INTERNATIONAL JOURNAL OF eBUSINESS AND eGOVERNMENT STUDIES
Vol: 15 **No:** 1 **Year:** 2023 **ISSN:** 2146-0744 (Online) (**pp.** 512-532) **DOI:** 10.34111/ijebeg.2023150123
Received: 29.06.2023 | **Accepted:** 03.10.2023 | **Published Online:** 18.11.2023

-RESEARCH ARTICLE-

IMPROVING PERFORMANCE OF COPRA TYPE CLASSIFICATION USING FEATURE EXTRACTION WITH K-NEAREST NEIGHBOUR ALGORITHM

Husni Teja Sukmana*

Department of Informatics Engineering,
Universitas Islam Negeri Syarif Hidayatullah, Jakarta, Indonesia
Email: husniteja@uinjkt.ac.id
https://orcid.org/0000-0002-0756-818X

Puspitasari

Department of Informatics Engineering,
Universitas Islam Negeri Syarif Hidayatullah, Jakarta, Indonesia
Email: puspita.sari19@mhs.uinjkt.ac.id
https://orcid.org/0000-0003-3158-0485

Aryajaya Alamsyah

Department of Informatics Engineering,
Universitas Islam Negeri Syarif Hidayatullah, Jakarta, Indonesia
Email: aryajayaalamsyah@gmail.com
https://orcid.org/0000-0002-5991-9746

Saepul Aripiyanto

Department of Informatics Engineering,
Universitas Islam Negeri Syarif Hidayatullah, Jakarta, Indonesia
Email: saepul.aripiyanto@uinjkt.ac.id
https://orcid.org/0000-0003-0979-3441

Lee Kyung Oh

Department of Computer Engineering,
Sun Moon University, South Korea
Email: leeko@sunmoon.ac.kr

*Corresponding Author Email: husniteja@uinjkt.ac.id

Citation (APA): Sukmana, H. T., Puspitasari, Alamsyah, A., Aripiyanto, S., Oh, L. K. (2023). Improving Performance of Copra Type Classification Using Feature Extraction with K-Nearest Neighbour Algorithm. *International Journal of eBusiness and eGovernment Studies, 15*(1), 512-532. DOI: 10.34109/ijebeg.2023150123

─Abstract─

Through observations and interviews with coconut farmers in the Indragiri Hilir, Bitung, and Palembang districts, it has been found that the assessment of copra quality continues to encounter various obstacles, such as demanding substantial labour, time, and expenses. This research focuses on categorising copra types by extracting features and utilising the K-nearest neighbour algorithm. The research involves a dataset from copra warehouses in Indragiri Hilir Regency, Riau Province, consisting of 613 digital images classified into three categories: edible, regular, and reject. When categorising copra types, feature extraction methods like colour, shape, and texture features are taken into account. According to the study findings, the model accuracy when all features are taken into account is 84%.

Keywords: Classification Algorithm, Copra Type, Feature Extraction, K-Nearest Neighbor.

INTRODUCTION

Kopra comes from the flesh of the coconut (Cocos nucifera) and is typically processed using traditional drying techniques, such as sun-drying or smoking (Aryanti, Nafiunisa, & Willis, 2016). Copra is categorised into three types depending on its quality and purpose: edible, regular, and reject copra (Yuni, Ekpa, & Oyekale, 2017; Zubair & Ullah, 2020; Zwing et al., 2020). Edible copra meets high standards of hygiene and food safety, making it a valuable raw material in the food and coconut oil industries (Yuni, Ekpa, & Oyekale, 2017). Regular copra is used in the cosmetics, soap, and other non-food products sectors (Yuni, Ekpa, & Oyekale, 2017). Reject copra does not meet quality standards for use in the food and beverage industry (Yuni, Ekpa, & Oyekale, 2017). According to findings from an interview with a copra merchant in Indragiri Hilir Regency, Riau Province, it has been discovered that copra types are typically classified manually by copra merchants relying on intuition. Identifying different copra types typically involves examining specific traits. As an illustration, edible copra typically measures less than 9 cm in diameter, has a bowl-like shape, lacks defects on the husk, is not hollow, does not have a brownish colour, is free from fungal growth, and has white flesh. On the other hand, standard copra measures over 9 cm in diameter and can vary in shape, while maintaining similar qualities to edible copra. Copra of the reject type emerges after the sorting process of Edible and Regular. This type typically exhibits signs of fungi, a brownish colour, and potential holes or cracks. Undoubtedly, this procedure demands a substantial investment of effort, time, and resources. Moreover, the varying viewpoints of each merchant when identifying copra types result in inconsistent classification. As a consequence, the efficiency of farmers' performance decreases, particularly when conducted on a large scale. Therefore, there is a necessity for the creation of a method that can automatically and precisely categorise various kinds of copra to aid in this procedure. Accurate and efficient categorization can assist in managing copra supplies according to quality standards and meeting the diverse requirements of copra users in different industrial sectors more effectively (Rahmawati & Gunawan, 2020).

Vol: 15 No: 1 Year: 2023 ISSN: 2146-0744 (Online) (pp. 512-532) DOI: 10.34111/ijebeg.2023150123

This research project is centred on categorising copra types through the utilisation of the KNN technique with a focus on feature extraction. This is based on using a dataset that includes digital images. Feature extraction is focused on obtaining distinct values from each image to enable the computer to recognise the objects within them (Woods et al., 2018). The feature extraction process utilises various fundamental techniques such as colour, shape, and texture features (Akmal, Munir, & Santoso, 2023; Dhanashree et al., 2016; Kandalkar et al., 2015; Satpute & Jagdale, 2016). This study examines colour features using the colour moment method, incorporating parameters like RGB, HSV, and grayscale (Abdullah et al., 2021; Laxmi & Kusumah, 2019; Malakar & Mukherjee, 2013; Shinde et al., 2015). In addition, shape features use region-based methods with area parameters and contour-based techniques with perimeter parameters (Dhanoa & Garg, 2016; Nagarjun et al., 2019; Tandel & Patel, 2016). Finally, texture features utilise the Grey Level Co-occurrence Matrix (GLCM) along with contrast, dissimilarity, homogeneity, energy, and correlation parameters from various sources (Ding, Xiao, & Weng, 2017; Humeau-Heurtier, 2019; Mutlaq & Abdulbaqia, 2022; Veronica, 2020). According to the research literature, choosing these characteristics has been shown to improve the performance of the classification model with digital image datasets (Andono & Nugraini, 2022; Arya, Mittal, & Singh, 2018; Herdiansyah et al., 2022; Kartika et al., 2021; Overbeek, Beelen, & Jansen, 2022; Sihombing et al., 2022; Sugiartha et al., 2018).

Table 1: Research Contribution of Copra Type Classification.

