INTERNATIONAL JOURNAL OF APPLIED ARTIFICIAL INTELLIGENCE AND ROBOTICS
E-ISSN: 3124-1212, DOI: 10.67745/ijaic.v2i2.26
Vol. 2 No. 2 2026 pp. 42-54

Facial Skin Color Classification Using Hybrid Features With The K-Nearest Neighbor (KNN) Method

Nurul Agustina[1], Radhiyatammardhiyyah[2*], Cut Aulia Safira Rahman[3]

[1,2] Information Technology and Computer Departement, Politeknik Negeri Lhokseumawe Lhokseumawe, Aceh, Indonesia
[3] Sekolah Vokasi Universitas Gadjah Mada, Yogyakarta, Indonesia

*Corresponding Author: radhiyatammardhiyyah @pnl.ac.id

Article info: Received 20/09/2026, Revised 10/10/2026, Accepted 17/10/2026

This is an open access article under the CC BY-SA license.

Abstract

Facial skin color classification can still be affected by differences in lighting, cameras, and image conditions, causing visual assessment to potentially produce inconsistent results. This study aims to develop a facial skin color classification system using a combination of color and texture features and the K-Nearest Neighbor (KNN) algorithm. The research data consist of primary and secondary datasets. The research stages include data collection, preprocessing, color feature extraction using Color Moments in the YCbCr color space, texture feature extraction using Gray Level Co-occurrence Matrix (GLCM), hybrid feature construction, KNN classification, and performance evaluation. The color and texture features are combined into a hybrid feature representation, which is then used as the input for KNN. K-value testing was conducted using several configurations to determine the optimal parameter. The results show that K = 3 achieved the highest performance in both dataset scenarios. The primary dataset achieved an accuracy of 90.61%, while the secondary dataset achieved an accuracy of 87.43%. Class-wise evaluation indicates that the model can recognize all three skin color categories with good performance. The results demonstrate that combining color and texture information can be an effective approach for facial skin color classification.

Keywords: facial skin color, hybrid features, Color Moments, YCbCr, GLCM, KNN

1. Introduction

The development of image processing technology has enabled facial skin color characteristics to be analyzed automatically. Kye and Lee demonstrated that skin color classification can be performed through clustering in several color spaces, including YCbCr, indicating that the selection of color representation is an important aspect of a classification system [1]. This automated approach is relevant because the color captured by a camera is influenced not only by skin characteristics but also by lighting conditions and image acquisition devices. Tadesse et al. developed STAR-ED to analyze skin tone using machine learning and emphasized the importance of more diverse skin color representations in data-driven systems [2]. Rejón Piña and Ma subsequently developed CASCo as an automated skin color measurement tool for social research, demonstrating that image-based measurements can be systematically applied when data collection and grouping processes are conducted in a controlled manner [3]. Both studies demonstrate that skin color classification can be performed computationally; however, the need for feature representations that appropriately capture the characteristics of the images remains an important issue.

In the image processing stage, the YCbCr color space is widely used to separate luminance and chrominance information. Apridiansyah et al. applied YCbCr for skin color segmentation in face detection and demonstrated that this representation can help separate skin color characteristics from images [4]. Jung et al. also demonstrated that skin tone mapping using optical and deep learning approaches can produce a more structured representation of skin color. [5]. Hamrani et al., through AIDA, developed an approach for skin color classification and segmentation, highlighting the importance of image processing and managing color variations in automated systems

[6]. Yu et al. used machine learning to classify skin phototypes based on optical measurements. The results indicate that skin tone characteristics can be learned through numerical features representing skin conditions. [7]. Ulrich et al. developed an artificial intelligence-based approach for skin tone analysis and demonstrated that automated methods can be used to classify skin color more consistently [8]. Nevertheless, approaches that rely on optical measurements or deep learning models may require greater hardware resources, larger datasets, and higher computational capacity compared with classification methods based on measurable features and simple classifiers.

The study most closely related to this research was conducted by Syahputra et al., who used Gray Level Co-occurrence Matrix (GLCM) and K-Nearest Neighbor (KNN) for skin tone classification in foundation shade matching. The study achieved an accuracy of 75.65%; however, its primary focus was on GLCM texture features, so color distribution information was not explicitly combined with statistical color features [9]. Putri et al. used KNN to predict color selection based on skin tone and demonstrated that skin color characteristics can be used as a basis for decision-making in recommendation applications [10]. In addition to texture features, Color Moments can summarize the color distribution of an image using statistical measures such as mean, standard deviation, and skewness. The use of statistical color features provides a more compact representation than using all pixel values [11]. Winarno et al. demonstrated that combining multiple feature groups can provide a more informative representation for classification, achieving an accuracy of 99.76% using hybrid features and an *Artificial Neural Network* (ANN) [12]. Nisa et al. also combined GLCM and Color Moments for image classification and achieved the best average accuracy of 78.87% using Support Vector Machine (SVM), while Putra et al. utilized GLCM and color features for machine learning-based classification and achieved the highest accuracy of 96% when color and texture features were combined using SVM [13], [14]. The findings of these studies provide a basis for concluding that color and texture information can complement each other.

From an evaluation perspective, Cabot and Ross explained that classification model assessment should use appropriate metrics so that model performance is not evaluated based on a single measure alone [15]. Vanacore et al. discussed classifier evaluation using a confusion matrix, while Miller et al. emphasized the importance of accuracy, precision, recall, and F1-score in evaluating model performance [16], [17]. Therefore, the evaluation in this study does not rely solely on accuracy but also includes class-wise metrics to examine the patterns of classification errors. Based on previous studies, there is still room to develop a facial skin color classification system that combines color and texture information, uses relatively simple features, and is evaluated using more than one dataset source. This study proposes a hybrid feature representation consisting of Color Moments in the YCbCr color space and GLCM texture features, followed by classification using KNN. The study also compares the performance of the primary and secondary datasets to determine the influence of data characteristics on model performance. Thus, this study focuses not only on achieving accuracy but also on analyzing the contribution of the features, the K value, and differences in the characteristics of the data sources.

