Improving Performance of Copra Type Classification using Feature Extraction with K-Nearest Neighbor Algorithm

**Puspitasari1, Husni Teja Sukmana2****, Aryajaya Alamsyah3, Saepul Aripiyanto4, Ammar Sufyan5**

1,2,3Department of Informatics Engineering, University of State Islamic Syarif Hidayatullah Jakarta

|  |  |  |
| --- | --- | --- |
| **Article Info** |  | ABSTRACT |
| ***Article history:***  Received Sep 3, 2019  Revised May 17, 2020  Accepted June 28, 2020 |  | The results of interviews with copra traders in Indragiri Hilir Regency, Riau Province, indicate that the classification of copra types is still being done manually, requiring significant labor, time, and costs. This research aims to classify copra types based on feature extraction using the K-nearest neighbor algorithm. The study utilizes a dataset from copra warehouses in Indragiri Hilir Regency, Riau Province, comprising 613 digital images categorized into three classes: edible, regular, and reject. Classifying copra types has considered feature extraction such as color, shape, and texture features. The findings of this research show that the model's accuracy, considering all features, resulted in an accuracy of 84%. |
| ***Keywords:***  Classification Algorithm  Copra Type  Feature Extraction  K-Nearest Neighbor |
| ***Corresponding Author:***  Husni Teja Sukmana,  Informatics Department, Faculty of Science & Technology, UIN Syarif Hidayatullah Jakarta  Jl. Ir. H. Juanda No 95, Ciputat Timur, Tangerang Selatan, Indonesia 15419  Email: husniteja@uinjkt.ac.id | | |

# INTRODUCTION

Kopra is derived from the flesh of the coconut (Cocos nucifera) processed conventionally through drying methods, either through sun-drying or smoking [1]. Copra is divided into three types based on its quality and use, namely edible, regular, and reject copra [2]–[4]. Edible copra meets hygiene and food safety standards and is used as a raw material in the food and coconut oil industries [2]. Regular copra is used in the cosmetics, soap, and other non-food products industries [2]. While reject copra does not meet quality standards for use in the food and beverage industry [2]. Based on the results of an interview with a copra merchant in Indragiri Hilir Regency, Riau Province, it has been revealed that the classification of copra types is generally still done manually by copra merchants based on intuition. Distinguishing between copra types is usually done by observing specific characteristics. For example, edible copra has a diameter of less than 9 cm, bowl-shaped, free from defects on the husk, not hollow, not brownish in color, free from fungal growth, and white in its flesh. Meanwhile, regular copra has a diameter of more than 9 cm and its shape does not have to be bowl-shaped, but other characteristics remain the same as edible copra. The reject type of copra appears after the sorting process of Edible and Regular, where copra with the reject type generally contains fungi, has a brownish color, and may have holes or cracks. This process certainly requires a significant amount of labor, time, and cost. Additionally, the differing perceptions of each merchant in identifying copra types make the classification inconsistent. This results in farmers performance becoming less efficient, especially when carried out on a large scale [3], [4]. Therefore, the development of a method that can automatically and accurately classify types of copra is needed to support this process. Accurate and efficient classification can help manage copra supplies with applicable quality standards and meet the diverse needs of copra users in various industrial sectors more precisely [5].

Some research on the classification of copra types has been carried out by Abdullah et al. [6], Lim et al. [7], Adang et al. [8], Marlis et al. [9], Lahay et al. [10]. The first study [6] created software for copra quality classification using the Nearest Mean Classifier (MNC) algorithm. The results of this study have an accuracy of 80.67%. The second study [7] made a conveyor belt machine to automatically detect coconut quality using the K-nearest neighbor (KNN) algorithm. The results of this study have an accuracy of 86.67%. The third research [8] made software for the classification of copra maturity using the Naïve Bayes (NB) algorithm. The results of this study have an accuracy of 91.12%. The fourth study [9] created a model for classifying white copra or edible copra quality using the KNN algorithm. The results of this study have an accuracy of 93.33%. Finally, the fifth study [10] created a model to determine the quality of copra using a fuzzy logic algorithm. The results of this study have an accuracy of 95%. Based on the literature study results, several algorithms have been used to classify copra types and coconut quality, such as MNC, KNN, NB, and fuzzy logic [6]-[10]. In this research, the copra type classification method uses KNN because this method tends to be easy to understand and implement, can be used for various types of data (numeric and categorical), is non-parametric so that it can be used for datasets that have multiple distributions, is resistant to noise and missing values [11]–[15]. However, this method also has several limitations, such as requiring optimal K parameter tuning, which is sensitive to feature selection [11]–[15]. Based on the results of literature studies, the KNN algorithm has proven effective and efficient for classifying digital images and tabular data in various case studies [16]–[22].

This study focuses on classifying copra types using the KNN method based on feature extraction. It is grounded in the utilization of a dataset comprising digital images. Feature extraction aims to obtain unique values from each image, allowing the objects within them to be recognized by the computer [23]. This feature extraction process adopts several fundamental techniques, including color, shape, and texture features [24]–[27]. In this research, color features utilize the color moment method with parameters such as color Red-Green-Blue (RGB), color Hue-Value-Saturation (HSV), and grayscale [28]–[31]. Furthermore, shape features employ region-based methods with area parameters and contour-based techniques with perimeter parameters [32]–[34]. Lastly, texture features use the Gray Level Co-occurrence Matrix (GLCM) with contrast, dissimilarity, homogeneity, energy, and correlation parameters [35]–[38]. Based on the study literature, selecting these features has proven effective in enhancing the classification model's performance with digital image datasets [39]–[45].

Therefore, this study aims to create a classification model for types of copra by considering feature extraction using the KNN algorithm. The contribution of this research lies in the use of feature extraction to improve accuracy. Subsequently, the research is expected to enhance the efficiency of copra selection for copra farmers in Indragiri District, Riau Province. The contributions of this study are presented in Table 1.

