JKPK (JURNAL KIMIA DAN PENDIDIKAN KIMIA), Vol. 10, No. 2, 2025 pp. 309-322
Chemistry Education Study Program, Universitas Sebelas Maret ISSN 2503-4146
https://jurnal.uns.ac.id/jkpk ISSN 2503-4154 (online)

SMARTPHONE-BASED DIGITAL IMAGE ANALYSIS FOR QUALITATIVE CLASSIFICATION OF FOOD DYES USING MACHINE LEARNING: EFFECTS OF COLOR SPACE AND LIGHTING CONDITIONS

Yozef Tjandra[1], Giovania Evangeline Halim[2], Jansen Briano[2], and Martin Tjahjono[2]*

[1]Calvin Institute of Technology, Department of IT and Big Data Analytics, Jakarta, Indonesia
[2]Calvin Institute of Technology, Department of Chemical and Food Processing, Jakarta, Indonesia

ARTICLE INFO

Keywords:
Colorimetry;
Color Space;
Machine Learning;
Qualitative Analysis
Food dyes;
Smartphone.

Article History:
Received: 2025-08-02
Accepted: 2025-08-23
Published: 2025-08-31
doi:10.20961/jkpk.v10i2.107520

©2025 The Authors. This open-access article is distributed under a (CC-BY-SA License)

ABSTRACT

Smartphone-based digital image analysis (DIA) has emerged as an affordable and accessible method for chemical analysis, particularly in colorimetry. While most existing studies have focused on quantitative applications, this study explores a machine learning–assisted DIA approach for the qualitative classification of synthetic food dyes. Digital images of nine food dyes solutions (Carmoisine, Sunset Yellow, Allura Red, Ponceau 4R, Tartrazine, Fast Green FCF, Brilliant Blue FCF, Quinoline Yellow WS, and Indigo Carmine), were captured under both controlled (closed) and open lighting conditions using a smartphone camera. The images were subsequently processed to extract color values in different color spaces, namely RGB, normalized RGB (rgb), HSL, and CIELAB. These values served as input features for a *k*-nearest neighbors (KNN) classifier trained to identify the dye present in each solution. The KNN model performed well on model solutions, with at least 86% accuracy across all color spaces and lighting conditions. To assess practical applicability, the classifier was also tested on seven commercial food and health products. The results show that HSL color space yielded the highest classification accuracy in the commercial sample testing, across both lighting setups, with the open condition consistently producing better performance. These findings demonstrate the potential use of smartphone-based DIA combined with machine learning for low-cost, portable, and reliable solutions for qualitative colorimetric analysis.

*Corresponding Author: martin.tjahjono@calvin.ac.id
How to cite: Y. Tjandra, G. E. Halim, J. Briano, and M. Tjahjono, "Smartphone-Based Digital Image Analysis for Qualitative Classification of Food Dyes Using Machine Learning: Effects of Color Space and Lighting Conditions," *Jurnal Kimia dan Pendidikan Kimia (JKPK)*, vol. 10, no. 2, pp. 309–322, 2025. [Online]. Available: https://doi.org/ 10.20961/jkpk.v10i2.107520

INTRODUCTION

Colorimetry is an analytical technique commonly used for the qualitative and quantitative determination of substances based on color information [1]. Standard colorimetric practices rely on specialized instruments such as spectrophotometers or colorimeters for analysis [2]. However, these specialized instruments are often costly, non-portable, and often require specially trained personnel to operate. As a result, recent trends have shifted in favor of more affordable and accessible analytical methods, with smartphone-based digital image analysis (DIA) gaining significant traction [3-6].

Smartphone-based DIA typically involves capturing images of a colored substance using a smartphone camera, extracting the relevant color values from the images, then analyzing them to determine the substrate concentrations [6]. Modern smartphones are equipped with high resolution cameras and advanced computational abilities, which allows their use for field or onsite chemical analysis [7-9]. Applications of smartphone-based DIA in colorimetry include food spoilage monitoring [8, 9], heavy metal analysis in drinking water [10, 11], point-of-care diagnostics [12-15], and soil analysis for pH and organic content [16]. The smartphone DIA methods from the aforementioned studies were able to provide analytical results within a few seconds with accuracy achieved more than 90%.

Despite extensive research on smartphone colorimetry, most studies have focused on quantitative analysis, whereas applications in qualitative testing remain relatively limited [17]. However, qualitative analysis is equally important for rapid screening, such as detecting banned synthetic dyes, distinguishing between visually similar dyes like Tartrazine and Quinoline Yellow, or verifying label claims in commercial drinks without requiring precise concentration data.