2. Methods

The dataset is an important component in model development and evaluation because the characteristics of the data sources can affect feature distributions and classification results. In this study, two data sources were used so that the model was not evaluated solely on images obtained through the same acquisition procedure. Comparing the two data sources also provides an overview of the method's robustness to variations in image characteristics. Before classification, the images were divided into training and testing data using an 80:20 ratio. This division was applied to each dataset scenario so that the evaluation results could be compared consistently. The primary dataset consisted of 450 images, with 150 images for each class: Kuning Langsat, Sawo Matang, and Coklat Gelap. The secondary dataset consisted of 927 images grouped into White, Olive, and Dark Brown. The primary dataset was validated to ensure the accuracy of the class labels before being used as research data.

Because color and texture features have different value ranges, standardization is required before distance measurement. Scaling techniques can affect classifier performance; therefore, the standardization parameters are calculated from the training data and applied consistently to the testing data [18]. After the features were prepared, classification was performed using KNN. The selection of the distance metric and K configuration needs to be evaluated because changes in both can affect the labels assigned based on the nearest neighbors. Previous studies have demonstrated the use of Euclidean Distance in KNN [19], in KNN for classification with feature evaluation [20], as well as the use of KNN in recommendation systems [21].

2.1 Research Model

The research model divides the classification process into a training phase and a testing phase. In the training phase, the data undergo preprocessing, color and texture feature extraction, and hybrid feature construction, after which the KNN model is stored. In the testing phase, the images are processed through the same stages, and the resulting hybrid features are used as input to the KNN model to determine the skin color class.

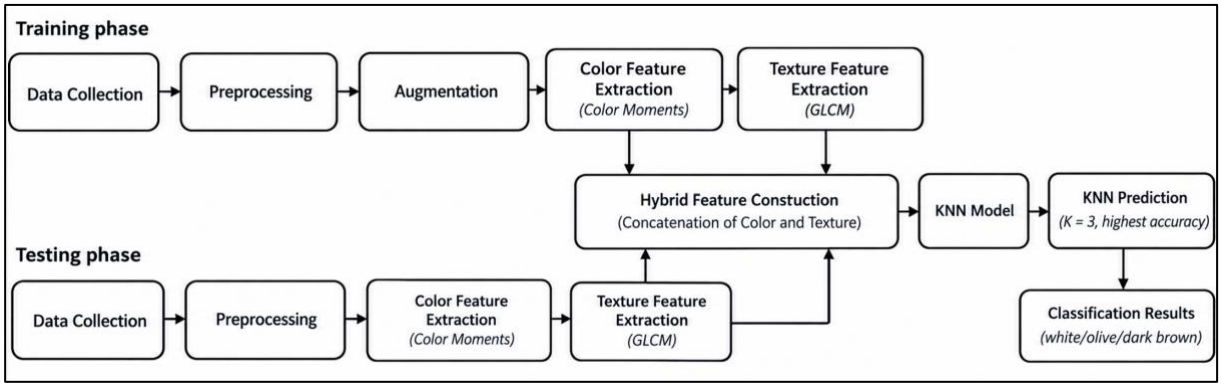


Figure 1. Research Model

During the testing phase, the classification results are mapped into three skin color categories, namely Kuning Langsat, Sawo Matang, and Coklat Gelap. This workflow serves as the basis for system implementation and model performance evaluation. Fig. 1 workflow of the proposed research methodology.

2.2 Dataset

The research data consisted of primary and secondary datasets. The use of two data sources was intended to evaluate the model's performance on images with different characteristics.

a. Primary Dataset

The primary dataset was obtained by capturing facial images of students from the Informatics Engineering Study Program using a Canon EOS 250D camera with a Canon EF-S 18–55 mm lens. Supporting images for certain categories were also obtained from Pinterest. Before being used, the primary dataset was validated by Dr. Risa Fitria to ensure the classification of skin colors into three classes. Fig. 2 presents examples of images from the primary dataset used in this study, while Table 1 shows the distribution of images across the respective classes.



(a) (b) (C)

Figure 2. Examples of the Primary Dataset: (a) Kuning Langsat Class, (b) Sawo Matang Class, and (c) Coklat Gelap Class

Table 1. Primary Dataset Distribution

Dataset	Class	Number
Primary	Kuning Langsat	150
Primary	Sawo Matang	150
Primary	Coklat Gelap	150
Primary	Total	450

b. Secondary Dataset

The secondary dataset was obtained from the Skin Tone Classification Dataset available on the Kaggle platform. The dataset had been categorized into three skin color categories and was used as comparative data in this study. Fig. 3 presents examples of images from the primary dataset used in this study, while Table 2 shows the distribution of images across the respective classes.



(a) (b) (c)

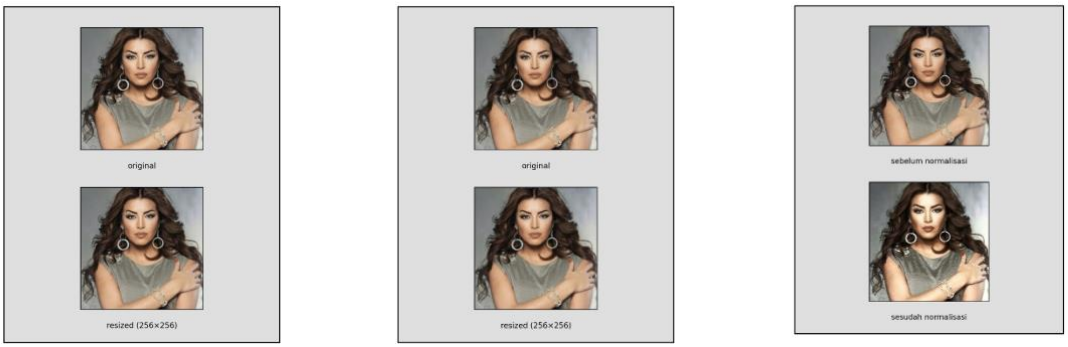
Figure 3. Examples of the Secondary Dataset: (a) White Class, (b) Olive Class, and (c) Dark Brown Class