Table 1. Research contribution of copra type classification

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| References | Feature extractions | | | Algorithms | Results |
| Color features | Shape features | Texture features |
| Abdullah et al. [5] | - Color RGB | x | - Contras  - Homogeneity  - Energy  - Correlation | Nearest Mean Classifier | Accuracy = 74% |
| Lim et al. [6] | - Color RGB | x | x | K-Nearest Neighbor | Accuracy = 86% |
| Adang et al. [7] | - Color RGB  - Color HIS | x | x | Naïve Bayes | Accuracy = 91% |
| Marlis et al. [8] | - Color RGB | - Area  - Perimeter | x | K-Nearest Neighbor | Accuracy = 93% |
| Lahay et al. [9] | x | x | x | Fuzzy Logic | Accuracy = 95% |
| This research | - Color RGB  - Color HSV  - Greyscale | - Area  - Perimeter | - Contrast  - Dissimilarity  - Homogeneity  - Energy  - Correlation | K-Nearest Neighbor | Accuracy = 84% |

# METHOD

The research was conducted with several stages that followed systematic rules. Some steps of research such as data acquisition, data preprocessing, feature extraction, feature scaling, distribution of training data and test data, implementation of the knn algorithm, and evaluation of the classification model. The research stages are conducted as shown in Figure 1.

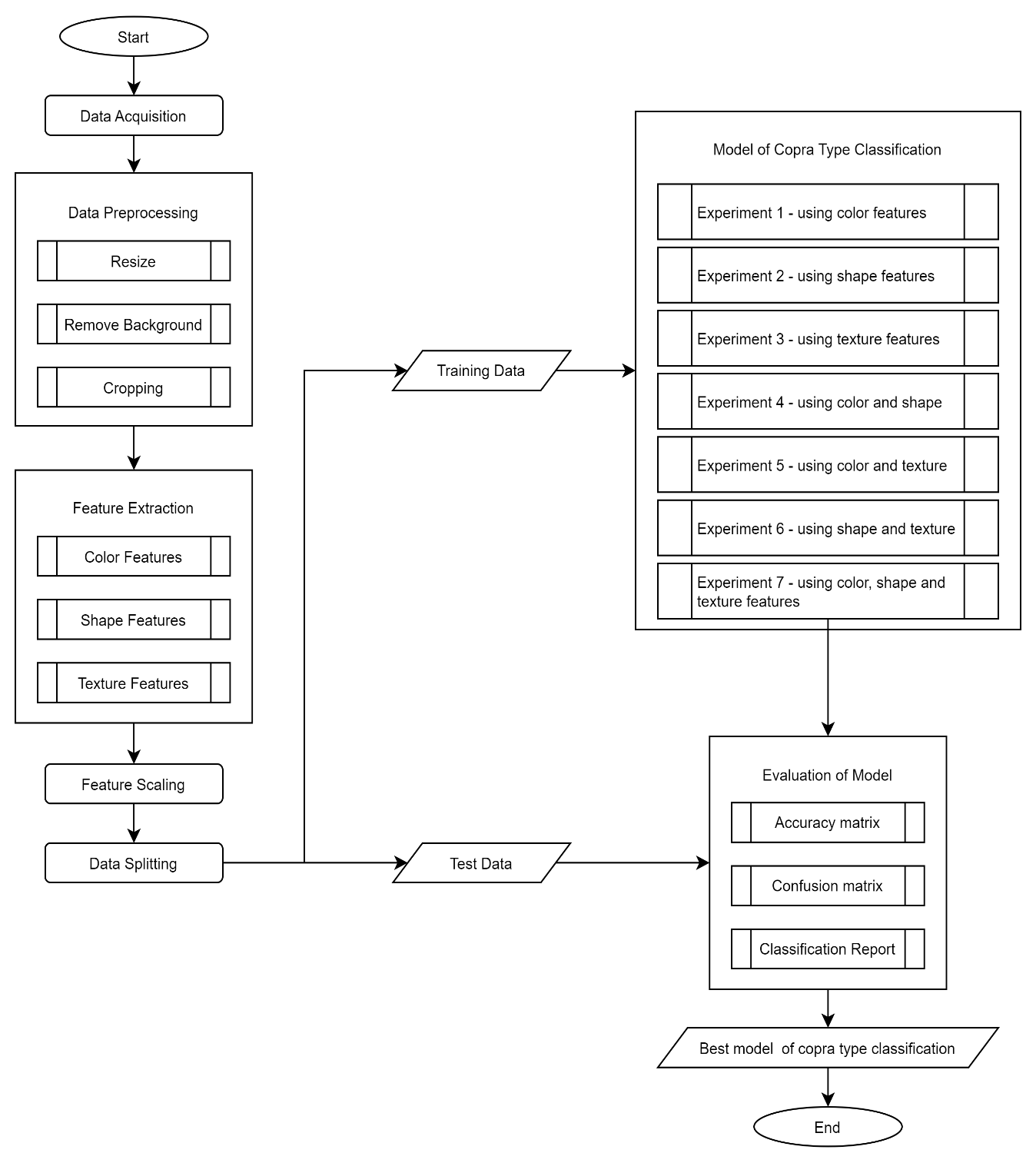


Figure 1. The research stages of copra type classification model

## 2.1. Data Acquisition

The research data was obtained from the coconut copra warehouse in Indragiri Hilir Regency, Riau Province. This research data, which can be referred to as a dataset, consists of three types of non-shriveled copra, namely edible, regular, and reject. Each type of copra has 613 image data, making the total dataset reach 1,839 image data. The image acquisition process was carried out using a smartphone camera to obtain digital images. The results of data acquisition are as shown in Figure 2.