Traditional colorimetric methods measure color intensity against a predefined color–analyte constant [18]. This optimizes them for tracking changes in substrate concentration based on color intensity, but not for classifying the substance type. For example, dyes like Allura Red and Carmoisine may produce similar color values at certain concentrations, making it difficult to distinguish them using conventional approaches. Additionally, color signal values are often dependent on substrate concentration, which limits their ability to identify specific compounds by color information alone [19]. Machine learning (ML) techniques can help overcome this limitation by learning multi-dimensional patterns in color space, enabling classification that remains robust across a range of concentrations.

Machine learning (ML) has been increasingly integrated into the smartphone-based DIA technique to improve the reliability and precision of the method [8, 20–22]. The ML models most commonly used in smartphone colorimetry are Support Vector Machines (SVM), Artificial Neural Network (ANN), and Convolutional Neural Network (CNN), with size of datasets typically around 1200. Training the ML models on the image datasets can help compensate for the variations in lighting environment and camera setting [23]. Furthermore, deep learning networks can automate the feature extraction and classification processes, which could substantially improve the specificity and sensitivity of the colorimetric analysis [24, 25].

In this study, we explore the utility of smartphone-based DIA for the qualitative identification of synthetic food dyes in liquid solutions, assisted by machine learning. This work is among the first to apply smartphone-based DIA for synthetic food coloring analysis, combined with ML classification across different color spaces and lighting conditions. Digital images of nine food dye solutions (including carmoisine, allura red, brilliant blue, and tartrazine), were captured under both controlled (closed) and open lighting conditions using a smartphone camera. These dyes were chosen for both their market prevalence in commercial food and beverage products, and their regulatory significance, since several of them have been re-evaluated for potential health risks.

The images of the dye solutions were analyzed to extract color values in four different color spaces, namely RGB, normalized RGB (rgb), HSL, and CIELAB. For food dye classification, the *k*-nearest neighbors (KNN) algorithm was employed, specifically chosen for its simplicity in data processing. Since the analyzed data consist primarily of three-dimensional color values, additional feature engineering or preprocessing was unnecessary. Consequently, more complex models, such as Support Vector Machines (SVM) or deep learning, were not employed to avoid excessive classification bias. Overall, this study demonstrates that ML-supported DIA can facilitate the qualitative analysis of food dyes in commercial products, providing a low-cost, rapid, and on-site screening solution for real-world food safety applications.

METHODS

1. Materials

Nine synthetic food dyes: Carmoisine (C, CI 14720, 94.5%), Sunset Yellow FCF (SY, CI 15985, 89.6%), Allura Red (AR, CI 16035, 90.3%), Ponceau 4R (P, CI 16255, 88.4%), Tartrazine (T, CI 19140, 90.9%), Fast Green FCF (FG, CI 42053, 92.2%), Brilliant Blue FCF (BB, CI 42090, 90.0%), Quinoline Yellow WS (QY, CI 47005, 72.1%), and Indigo Carmine (IC, CI 73015, 90.4%) were investigated in this study. The nine dyes were selected due to their prevalence in commercial food products. Industrial-grade dyes were used instead of analytical grade alternatives to better reflect real world conditions. All dyes were stored in a dark chemical cabinet at room temperature prior to use.

Seven commercial products, labeled S1–S7, each containing one of the nine evaluated food dyes, were also used to validate the qualitative analysis across various color spaces. Details of these products are presented in Table 1. All commercial samples were purchased from local convenience stores in Jakarta, Indonesia.

Table 1. Seven commercial samples (S1–S7) and their corresponding dye constituents.

Sample	Product Type	Dye Constituent*
S1	Mouthwash A	Fast Green
S2	Mouthwash B	Fast Green
S3	Mouthwash C	Brilliant Blue
S4	Blueberry Soda	Brilliant Blue
S5	Energy Drink	Tartrazine
S6	Cherry Soda	Allura Red
S7	Honey-Flavored Drink	Quinoline Yellow

*As stated in the product label.

The commercial samples used in this study were transparent liquid products which contain only one type of coloring and represent dye variants subject of this study. These products required minimal preprocessing and were selected for their suitability for direct analysis.