Tabel 2. Secondary Dataset Distribution

Dataset	Class	Number
Secondary	White	325
Secondary	Olive	355
Secondary	Dark Brown	247
Secondary	Total	927

Preprocessing aims to standardize the input and focus the analysis on the relevant facial region. This stage includes resizing, determining the Region of Interest (ROI), applying Gaussian Blur, and normalization. Recent reviews of facial image preprocessing indicate that size standardization, image quality enhancement, and noise reduction are commonly applied before the feature extraction process. [22]. Image processing approaches have also been applied in image-based skin classification studies to obtain regions that are more relevant to the object of analysis [23], [24]. Dataset characteristics are an important consideration because images obtained from different sources may vary in lighting, background, pose, camera quality, and color composition. These differences may result in different feature distributions. Therefore, preprocessing in this study is not only considered a technical stage but also serves to ensure that the information passed to the feature extraction stage remains focused on the face. Facial image preprocessing approaches generally include size standardization, noise reduction, and focusing on the object region being analyzed.

The primary dataset consisted of 450 images with an equal number of images across the three classes, while the secondary dataset consisted of 927 images distributed into the White, Olive, and Dark Brown classes. The primary dataset was first validated to ensure the appropriateness of the skin color labels, while the secondary dataset was used as a comparison based on a publicly available source. Both datasets were divided using an 80:20 ratio through a stratified split to maintain the class proportions in the training and testing data. This design enables the model performance to be evaluated across two different data source characteristics.



(a) (b) (c)

Figure 4. Examples of Image Preprocessing Results : (a) Resizing, (b) ROI, and (c) Normalization

2.3 Image Preprocessing

Preprocessing is performed to standardize the input and reduce the effects of image variations before feature extraction. Recent literature on facial image preprocessing identifies normalization, noise reduction, illumination adjustment, and augmentation as important steps for improving image consistency. In this study, resizing, Region of Interest (ROI) determination, Gaussian Blur, and normalization were performed. Resizing standardizes the image dimensions; ROI focuses the analysis on the facial region; Gaussian Blur reduces noise; while normalization helps equalize the range of pixel values [25]. Preprocessing produces images that are more focused on the object of interest, so that the extracted features are less influenced by non-skin regions. This stage is also consistent with the use of ROI approaches in skin and facial image analysis. Color feature extraction aims to transform color information in the images into numerical attributes that can be processed by the classifier. In this study, the YCbCr color space was used to separate luminance and chrominance components, and each channel was summarized using Color Moments. Three statistical measures were used: mean to describe the average value, standard deviation to describe the spread of the values, and skewness to describe the asymmetry of the distribution. The use of statistical color features as a compact representation is consistent with the application of Color Moments in image classification studies. Fig. 4. show Examples of the image preprocessing results, illustrating the transformations applied to the input images before feature extraction and classification.

2.4 Color Feature Extraction

Color features are used to represent the color distribution characteristics of facial images. YCbCr was selected because it separates luminance in the Y channel from chrominance in the Cr and Cb channels [26]. Three Color Moments, namely mean, standard deviation, and skewness, were calculated for each channel. Color Moments are statistical measures that summarize color distributions and can therefore be used as a feature vector. The mean was calculated using Equation (1), the standard deviation using Equation (2), and the skewness using Equation (3). With N representing the number of pixels and x_i representing the value of the i-th pixel, these three statistical measures were calculated for the Y, Cr, and Cb channels [13].

$$\mu = (1/N)\ \Sigma_{i=1}{}^{N}\ x_i \tag{1}$$

$$\sigma = \sqrt{[(1/N)\ \Sigma_{i=1}{}^{N}\ (x_i - \mu)^2]} \tag{2}$$

$$\gamma = \sqrt[3]{[(1/N)\ \Sigma_{i=1}{}^{N}\ (x_i - \mu)^3]} \tag{3}$$

Table 3. Examples of Color Feature Extraction Results for Each Class

Class	Feature	Value
Kuning Langsat	Y Mean	141,5838
	Y Std	47,6512
	Y Skew	−0,6682
	Cr Mean	144,6036
	Cr Std	7,7620
	Cr Skew	−0,0940
	Cb Mean	116,4071
	Cb Std	5,8782
	Cb Skew	0,9473
Sawo Matang	Y Mean	131,6843
	Y Std	54,5324
	Y Skew	−0,3612
	Cr Mean	143,9040
	Cr Std	10,4764
	Cr Skew	−0,1698
	Cb Mean	115,2166
	Cb Std	9,4499
	Cb Skew	0,5787
Coklat Gelap	Y Mean	65,1488
	Y Std	51,3386
	Y Skew	0,5543
	Cr Mean	135,3968
	Cr Std	7,5107
	Cr Skew	1,3410
	Cb Mean	127,5548
	Cb Std	5,3756
	Cb Skew	1,5771

These results show that each image produces nine color features. The differences in statistical values among the

classes indicate that the color distributions in the Y, Cr, and Cb channels have characteristics that can be used to distinguish between classes. GLCM is used to represent the spatial relationships between pixels based on their grayscale levels. The resulting matrix is then summarized into contrast, correlation, energy, and homogeneity. These four measures provide different information regarding local variation, pixel relationship regularity, distribution concentration, and texture uniformity. The use of GLCM in combination with color features has been applied in several image classification studies, making this approach relevant as a complement to color information. Color Moments were selected because these features summarize the distribution of pixel values using first-, second-, and third-order statistics. The mean provides an overview of the average channel value, the standard deviation describes the dispersion of values around the mean, while skewness describes the asymmetry of the distribution. These three statistical measures were calculated for the Y, Cr, and Cb channels, resulting in nine color attributes. In the context of this study, these nine attributes were used to capture skin tone and color distribution information within the classification feature space.