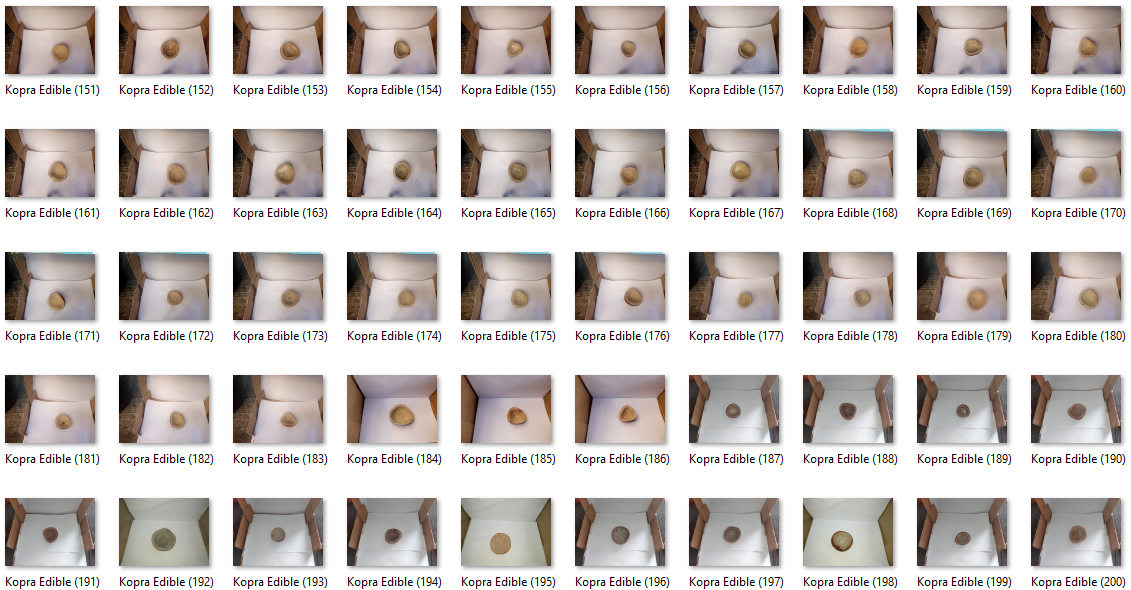


Figure 2. Dataset of copra type from Indragiri Hilir Regency, Riau Province

## 2.2. Data Preprocessing

Data preprocessing is conducted to transform unstructured data into a more structured format according to the research needs [12], [46]. In this study, image processing is conducted with resize image, remove background and cropping image [47]–[49]. The resizing process is carried out to transform the dimensions of the images, which were previously varied and large, into uniform and smaller sizes. Remove background is performed to eliminate irrelevant or distracting backgrounds in an image. Cropping is done to ensure that only relevant parts of the image are displayed. The results of data preprocessing are shown in Figure 3.

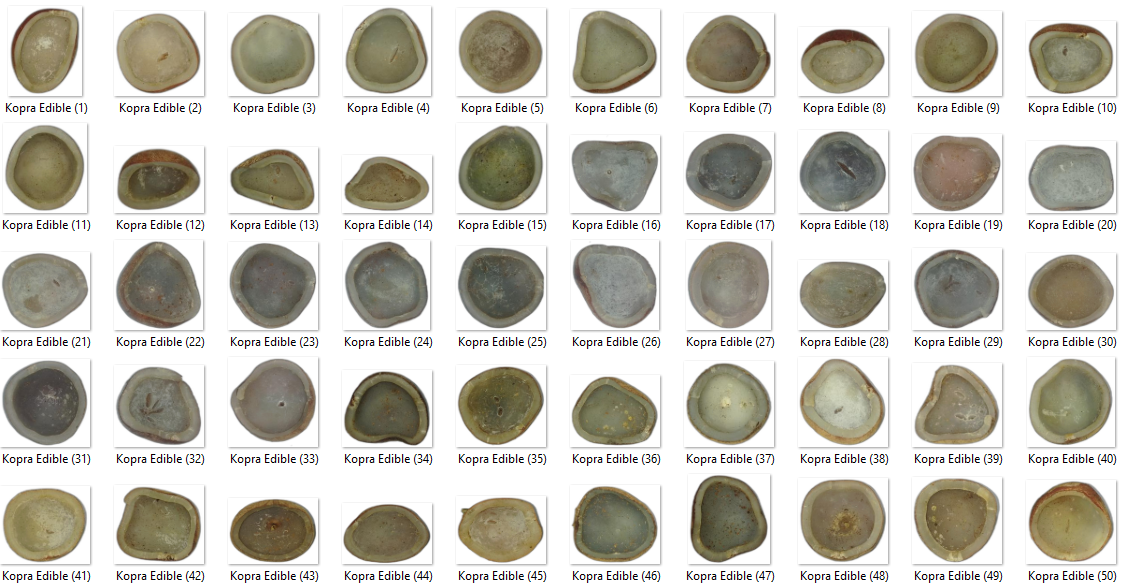


Figure 3. Result of preprocessing data with resize image, remove background and cropping image

## 2.3. Feature Extraction

Feature extraction is a process of identifying and extracting crucial information from a digital image to form simpler features while considering the characteristics of the digital image [23]. In this study, the benefit of feature extraction is simplifying the classification process of copra types. Feature extraction techniques include color, shape, and texture features. Color features help identify digital images based on color distribution, color scheme, and specific color characteristics in each pixel of the digital image [24]–[27]. Shape features aid in identifying digital images based on geometric information such as size, object area, the length of lines surrounding the object, and object aspect ratio [28]–[31]. Texture features involve extracting features from a digital image based on specific patterns formed from pixel arrangements to differentiate one object from another [32]–[34]. In this research, color, shape, and texture features utilize various parameters, as depicted in Figure 4 [23]-[34].

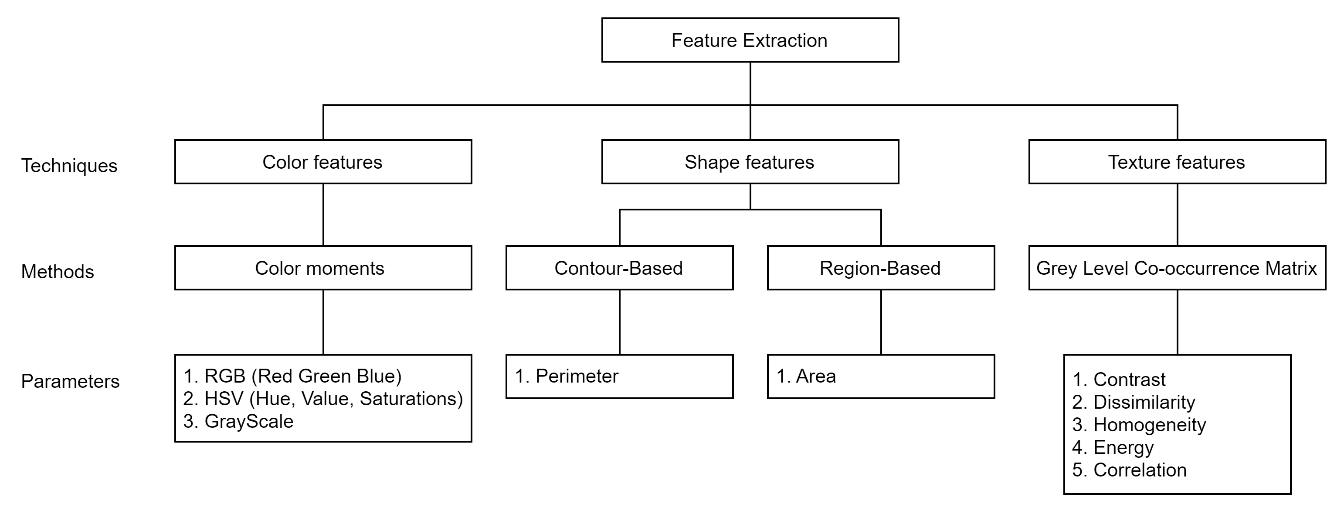


Figure 4. Result of feature extraction on copra type classification

## 2.4. Feature Scaling

Feature scaling is a process of normalizing features so that all features have similar values. The purpose of feature scaling is to standardize all features to the same scale. This is crucial for several machine learning algorithms because they can perform more accurately when features are on the same scale [50]. In this research, feature scaling employs the min-max scaler method with a value range from 0 to 1. The formula for the min-max scaler is as shown in Equation 1 [50].