2. Preparation of Standard Solutions and Samples

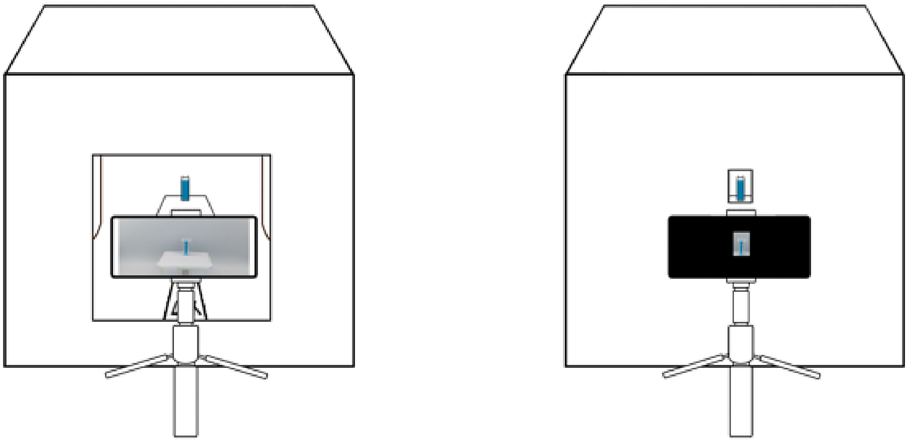
To prepare the stock solutions of each dye, 100 mg of dye powder was accurately weighed using an analytical balance (Sartorius Entris 224i-1S) and dissolved in ultrapure water (Adrona B30 HPLC system, Riga, Latvia) in a 100 mL volumetric flask (Iwaki, Japan), yielding a final concentration of approximately 1000 mg/L. Five standard solutions for each dye, with the concentration range of approximately 5-50 mg/L, were prepared by pipetting appropriate volumes of the stock solution with a 1000 µL micropipette (Corning, New York) into separate 100 mL Grade A volumetric flasks (Iwaki, Japan), and then diluting to volume with ultrapure water. The concentrations of solutions were based on mass without purity adjustment.

Commercial samples S4 and S6, which contained soda, were degassed before analysis to prevent interference with the color measurements. Degassing was carried out by heating the solution to 60°C and stirring for approximately 2–3 minutes until no bubbles were observed. The other commercial samples required no prior treatment and were analyzed directly.

3. Data Collection

Each standard and commercial sample solution was placed in a 1.5 mL polystyrene cuvette (Kartell, Italy) with a 10 mm path length for digital image analysis (DIA). The cuvette was positioned inside a Puluz PU5060 photo box (40 × 40 × 40 cm) with a white background and illuminated by two 30W, 5500K LED light strips. The lighting was adjusted to a uniform intensity of 500 ± 5 lux using an AS803 lux meter to ensure consistent illumination conditions.

Images taken in JPG format were captured using a smartphone (OPPO F11; 48 MP camera, f/1.8 aperture, wide-angle lens, 1/2.0" sensor size, 0.8 µm pixel size) positioned 30 cm from the cuvette. The camera operated with automatic settings and no flash. Five images for every concentration of the dye standard standards were taken, thus producing a balanced dataset across all variants of food dyes. The experimental setup followed the procedure described in a previous study [26]. In this work, the experiments were conducted under two different setups, namely closed (Figure 1a) and open lighting conditions (Figure 1b). The open setup allowed ambient environmental light to enter, introducing lighting noise into the measurements. Data from both setups were used to evaluate the sensitivity and utility of the DIA measurements.



a. b.

Figure 1. (a) Open and (b) closed setups for the DIA measurements

A total of 1368 images were collected: 1228 images of standard solutions (comprising 9 different dyes, 5 concentration levels, triplicate sample solutions, 4-5 images per sample, and 2 lighting setups) and 140 images of commercial samples (comprising 7 product samples, duplicate sample solutions, 5 images per sample, and 2 lighting setups). Images of the standard solutions were used to construct the training and testing datasets for the machine learning models, whereas images of each commercial sample were averaged and directly analyzed for classification. All images were transferred to a computer, and the average red, green, and blue (RGB) values were extracted from a region of interest (ROI) of approximately 1000 pixels using Adobe Photoshop CS6. The ROI for each picture was taken from the center of the cuvette.

4. Computational

4.1. Color Space Models

Color is an important descriptor in image analysis, and selecting an appropriate color space is necessary for accurate image representation. The effectiveness of many image processing algorithms often depends on the choice of a suitable color space [27].

In this study, RGB values obtained from digital images were transformed into other color spaces, namely normalized RGB (rgb), HSL, and CIELAB. These transformations were performed using equations previously described by Chavolla et al. [28]. Transformation formulas are provided in Section A of the Supplementary Material.

Normalized RGB (rgb) is a variant of RGB color space, first developed for the purpose of reducing color effects from lighting changes. This color space expresses the proportion of red, green, and blue in an object to a total of 100%. This color space is capable of minimizing color variations caused by shadow or irregular lighting intensity, but this effect may result in a loss of contrast [28].

HSL color space (Hue, Saturation, Lightness) uses a cylindrical coordinate system and separates the color gamut into more visually intuitive dimensions. CIELAB consists of lightness channel (L) and two chromatic channels (a* and b*). This color space represents a wider range of colors than the RGB space and is typically used to enhance and analyze color images [28].

The RGB color model parameters for all images taken in this study are made publicly available as an open access resource on Kaggle [29].

4.2. KNN Classification

The *k*-nearest neighbors (KNN) algorithm was utilized to classify the food dyes based on their color features. KNN is a non-parametric, instance-based learning algorithm that assigns a label to a test sample by identifying the most frequent class among its *k* nearest neighbors in the training set. In this study, the number of neighbors (*k*) was set to 5, and the euclidean distance metric was used. These parameters were heuristically selected to consider a balance between bias and variance.