References	Feature extractions			Algorithms	Results
	Color Features	Shape Features	Texture Features		
(Abdullah et al., 2017)	- Color RGB	x	- Contras - Homogeneity - Energy - Correlation	Nearest Mean Classifier	Accuracy = 74%
(Lim, Jeong, & Choi, 2019)	- Color RGB	x	x	K-Nearest Neighbor	Accuracy = 86%
(Adang et al., 2020)	- Color RGB - Color HIS	x	x	Naïve Bayes	Accuracy = 91%
(Marni & Marlis, 2021)	- Color RGB	- Area - Perimeter	x	K-Nearest Neighbor	Accuracy = 93%
(Lahey, 2023)	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%

Vol: 15 **No:** 1 **Year:** 2023 **ISSN:** 2146-0744 (Online) (**pp.** 512-532) **DOI:** 10.34111/ijebeg.2023150123

This study aims to develop a classification model for different types of copra by focusing on feature extraction with the KNN algorithm. This research's contribution is in utilising feature extraction to enhance accuracy. Following this, the study aims to improve the effectiveness of copra selection for copra farmers in Indragiri District, Riau Province. The findings of this study are outlined in Table 1.

LITERATURE REVIEW

Several studies have been conducted on the classification of copra types by various researchers (Abdullah et al., 2017; Adang et al., 2020; Lahey, 2023; Lim, Jeong, & Choi, 2019; Marni & Marlis, 2021). In a study conducted by Abdullah et al. (2017), software was developed for classifying copra quality using the Nearest Mean Classifier (MNC) algorithm. This study yielded an accuracy rate of 80.67%. In a study conducted by Lim, Jeong, & Choi (2019), a conveyor belt machine was developed to automatically identify coconut quality by utilising the K-nearest neighbour (KNN) algorithm. This study's findings show an accuracy rate of 86.67%. In a recent study by Adang et al. (2020), a software was developed to classify copra maturity utilising the Naïve Bayes (NB) algorithm. This study's results show an accuracy rate of 91.12%. In a recent study by Marni & Marlis (2021) and colleagues, a model was developed to classify the quality of white copra or edible copra using the KNN algorithm. This study yielded an accuracy rate of 93.33%. At last, the fifth study (Lahey, 2023) developed a model for assessing the quality of copra through a fuzzy logic algorithm. This study's findings boast a 95% accuracy rate. After reviewing the literature, various algorithms have been employed to categorise copra types and assess coconut quality, including MNC, KNN, NB, and fuzzy logic (Abdullah et al., 2017; Adang et al., 2020; Lahey, 2023; Lim, Jeong, & Choi, 2019; Marni & Marlis, 2021).

RESEARCH METHODOLOGY

The study was carried out through multiple stages that adhered to systematic guidelines. Several research steps involve data acquisition, data preprocessing, feature extraction, feature scaling, distributing training and test data, implementing the knn algorithm, and evaluating the classification model. The research stages are illustrated in Figure 1.

Data Acquisition

The research data is sourced from the coconut copra warehouse in Indragiri Hilir Regency, Riau Province. This dataset includes three types of copra: edible copra, regular copra, and reject copra. Every variety of copra includes 613 image data, totaling a dataset of 1,839. Image acquisition was conducted using a smartphone camera to capture digital images. The results of the acquisition are displayed in Figure 2**.**

Vol: 15 **No:** 1 **Year:** 2023 **ISSN:** 2146-0744 (Online) (**pp.** 512-532) **DOI:** 10.34111/ijebeg.2023150123

Data Preprocessing

Data preprocessing involves converting unstructured data into a structured format based on research requirements (Han et al., 2012; Yanto, Subali, & Suyanto, 2019). The study involves conducting image processing tasks such as resizing, background removal, and cropping based on previous research (Furht et al., 2018; Sundararajan & Sarwat, 2017; Thanki, Kothari, & Borra, 2021). Images are resized to standardise their dimensions from varied and large to uniform and smaller sizes (Furht et al., 2018; Sundararajan & Sarwat, 2017; Thanki, Kothari, & Borra, 2021). Eliminating background is a common practice in image editing to get rid of any distracting or irrelevant elements (Furht et al., 2018; Sundararajan & Sarwat, 2017; Thanki, Kothari, & Borra, 2021). Image cropping focuses on displaying only the essential parts of the image. Figure 3 displays the outcomes of data preprocessing.

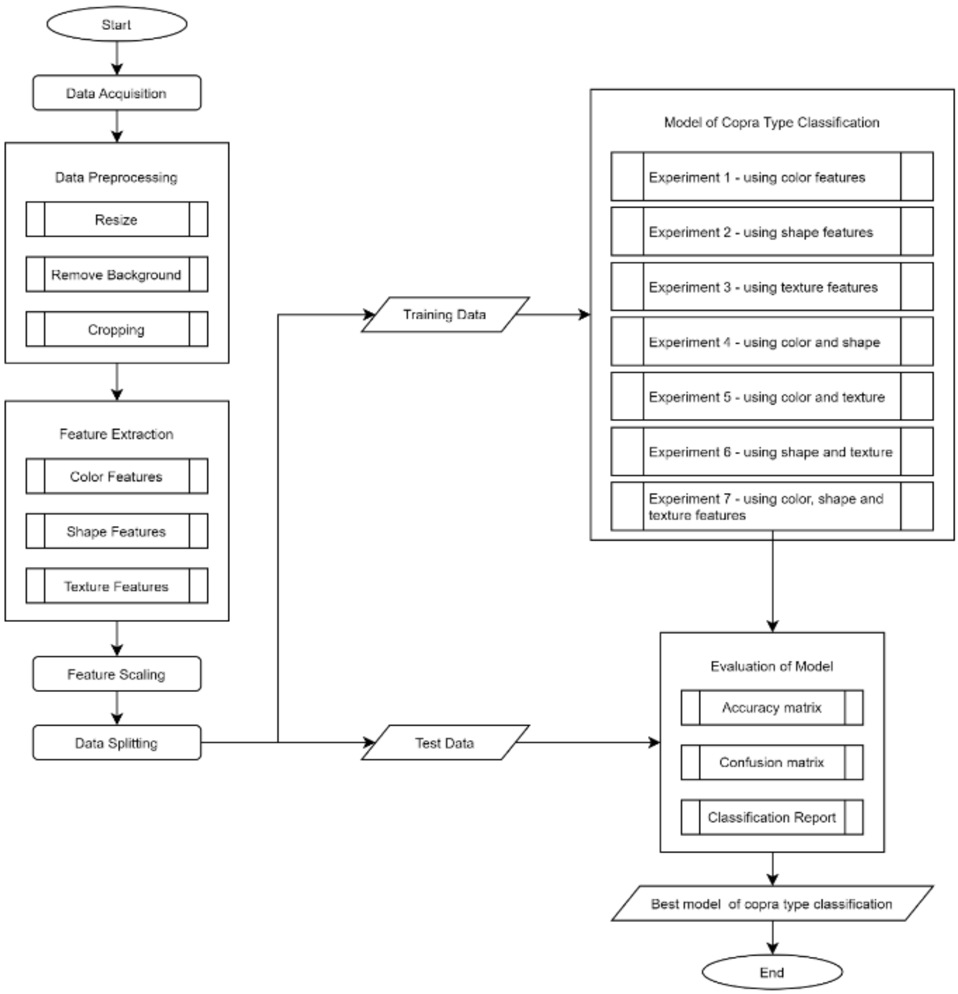

Figure 1: The Research Stages of Copra Type Classification Model.