|  |  |  |  |
| --- | --- | --- | --- |
|  | | | (1) |
| Description: | | |  |
| *x* | = | value of feature |  |
| *a* | = | value range 0 |  |
| *b* | = | value range 1 |  |
| *min(x)* | = | minimum value in the dataset |  |
| *max(x)* | = | maximum value in the dataset |  |

## 2.5. Data Splitting

Data splitting is dividing a dataset into two or more subsets of data. In this study, data division is categorized into training data and testing data. The training data is used to train the model for classifying types of copra, distinguishing between edible, regular, and rejected copra. Meanwhile, the testing data is employed to evaluate the performance of the copra classification model. In this research, data division was carried out using split validation method, with a percentage of 90% for training data and 10% for testing data.

## 2.6. Implementation of KNN algorithm

K-nearest neighbors (KNN) is one of the non-parametric classification algorithms that work by classifying an object in the test data based on the majority class of several k nearest neighbors in the training data [51]. In the KNN algorithm, it is essential to consider the value of K because a smaller K value makes the classification model vulnerable to noisy data. In contrast, a more considerable K value makes the classification model more resistant to noise but may introduce bias in the boundaries between classes [51]–[53]. In this research, determining the value of K does not utilize specific methods like the elbow method, silhouette coefficient, Calinski-Harabasz index, and Davies-Bouldin index because the determination of K values is based on the types of copra, namely edible, regular, and rejected, resulting in a value of K being 3. The KNN algorithm utilizes the concept of proximity between data points to perform data classification. In this study, calculating the distance between data points uses the Euclidean distance formula, as shown in equation 2 [51]–[53]. The KNN algorithm operates as depicted in pseudocode 1 [51].

|  |  |  |  |
| --- | --- | --- | --- |
|  | | | (2) |
| Description: | | |  |
|  | = | Distance between object x and object y. | |
|  | = | The value of training data object a in the i-th variable. | |
|  | = | The value of the testing data object b in the i-th variable. | |
| *n* | = | The number of independent variables. | |

|  |  |  |
| --- | --- | --- |
| **Algorithm 1**: K-nearest neighbors | | |
| **Input :** k, DataSet, TestData | | |
| **Output :** Class | | |
| 1 |  | |
| 2 |  | |
| 3 |  | |
| 4 |  | |
| 5 |  |  |
| 6 |  |  |
| 7 |  | |
| 8 |  | |
| 9 |  | |
| 10 |  |  |
| 11 |  | |
| 12 |  | |
| 13 |  | |

Determination of copra, including edible or regular copra or reject, based on the extraction of color, shape, and texture features. Then, the copra-type classification process uses the KNN algorithm. Making a copra-type classification model using several experiments aims to determine whether using color, shape, and texture features can increase the accuracy of the classification model when using the KNN algorithm. Some of these experiments are shown in Table 2.

Table 2. Some experiments on copra type classification

|  |  |  |
| --- | --- | --- |
| Experiments | Parameters | Objective |
| 1 | Mean-R, mean-G, mean-B, mean-H, mean-S, mean-V, greyscale. | It is knowing the accuracy of the classification model when using color feature. |
| 2 | Area, perimeter. | It is knowing the accuracy of the classification model when using shape feature. |

Table 2. Some experiments on copra type classification (*continued*)

|  |  |  |
| --- | --- | --- |
| Experiments | Parameters | Objective |
| 3 | Contrast, dissimilarity, homogeneity, energy, correlation. | It is knowing the accuracy of the classification model when using texture feature. |
| 4 | Mean-R, mean-G, mean-B, mean-H, mean-S, mean-V, greyscale, area, perimeter | It is knowing the accuracy of the classification model when using color and shape features. |
| 5 | Mean-R, mean-G, mean-B, mean-H, mean-S, mean-V, greyscale, contrast, dissimilarity, homogeneity, energy, correlation. | It is knowing the accuracy of the classification model when using color and texture features. |
| 6 | Area, perimeter, contrast, dissimilarity, homogeneity, energy, correlation. | It is knowing the accuracy of the classification model when using shape and texture features. |
| 7 | Mean-R, mean-G, mean-B, mean-H, mean-S, mean-V, greyscale, area, perimeter, contrast, dissimilarity, homogeneity, energy, correlation. | It is knowing the accuracy of the classification model when using color, shape, and texture features. |

## 2.7. Model Evaluation

Model evaluation is the process of measuring the performance of the copra-type classification model. The purpose of model evaluation is to assess how well the classification model performs and to measure the extent to which the classification model can generalize the data [11]–[14]. In this research, model evaluation employs the confusion matrix method to determine accuracy, precision, recall, and f1-score. The formulas for calculating accuracy, precision, recall, and f1-score are illustrated in equations 3 to 6 [11]–[14].

|  |  |  |  |
| --- | --- | --- | --- |
| *Accuracy* | |  | (3) |
| *Recall* | |  | (4) |
| *Precision* | |  | (5) |
| *F1-Score* | |  | (6) |
| Description: | | | | |
| TP | = | True positive | | |
| TN | = | True negative | | |
| FP | = | False positive | | |
| FN | = | False negative | | |

# RESULTS AND DISCUSSION

## 3.1. Implementation of Feature Extraction

In this research, feature extraction involves color, shape, and texture features. Each feature has different parameters, as seen in Figure 4. The extracted color, shape, and texture features have different value ranges, thus employing feature scaling to adjust the value range to 0 to 1 using Equation 1. The purpose of feature extraction and feature scaling is to enable the classification model to identify coconut types with reasonably good accuracy. The results of the feature extraction implementation are presented in Table 3.