Figure 2 illustrates the working principle of the *k*-nearest neighbors (KNN) algorithm within a two-dimensional feature space. In this example, two classes of training data are shown, namely blue circles (●) and orange crosses (×). A purple triangle (▲) denotes a new, unlabeled test instance. To classify this point, the algorithm identifies its five nearest neighbors, indicated by the points within the dashed circle. Among these neighbors, three belong to the blue circle class and two to the orange cross class. As the majority class is the blue circle, the test instance is accordingly assigned to that class.

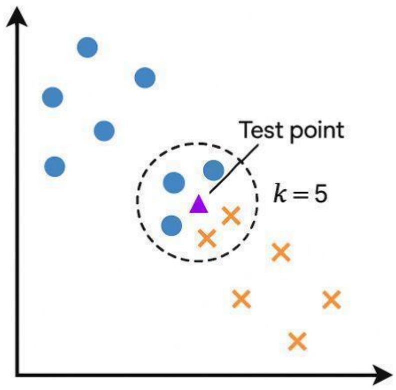


Figure 2. Visualization of the KNN (*k*=5) algorithm.

The dataset was subsequently randomly partitioned (using random seed 103) into training and testing subsets using an 80:20 split ratio, to ensure robust model training and unbiased evaluation. This means that 80% of the samples were used to train the classification model, while the remaining data were reserved for evaluating its performance. Since the dataset contains generally balanced numbers of instances among all dye labels, the uniformly random train-test split would provide each data partition with all possible dye labels with high probability.

Moreover, separate KNN classifiers were trained and tested for each combination of color space representation and sampling lighting conditions (open and closed setups) to evaluate how both aspects influence classification accuracy. All computations and modeling procedures were carried out using the *scikit learn* library version 1.5.1 [30] in Python 3.12.3.

The performance of the KNN classifiers was evaluated using standard classification metrics, namely accuracy and F1-score. Accuracy metric measures the proportion of correctly classified instances. Meanwhile, F1-score measures the harmonic mean of precision (the proportion of true positive predictions among all predicted positives) and recall (the proportion of true positive predictions among all actual positives). The latter metric is particularly useful for further understanding of each class's performance. These metric values range from 0% to 100%, with values closer to 100% indicating better classification performance. All metrics were calculated in this study for each food dye in every color space representation and lighting sampling condition, to facilitate comparative analysis and to determine the most robust color model for food dye classification.

Moreover, we provide a confusion matrix for each classification model in the Supplementary Material (Section B) to further assess misclassification. A confusion matrix is a tabular representation that compares the actual class labels with the predicted labels generated by a model. Rows correspond to the true classes, while columns correspond to the predicted classes, allowing clear identification of both correct predictions and misclassifications. This makes the confusion matrix a valuable tool for evaluating model performance, as it reveals not only overall accuracy but also specific error patterns, such as which classes are most frequently confused

RESULT AND DISCUSSION

1. Classification Results

The performance of the *k*-nearest neighbors (KNN) classifier was evaluated across all color spaces, namely RGB, rgb, HSL, and CIELAB, using model standard solutions containing food dyes. The overall classification accuracy for both training and testing datasets are summarized in Table 2.

Across all color spaces, the KNN model successfully achieved strong classification results, with a minimum accuracy of 86% across all color spaces and lighting conditions. This high accuracy is likely due to the simplicity of the model sample solutions, which contain no interfering constituents. As shown in Table 2, the test accuracy is slightly lower than the training accuracy, as the model encounters new color values in the test set, which are previously unseen during training. The most notable decrease in accuracy was observed in the RGB color space under the closed lighting condition, decreasing from 98.84% on the training set to 88.89% on the test set. Further, a simple one-sided paired *t*-test is conducted to provide statistical justification using eight pairs of train-test accuracy values among the four-color spaces and two lighting conditions. As the test yields a *p*-value of 0.9848, this suggests that there is no statistically significant decrease between the models' performance on the training and test dataset. Hence, this considerably negligible performance drop between training and test dataset could affirm that the model avoids both overfitting and underfitting issues.

Table 2. Training and test classification accuracy on different color spaces and lighting conditions (closed and open setups).

Color Space	Training Dataset Accuracy		Test Dataset Accuracy	
	Closed	Open	Closed	Open
CIELAB	0.9977	0.9982	0.9444	0.9928
HSL	0.9536	0.9365	0.8611	0.9565
RGB	0.9884	0.9927	0.8889	0.9710
rgb	0.9884	0.9964	0.9167	0.9855

A comparison of test results between two different lighting setups in Table 2 shows that images analyzed under the open setup consistently resulted in higher classification accuracy across all color spaces (*t*-test, p = 0.02354). This comparison indicates that ambient lighting conditions support the model's ability to generalize, likely due to increased variability in training data (see section C in Supplementary Material). This suggests that the open setup is more suitable for dye classification analysis. The presence of lighting noise in the open setup may have exposed the model to a wider range of conditions, leading to improving its generalization and robustness in classification.

