j.jafr.2021.100210.

[2] G. Ren, N. Gan, Y. Song, J. Ning, and Z. Zhang, “Evaluating Congou black tea quality using a lab-made computer vision system coupled with morphological features and chemometrics,” *Microchem. J.*, vol. 160, p. 105600, 2021.

[3] R. Selvanarayanan, S. Rajendran, and Y. Alotaibi, “Early Detection of Colletotrichum Kahawae Disease in Coffee Cherry Based on Computer Vision Techniques,” *Comput. Model. Eng. Sci.*, vol. 139, no. 1, pp. 759–782, 2024, doi: 10.32604/cmes.2023.044084.

[4] D. S. Leme, S. A. da Silva, B. H. G. Barbosa, F. M. Borém, and R. G. F. A. Pereira, “Recognition of coffee roasting degree using a computer vision system,” *Comput. Electron. Agric.*, vol. 156, pp. 312–317, Jan. 2019, doi: 10.1016/j.compag.2018.11.029.

[5] H. Zhou, Y. Xin, and S. Li, “A diabetes prediction model based on Boruta feature selection and ensemble learning,” *BMC Bioinformatics*, vol. 24, no. 1, p. 224, 2023.

[6] D. Theng and K. K. Bhoyar, “Feature selection techniques for machine learning: a survey of more than two decades of research,” *Knowl. Inf. Syst.*, vol. 66, no. 3, pp. 1575–1637, 2024.

[7] M. A. Khan *et al.*, “Multimodal brain tumor classification using deep learning and robust feature selection: A machine learning application for radiologists,” *Diagnostics*, vol. 10, no. 8, p. 565, 2020.

[8] A. P. M. S. Hamdard and A. P. H. Lodin, “Effect of Feature Selection on the Accuracy of Machine Learning Model,” *Int. J. Multidiscip. Res. Anal.*, vol. 06, no. 09, Sep. 2023, doi: 10.47191/ijmra/v6-i9-66.

[9] A. Newaz and S. Muhtadi, “Performance Improvement of Heart Disease Prediction by Identifying Optimal Feature Sets Using Feature Selection Technique,” in *2021 International Conference on Information Technology (ICIT)*, Jul. 2021, pp. 446–450, doi: 10.1109/ICIT52682.2021.9491739.

[10] S. Widyaningtyas, M. Arwani, S. Sucipto, and Y. Hendrawan, “Filter Feature Selection for Detecting Mixture, Total Phenol, and pH of Civet Coffee,” *Int. J. Artif. Intell. Robot.*, vol. 6, no. 2, pp. 75–82, 2024.

[11] A. Darwish, “Bio-inspired computing: Algorithms review, deep analysis, and the scope of applications,” *Futur. Comput. Informatics J.*, vol. 3, no. 2, pp. 231–246, 2018.

[12] E. T. Yasin *et al.*, “Optimized feature selection using gray wolf and particle swarm algorithms for corn seed image classification,” *J. Food Compos. Anal.*, vol. 145, p. 107738, Sep. 2025, doi: 10.1016/j.jfca.2025.107738.

[13] X. Song, Y. Zhang, D. Gong, and X. Sun, “Feature selection using bare-bones particle swarm optimization with mutual information,” *Pattern Recognit.*, vol. 112, p. 107804, 2021.

[14] H. Mohammadzadeh and F. S. Gharehchopogh, “A novel hybrid whale optimization algorithm with flower pollination algorithm for feature selection: Case study Email spam detection,” *Comput. Intell.*, vol. 37, no. 1, pp. 176–209, 2021.

[15] M. Ansari Shiri and N. Mansouri, “Hybrid Filter-Wrapper Feature Selection using Modified Flower Pollination Algorithm,” *Comput. Knowl. Eng.*, vol. 8, no. 2, pp. 55–74, 2025.

[16] M. Barhoush, B. H. Abed-alguni, and N. E. A. Al-qudah, “Improved discrete salp swarm algorithm using exploration and exploitation techniques for feature selection in intrusion detection systems,” *J. Supercomput.*, vol. 79, no. 18, pp. 21265–21309, Dec. 2023, doi: 10.1007/s11227-023-05444-4.

[17] H. Zamani, “Evolutionary salp swarm algorithm with multi-search strategies and advanced memory mechanism for solving global optimization and complex engineering problems,” *Sci. Rep.*, vol. 15, no. 1, p. 33934, Sep. 2025, doi: 10.1038/s41598-025-09345-9.

[18] A. E. Hegazy, M. A. Makhlouf, and G. S. El-Tawel, “Improved salp swarm algorithm for feature selection,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 32, no. 3, pp. 335–344, Mar. 2020, doi: 10.1016/j.jksuci.2018.06.003.

[19] C. Shen and K. Zhang, “Two-stage improved Grey Wolf optimization algorithm for feature selection on high-dimensional classification,” *Complex Intell. Syst.*, vol. 8, no. 4, pp. 2769–2789, 2022.

[20] S. Widyaningtyas and M. Arwani, “Color Feature Selection Optimized with Bio inspired Algorithms in Classify Purity of Palm Civet Coffee,” *J. Teknol. Inf. dan Terap.*, vol. 12, no. 1, 2025.