Table 3. Results of feature extractions and feature selections of copra type classification

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 | X11 | X12 | X13 | X14 | X15 | Y |
| 0,50 | 0,56 | 0,59 | 0,12 | 0,50 | 0,16 | 0,54 | 0,47 | 0,04 | 0,16 | 0,06 | 0,08 | 0,76 | 0,18 | 0,95 | 1 |
| 0,82 | 0,85 | 0,83 | 0,12 | 0,82 | 0,26 | 0,84 | 0,84 | 0,10 | 0,09 | 0,16 | 0,24 | 0,43 | 0,24 | 0,97 | 1 |
| 0,64 | 0,75 | 0,72 | 0,18 | 0,65 | 0,21 | 0,71 | 0,67 | 0,10 | 0,19 | 0,11 | 0,19 | 0,48 | 0,20 | 0,97 | 1 |

Note: X1=Mean-R, X2=Mean=G, X3=Mean-G, X4=Mean-H, X5=Mean-S, X6=Mean-V, X7=Mean-Grey, X8=Standard-deviation, X9=Area, X10=Perimeter, X11=Contrast, X12=Dissimilarity, X13=Homogeneity, X14=Energym X15=Correlation, Y=Copra-type.

Table 3. Results of feature extractions and feature selections of copra type classification (*continued*)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 | X11 | X12 | X13 | X14 | X15 | Y |
| 0,73 | 0,77 | 0,71 | 0,14 | 0,73 | 0,33 | 0,75 | 0,65 | 0,13 | 0,11 | 0,19 | 0,32 | 0,25 | 0,12 | 0,93 | 2 |
| 0,78 | 0,86 | 0,87 | 0,16 | 0,78 | 0,17 | 0,84 | 0,75 | 0,19 | 0,12 | 0,27 | 0,41 | 0,28 | 0,19 | 0,92 | 2 |
| … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … |
| 0,27 | 0,17 | 0,16 | 0,03 | 0,27 | 0,65 | 0,20 | 0,29 | 0,01 | 0,07 | 0,23 | 0,29 | 0,59 | 0,35 | 0,68 | 3 |
| 0,58 | 0,57 | 0,44 | 0,13 | 0,58 | 0,56 | 0,56 | 0,48 | 0,02 | 0,05 | 0,22 | 0,36 | 0,24 | 0,07 | 0,85 | 3 |

Note: X1=Mean-R, X2=Mean=G, X3=Mean-G, X4=Mean-H, X5=Mean-S, X6=Mean-V, X7=Mean-Grey, X8=Standard-deviation, X9=Area, X10=Perimeter, X11=Contrast, X12=Dissimilarity, X13=Homogeneity, X14=Energym X15=Correlation, Y=Copra-type.

## 3.2. Implementation of copra type classification on KNN algorithm

The classification model of coconut types employs the KNN algorithm. The development of this classification model involves seven experiments, as shown in Table 2. These seven experiments aim to determine whether performance improves when the classification model uses feature extraction and the KNN algorithm. The performance measurement of the classification model is conducted using a confusion matrix to obtain accuracy, precision, recall, and f1-score values. The results of the confusion matrix for the coconut-type classification model are displayed in Figure 5.

|  |  |
| --- | --- |
|  |  |
| Experiment 1 | Experiment 2 |
|  |  |
| Experiment 3 | Experiment 4 |
|  |  |
| Experiment 5 | Experiment 6 |
|  | |
| Experiment 7 | |

Figure 5. Results of confusion matrix for copra type classification

Based on the confusion matrix, all true positive, true negative, false positive, and false negative values have been identified, enabling the calculation of accuracy, precision, recall, and f1-score using equations 3 and 4. Therefore, the accuracy, precision, recall, and f1-score results for the coconut-type classification model are shown in Table 4 and Figure 6.

Table 4 Result of model evaluation of copra type classification

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Confusion matrix | Experiments | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Accuracy | 71,7 | 54,4 | 79,9 | 72,8 | 83,7 | 81,0 | 84,2 |
| Precision | 71,0 | 54,8 | 79,3 | 72,9 | 83,2 | 80,3 | 84,1 |
| Recall | 70,7 | 53,1 | 79,0 | 71,9 | 83,0 | 80,1 | 83,7 |
| F1-Score | 70,6 | 52,9 | 79,1 | 72,1 | 83,0 | 80,1 | 83,7 |

|  |  |
| --- | --- |
|  |  |
|  |  |
| Figure 6. Result of model evaluation of copra type classification | |

Based on Table 4 and Figure 6, it is evident that the study on classifying types of copra using feature extraction with the KNN algorithm has proven successful. This research has yielded several interesting facts and findings. The result of this study indicates that the copra-type classification model achieves an average accuracy of 75%. When the classification model uses color features (experiment 1), it attains an accuracy of 72%. Shape features (experiment 2) result in an accuracy of 54%, while texture features (experiment 3) yield an accuracy of 79%. These experiments show that using color features decreases accuracy by 3%, shape features decrease accuracy by 21%, whereas texture features can increase accuracy by 4%. This experimentation confirms that texture features are the most effective for the copra-type classification.

Next, when the copra-type classification model combines two features, such as color with shape (experiment 4), color with texture (experiment 5), and shape with texture (experiment 6), there tends to be an improvement in the model's performance. These experiments show that Experiment 4 produces an accuracy of 72%, decreasing accuracy by 2%. Experiment 5 yields an accuracy of 83%, increasing accuracy by 8%, and Experiment 6 results in an accuracy of 80%, increasing accuracy by 5%. These experiments revealed that combining color and shape features can reduce the classification model error from 3% to 2%. While color or shape features tend to influence the model's performance, incorporating texture features alongside color or shape can increase accuracy by 10% and 7%. Finally, when the copra-type classification model integrates all features (experiment 7), it achieves an accuracy of 84%, thereby increasing accuracy by 9%.

## 3.3. Analysis of correlation between features and labels

Based on the classification results of coconut types using feature extraction and KNN algorithm, the shape feature significantly influences the classification model's performance. This might occur because when the classification model uses texture features, it achieves an accuracy of 79%, color features result in an accuracy of 71%, and shape features yield an accuracy of 54%. These accuracy results were further tested statistically using a correlation test between the features utilized and the type of copra. The initial testing employed a bivariate correlation test to observe feature extraction parameters regarding copra types, as shown in Table 5. The bivariate correlation test utilized the Pearson correlation method. The test results reveal that texture features exhibit a reasonably good correlation with copra types, with the highest correlation value being 0,6 and the lowest being 0,4. On the other hand, color features show a less robust correlation with copra types, as the highest correlation value is 0,6 and the lowest is 0,1. Lastly, shape features exhibit no correlation with copra types, with the highest correlation value at 0,1 and the lowest at 0. Therefore, it is evident that texture features prove to be the most effective features for classifying copra types.