Analyzing metrics of F1-scores (Table 3) offers useful insights into class-specific food dye performance. As shown in Table 3, all dyes show high macro average F1-scores across all color spaces (i.e. above 90%) when analyzed in open setups. This indicates that the classification model can successfully distinguish between the dyes in these model solutions. However, the macro average F1-scores of dyes obtained under the closed setup indicate that some dyes are relatively more difficult to differentiate. Specifically, Allura Red (AR), Brilliant Blue (BB), and Fast Green (FG) show average F1-score below 90%. In addition, examination of the confusion matrices (see in Section B of the Supplementary Material) further highlights the specific challenges faced by the classifiers. Among all dyes, Brilliant Blue (BB) and Fast Green (FG) consistently emerge as the most difficult to distinguish, with many models frequently interchanging prediction results between these two classes. This suggests a high degree of visual similarity between their color profiles, leading to systematic misclassifications. A smaller number of errors were also observed for Allura Red (AR), although these occurred less frequently. In contrast, Carmoisine (C) and Quinoline Yellow (QY) achieved perfect F1-scores across all conditions, implying highly reliable classification.

Table 3. F-1 score for each dye class in the KNN classification.

Color Space	Setup	Dye label									
		AR	BB	C	FG	IC	P	QY	SY	T	Macro Average
CIELAB	Closed	1.00	0.73	1.00	0.84	1.00	1.00	1.00	1.00	1.00	0.95
	Open	1.00	0.96	1.00	0.97	1.00	1.00	1.00	1.00	1.00	0.99
HSL	Closed	0.54	0.87	1.00	0.92	0.67	1.00	1.00	1.00	1.00	0.89
	Open	0.89	1.00	1.00	1.00	0.89	0.92	1.00	0.94	1.00	0.96
RGB	Closed	1.00	0.56	1.00	0.68	1.00	1.00	1.00	1.00	0.82	0.90
	Open	0.96	0.89	1.00	0.89	0.96	1.00	1.00	1.00	1.00	0.97
rgb	Closed	0.97	0.67	1.00	0.78	0.96	1.00	1.00	1.00	1.00	0.93
	Open	0.93	1.00	1.00	1.00	0.92	1.00	1.00	1.00	1.00	0.98
Macro Average	Closed	0.88	0.70	1.00	0.81	0.91	1.00	1.00	1.00	0.95	0.92
	Open	0.95	0.96	1.00	0.96	0.94	0.98	1.00	0.98	1.00	0.98

2. Classification of Commercial Samples

To evaluate the utility of the DIA-based classification model in more complex settings, seven commercial food and health products (S1–S7) were analyzed. Digital images of the sample solutions were taken under two lighting conditions (open and closed setups). The sample images are provided in Figure 3. Subsequently, these images were directly analyzed using four color spaces (RGB, rgb, HSL, and CIELAB), to determine the most suitable representation for accurate qualitative dye identification. It is important to note that these samples are compositionally richer than the previous model solutions, due to the presence of additional constituents. Furthermore, the dyes concentrations in these commercial samples may differ from or fall outside the concentration range of dyes used in the training classification models.

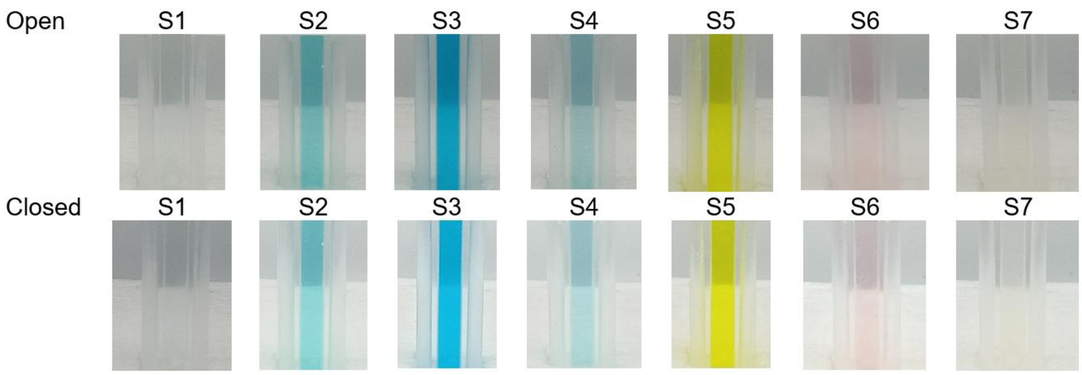


Figure 3. Commercial sample images measured under open and closed setups in the DIA measurements.