The extraction results show differences in characteristics among the classes. In the Kuning Langsat class, the mean value of the Y channel is higher than that of the Coklat Gelap class, while the skewness values in several channels also show different patterns. These differences are not used as a single classification rule but serve as numerical information that is learned together with the texture features. Thus, Color Moments function as a statistical summary of color information rather than as a direct class determinant. Table 3. show the examples of color feature extraction results for each class.

2.5 Texture Feature Extraction

GLCM-based studies on skin images have shown that contrast, correlation, energy, and homogeneity can be used as texture features. The combination of texture and color features has also been applied in image classification. In this study, the images were converted to grayscale, and the GLCM was constructed using a distance of one pixel and four orientations, namely 0°, 45°, 90°, and 135°. Four parameters were calculated for each orientation, resulting in 16 texture features [26]. Table 4 presents examples of the GLCM-based texture feature extraction results obtained from the preprocessed images.

$$\text{Contrast} = \Sigma_i \Sigma_j \, (i - j)^2 \, P(i,j) \tag{4}$$

$$\text{Correlation} = \Sigma_i \Sigma_j \, [(i-\mu_i)(j-\mu_j)P(i,j)]/(\sigma_i \sigma_j) \tag{5}$$

$$\text{Energy} = \Sigma_i \Sigma_j \, P(i,j)^2 \tag{6}$$

$$\text{Homogeneity} = \Sigma_i \Sigma_j \, P(i,j)/(1 + |i - j|) \tag{7}$$

Table 4. Texture Feature Extraction Results

Angle	Distance	Contrast	Correlation	Energy	Homogeneity
0°	1	0,883	0,979	0,030	0,805
45°	1	1,665	0,960	0,024	0,743
90°	1	1,050	0,975	0,029	0,791
135°	1	1,618	0,961	0,024	0,743

Hybrid feature construction was performed after the color and texture features had been obtained separately. The color features produced nine attributes, while GLCM produced 16 attributes from four orientations. The two feature groups were then horizontally concatenated, resulting in a representation of 25 attributes for each image. The combination of multiple feature groups was intended to preserve complementary information before the classification process and is consistent with the hybrid feature approach used in image classification studies. In texture extraction, the preprocessed images were converted to grayscale so that the relationships between pixels could be analyzed based on grayscale levels. The GLCM was constructed using a distance of one pixel and four orientations, namely 0°, 45°, 90°, and 135°. The use of multiple orientations was intended to ensure that spatial relationship information was not dependent on a single direction. Contrast, correlation, energy, and homogeneity were calculated from each matrix, resulting in 16 texture features.

The example results show that the contrast values vary across orientations, while the correlation values are relatively high in all directions. Energy and homogeneity provide information regarding the regularity and uniformity of patterns in the analyzed region. These results indicate that GLCM provides information that differs from Color Moments. While color features primarily summarize intensity distributions across color channels, GLCM features describe local spatial relationships. This difference in the nature of the information provides the basis for combining the two feature groups.

2.6 Hybrid Feature Construction

The nine color features and 16 texture features were combined to form 25 hybrid features. The resulting features were normalized using StandardScaler. [27]. KNN was then applied using the Euclidean distance metric. The K values tested were 3, 5, 7, and 9. The model determined the class based on the majority label among the nearest neighbors.

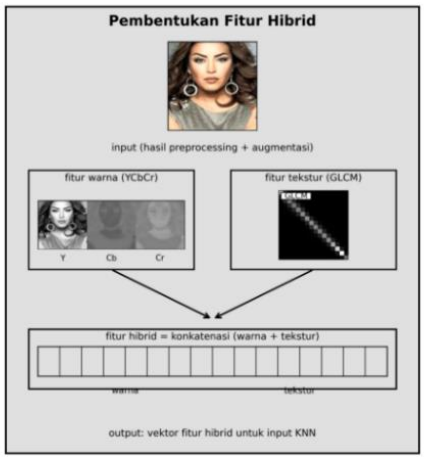


Figure 5. Visualization of Hybrid Feature Construction

$$\text{F_hybrid} = [\text{F_color} \mid \text{F_texture}] \tag{8}$$

Table 5. Hybrid Feature Composition

Feature Type	Method	Number
Color Features	Color Moments–YCbCr	9
Texture Features	GLCM	16
Total Hybrid Features	Concatenation	25

The hybrid feature construction resulted in 25 numerical attributes, consisting of nine color features and 16 texture features. This stage ends with the formation of the feature representation, while KNN classification is described separately in the following subsection. Before calculating the distance, all feature attributes were standardized using StandardScaler so that features with different scales would not dominate the calculation. The selection of a scaling technique can affect classifier performance; therefore, the standardization process needs to be applied consistently to the training and testing data. After standardization, KNN determines the class based on the proximity of the test data to the training data. Euclidean Distance was used as the distance measure because it is one of the commonly applied distance measures in KNN. The K value was determined by comparing several configurations, while the use of KNN for classification and recommendation tasks has also been applied in previous studies.

Hybrid feature construction is the stage of combining feature representations, not the classification stage. The nine color features and 16 texture features are concatenated sequentially into a single vector containing 25 attributes. Mathematically, the combination can be expressed as $F_{\text{hybrid}} = [F_{\text{color}} \mid F_{\text{texture}}]$. This process preserves all attributes obtained from the two feature extraction methods, allowing color and texture information to be available simultaneously during the classification stage. Fig 5. show visualization of hybrid feature construction visualization of hybrid feature construction and Table 5 show the hybrid feature composition.