Vol: 15 **No:** 1 **Year:** 2023 **ISSN:** 2146-0744 (Online) (**pp.** 512-532) **DOI:** 10.34111/ijebeg.2023150123

Figure 2: Dataset of Copra Type from Indragiri Hilir Regency, Riau Province.

Figure 3: Result of Preprocessing Data with Resize Image, Remove Background and Cropping Image.

Feature Extraction

Identifying and extracting essential information from a digital image to create simpler features involves considering the image's characteristics (González Fernández et al., 2018). This study highlights how feature extraction simplifies the classification process of copra types. Colour, shape, and texture features are commonly used in feature extraction techniques. Colour features are crucial for distinguishing digital images by analysing colour distribution, colour scheme, and unique colour characteristics in individual pixels of the digital image (Akmal, Munir, & Santoso, 2023; Dhanashree et al., 2016; Kandalkar et al., 2015; Satpute & Jagdale, 2016). Geometric information such as size, object area, the length of lines surrounding the object, and object aspect ratio are crucial in identifying digital images (Abdullah et al., 2021; Laxmi & Kusumah, 2019; Malakar & Mukherjee, 2013; Shinde et al., 2015). Features are extracted from a digital image by analysing pixel patterns to distinguish between different objects (Dhanoa & Garg, 2016; Nagarjun et al., 2019; Tandel & Patel, 2016). This study examines colour, shape, and texture features using different parameters, as shown in Figure 4.

Vol: 15 **No:** 1 **Year:** 2023 **ISSN:** 2146-0744 (Online) (**pp.** 512-532) **DOI:** 10.34111/ijebeg.2023150123

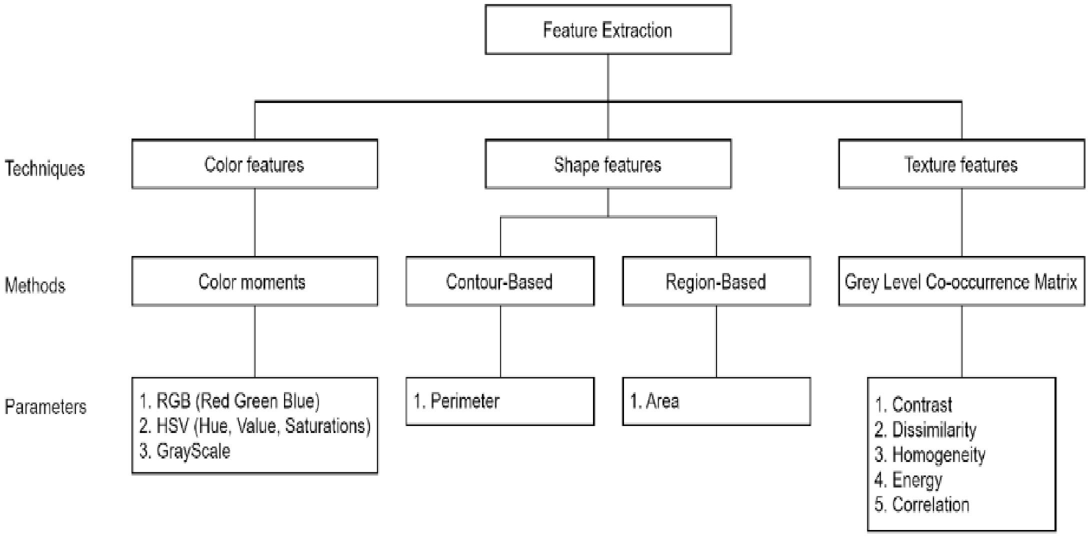

Figure 4: Result of Feature Extraction on Copra Type Classification.

Feature Scaling

Normalising features is a method used to ensure that all features have similar values. Standardising all features to the same scale is the main objective of feature scaling. It is essential for various machine learning algorithms as they can achieve higher accuracy when features are standardised. For this study, feature scaling is done using the min-max scaler method with a value range of 0 to 1. The formula for the min-max scaler as presented in Equation 1 by Brownlee et al. (2020).

$$Normalization = a + \frac{\left(x - min\,(x)\right) * (b - a)}{max(x) - \min\,(x)} \tag{1}$$

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

Data Splitting

Splitting data involves dividing a dataset into two or more subsets of data. In this research, data is divided into training data and testing data. The data for training is utilised to educate the model on categorising different types of copra, such as edible, regular, and rejected copra. At the same time, the testing data is used to assess the effectiveness of the copra classification model. For this study, the data was divided using a split validation technique, allocating 90% for training and 10% for testing.

Vol: 15 **No:** 1 **Year:** 2023 **ISSN:** 2146-0744 (Online) (**pp.** 512-532) **DOI:** 10.34111/ijebeg.2023150123

Implementation of KNN Algorithm

K-nearest neighbours (KNN) is a non-parametric classification algorithm that classifies an object in the test data by considering the majority class of the k nearest neighbours in the training data (Purnama et al., 2019). When using the KNN algorithm, it's crucial to take into account the value of K since a lower K value can make the classification model more susceptible to noisy data. On the other hand, a higher K value enhances the classification model's ability to handle noise, although it could potentially lead to bias in the class boundaries (Primartha et al., 2023; Purnama et al., 2019; Purnama, Stiawan, & Budiarto, 2018).