[21] S. Widyaningtyas, S. Arwani, Muhammad, Sucipto, and Y. Hendrawan, “Improving the accuracy of green bean palm civet coffee purity classification using wrapper feature selection,” *Coffee Sci. 1984-3909*, vol. 20, pp. e202277–e202277, 2025.

[22] A. Solak, A. Onat, and O. Kilinç, “An Improved Bare Bones Particle Swarm Optimization Algorithm Based on Sequential Update Mechanism and a Modified Structure,” *IEEE Access*, vol. 13, pp. 4789–

4814, 2025, doi: 10.1109/ACCESS.2025.3525603.
[23] J. Guo and Y. Sato, “A Dynamic Reconstruction Bare Bones Particle Swarm Optimization Algorithm,” in *2018 IEEE Congress on Evolutionary Computation (CEC)*, Jul. 2018, pp. 1–6, doi: 10.1109/CEC.2018.8477883.
[24] J. Guo and Y. Sato, “A hierarchical bare bones particle swarm optimization algorithm,” in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Oct. 2017, pp. 1936–1941, doi: 10.1109/SMC.2017.8122901.
[25] M. K. Y. Shambour, A. A. Abusnaina, and A. I. Alsalibi, “Modified global flower pollination algorithm and its application for optimization problems,” *Interdiscip. Sci. Comput. Life Sci.*, vol. 11, no. 3, pp. 496–507, 2019.
[26] V. Kansal and J. S. Dhillon, “Emended salp swarm algorithm for multiobjective electric power dispatch problem,” *Appl. Soft Comput.*, vol. 90, p. 106172, May 2020, doi: 10.1016/j.asoc.2020.106172.
[27] H. Faris, S. Mirjalili, I. Aljarah, M. Mafarja, and A. A. Heidari, “Salp Swarm Algorithm: Theory, Literature Review, and Application in Extreme Learning Machines,” 2020, pp. 185–199.
[28] S. Mirjalili, A. H. Gandomi, S. Z. Mirjalili, S. Saremi, H. Faris, and S. M. Mirjalili, “Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems,” *Adv. Eng. Softw.*, vol. 114, pp. 163–191, Dec. 2017, doi: 10.1016/j.advengsoft.2017.07.002.
[29] M. Al-Shabi, C. Ghenai, M. Bettayeb, F. Faraz Ahmad, and M. El Haj Assad, “Estimating PV models using multi-group salp swarm algorithm,” *IAES Int. J. Artif. Intell.*, vol. 10, no. 2, p. 398, Jun. 2021, doi: 10.11591/ijai.v10.i2.pp398-406.
[30] S. Ben Chaabane, A. Belazi, S. Kharbech, A. Bouallegue, and L. Clavier, “Improved Salp Swarm Optimization Algorithm: Application in Feature Weighting for Blind Modulation Identification,” *Electronics*, vol. 10, no. 16, p. 2002, Aug. 2021, doi: 10.3390/electronics10162002.
[31] L. I. Wong, M. H. Sulaiman, M. R. Mohamed, and M. S. Hong, “Grey Wolf Optimizer for solving economic dispatch problems,” in *2014 IEEE International Conference on Power and Energy (PECon)*, Dec. 2014, pp. 150–154, doi: 10.1109/PECON.2014.7062431.
[32] P. Cunningham and S. J. Delany, “K-nearest neighbour classifiers-a tutorial,” *ACM Comput. Surv.*, vol. 54, no. 6, pp. 1–25, 2021.
[33] C. Bai, T. Xiao, Y. Chen, H. Wang, F. Zhang, and X. Gao, “Faster-LIO: Lightweight tightly coupled LiDAR-inertial odometry using parallel sparse incremental voxels,” *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 4861–4868, 2022.
[34] U. Pujianto, A. P. Wibawa, and M. I. Akbar, “K-Nearest Neighbor (K-NN) based missing data imputation,” in *2019 5th International Conference on Science in Information Technology (ICSITech)*, 2019, pp. 83–88.
[35] M. Afshar and H. Usefi, “Optimizing feature selection methods by removing irrelevant features using sparse least squares,” *Expert Syst. Appl.*, vol. 200, p. 116928, 2022.
[36] F. F. L. dos Santos, J. T. F. Rosas, R. N. Martins, G. de M. Araújo, L. de A. Viana, and J. de P. Gonçalves, “Quality assessment of coffee beans through computer vision and machine learning algorithms,” 2020.
[37] M. García, J. E. Candelo-Becerra, and F. E. Hoyos, “Quality and defect inspection of green coffee beans using a computer vision system,” *Appl. Sci.*, vol. 9, no. 19, p. 4195, 2019.
[38] N. O. Adiwijaya and R. Sarno, “Specialty Coffees Classification Utilizes Feature Selection and Machine Learning,” in *2nd International Conference on Neural Networks and Machine Learning 2023 (ICNNML 2023)*, 2024, pp. 94–101.
[39] S. WIDYANINGTYAS and M. ARWANI, “Color Feature Selection Optimized with Bio-Inspired Algorithms in Classify Purity of Luwak Coffee,” 2025.
[40] M. Fahmy Amin, “Confusion Matrix in Binary Classification Problems: A Step-by-Step Tutorial,” *J. Eng. Res.*, vol. 6, no. 5, pp. 0–0, Dec. 2022, doi: 10.21608/erjeng.2022.274526.