This composition provides a clear division between the sources of information in the final vector. The first nine attributes are derived from Color Moments in the YCbCr color space, while the following 16 attributes are derived from GLCM. Once the vector is formed, standardization is then performed as a preparation for distance measurement. This separation of stages is important so that the formation of the feature representation can be analyzed independently before the class decision is determined by KNN.

2.7 K-Nearest Neighbor Classification

KNN is a proximity-based classification method that determines the class of test data based on the majority label among a specified number of nearest neighbors in the training data. In this study, the hybrid features were first normalized using StandardScaler so that differences in feature scales would not dominate the distance calculation. Studies on feature scaling have shown that the choice of normalization technique can affect classification performance [18].

$$z = (x - \mu_train) / \sigma_train \quad (9)$$

The distance between data points was calculated using Euclidean Distance. The use of Euclidean Distance in KNN has also been applied in KNN-based classification studies involving distance comparisons [9].

$$d(X,Y) = \sqrt{[\Sigma_{i=1}^{n} (X_i - Y_i)^2]} \quad (10)$$

The K values tested were 3, 5, 7, and 9. After the distances were sorted from the smallest to the largest, the labels of the K nearest neighbors were counted, and the class with the highest number of votes was selected as the prediction. KNN has also been used in recommendation-based and image classification studies. Recent studies have discussed facial image preprocessing involving normalization, noise reduction, and image standardization. ROI approaches and color analysis in images have also been applied in machine learning-based studies. Research on skin type classification using image processing and deep learning has shown that image quality and feature representation can affect classification results. Random Forest-based research on skin tone and color recommendation has also demonstrated the use of automated classification for facial images [28].

KNN determines the class of test data based on its proximity to the training data. Since the hybrid vector contains attributes with different scales, StandardScaler was used, with the mean and standard deviation parameters calculated from the training data. This transformation is presented in Equation (9) and aims to prevent attributes with larger value ranges from dominating the distance calculation. After standardization, the Euclidean distance was calculated between the test data and training data. A smaller distance indicates greater proximity in the feature space.

Four K configurations were tested, namely 3, 5, 7, and 9. For each configuration, the distances were sorted in ascending order, and the labels of the K nearest neighbors were used in the majority voting process. The best K value was determined based on the accuracy obtained on the test data. Testing multiple K values is necessary because the number of neighbors can influence the model's decision: smaller K values place greater emphasis on the nearest neighbors, whereas larger K values involve more data points in the decision-making process.

2.8 Testing Scenarios and Parameters

The evaluation was conducted under two scenarios, namely the primary dataset and the secondary dataset. Each scenario used the same training and testing data split so that differences in the results primarily reflected the characteristics of the datasets. The K values tested were 3, 5, 7, and 9. Each K value was evaluated using accuracy, precision, recall, and F1-score so that the best configuration could be determined based on overall performance. In the classification process, the features from the training data were used as the basis for building the KNN model, while the test data were used to measure its generalization ability. StandardScaler was fitted using the training data and then applied to the test data to avoid differences in feature scaling between the two data groups. This evaluation procedure follows the principle of model evaluation, which separates the learning and testing processes to minimize measurement bias. Table 6 present the research parameters.

Table 6. Research Parameters

Parameter	Value
Classification Method	K-Nearest Neighbor (KNN)
Distance Measurement Method	Euclidean Distance
K Values Tested	3, 5, 7, 9
Input Data	25 Hybrid Features
Output Data	Facial skin color classes

2.9 Model Evaluation

The model was evaluated using accuracy, precision, recall, and F1-score. Accuracy was used to measure the proportion of correct predictions among all test data. Precision indicates the correctness of predictions for a particular class, recall indicates the model's ability to identify data that truly belong to a particular class, while F1-score provides a combined measure of precision and recall. Per-class evaluation was performed so that model performance was not assessed solely based on overall accuracy [9] [17].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (11)$$

$$Precision = \frac{TP}{TP+FP} \quad (12)$$

$$Recall = \frac{TP}{TP+FN} \quad (13)$$

$$F1 = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (14)$$

Testing was conducted for each K value across the two dataset scenarios. The best K value was determined based on the highest accuracy and was subsequently used for per-class performance evaluation and comparison of performance between the datasets.

3. Result and Discussions

3.1 K Value Testing Results

K value testing was conducted to determine the effect of the number of neighbors on classification performance. On the secondary dataset, K = 3 achieved an accuracy of 87.43%, while K = 5, 7, and 9 achieved accuracies of 86.10%, 84.19%, and 82.48%, respectively. On the primary dataset, K = 3 achieved an accuracy of 90.61%, while K = 5, 7, and 9 achieved accuracies of 89.20%, 85.92%, and 83.57%, respectively. Table 7 present the testing results on the secondary dataset based on different K values.

Table 7. Testing Results on the Secondary Dataset Based on Different K Values

K	Secondary Accuracy	Precision	Recall
3	87,43%	88%	87%
5	86,10%	88%	86%
7	84,19%	86%	84%
9	82,48%	84%	82%

K = 3 achieved the highest accuracy on both datasets. The decrease in performance as the K value increased indicates that, given the characteristics of the data in this study, using the three nearest neighbors provided more appropriate classification decisions than using a larger number of neighbors. Therefore, K = 3 was subsequently used as the optimal parameter for evaluating both scenarios. Table 8 present the testing results on the primary dataset based on different K values.

Table 8. Testing Results on the Primary Dataset Based on Different K Values

K	Primary Accuracy	Precision	Recall
3	90,61%	91%	91%
5	89,20%	89%	89%
7	85,92%	86%	86%
9	83,57%	85%	83%

3.2 Accuracy Comparasion Between Dataset Scenarios

The primary dataset achieved an accuracy of 90.61%, which was higher than the secondary dataset at 87.43%. This difference may be associated with variations in data source characteristics, image quality, lighting conditions, and object diversity. These results indicate that the same model can produce different performance when the characteristics of the data distribution change. Confusion matrix analysis was used to identify the classes that were most frequently classified correctly and those that were still misclassified. In the three-class classification, the matrix shows the number of predictions for each combination of actual and predicted labels. Per-class analysis is necessary because overall accuracy may conceal differences in performance among classes. Table 9 present performance comparison based on dataset scenarios.