Table 5. Result of bivariate correlation of feature extractions with copra types

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Color Features | | | | | | | | Shape | | Texture | | | | |  |
| X1 | 1,0 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| X2 | 0,9 | 1,0 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| X3 | 0,8 | 0,9 | 1,0 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| X4 | 0,1 | 0,2 | 0,2 | 1,0 |  |  |  |  |  |  |  |  |  |  |  |  |
| X5 | 1,0 | 0,9 | 0,8 | 0,1 | 1,0 |  |  |  |  |  |  |  |  |  |  |  |
| X6 | -0,1 | -0,3 | -0,6 | -0,3 | -0,1 | 1,0 |  |  |  |  |  |  |  |  |  |  |
| X7 | 1,0 | 1,0 | 0,9 | 0,2 | 1,0 | -0,3 | 1,0 |  |  |  |  |  |  |  |  |  |
| X8 | 0,9 | 0,9 | 0,8 | 0,1 | 0,9 | -0,2 | 0,9 | 1,0 |  |  |  |  |  |  |  |  |
| X9 | 0,7 | 0,7 | 0,6 | 0,1 | 0,7 | -0,1 | 0,7 | 0,7 | 1,0 |  |  |  |  |  |  |  |
| X10 | 0,5 | 0,5 | 0,4 | 0,1 | 0,5 | 0,0 | 0,5 | 0,6 | 0,6 | 1,0 |  |  |  |  |  |  |
| X11 | 0,1 | 0,0 | -0,1 | 0,0 | 0,1 | 0,4 | 0,0 | 0,2 | 0,2 | 0,4 | 1,0 |  |  |  |  |  |
| X12 | 0,0 | -0,1 | -0,2 | 0,0 | 0,0 | 0,4 | -0,1 | 0,1 | 0,1 | 0,3 | 1,0 | 1,0 |  |  |  |  |
| X13 | 0,0 | 0,0 | 0,1 | -0,1 | 0,0 | -0,3 | 0,0 | 0,0 | -0,1 | -0,2 | -0,7 | -0,8 | 1,0 |  |  |  |
| X14 | -0,4 | -0,4 | -0,4 | -0,2 | -0,4 | 0,0 | -0,5 | -0,1 | -0,1 | -0,1 | 0,1 | 0,1 | 0,3 | 1,0 |  |  |
| X15 | 0,5 | 0,5 | 0,6 | 0,0 | 0,5 | -0,4 | 0,6 | 0,5 | 0,2 | 0,1 | -0,7 | -0,7 | 0,5 | -0,2 | 1,0 |  |
| Y | -0,5 | -0,5 | **-0,6** | -0,1 | -0,5 | 0,4 | -0,5 | -0,3 | **-0,1** | 0,0 | 0,5 | **0,6** | -0,5 | 0,4 | **-0,6** | 1,0 |
|  | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 | X11 | X12 | X13 | X14 | X15 | Y |

Note: X1=Mean-R, X2=Mean=G, X3=Mean-G, X4=Mean-H, X5=Mean-S, X6=Mean-V, X7=Mean-Grey, X8=Standard-deviation, X9=Area, X10=Perimeter, X11=Contrast, X12=Dissimilarity, X13=Homogeneity, X14=Energym X15=Correlation, Y=Copra-type.

The evidence from the experiments in this research was also validated using multivariate correlation testing. Multivariate correlation testing is beneficial in substantiating the significance of this research's contributions. The multivariate correlation testing employed the ordinary least squares regression method, as shown in Table 6. These test results show that feature extraction in each experiment significantly affects the accuracy of the copra-type classification model. These test outcomes affirmed the effectiveness of each experiment, as they exhibited strong correlations, ranging from 0,86 to a minimum of 0,15. Furthermore, each experiment proved suitable for distinguishing between types of copra due to the moderately strong relationship between the variation in data and the scattered determination values, ranging from 0,74 to a minimum of 0,02.

At first glance, the second experiment displayed no correlation and determination concerning the classification of copra types. However, when combining the second experiment with the third in the sixth trial, the model's accuracy still demonstrated good performance, with a correlation value of 0,81 and a determination of 0,65. Subsequently, combining all features or conducting the seventh experiment proved most effective, with a correlation value of 0,86 and a determination of 0,74.

Table 6. Result of multivariate correlations and determinations of feature extractions with copra types

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Coefficient | Experiment | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Correlation | 0,73 | 0,15 | 0,81 | 0,75 | 0,85 | 0,81 | 0,86 |
| Determination | 0,52 | 0,02 | 0,65 | 0,56 | 0,73 | 0,65 | 0,74 |

1. **CONCLUSION**

In this study, classifying copra types while considering the extraction of color, shape, and texture features using the k-nearest neighbor method has enhanced accuracy. Extracted color features encompass RGB, HSV, and grayscale; shape features include area and perimeter; texture features consist of contrast, dissimilarity, homogeneity, energy, and correlation. Research findings indicate that copra-type classification, considering feature extraction, yields good accuracy with an average accuracy rate of 75%. Furthermore, the study results show that texture features significantly influence accuracy, reaching 79% in the testing phase. The research also demonstrates that the amalgamation of all elements (color, shape, and texture) yields the highest accuracy at 84%. Thus, the conclusion drawn from this research is that considering feature extraction in copra-type classification can increase accuracy by up to 9%.

# REFERENCES

[1] G. Mardiatmoko and M. Ariyanti, *Coconut plants (Cocos nucifera L)*. Ambon: Publishing agency of the faculty of agriculture, Patimura University., 2011.

[2] T. H. Yuni, *Feasibility of Integrated Coconut Industry*. Yogyakarta: Publisher Sambudra Biru, 2017.

[3] K. R. A. Zwingly, T. F. Lolowang, and L. R. J. Pangemanan, “Analysis of Production Factors Affecting Copra Production in West Tomohon District,” *Agri-Sosioekonomi*, vol. 14, no. 3, p. 17, 2018, doi: 10.35791/agrsosek.14.3.2018.21531.

[4] A. Zubair, N. Hamzah, and M. Rusdi, “Quality Improvement of Copra Through the Implementation of White Copra Drying Oven,” *Second Int. Conf. Food Agric.*, pp. 104–112, 2020.