For this study, the value of K is determined based on the different types of copra - edible, regular, and rejected - resulting in a K value of 3. Specific methods such as the elbow method, silhouette coefficient, Calinski-Harabasz index, and Davies-Bouldin index are not used for this purpose. The KNN algorithm leverages the concept of closeness between data points to carry out data classification. This study involves the use of the Euclidean distance formula (Primartha et al., 2023; Purnama et al., 2019; Purnama, Stiawan, & Budiarto, 2018) to calculate the distance between data points, as referenced in the provided sources, and the KNN algorithm as demonstrated in the pseudocode (Purnama et al., 2019).

$$d(a,b) = \sqrt{\sum_{i=1}^{n} (x_{ai} - y_{bi})^2} \tag{2}$$

He Description:

$d(a,b)$ = Distance between object x and object y.
x_{ai} = The value of training data object a in the i-th variable.
y_{bi} = 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

$D_j \leftarrow DataSet$
$X \leftarrow TestData$
$N \leftarrow length\ of\ D_j$
for $i = 1 \rightarrow (N-1)$ **do**
    $d \leftarrow Euclidean\ distance\ (X, D_j)$
    $List[i] \leftarrow d$
end
$Sorting\ List[i]$
for $i = 1 \rightarrow k$ **do**
    $Class[k] \leftarrow label\ of\ List[i]$
end
$ClassData \leftarrow calculate\ majority\ Class[k]$
return $Class$

Vol: 15 **No:** 1 **Year:** 2023 **ISSN:** 2146-0744 (Online) (**pp.** 512-532) **DOI:** 10.34111/ijebeg.2023150123

Identifying different types of copra by analysing colour, shape, and texture features. Next, the copra-type classification process involves utilising the KNN algorithm. Developing a classification model through multiple experiments to assess the impact of colour, shape, and texture features on enhancing the accuracy of the model with the KNN algorithm. Several of these experiments are displayed 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.
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.

Model Evaluation

Assessing the model involves determining how well the copra-type classification model performs. Assessing the performance of a classification model involves evaluating how well it performs and its ability to generalise the data, as discussed by various

INTERNATIONAL JOURNAL OF eBUSINESS AND eGOVERNMENT STUDIES
Vol: 15 **No:** 1 **Year:** 2023 **ISSN:** 2146-0744 (Online) (**pp.** 512-532) **DOI:** 10.34111/ijebeg.2023150123

authors (Abhishek, Karthik, & Sneha, 2022; Agarwal et al., 2021; Andreas et al., 2016; Suyanto et al., 2021). For this study, the evaluation of the model utilises the confusion matrix technique to calculate accuracy, precision, recall, and f1-score. Equations 3 to 6 illustrate the formulas for calculating accuracy, precision, recall, and f1-score (Abhishek, Karthik, & Sneha, 2022; Agarwal et al., 2021; Andreas et al., 2016; Suyanto et al., 2021).

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{3}$$

$$\text{Recall} = \frac{(TP)}{TP + FN} \tag{4}$$

$$\text{Precision} = \frac{(TP)}{(TP + FP)} \tag{5}$$

$$\text{F1-Score} = \frac{(2 * Precision * Recall)}{Precision + Recall} \tag{6}$$

Description:

TP = True positive
TN = True negative
FP = False positive
FN = False negative

RESULTS AND DISCUSSION

Implementation of Feature Extraction

Feature extraction in this study includes colour, shape, and texture features. Every feature comes with various parameters, as illustrated in Figure 4. After extracting colour, shape, and texture features, it is necessary to adjust their value ranges to 0 to 1 through feature scaling using Equation 1. Feature extraction and feature scaling are essential for helping the classification model accurately identify different types of coconuts. Here are the findings from the feature extraction implementation, as shown in Table 3 and Table 4.

Vol: 15 **No:** 1 **Year:** 2023 **ISSN:** 2146-0744 (Online) (**pp.** 512-532) **DOI:** 10.34111/ijebeg.2023150123

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
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.

Implementation of Copra Type Classification on KNN Algorithm

The coconut types are classified using the KNN algorithm in the model. The creation of this classification model includes seven experiments, as illustrated in Table 2. This study involves seven experiments that seek to investigate if performance is enhanced by utilising feature extraction along with the KNN algorithm in the classification model. Measuring the performance of the classification model involves using a confusion matrix to calculate accuracy, precision, recall, and f1-score values. The confusion matrix results for the coconut-type classification model can be found in Figure 5.

From the confusion matrix, the true positive, true negative, false positive, and false negative values have been identified, allowing for the calculation of accuracy, precision, recall, and f1-score using equations 3 and 4. The accuracy, precision, recall, and f1-score results for the coconut-type classification model can be found in Table 5 and Figure 6.

Vol: 15 **No:** 1 **Year:** 2023 **ISSN:** 2146-0744 (Online) (**pp.** 512-532) **DOI:** 10.34111/ijebeg.2023150123