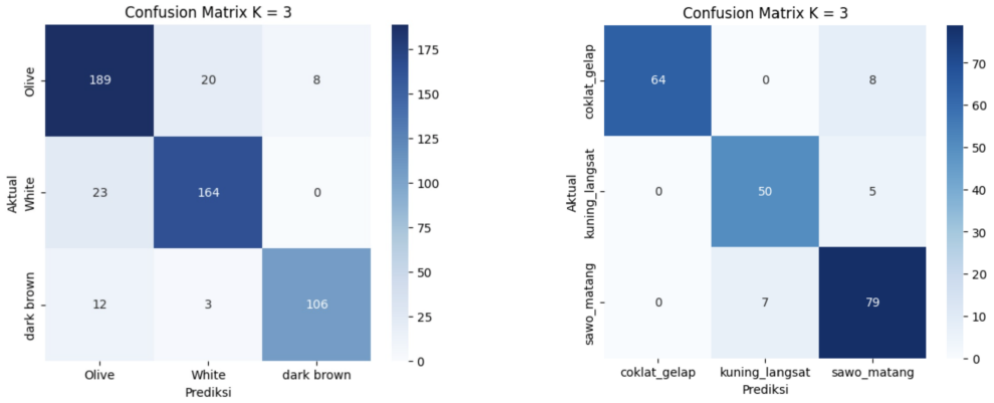
Table 9. Performance Comparison Based on Dataset Scenarios

Dataset	Best K	Accuracy	Precision	Recall
Secondary	3	87,43%	88%	87%
Primary	3	90,61%	91%	91%

3.3 Confusion Matrix and Per-Class Evaluation

The confusion matrix was used to examine the relationship between actual and predicted labels. From the matrix,

accuracy, precision, recall, and F1-score can be calculated. Classification evaluation literature emphasizes that a confusion matrix helps identify errors in each class rather than evaluating overall performance alone. Per-class evaluation was conducted using the KNN model with K = 3. On the secondary dataset, the highest precision was achieved by the Dark Brown class at 93%, while the Olive class had a precision of 84%. On the primary dataset, the Coklat Gelap class achieved a precision of 100%, while the highest recall was obtained by the Sawo Matang class at 92%.



(a) (b)

Figure 6. Confusion Matrix: (a) Secondary Dataset and (b) Primary Dataset

Table 10. Per-Class Evaluation on Both Datasets

Dataset	Class	Precision	Recall	F1-Score
Secondary	Olive	84%	87%	86%
	White	88%	88%	88%
	Dark Brown	93%	88%	90%
	Rata-rata	88%	87%	87%
Primary	Coklat Gelap	100%	89%	94%
	Kuning Langsat	88%	91%	89%
	Sawo Matang	86%	92%	89%
	Rata-rata	91%	91%	91%

The evaluation results show that the model has a relatively good ability to recognize each class. In the primary dataset, a precision of 100% for the Coklat Gelap class indicates that all images predicted as this class corresponded to the actual labels. Meanwhile, a recall of 92% for the Sawo Matang class indicates that most images belonging to this class were correctly recognized by the model. In the secondary dataset, the highest F1-score was achieved by the Dark Brown class at 90%. Fig. 6. Displays the confusion matrices of the classification model on the primary and secondary datasets. Table 10 presents the per-class performance of the classification model on both datasets, including precision, recall, and F1-score.

3.4 Comparison With Previous Studies

In previous studies, the use of GLCM and KNN achieved an accuracy of 75.65%. This study achieved accuracies of 87.43% on the secondary dataset and 90.61% on the primary dataset. These differences indicate that the addition of color features through Color Moments in the YCbCr color space can provide additional information that complements the texture features obtained from GLCM. However, this comparison should be interpreted while considering differences in experimental settings and data characteristics.

Table 11. Comparison of Results with Previous Studies

Study	Method	Features	Class	Accuracy
Syahputra dkk. (2025)	KNN	GLCM	3	75,65%
This study - Secondary	KNN	Color Moments (YCbCr)+GLCM	3	87,43%
This study - Primary	KNN	Color Moments (YCbCr)+GLCM	3	90,61%

3.5 Discussion

The results show that the combination of color and texture features can provide a sufficiently informative

representation for distinguishing three skin color categories. The use of YCbCr helps separate luminance and chrominance information, while Color Moments summarize the statistical distribution across the three channels. These findings are consistent with studies that use YCbCr for skin analysis, as well as studies that employ color statistics as features. GLCM provides additional information regarding local patterns through contrast, correlation, energy, and homogeneity. GLCM-based approaches for skin images have also been employed in previous studies, while the combination of color and texture features has demonstrated benefits in image classification. Therefore, the hybrid features used in this study do not rely solely on color intensity but also consider the local structure of the image. The higher performance of the primary dataset compared with the secondary dataset indicates that data quality and distribution affect classification results. Lighting conditions, camera variations, pose, background, and class composition may lead to differences in feature distributions. Studies on facial image preprocessing have also emphasized that normalization, noise reduction, and image standardization are important for maintaining input quality.