A summary of the classification results is presented in Table 4. As shown in Table 4, applying the KNN method to commercial samples resulted in lower accuracy than when used on model solution samples. This is likely influenced by the presence of other constituents, such as sugar, acids, and artificial flavouring, that might interfere with colorimetric measurement [5, 31].

Table 4. Classification performance of RGB, rgb, HSL, and CIELAB color spaces on commercial products.

Sample*	Dye Constituent	Setup	Correct Prediction			
			RGB	**rgb**	**HSL**	**CIELAB**
S1	Fast Green	Closed			√	
S2	Fast Green	Closed			√	
S3	Brilliant Blue	Closed	√	√	√	√
S4	Brilliant Blue	Closed				
S5	Tartrazine	Closed	√		√	√
S6	Allura Red	Closed				
S7	Quinoline Yellow	Closed	√	√	√	√
			42.9%	28.6%	**71.5%**	42.9%
S1	Fast Green	Open			√	
S2	Fast Green	Open			√	√
S3	Brilliant Blue	Open	√	√	√	√
S4	Brilliant Blue	Open	√	√		√
S5	Tartrazine	Open	√	√	√	√
S6	Allura Red	Open			√	
S7	Quinoline Yellow	Open	√	√	√	√
			57.1%	57.1%	**85.7%**	71.5%

Among the color spaces explored, the HSL color space achieved the most robust classification performance and could outperform other color spaces. It successfully identified the dye constituents in five out of seven samples (71.4%) under the closed setup, and in six out of seven samples (85.7%) under the open setup.

Two factors may account for HSL's superior performance. First, HSL color space effectively decouples color information (Hue) from brightness and saturation, which helps to minimize interference from lighting variability [32]. Secondly, its cylindrical representation provides clearer boundaries between color classes, facilitating better dye color classification by the KNN algorithm [28]. However, this explanation should be regarded as qualitative and limited by the lack of quantitative justification.

Similar to the results of the previous standard model solution testing (Section 4.1), the open setup consistently achieved better classification results than the closed setup. This may be attributed to the broader histogram RGB channel (see in Section C in Supplementary Material) and the smartphone camera’s automatic settings (i,e., white balance, ISO). Such conditions may better align with the variability encountered during training, improving generalization. In contrast, the uniformity of the closed setup limits variability, making the model more sensitive to deviations and requiring stricter conditions to maintain classification performance.

In general, this study underscores the utility of the DIA approach combined with machine learning for qualitative classification of food dyes. These results highlight the method's potential used in routine quality control and regulatory monitoring, where environmental variability is unavoidable. The demonstrated robustness of the HSL color space and the open setup further supports the applicability of smartphone-based image analysis in low resource and field-based settings.

3. Challenges and Opportunities of DIA

This study demonstrates that the application of smartphone-based digital image analysis (DIA) in combination with machine learning algorithms supports its use for the qualitative classification of food dyes in transparent liquid samples. This technique shows promise for implementation in limited resource environments, such as small laboratories. The portability of this approach can support onsite applications, including food quality monitoring and regulatory inspections. In educational settings, the affordability and accessibility of the smartphone-based DIA facilitates learners to engage with colorimetry techniques.

Despite the above-mentioned advantages, the DIA model still poses several challenges. First, the current approach was validated strictly using relatively simple solutions containing only a single type of dye. However, commercial products often contain more than one type of dye. Furthermore, food and health products may contain complex constituents such as proteins, acids, or alcohols, which may interact with the food dye, that could alter its color properties. This led to lower classification accuracy compared to when the approach was applied to standard solutions.

To mitigate these limitations, future work should expand training data beyond single-dye standard solutions to include samples that reflect real commercial products, including mixtures of multiple dyes, and a wide range of pH, alcohol, and protein contents. Additionally, standardized sample preparation, such as buffering, controlled dilution, or inclusion of internal standards, can minimize matrix effects from other constituents. It is also beneficial to explore the influence of varying smartphone hardware and software specifications on classification performance, as such factors could affect consistency and reliability in real-world applications.

Nevertheless, the result of this study can provide a foundational step toward reliable and low-cost colorimetric analysis using smartphones. Future research on smartphone-based DIA can focus on the development of machine learning models for simultaneous quantitative and qualitative analysis of analytes. With this development, the machine learning-assisted DIA approach has potential for wider use across fields such as food science, healthcare, and environmental monitoring.

CONCLUSION

This study demonstrates the qualitative classification of food dyes using smartphone-based digital image analysis (DIA) with the k-nearest neighbors (KNN) algorithm. The open lighting setup consistently produced higher classification accuracy, suggesting that ambient lighting improves the model's generalizability. The KNN model performed well on model solutions, with at least 86% accuracy across all color spaces and lighting conditions. Among the color spaces, HSL showed the best performance on commercial samples because it separates color components into segments that are more intuitive for qualitative classification. The results highlight that both color space selection and lighting conditions are essential factors and could significantly affect classification performance.