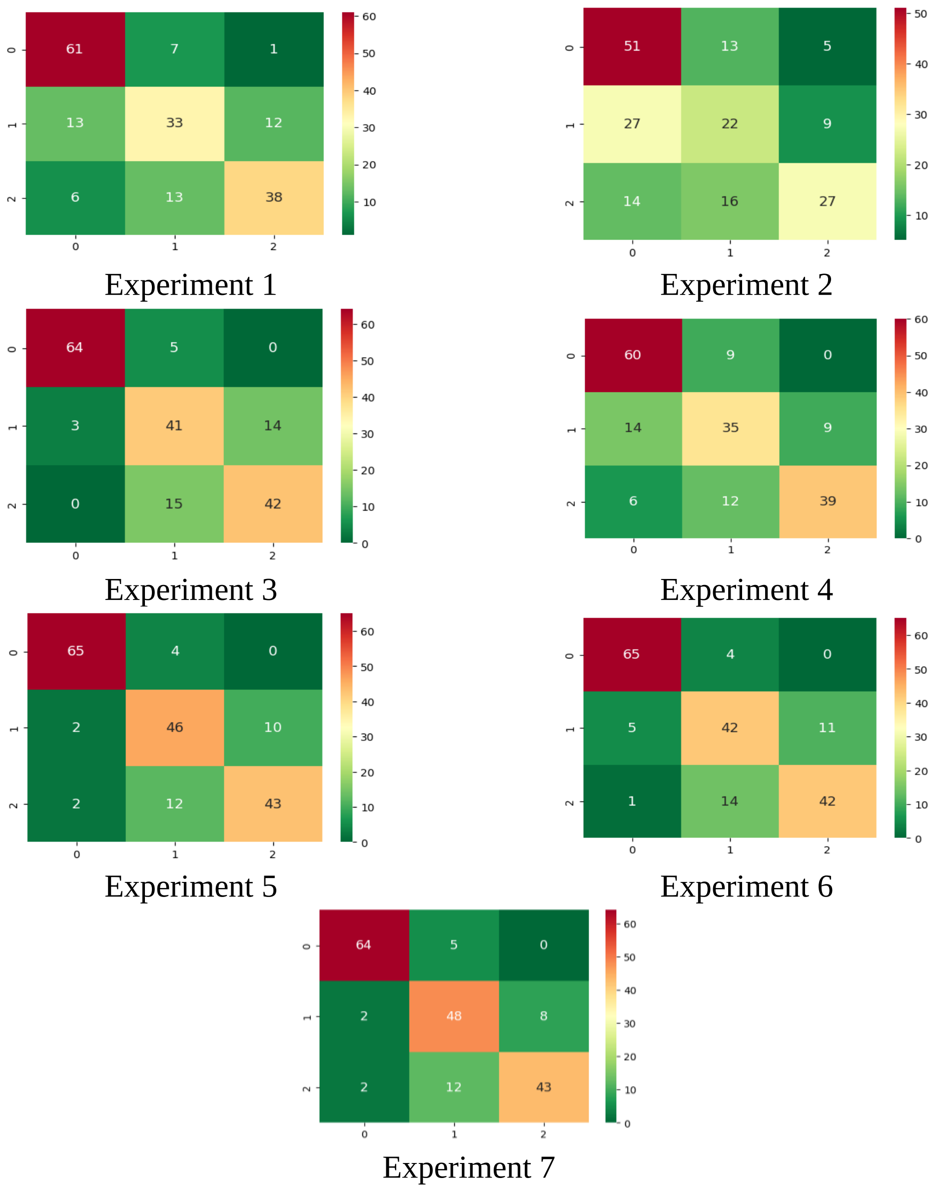

Figure 5. Results of Confusion Matrix for Copra Type Classification.

Table 5: 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

Vol: 15 **No:** 1 **Year:** 2023 **ISSN:** 2146-0744 (Online) (**pp.** 512-532) **DOI:** 10.34111/ijebeg.2023150123

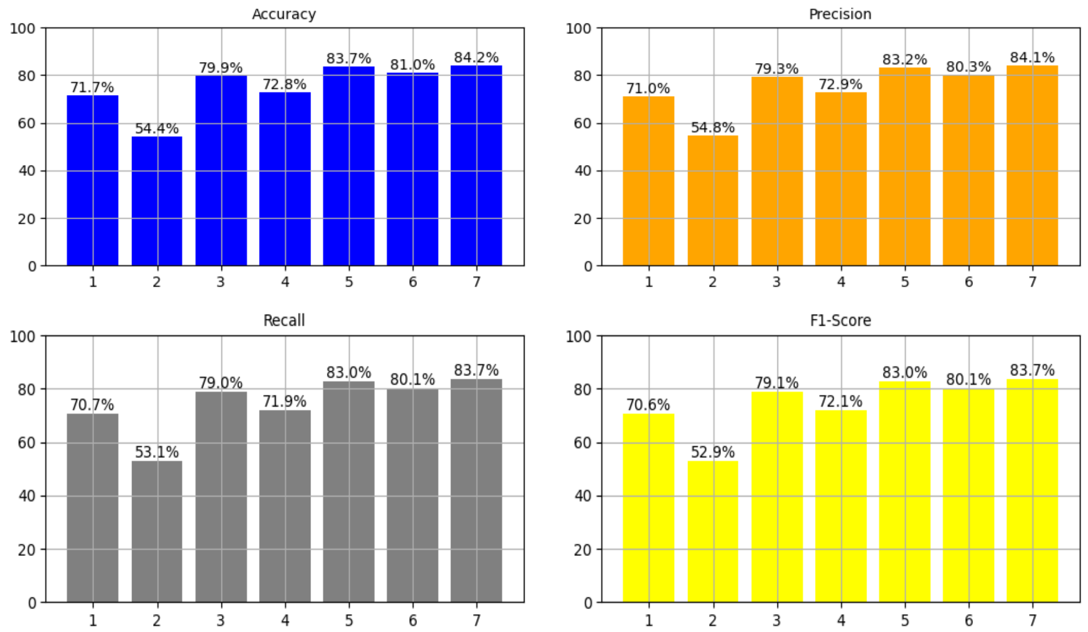

Figure 6: Result of Model Evaluation of Copra Type Classification.

According to Table 5 and Figure 6, the research on categorising copra types through feature extraction with the KNN algorithm has demonstrated success. The study has produced numerous intriguing facts and discoveries. The study findings show that the copra-type classification model attains an average accuracy of 75%. When utilising colour features in the classification model (experiment 1), it achieves an accuracy of 72%. In Experiment 2, shape features achieved an accuracy of 54%, whereas Experiment 3 showed texture features with an accuracy of 79%. The experiments indicate that colour features result in a 3% decrease in accuracy, shape features lead to a 21% decrease in accuracy, and texture features can result in a 4% increase in accuracy. This study confirms that texture features are the most effective for classifying copra types.

In subsequent experiments, combining two features in the copra-type classification model, like colour with shape (experiment 4), colour with texture (experiment 5), and shape with texture (experiment 6), leads to enhanced model performance. The experiments indicate that Experiment 4 yields an accuracy of 72%, resulting in a 2% decrease in accuracy. Experiment 5 achieved an accuracy of 83%, showing an increase of 8%, while Experiment 6 achieved an accuracy of 80%, with an increase of 5%. The experiments demonstrated that integrating colour and shape features can decrease the classification model error from 3% to 2%. Colour or shape features impact the model's performance, but adding texture features can boost accuracy by 10% and 7% respectively. Upon integrating all features in the copra-type classification model (experiment 7), the accuracy reaches 84%, marking a 9% increase in accuracy.

Vol: 15 **No:** 1 **Year:** 2023 **ISSN:** 2146-0744 (Online) (**pp.** 512-532) **DOI:** 10.34111/ijebeg.2023150123

Analysis of Correlation between Features and Labels

The results of classifying coconut types indicate that the shape feature has a substantial impact on the performance of the classification model when using feature extraction and the KNN algorithm. One possible reason for this discrepancy is that the classification model's utilisation of texture features leads to an accuracy of 79%, while colour features and shape features result in accuracies of 71% and 54%, respectively. Statistical analysis was conducted to test the accuracy results by correlating the features used with the type of copra. Initial testing utilised a bivariate correlation test to examine feature extraction parameters related to copra types, as depicted in Table 6. The Pearson correlation method was used for the bivariate correlation test. The test results indicate that texture features show a strong correlation with copra types, ranging from 0.4 to 0.6. Colour features exhibit a weaker correlation with copra types, with correlation values ranging from 0.1 to 0.6. Shape features show no correlation with copra types, with the highest correlation value being 0.1 and the lowest being 0. Thus, texture features are the most effective for classifying copra types.

Table 6: 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.

Vol: 15 **No:** 1 **Year:** 2023 **ISSN:** 2146-0744 (Online) (**pp.** 512-532) **DOI:** 10.34111/ijebeg.2023150123

The research experiments were validated through multivariate correlation testing. Utilising multivariate correlation testing is valuable for establishing the importance of this research's findings. The study utilised ordinary least squares regression for multivariate correlation testing, as displayed in Table 7. The test results indicate that feature extraction has a significant impact on the accuracy of the copra-type classification model in each experiment. The test results confirmed the efficacy of each experiment, showing strong correlations between 0.86 and 0.15. Moreover, each experiment was effective in differentiating between types of copra based on the significant correlation between data variation and determination values, which ranged from 0.74 to 0.02.