The results of this study also differ in approach from studies that use CNNs for skin tone classification and color recommendation. Deep learning approaches can learn representations directly from images but generally require larger datasets and greater computational resources. CNN-based studies on skin tone classification have demonstrated the potential of this approach for color recommendation. In contrast, this study employs hybrid features and KNN so that the relationship between feature characteristics and classification results can be analyzed more directly. Overall, K = 3 was the best-performing configuration for both datasets. These results indicate that a simple model can achieve good performance when the feature representation contains relevant information. This finding is consistent with previous KNN studies showing that the selection of the K value and distance metric can affect classification performance, while the use of KNN for recommendation based on object characteristics has also been applied in other studies. In terms of feature characteristics, Color Moments provide a statistical summary of the three YCbCr channels, while GLCM provides information about texture patterns. The combination of these two approaches ensures that the image representation does not rely solely on intensity or chrominance but also considers local relationships between pixels. This approach is consistent with studies demonstrating the benefits of combining color and texture features for image classification. The difference in performance between the primary and secondary datasets should also be understood as part of the research findings, rather than merely as a difference in accuracy values. Datasets with different acquisition conditions, camera quality, lighting conditions, and subject characteristics may produce different feature distributions.

Therefore, testing on two data sources provides additional information regarding the consistency of the method and helps identify the conditions under which the model performs classification more easily or with greater difficulty. From an implementation perspective, the use of KNN offers the advantage that classification can be performed based on feature similarity without requiring a complex model training process. However, KNN remains sensitive to feature scaling and data distribution. Therefore, feature standardization and K value selection are important components of the experimental design. This is consistent with studies showing that scaling can affect classification performance and that different distance configurations can produce different results in KNN. Accuracy represents the proportion of correct predictions among all data, while precision indicates the correctness of predictions when the model assigns a particular class label. Recall represents the model's ability to identify data that actually belong to a given class, while F1-score is a combined measure of precision and recall. The use of these multiple evaluation metrics provides a more comprehensive view of model performance than using accuracy alone.

4. Conclusion

This study developed a facial skin color classification system using hybrid features consisting of nine color features extracted using Color Moments in the YCbCr color space and 16 texture features extracted using GLCM, resulting in a total of 25 features. These features were used as inputs to the KNN method with Euclidean Distance. Testing with K values of 3, 5, 7, and 9 showed that K = 3 provided the best performance in both dataset scenarios. The primary dataset achieved an accuracy of 90.61%, with average precision, recall, and F1-score values of 91%, while the secondary dataset achieved an accuracy of 87.43%, with precision of 88%, recall of 87%, and F1-score of 87%. The differences in results between the datasets indicate that data characteristics affect classification performance. The primary dataset achieved higher performance, while the secondary dataset had greater data variation. Future studies may expand the number and diversity of images, compare KNN with other methods such as SVM or Random Forest, and evaluate additional features such as LBP or HOG.

5. Acknowledgments

The authors would like to express their gratitude to the academic supervisors, examiners, and the Informatics Engineering Study Program, Department of Information Technology and Computer, Politeknik Negeri Lhokseumawe, for their guidance and support throughout this research.

References

[1] S. Kye and O. Lee, "Skin color classification of Koreans using clustering," *Skin Research and Technology*, vol.