[5] R. R. Rachmawati and E. Gunawan, “The Role of Millennial Farmers in Supporting Agricultural Exports in Indonesia,” *Agroecon. Res. forum*, vol. 38, no. 1, p. 67, 2020, doi: 10.21082/fae.v38n1.2020.67-87.

[6] A. Abdullah, U. Usman, and M. Efendi, “Copra Quality Classification System Based on Color and Texture Using Nearest Mean Classifier (NMC) Method,” *J. Inf. Technol. Comput. Sci.*, vol. 4, no. 4, pp. 297–303, 2017, doi: 10.25126/jtiik.201744479.

[7] T. C. Lim, J. O. Torregosa, A. R. A. Pescadero, and R. S. Pangantihon, “De-husked Coconut Quality Evaluation using Image Processing and Machine Learning Techniques,” *ACM Int. Conf. Proceeding Ser.*, pp. 28–33, 2019, doi: 10.1145/3383783.3383789.

[8] Y. Adang, A. Rabi, and R. Arifuddin, “Classification of Copra Maturity Level Using the naïve bayes Method,” *Cyclotron*, vol. 3, no. 1, 2020, doi: 10.30651/cl.v3i1.4307.

[9] R. Rahayu Marlis, Abdullah, and F. Yunita, “White Copra Quality Prediction System Using k-Nearest Neighbor (k-NN),” *Sist. J. Sist. Inf.*, vol. 10, no. 2, pp. 290–299, 2021, [Online]. Available: http://sistemasi.ftik.unisi.ac.id

[10] I. H. Lahay, H. Hasanuddin, J. D. Giu, and M. G. Bawole, “Determination of Copra Grade by Application of Fuzzy Logic Method,” *Jambura J. Electr. Electron. Eng.*, vol. 5, no. 1, pp. 122–129, 2023, doi: 10.37905/jjeee.v5i1.17073.

[11] Andreas, *Introduction to Machine Learning with Python*. 2016. doi: 10.2174/97898151244221230101.

[12] Suyanto, *Data Mining of Classification and Clustering Revised Edition*. Bandung: Informatika, 2019.

[13] R. Karthik and S. Abhishek, *Machine Learning Using TIL*, vol. 321. 2019. doi: 10.3233/FAIA200024.

[14] Tanay Agrawal, *Hyperparameter Optimization in Machine Learning: Make Your Machine Learning and Deep Learning Models More Efficient*. 2021. doi: 10.1007/978-1-4842-6579-6.

[15] Suyanto, *Basic and Advanced Machine Learning*. Bandung: Informatika, 2023.

[16] S. R. Raysyah, A. Veri, and I. M. Dadang, “Classification of Coffee Fruit Maturity Level Based on Color Detection Using Knn and Pca Method,” *JSiI (Journal Inf. Syst.*, vol. 8, no. 2, pp. 88–95, 2021, doi: 10.30656/jsii.v8i2.3638.

[17] N. Khairina, T. T. S. Sibarani, R. Muliono, Z. Sembiring, and M. Muhathir, “Identification of Pneumonia using The K-Nearest Neighbors Method using HOG Fitur Feature Extraction,” *J. Informatics Telecommun. Eng.*, vol. 5, no. 2, pp. 562–568, 2022, doi: 10.31289/jite.v5i2.6216.

[18] S. Suharyana, F. Anwar, A. C. Dewi, M. Yunianto, U. Salamah, and R. Chai, “Pneumonia Classification Based on GLCM Features Extraction using K-Nearest Neighbor,” *Indones. J. Appl. Phys.*, vol. 13, no. 2, p. 325, 2023, doi: 10.13057/ijap.v13i2.77120.

[19] R. Sari, “Sentiment Analysis on World Fantasy Attraction Reviews Using K-Nearest Neighbor (K-NN) Algorithm,” *Evol. J. Sci. Manag.*, vol. 8, no. 1, pp. 10–17, 2020, doi: 10.31294/evolusi.v8i1.7371.

[20] A. C. Khotimah and E. Utami, “Comparison of Naïve Bayes Classifier, K-Nearest Neighbor and Support Vector Machine Algorithms in Classification of Individual Characters in Twitter Accounts,” *J. Tek. Inform.*, vol. 3, no. 3, pp. 673–680, 2022, [Online]. Available: http://jutif.if.unsoed.ac.id/index.php/jurnal/article/view/254

[21] R. Setiawan and A. Triayudi, “Web-based Classification of Toddler Nutritional Status Using Naïve Bayes and K-Nearest Neighbor,” *J. Media Inform. Budidarma*, vol. 6, no. 2, p. 777, 2022, doi: 10.30865/mib.v6i2.3566.

[22] A. S. Miha Djami, N. W. Utami, and A. A. I. I. Paramitha, “The Prediction Of Product Sales Level Using K-Nearest Neighbor and Naive Bayes Algorithms (Case Study : PT Kotamas Bali),” *J. Pilar Nusa Mandiri*, vol. 19, no. 2, pp. 77–84, 2023, doi: 10.33480/pilar.v19i2.4420.

[23] R. C. Gonzalez and R. E. Woods, *4TH EDITION Digital image processing*. 2018.

[24] N. Kandalkar, A. Mani, G. Pandey, and J. Soni, “Content Based Image Retrieval using Color , Shape and Texture Extraction Techniques,” *Int. J. Eng. Tech. Res. IJETR*, vol. 3, no. 5, pp. 74–79, 2015.

[25] M. Dhanashree *et al.*, “Color, Shape and Texture feature extraction for Content Based Image Retrieval System: A Study,” *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 5, no. 4, pp. 303–306, 2016, doi: 10.17148/IJARCCE.2016.5477.

[26] M. R. Satpute and S. MJagdale, “Color, Size, Volume, Shape and Texture Feature Extraction Techniques for Fruits: A Review,” *Int. Res. J. Eng. Technol.*, no. 2010, pp. 2395–56, 2016.

[27] Akmal, R. Munir, and J. Santoso, “Automatic Weight of Color, Texture, and Shape Features in Content-Based Image Retrieval Using Artificial Neural Network,” *Int. J. Informatics Vis.*, vol. 7, no. 3, pp. 665–672, 2023, doi: 10.30630/joiv.7.3.1184.