Upon initial inspection, the second experiment showed no clear relationship or certainty in categorising copra types. When the second experiment was combined with the third in the sixth trial, the model's accuracy remained strong, showing a correlation value of 0.81 and a determination of 0.65. Afterward, the most effective approach was to combine all features or carry out the seventh experiment, yielding a correlation value of 0.86 and a determination of 0.74.

Table 7: Result of Model Evaluation of Copra Type Classification.

Coefficient	Experiments						
	1	2	3	4	5	6	7
Correlation	71,7	54,4	79,9	72,8	83,7	81,0	84,2
Determination	71,0	54,8	79,3	72,9	83,2	80,3	84,1

DISCUSSION

The copra type classification method utilises KNN due to its ease of understanding and implementation, versatility with different data types, non-parametric nature suitable for datasets with diverse distributions, and resilience to noise and missing values (Abhishek, Karthik, & Sneha, 2022; Agarwal et al., 2021; Andreas et al., 2016; Suyanto et al., 2023; Suyanto et al., 2021). Nevertheless, this approach presents various constraints, including the necessity for precise K parameter adjustment, which is influenced by feature selection (Abhishek, Karthik, & Sneha, 2022; Agarwal et al., 2021; Andreas et al., 2016; Suyanto et al., 2023; Suyanto et al., 2021). The KNN algorithm has demonstrated effectiveness and efficiency in classifying digital images and tabular data across different case studies, as indicated by literature research (Jami et al., 2023; Khairina & Games, 2022; Khotimah, Utami, & Listyanto, 2022; Raysyah, Arinal, & Mulyana, 2021; Sari & Susilo, 2020; Suharyana et al., 2023; Triayudi, Suparman, & Andrianingsih, 2022).

CONCLUSION AND SUGGESTION

Conclusion

The study improved accuracy by classifying copra types based on colour, shape, and texture features using the k-nearest neighbour method. The colour features extracted are RGB, HSV, and grayscale. Shape features include area and perimeter, while texture features consist of contrast, dissimilarity, homogeneity, energy, and correlation. Research findings suggest that the copra-type classification, when factoring in feature extraction, achieves a commendable average accuracy rate of 75%. Moreover, the study findings indicate that texture features have a substantial impact on accuracy, achieving 79% during the testing phase. The study shows that combining all elements (colour, shape, and texture) results in the highest accuracy of 84%. Therefore, the research concludes that incorporating feature extraction in copra-type classification can enhance accuracy by up to 9%.

Suggestions

Future researchers should consider investigating the classification of copra types. Deep learning models have shown success in image recognition tasks, improving accuracy through automatic learning of hierarchical features. He furthered his research by examining alterations in copra characteristics over time. The recommendations are intended to inspire upcoming researchers to delve into sophisticated methods and aspects for categorising copra types, potentially expanding the limits of precision and practicality.

REFERENCES

Abdullah, A. R., Too, J., Zawawi, T. T., Saad, N. M., & Musa, H. (2017). Classification of EMG signal based on time domain and frequency domain features. *International Journal of Human and Technology Interaction, 1*(1), 25-30. Retrieved from https://journal.utem.edu.my/index.php/ijhati/article/view/2840

Abdullah, M. N., Shukran, M. A. M., Isa, M. R. M., Ahmad, N. S. M., Khairuddin, M. A., Yunus, M. S. F. M., & Ahmad, F. (2021). Colour features extraction techniques and approaches for content-based image retrieval (CBIR) system. *Journal of Materials Science and Chemical Engineering, 9*(7), 29-34. doi: https://doi.org/10.4236/msce.2021.97003

Abhishek, S. R., Karthik, P. B. H., & Sneha, N. (2022). An emerging treatment technology: Exploring deep learning and computer vision approach in revealing biosynthesized nanoparticle size for optimization studies. In *Contaminants of Emerging Concerns and Reigning Removal Technologies* (pp. 403-428). CRC Press. doi: https://doi.org/10.1201/9781003247869-20

Adang, L. A., Pizzino, A., Malhotra, A., Dubbs, H., Williams, C., Sherbini, O., et al. (2020). Phenotypic and imaging spectrum associated with WDR45. *Pediatric Neurology, 109*, 56-62. doi: https://doi.org/10.1016/j.pediatrneurol.2020.03.005

Vol: 15 **No:** 1 **Year:** 2023 **ISSN:** 2146-0744 (Online) (**pp.** 512-532) **DOI:** 10.34111/ijebeg.2023150123

Agarwal, N., Sondhi, A., Chopra, K., & Singh, G. (2021). Transfer learning: Survey and classification. *Smart Innovations in Communication and Computational Sciences: Proceedings of ICSICCS 2020, 1168*, 145-155. doi: https://doi.org/10.1007/978-981-15-5345-5_13

Akmal, A., Munir, R., & Santoso, J. (2023). Automatic Weight of Color, Texture, and Shape Features in Content-Based Image Retrieval Using Artificial Neural Network. *JOIV: International Journal on Informatics Visualization, 7*(3), 665-672. doi: http://dx.doi.org/10.30630/joiv.7.3.1184

Andono, P. N., & Nugraini, S. H. (2022). Texture feature extraction in grape image classification using k-nearest neighbor. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), 6*(5), 768-775. doi: https://doi.org/10.29207/resti.v6i5.4137

Andreas, J., Rohrbach, M., Darrell, T., & Klein, D. (2016). Learning to compose neural networks for question answering. *arXiv preprint arXiv:1601.01705*. doi: https://doi.org/10.48550/arXiv.1601.01705

Arya, M., Mittal, N., & Singh, G. (2018). Texture-based feature extraction of smear images for the detection of cervical cancer. *IET Computer Vision, 12*(8), 1049-1059. doi: https://doi.org/10.1049/ietcvi.2018.5349

Aryanti, N., Nafiunisa, A., & Willis, F. M. (2016). Extraction and characterization of chlorophyll from suji leaves (Pleomele angustifolia) as a natural food coloring. *Journal of Food Technology Applications, 5*(4), 129-135. doi: http://dx.doi.org/10.17728/jatp.196

Brownlee, A. E., Wright, J. A., He, M., Lee, T., & McMenemy, P. (2020). A novel encoding for separable large-scale multi-objective problems and its application to the optimisation of housing stock improvements. *Applied Soft Computing, 96*, 106650. doi: https://doi.org/10.1016/j.asoc.2020.106650