28, no. 6, pp. 796–803, 2022, doi: 10.1111/srt.13201.
[2] G. A. Tadesse *et al.*, "Skin Tone Analysis for Representation in Educational Materials (STAR-ED) using machine learning," *NPJ Digit. Med.*, vol. 6, p. 151, 2023, doi: 10.1038/s41746-023-00881-0.
[3] R. A. Rejón Piña and C. Ma, "Classification Algorithm for Skin Color (CASCo): A new tool to measure skin color in social science research," *Soc. Sci. Q.*, vol. 104, no. 2, pp. 168–179, 2023, doi: 10.1111/ssqu.13242.
[4] Y. Apridiansyah and E. D. Putra, "Segmentasi Warna Kulit Menggunakan Ruang Warna YCBCR Untuk Deteksi Wajah Manusia," *Jurnal Media Infotama*, vol. 19, no. 1, pp. 205–210, 2023, doi: 10.37676/jmi.v19i1.3808.
[5] G. Jung, S. Kim, and S. Yoo, "Skin Tone Analysis Through Skin Tone Map Generation With Optical Approach and Deep Learning," *Skin Research and Technology*, vol. 30, no. 10, p. e70088, 2024, doi: 10.1111/srt.70088.
[6] A. Hamrani *et al.*, "AI Dermatochroma Analytica (AIDA): Smart Technology for Robust Skin Color Classification and Segmentation," *Cosmetics*, vol. 11, no. 6, p. 218, 2024, doi: 10.3390/cosmetics11060218.
[7] X. Yu, K. G. Ong, and M. A. McGeehan, "Skin Phototype Classification with Machine Learning Based on Broadband Optical Measurements," *Sensors*, vol. 24, no. 22, p. 7397, 2024, doi: 10.3390/s24227397.
[8] P. Ulrich, A. Zink, T. Biedermann, and S. Sitaru, "Beyond Fitzpatrick: automated artificial intelligence-based skin tone analysis in dermatological patients," *NPJ Digit. Med.*, vol. 8, no. 1, p. 378, 2025, doi: 10.1038/s41746-025-01770-4.
[9] M. R. Syahputra *et al.*, "Accurate Skin Tone Classification for Foundation Shade Matching using GLCM Features-K-Nearest Neighbor Algorithm," *Jurnal Teknik Informatika (JUTIF)*, vol. 6, no. 5, pp. 3558–3571, 2025, doi: 10.52436/1.jutif.2025.6.5.4723.
[10] A. D. Putri and S. Adrianto, "Prediksi Pemilihan Warna Hijab Berdasarkan Tone Kulit Menggunakan Algoritma K-Nearest Neighbor (KNN)," *Jurnal Indonesia: Manajemen Informatika dan Komunikasi*, vol. 6, no. 3, pp. 1742–1755, 2025, doi: 10.63447/jimik.v6i3.1580.
[11] S.-Z. Shi, T. Sheng, Y.-Y. Wang, K.-K. Zhang, S. Liu, and H.-W. Gao, "Color-deconvolution-based feature image extraction and application in water quality analysis," *Analytical Methods*, vol. 14, pp. 4912–4920, 2022, doi: 10.1039/D2AY01600A.
[12] E. Winarno, A. Septiarini, W. Hadikurniawati, and H. Hamdani, "The Hybrid Features and Supervised Learning for Batik Pattern Classification," *Journal on Computing and Cultural Heritage*, vol. 17, no. 2, pp. 1–13, 2024, doi: 10.1145/3631131.
[13] I. Z. Nisa, S. N. Endah, P. S. Sasongko, R. Kusumaningrum, K. Khadijah, and R. Rismiyati, "Klasifikasi Citra Sampah Menggunakan Support Vector Machine dengan Ekstraksi Fitur Gray Level Co-Occurrence Matrix dan Color Moments," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 9, no. 5, pp. 921–930, 2022, doi: 10.25126/jtiik.2022954868.
[14] L. S. A. Putra, V. A. Gunawan, and A. S. Saragih, "Detection of Actinic Keratosis Skin Cancer Using Gray Level Co-occurrence Matrix Texture Extraction and Color Extraction With Support Vector Machine Classification," *TEKNIK*, vol. 44, no. 2, pp. 158–166, 2023, doi: 10.14710/teknik.v44i2.44895.
[15] J. H. Cabot and E. G. Ross, "Evaluating prediction model performance," *Surgery*, vol. 174, no. 3, pp. 723–726, 2023, doi: 10.1016/j.surg.2023.05.023.
[16] A. Vanacore, M. S. Pellegrino, and A. Ciardiello, "Fair evaluation of classifier predictive performance based on binary confusion matrix," *Comput. Stat.*, vol. 39, pp. 363–383, 2024, doi: 10.1007/s00180-022-01301-9.
[17] C. Miller, T. Portlock, D. M. Nyaga, and J. M. O'Sullivan, "A review of model evaluation metrics for machine learning in genetics and genomics," *Frontiers in Bioinformatics*, vol. 4, p. 1457619, 2024, doi: 10.3389/fbinf.2024.1457619.
[18] L. B. V de Amorim, G. D. C. Cavalcanti, and R. M. O. Cruz, "The choice of scaling technique matters for classification performance," *Appl. Soft Comput.*, vol. 133, p. 109924, 2023, doi: 10.1016/j.asoc.2022.109924.
[19] Salsabila, S. Martha, and W. Andani, "Komparasi Algoritma K-Nearest Neighbor dengan Euclidean Distance dan Manhattan Distance untuk Klasifikasi Stunting Balita," *BIMASTER*, vol. 13, no. 2, pp. 285–292, 2024, doi: 10.26418/bbimst.v13i2.77245.
[20] S. I. P. Sari, W. J. Pranoto, and N. A. Verdikha, "Analisis Pengaruh Gain Ratio Untuk Algoritma K-Nearest Neighbor Pada Klasifikasi Data Banjir di Kota Samarinda," *Jurnal Sains Komputer dan Teknologi Informasi*, vol. 6, no. 1, pp. 54–59, 2023, doi: 10.33084/jsakti.v6i1.5472.
[21] T. M. S. Mulyana, "Implementasi Algoritma K-Nearest Neighbor Untuk Rekomendasi Smartphone," *PETIR*, vol. 16, no. 1, 2023, doi: 10.33322/petir.v16i1.1707.
[22] R. Soekarta and K. R. Ku-Mahamud, "Recent Facial Image Preprocessing Techniques: A Review," *Engineering Proceedings*, vol. 84, no. 1, p. 39, 2025, doi: 10.3390/engproc2025084039.
[23] N. Nurapipah and S. S. Yuliana, "Skin Tone Classification in Digital Images Using CNN For Make-Up and Color Recommendation," *Journal of Intelligent Systems Technology and Informatics*, vol. 1, no. 3, pp. 78–85, 2025, doi: 10.64878/jistics.v1i3.29.
[24] S. Saiwaeo, S. Arwatchananukul, L. Mungmai, W. Preedalikit, and N. Aunsri, "Human skin type classification using image processing and deep learning approaches," *Heliyon*, vol. 9, no. 11, p. e21176, 2023, doi: 10.1016/j.heliyon.2023.e21176.
[25] A. Pandey, M. S. Teja, P. Sahare, V. Kamble, M. Parate, and M. F. Hashmi, "Skin cancer classification using

non-local means denoising and sparse dictionary learning based CNN," *Journal of Electrical Systems and Information Technology*, vol. 11, no. 36, 2024, doi: 10.1186/s43067-024-00162-0.

[26] S. Kapoor and A. Narayanan, "Leakage and the reproducibility crisis in machine-learning-based science," *Patterns*, vol. 4, no. 9, p. 100804, 2023, doi: 10.1016/j.patter.2023.100804.

[27] C. I. Purba, Alrizal, and Y. Fendriani, "Klasifikasi Kanker Kulit dari Citra Dermoskopi Menggunakan Fitur Gray Level Co-occurrence Matrix (GLCM) dengan Algoritma Machine Learning," *JFT: Jurnal Fisika dan Terapannya*, vol. 12, no. 1, pp. 30–44, 2025, doi: 10.24252/jft.v12i1.56651.

[28] S. Mustafa, A. Jaffar, M. W. Iqbal, A. Abubakar, A. S. Alshahrani, and A. Alghamdi, "Hybrid Color Texture Features Classification Through ANN for Melanoma," *Intelligent Automation & Soft Computing*, vol. 35, no. 2, pp. 2205–2218, 2023, doi: 10.32604/iasc.2023.029549.

[29] F.-T.-Z. Khanam, A. Al-Naji, A. G. Perera, D. Wang, and J. Chahl, "Non-invasive and non-contact automatic jaundice detection of infants based on random forest," *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.*, vol. 11, no. 6, pp. 2516–2529, 2023, doi: 10.1080/21681163.2023.2244601.