These findings also show that the machine learning-assisted DIA is a practical and affordable method for colorimetric analysis, especially in settings with limited resources like small laboratories, field testing sites, or classrooms. With further development, this method could potentially become a simple and reliable tool for broader analysis in areas such as food science, healthcare, and environmental monitoring.

ACKNOWLEDGEMENT

This work is supported by CIT Research Grant.

REFERENCES

[1] Y. Ohno, "CIE fundamentals for color measurements," in IS&T NIP16 International Conference on Digital Printing Technologies, pp. 540–545, 2000.

[2] V. Kılıç, N. Horzum, and M. E. Solmaz, "From sophisticated analysis to colorimetric determination: Smartphone spectrometers and colorimetry," in Color Detection, IntechOpen, 2018, doi: 10.5772/intechopen.82227

[3] G. M. Fernandes et al., "Novel approaches for colorimetric measurements in analytical chemistry–a review," Anal. Chim. Acta, vol. 1135, pp. 187–203, 2020. doi: 10.1016/j.aca.2020.07.030

[4] F. C. Böck et al., "PhotoMetrix and colorimetric image analysis using smartphones," J. Chemom., vol. 34, no. 12, p. e3251, 2020. doi: 10.1002/cem.3251

[5] M. Saadati, "Smartphone-Based Digital Image Analysis for Determination of Some Food Dyes in Commercial Products," Food Anal. Methods, vol. 14, no. 11, pp. 2367-2374, 2021. doi: 10.1007/s12161-021-02059-4

[6] Y. Fan, J. Li, Y. Guo, L. Xie, and G. Zhang, "Digital image colorimetry on smartphone for chemical analysis: A review," Measurement, vol. 171, p. 108829, 2021. doi: 10.1016/j.measurement.2020.108829

[7] S. Erdemir and S. Malkondu, "On-site and low-cost detection of cyanide by simple colorimetric and fluorogenic sensors: smartphone and test strip applications," Talanta, vol. 207, p. 120278, 2020. doi: 10.1016/j.talanta.2019.120278

[8] V. Doğan, M. Evliya, L. N. Kahyaoglu, and V. Kılıç, "On-site colorimetric food spoilage monitoring with smartphone embedded machine learning," Talanta, vol. 266, p. 125021, 2024. doi: 10.1016/j.talanta.2023.125021

[9] W. Fu, X. Fu, Z. Li, Z. Liu, and X. Li, "Advances in smartphone assisted sensors for on-site detection of food safety based on fluorescence on-off-on mode: A review," Chem. Eng. J.,

vol. 481, p. 151225, 2024. doi: 10.1016/j.cej.2024.151225

[10] H. Nguyen et al., "Smartphone nanocolorimetry for on-demand lead detection and quantitation in drinking water," Anal. Chem., vol. 90, no. 19, pp. 11517–11522, 2018. doi: 10.1021/acs.analchem.8b02808

[11] P. Jarujamrus et al., "Use of a smartphone as a colorimetric analyzer in paper-based devices for sensitive and selective determination of mercury in water samples," Anal. Sci., vol. 34, no. 1, pp. 75–81, 2018.

[12] H.-C. Wang et al., "Development and clinical trial of a smartphone-based colorimetric detection system for self-monitoring of blood glucose," Biomed. Opt. Express, vol. 11, no. 4, pp. 2166–2177, 2020. doi: 10.1364/BOE.389638

[13] T. T. Wang, C. K. Lio, H. Huang, R. Y. Wang, H. Zhou, P. Luo, and L. S. Qing, "A feasible image-based colorimetric assay using a smartphone RGB camera for point-of-care monitoring of diabetes," Talanta, vol. 206, p. 120211, 2020. doi: 10.1016/j.talanta.2019.120211

[14] S. K. Biswas, A. Bairagi, S. Nag, A. Bandopadhyay, I. Banerjee, A. Mondal, and S. Chakraborty, "Nucleic acid based point-of-care diagnostic technology for infectious disease detection using machine learning empowered smartphone-interfaced quantitative colorimetry," Int. J. Biol. Macromol., vol. 253, p. 127137, Dec. 2023. doi: 10.1016/j.ijbiomac.2023.127137

[15] S. Chunta, P. Jarujamrus, A. Prakobkij, S. Khongwichit, N. Ditcharoen, S. Pencharee, and M. Amatatongchai, "Point-of-care blood tests using a smartphone-based colorimetric analyzer for health check-up," Microchim. Acta, vol. 191, no. 7, p. 402, 2024. doi: 10.1007/s00604-024-06463-5

[16] M. Tobiszewski and C. Vakh, "Analytical applications of smartphones for agricultural soil analysis," Anal. Bioanal. Chem., vol. 415, no. 18, pp. 3703–3715, 2023. doi: 10.1007/s00216-023-04558-1