Dhanashree, S., Priyanka, M., Manisha, K., & Vilasrao, K. (2016). Molecularly imprinted polymers: Novel discovery for drug delivery. *Current Drug Delivery, 13*(5), 632-645. doi: https://doi.org/10.2174/1567201813666160101120238

Dhanoa, J. S., & Garg, A. (2016). A novel technique for shape feature extraction using content based image retrieval. *MATEC Web of Conferences, 57*, 02002. doi: https://doi.org/10.1051/matecconf/20165702002

Ding, K., Xiao, L., & Weng, G. (2017). Active contours driven by region-scalable fitting and optimized Laplacian of Gaussian energy for image segmentation. *Signal Processing, 134*, 224-233. doi: https://doi.org/10.1016/j.sigpro.2016.12.021

Furht, B., Akar, E., Andrews, W. A., Furht, B., Akar, E., & Andrews, W. A. (2018). Introduction to digital imaging. *Digital Image Processing: Practical Approach*, 1-5. doi: https://doi.org/10.1007/978-3-319-96634-2_1

González Fernández, E., Sandoval Orozco, A. L., Garcia Villalba, L. J., & Hernandez-Castro, J. (2018). Digital image tamper detection technique based on spectrum analysis of CFA artifacts. *Sensors, 18*(9), 2804. doi: https://doi.org/10.3390/s18092804

Han, H.-J., Ryu, J.-H., Cho, S., Park, Y.-J., & Ahn, Y.-H. (2012). Overview of geostationary ocean color imager (GOCI) and GOCI data processing system (GDPS). *Ocean Science Journal*, *47*(3), 223-233. doi: https://doi.org/10.1007/s12601-012-0024-4

Herdiansyah, M. I., Antoni, D., Akbar, M., Cholil, W., & Syaputra, H. (2022). Use of Blockchain for Designing Digital Documents in Public Services. In *2022 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)* (pp. 55-60). IEEE. doi: https://doi.org/10.1109/ICIMCIS56303.2022.10017661

Humeau-Heurtier, A. (2019). Texture feature extraction methods: A survey. *IEEE Access*, *7*, 8975-9000. doi: https://doi.org/10.1109/ACCESS.2018.2890743

Jami, S. I., Munir, S., Malick, R. A. S., Ahmed, G., Khan, S., & Rodrigues, J. J. (2023). An integrated approach: using knowledge graph and network analysis for harnessing digital advertisement. *Multimedia Tools and Applications*, *82*(6), 8883-8898. doi: https://doi.org/10.1007/s11042-021-11856-2

Kandalkar, N., Mani, A., Pandey, G., & Soni, J. (2015). Content based image retrieval using color, shape and texture extraction techniques. *International Journal of Engineering and Technical Research*, *3*(5), 74-79. doi: https://doi.org/10.1007/s13369-020-04384-y

Kartika, D. S. Y., Maulana, H., Wulansari, A., Safitri, E. M., & Muttaqin, F. (2021). Digital Image Segmentation Resulting from X-Rays of Covid Patients using K-Means and Extraction Features Method. *IJCONSIST Journals*, *3*(1), 25-28. doi: https://doi.org/10.33005/ijconsist.v3i1.55

Khairina, F., & Games, D. (2022). The Influence of Talent Management Practices on Employee Performance: The Mediating Role of Employee Engagement and Employee Job Satisfaction.(Case Study at PT Bank Negara Indonesia (Persero) Tbk Regional Office 02). *Enrichment: Journal of Management*, *12*(4), 2879-2892. doi: https://doi.org/10.35335/enrichment.v12i4.748

Khotimah, K., Utami, S. N., & Listyanto, M. E. D. (2022). Cultivation techniques and feasibility of hydroponic lettuce farming by utilizing yards in Hidroponik Media. *Online Journal of Tropical Agriculture*, *9*(3), 224-231. doi: https://doi.org/10.32734/jpt.v9i3.10775

Lahey, M. (2023). Interoperable Findability: Investigating an Information Ecology of Graded Feedback in Higher Education. In *International Conference on Human-Computer Interaction* (pp. 314-324). Springer. doi: https://doi.org/10.1007/978-3-031-35572-1_21

Laxmi, G. F., & Kusumah, F. S. F. (2019). Region of interest and color moment method for freshwater fish identification. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, *17*(3), 1432-1438. doi: http://doi.org/10.12928/telkomnika.v17i3.11749

Lim, C.-G., Jeong, Y.-S., & Choi, H.-J. (2019). Survey of temporal information extraction. *Journal of Information Processing Systems*, *15*(4), 931-956. doi: https://doi.org/10.3745/JIPS.04.0129

Vol: 15 **No:** 1 **Year:** 2023 **ISSN:** 2146-0744 (Online) (**pp.** 512-532) **DOI:** 10.34111/ijebeg.2023150123

Malakar, A., & Mukherjee, J. (2013). Image clustering using color moments, histogram, edge and K-means clustering. *International Journal of Science and Research, 2*(1), 532-537. Retrieved from https://www.ijsr.net/archive/v2i1/IJSRON2013308.pdf

Marni, L., & Marlis, T. A. (2021). Empowerment on Jumantik and Planting Mosquito Repellent Plants in the Context of Dengue Vector Control. *Central Asian Journal of Medical and Natural Science, 2*(5), 409-418. Retrieved from https://cajmns.centralasianstudies.org/index.php/CAJMNS/article/view/405/380

Mutlaq, M. A., & Abdulbaqia, H. S. (2022). A Review of Image Segmentation Methods in Brain Tumor. *Journal of Al-Qadisiyah for Computer Science and Mathematics, 14*(3), 1-9. doi: https://doi.org/10.29304/jqcm.2022.14.3.981

Nagarjun, T. N., Rathod, K., Chandana, V., & Kashyap, D. V. (2019). Object Recognition using Shape Features. *International Journal of Engineering Research & Technology, 8*(3), 139-142. doi: https://doi.org/10.17577/IJERTV8IS030094

Overbeek, S., Beelen, L., & Jansen, S. (2022). Are you of value to me? A partner selection reference method for software ecosystem orchestrators. *Science of Computer Programming, 214*, 102733. doi: https://doi.org/10.1016/j.scico.2021.102733

Primartha, R., Desiani, A., Andhini, S. P., Yahdin, S., Arhami, M., Arianda, D. F., & Eliyati, N. (2023). Combination of artificial neural networks and particle swarm optimization on air quality prediction. *AIP Conference Proceedings, 2913*(1), 030014. doi: https://doi.org/10.1063/5.0172001