[28] A. Malakar and J. Mukherjee, “Image Clustering using Color Moments, Histogram, Edge and K-means Clustering,” *Int. J. Sci. Res.*, vol. 2, no. 1, pp. 2319–7064, 2013, [Online]. Available: www.ijsr.net

[29] S. R. Shinde, S. Sabale, S. Kulkarni, and D. Bhatia, “Experiments on content based image classification using Color feature extraction,” *Proc. - 2015 Int. Conf. Commun. Inf. Comput. Technol. ICCICT 2015*, 2015, doi: 10.1109/ICCICT.2015.7045737.

[30] G. F. Laxmi and F. S. F. Kusumah, “Region of interest and color moment method for freshwater fish identification,” *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 17, no. 3, pp. 1432–1438, 2019, doi: 10.12928/TELKOMNIKA.V17I3.11749.

[31] M. N. Abdullah *et al.*, “Colour Features Extraction Techniques and Approaches for Content-Based Image Retrieval (CBIR) System,” *J. Mater. Sci. Chem. Eng.*, vol. 09, no. 07, pp. 29–34, 2021, doi: 10.4236/msce.2021.97003.

[32] M. Narottambhai and P. Tandel, “A Survey on Feature Extraction Techniques for Shape based Object Recognition,” *Int. J. Comput. Appl.*, vol. 137, no. 6, pp. 16–20, 2016, doi: 10.5120/ijca2016908782.

[33] J. S. Dhanoa and A. Garg, “A Novel Technique for Shape Feature Extraction Using Content Based Image Retrieval,” *MATEC Web Conf.*, vol. 57, 2016, doi: 10.1051/matecconf/20165702002.

[34] T. Nagarjun, K. Rathod, V. Chandana, and D. Kashyap, “IJERT-Object Recognition using Shape Features,” *Int. J. Eng. Res. Technol.*, vol. 8, no. 3, pp. 139–142, 2019.

[35] L. Veronica, I. H. Al Amin, B. Hartono, and T. Kristianto, “Texture Feature Extraction Using GLCM Matrix on Images with Variation of Object Direction,” *Pros. SENDI\_U*, pp. 978–979, 2019.

[36] A. Humeau-Heurtier, “Texture feature extraction methods: A survey,” *IEEE Access*, vol. 7, pp. 8975–9000, 2019, doi: 10.1109/ACCESS.2018.2890743.

[37] X. Ding, “Texture Feature Extraction Research Based on GLCM-CLBP Algorithm,” vol. 76, no. Emim, pp. 167–171, 2017, doi: 10.2991/emim-17.2017.36.

[38] W. K. Mutlag, S. K. Ali, Z. M. Aydam, and B. H. Taher, “Feature Extraction Methods: A Review,” *J. Phys. Conf. Ser.*, vol. 1591, no. 1, 2020, doi: 10.1088/1742-6596/1591/1/012028.

[39] M. Arya, N. Mittal, and G. Singh, “Texture-based feature extraction of smear images for the detection of cervical cancer,” *IET Comput. Vis.*, vol. 12, no. 8, pp. 1049–1059, 2018, doi: 10.1049/iet-cvi.2018.5349.

[40] Pulung Nurtantio Andono and S. H. Nugraini, “Texture Feature Extraction in Grape Image Classification Using K-Nearest Neighbor,” *J. RESTI (Systems Eng. Technol. Information)*, vol. 6, no. 5, pp. 768–775, 2022, doi: 10.29207/resti.v6i5.4137.

[41] I. G. R. A. Sugiartha, M. Sudarma, and I. M. O. Widyantara, “Color, Texture and Shape Feature Extraction for Clustered-Based Retrieval of Images (CLUE),” *Sci. Mag. Electr. Technol.*, vol. 16, no. 1, p. 85, 2016, doi: 10.24843/mite.1601.12.

[42] D. Satria Yudha Kartika and H. Maulana, “Preprocessing and normalization on butterfly dataset for color, shape and texture feature extraction,” *J. Comput. Electron. Telecommun.*, vol. 1, no. 2, pp. 1–8, 2021, doi: 10.52435/complete.v1i2.76.

[43] Y. R. Kaesmetan and M. V. Overbeek, “Digital Image Processing using Texture Features Extraction of Local Seeds in Nekbaun Village with Color Moment, Gray Level Co Occurance Matrix, and k-Nearest Neighbor,” *Ultim. J. Tek. Inform.*, vol. 13, no. 2, pp. 81–88, 2022, doi: 10.31937/ti.v13i2.2038.

[44] A. Herdiansah, R. I. Borman, D. Nurnaningsih, A. A. J. Sinlae, and R. R. Al Hakim, “Herbal Leaf Image Classification Using Backpropagation Neural Networks Based on Shape Feature Extraction,” *JURIKOM (Journal Comput. Res.*, vol. 9, no. 2, p. 388, 2022, doi: 10.30865/jurikom.v9i2.4066.

[45] Y. F. Sihombing, A. Septiarini, A. H. Kridalaksana, and N. Puspitasari, “Chili Classification Using Shape and Color Features Based on Image Processing,” *Sci. J. Informatics*, vol. 9, no. 1, pp. 42–50, 2022, doi: 10.15294/sji.v9i1.33658.

[46] Jiawei Han, Micheline Kamber, and Jian Pei, *Data Mining: Concepts and Techniques*. Elsevier, 2012.

[47] D. Sundararajan, *Digital image processing: A signal processing and algorithmic approach*. 2017. doi: 10.1007/978-981-10-6113-4.

[48] B. Furht, E. Akar, and W. A. Andrews, *Digital Image Processing: Practical Approach*. 2018. [Online]. Available: http://www.springer.com/series/10028

[49] R. M. Thanki and A. M. Kothari, *Digital image processing using SCILAB*. 2018. doi: 10.1007/978-3-319-89533-8.

[50] J. Brownlee, “Data Preparation for Machine Learning,” *Proc. - 2nd Int. Conf. Informatics, Multimedia, Cyber, Inf. Syst. ICIMCIS 2020*, pp. 284–289, 2020, doi: 10.1109/ICIMCIS51567.2020.9354273.

[51] B. Purnama, *Concepts and Practicum with Example Exercises Based on R and Python*. Bandung: Informatika, 2019.

[52] S. Pramana, Budi Yuniarto, Siti Mariyah, Ibnu Santoso, and Rani Nooraeni., *Data Mining with R*. Bogor: In Media, 2018.

[53] R. Primartha, *Machine Learning algorithms*. Bandung: Informatika, 2021.