[17] L. Ciaccheri, B. Adinolfi, A. A. Mencaglia, and A. G. Mignani, "Smartphone-Enabled Colorimetry," Sensors (Basel), vol. 23, no. 12, p. 5559, Jun. 14, 2023, doi: 10.3390/s23125559

[18] M. Joh, S. Kumaran, Y. Shin, H. Cha, E. Oh, K. H. Lee, and H.-J. Choi, "An ensemble model of machine learning regression techniques and color spaces integrated with a color sensor: application to color-changing biochemical assays," RSC Adv., vol. 15, pp. 1754–1765, 2025, doi: 10.1039/D4RA07510B. doi: 10.1039/D4RA07510B

[19] A. A. Abdella and E. A. Elshenawy, "A spatial hue smartphone-based colorimetric detection and discrimination of carmine and carminic acid in food products based on differential adsorptivity," Talanta, vol. 282, Art. no. 127053, 2025, doi: 10.1016/j.talanta.2024.127053

[20] A. Y. Mutlu, V. Kılıç, G. K. Özdemir, A. Bayram, N. Horzum, and M. E. Solmaz, "Smartphone-based colorimetric detection via machine learning," Analyst, vol. 142, no. 13, pp. 2434–2441, 2017. doi: 10.1039/C7AN00741H

[21] M. E. Solmaz, A. Y. Mutlu, G. Alankus, V. Kılıç, A. Bayram, and N. Horzum, "Quantifying colorimetric tests using a smartphone app based on machine learning classifiers," Sens. Actuators B

Chem., vol. 255, pp. 1967–1973, 2018. doi: 10.1016/j.snb.2017.08.220

[22] S. Bhatt, S. Kumar, M. K. Gupta, S. K. Datta, and S. K. Dubey, "Colorimetry-based and smartphone-assisted machine-learning model for quantification of urinary albumin," Meas. Sci. Technol., vol. 35, no. 1, p. 015030, 2023. doi: 10.1088/1361-6501/acfd4c

[23] R. Thakur, S. Bhatt, and S. K. Dubey, "Effect of smartphone camera settings in colorimetric measurements under controlled illumination," in Proc. 14th Int. Congr. Image Signal Process., BioMedical Eng. Informatics (CISP-BMEI), Oct. 2021, doi: 10.1109/CISP-BMEI53629.2021.9624389

[24] B. Khanal, P. Pokhrel, B. Khanal, and B. Giri, "Machine-learning-assisted analysis of colorimetric assays on paper analytical devices," ACS Omega, vol. 6, no. 49, pp. 33837–33845, Dec. 2021. doi: https://pubs.acs.org/doi/10.1021/acsomega.1c05086

[25] S. Duan, T. Cai, J. Zhu, X. Yang, E. G. Lim, K. Huang, K. Hoettges et al., "Deep learning-assisted ultra-accurate smartphone testing of paper-based colorimetric ELISA assays," Anal. Chim. Acta, vol. 1248, p. 340868, 2023. doi: 10.1016/j.aca.2023.340868

[26] G. E. Halim and M. Tjahjono, "Evaluation of color models for quantitative determination of food dyes using smartphone-based digital image analysis," JKPK (Jurnal Kimia dan Pendidikan Kimia), vol. 8, no. 1, pp. 86–100, 2023. Doi: 10.20961/jkpk.v8i1.72928 https://jurnal.uns.ac.id/jkpk/article/view/72928/40549

[27] M. A. Ansari and D. K. Singh, "Significance of Color Spaces and Their Selection for Image Processing: A Survey," Recent Adv. Comput. Sci. Commun., vol. 15, no. 7, pp. 946–956, Sep. 2022. doi: 10.2174/266625581466621030815210 8

[28] E. Chavolla, D. Zaldivar, E. Cuevas, and M. A. Perez, "Color spaces advantages and disadvantages in image color clustering segmentation," in Advances in Soft Computing and Machine Learning in Image Processing, 2018, pp. 3–22. doi: 10.1007/978-3-319-63754-9_1

[29] Y. Tjandra, G. Halim, J. Briano, and M. Tjahjono, "Food Dyes RGB data for Digital Image Colorimetry," Kaggle, Dataset, 2025. [Online]. Available: 10.34740/KAGGLE/DSV/10386234

[30] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825–2830, 2011.

[31] J. Park, Y. S. Cho, D. W. Seo, and J. Y. Choi, "An update on the sample preparation and analytical methods for synthetic food colorants in food products," Food Chem., vol. 459, p. 140333, 2024, doi: 10.1016/j.foodchem.2024.140333

[32] R. Hamdini, N. Diffellah, and A. Namane, "Color based object categorization using histograms of oriented hue and saturation," Trait. Signal, vol. 38, no. 5, pp. 1293–1307, Oct. 2021, doi: 10.18280/ts.380504