Purnama, B., Kurniabudi, K., Sharipuddin, S., Darmawijoyo, D., Stiawan, D., Samsuryadi, S., et al. (2019). Network anomaly detection research: a survey. *Indonesian Journal of Electrical Engineering and Informatics (IJEEI), 7*(1), 37-50. doi: https://doi.org/10.52549/ijeei.v7i1.773

Purnama, B., Stiawan, D., & Budiarto, R. (2018). Preprocessing and framework for unsupervised anomaly detection in IoT: work on progress. In *2018 International conference on electrical engineering and computer science (ICECOS)* (pp. 345-350). IEEE. doi: https://doi.org/10.1109/ICECOS.2018.8605231

Rahmawati, Z., & Gunawan, J. (2020). The Relationship between Job-related Factors, Work-life Balance and Job Satisfaction in Millennial Workers. *ITS Journal of Science and Art, 8*(2), D418-D423. doi: https://doi.org/10.12962/j23373520.v8i2.47782

Raysyah, S., Arinal, V., & Mulyana, D. I. (2021). Classification of coffee fruit maturity level based on color detection using knn and pca methods. *JSiI (Journal of Information Systems), 8*(2), 88-95. doi: https://doi.org/10.30656/jsii.v8i2.3638

Sari, R. F., & Susilo, B. (2020). Intrusion detection in IoT networks using deep learning algorithm. *Information, 11*(5), 279. doi: https://doi.org/10.3390/info11050279

Satpute, M. R., & Jagdale, S. M. (2016). Color, size, volume, shape and texture feature extraction techniques for fruits: a review. *International Research Journal of Engineering and Technology, 3*(4), 2395-2356. doi: https://doi.org/10.5120/ijca2015906780

Shinde, S. R., Sabale, S., Kulkarni, S., & Bhatia, D. (2015). Experiments on content based image classification using Color feature extraction. In *2015 international conference on communication, Information & Computing Technology (ICCICT)* (pp. 1-6). IEEE. doi: https://doi.org/10.1109/ICCICT.2015.7045737

Sihombing, Y. F., Septiarini, A., Kridalaksana, A. H., & Puspitasari, N. (2022). Chili Classification Using Shape and Color Features Based on Image Processing. *Scientific Journal of Informatics, 9*(1), 42-50. doi: https://doi.org/10.15294/sji.v9i1.33658

Sugiartha, N., Sugina, I. M., Putra, I. D. G. A. T., Indraswara, M. A., & Suryani, L. I. D. (2018). Development of an arduino-based data acquisition device for monitoring solar PV system parameters. In *International Conference on Science and Technology (ICST 2018)* (Vol. 1, pp. 995-999). Atlantis Press. doi: https://doi.org/10.2991/icst-18.2018.201

Suharyana, S., Anwar, F., Dewi, A. C., Yunianto, M., Salamah, U., & Chai, R. (2023). Pneumonia Classification Based on GLCM Features Extraction using K-Nearest Neighbor. *Indonesian Journal of Applied Physics, 13*(2), 325-338. Retrieved from https://jurnal.uns.ac.id/ijap/article/view/77120

Sundararajan, A., & Sarwat, A. I. (2017). Roadmap to prepare distribution grid-tied photovoltaic site data for performance monitoring. In *2017 International Conference on Big Data, IoT and Data Science (BID)* (pp. 110-115). IEEE. doi: https://doi.org/10.1109/BID.2017.8336582

Suyanto, E. R., Sofi, M., Lumantarna, E., & Aye, L. (2023). Comparison of waste photovoltaic panel processing alternatives in Australia. *Journal of Cleaner Production, 418*, 138128. doi: https://doi.org/10.1016/j.jclepro.2023.138128

Suyanto, S., Wibowo, A. T., Al Faraby, S., Saadah, S., & Rismala, R. (2021). Evolutionary rao algorithm. *Journal of Computational Science, 53*, 101368. doi: https://doi.org/10.1016/j.jocs.2021.101368

Tandel, P., & Patel, M. N. (2016). A survey on feature extraction techniques for shape based object recognition. *International Journal of Computer Applications, 137*(6), 16-20. doi: https://doi.org/10.5120/ijca2016908782

Thanki, R., Kothari, A., & Borra, S. (2021). Fragile watermarking framework for tamper detection of color biometric images. *International Journal of Digital Crime and Forensics, 13*(2), 35-56. doi: https://doi.org/10.4018/IJDCF.2021030103

Triayudi, A., Suparman, K., & Andrianingsih, A. (2022). Rancang Bangun Marketplace pada UMKM Terimbas Pandemi Covid-19 Menggunakan Metodologi Pengembangan Waterfall dan Metode FIFO. *Journal of Information and Communication Technology, 6*(1), 7-19. doi: https://doi.org/10.35870/jtik.v6i1.384

Veronica, B. K. (2020). An effective neural network model for lung nodule detection in CT images with optimal fuzzy model. *Multimedia Tools and Applications, 79*(19), 14291-14311. doi: https://doi.org/10.1007/s11042-020-08618-x

Vol: 15 **No:** 1 **Year:** 2023 **ISSN:** 2146-0744 (Online) (**pp.** 512-532) **DOI:** 10.34111/ijebeg.2023150123

Woods, M., Gonzalez-Juarrero, M., Markman, D. W., Antolin, M. F., Bowen, R. A., Wheat, W. H., & Jackson, M. (2018). Yersinia pestis survival and replication in potential ameba reservoir. *Emerging Infectious Diseases*, *24*(2), 294-302. doi: https://doi.org/10.3201/eid2402.171065

Yanto, B. E., Subali, B., & Suyanto, S. (2019). Improving Students' Scientific Reasoning Skills through the Three Levels of Inquiry. *International Journal of Instruction*, *12*(4), 689-704. doi: https://doi.org/10.29333/iji.2019.12444a

Yuni, D. N., Ekpa, D., & Oyekale, A. S. (2017). Poverty decomposition for high and low users of climate smart agricultural techniques in Northwest Nigeria. *International Journal of Environment and Pollution*, *5*(2), 26-41. Retrieved from https://eajournals.org/ijepr/vol-5-issue-2-may-2017/poverty-decomposition-high-low-users-climate-smart-agricultural-techniques-northwest-nigeria

Zubair, M., & Ullah, A. (2020). Recent advances in protein derived bionanocomposites for food packaging applications. *Critical Reviews in Food Science and Nutrition*, *60*(3), 406-434. doi: https://doi.org/10.1080/10408398.2018.1534800

Zwing, N., Failmezger, H., Hibar, D. P., Cañamero, M., & Korski, K. (2020). Analysis of spatial organization of suppressive myeloid cells and effector T cells in colorectal cancer—a potential tool for discovering prognostic biomarkers in clinical research. *Frontiers in Immunology*, *11*, 550250. doi: https://doi.org/10.3389/fimmu.2020.550250