applied sciences

MDPI

Article

Optimal Color Space Selection for Vermicompost Nitrogen Classification: A Comparative Study Using the KNN Model

Panida Lorwongtragool and Suthisa Leasen *

Faculty of Science and Technology, Rajamangala University of Technology Suvarnabhumi, Nonthaburi 11000, Thailand; panida.l@rmutsb.ac.th

* Correspondence: suthisa.l@rmutsb.ac.th; Tel.: +66-(0)61-1656622

Abstract

This study presents a cost-effective and accurate method for assessing nitrogen concentration in vermicompost fertilizer using a low-cost TCS3200 color sensor and a K-Nearest Neighbors (KNN) machine learning model. The objective was to evaluate the performance of four different color spaces—RGB, Lab, LCh, and CMYK—identify the most effective feature representation for a multi-class classification task based on accuracy and theoretical robustness to ambient light variations. A total of 2400 data points were collected from a standard chemical test kit and processed. A rigorous 60-fold cross-validation approach was used to determine the optimal model hyperparameters and to ensure the robustness of the findings. The results demonstrate that the model trained on the LCh color space achieved the highest classification accuracy of 0.9708 with an optimal K-value of 6, significantly outperforming Lab (0.9688), RGB (0.9625), and CMYK (0.9583). A detailed analysis of the confusion matrix revealed that the model successfully classified the 'High' and 'Medium' nitrogen levels with near-perfect accuracy, while minor misclassifications occurred between the 'Low' and 'Trace' categories (5 Low $\longrightarrow$ Trace, 6 Trace $\longrightarrow$ Low). The proposed system offers a practical, robust, and accessible tool for precision agriculture, enabling farmers to make informed decisions regarding fertilization, and directly supporting sustainable agriculture and responsible resource management. The findings indicate that the LCh color space is highly effective for this application, providing a viable solution for the rapid and reliable assessment of vermicompost quality. Most importantly, this inexpensive, on-site system removes the need for costly, time-consuming laboratory analyses, giving farmers and compost users the instantaneous, accurate nitrogen data they need to maximize crop yield, optimize nutrient application, and significantly reduce input costs from overfertilization.

Keywords: vermicompost; nitrogen; classification; KNN; color sensor

check for updates

Academic Editors: Rabin Bhattarai and Soonho Hwang

Received: 26 September 2025
Revised: 25 October 2025
Accepted: 28 October 2025
Published: 29 October 2025

Citation: Lorwongtragool, P.; Leasen, S. Optimal Color Space Selection for Vermicompost Nitrogen Classification: A Comparative Study Using the KNN Model. *Appl. Sci.* **2025**, *15*, 11578. https://doi.org/10.3390/app152111578

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Organic farming is essential to sustainable agriculture, but managing soil nutrients is challenging because its reliance on organic matter and microorganisms often leads to nutritional deficiencies. Chemical fertilizers, conversely, have stricter application rules. Nevertheless, whether using chemical or organic fertilizers, excessive fertilization can harm both the soil and the plants. Therefore, ensuring plants receive an appropriate and long-term amount of nutrients requires knowing the nutritional content of fertilizers just as much as testing the soil. The measurement of nutrients in soil and fertilizers is still a

difficult, expensive, and time-consuming procedure that often requires sending samples to labs. Consequently, farmers cannot make timely soil amendment decisions.

1.1. Nitrogen Content Assessment in Organic Fertilizer

Nitrogen (N) is a vital macronutrient for plant growth and is a key component that contributes to agricultural yield. Vermicompost is a high-quality organic fertilizer rich in plant nutrients that are in a readily available form, such as nitrate and ammonium ions. Therefore, the accurate and timely assessment of nitrogen content in vermicompost is of utmost importance for precision agriculture, which aims to optimize fertilizer use and minimize environmental impact.

Traditional methods for assessing nitrogen levels often rely on laboratory analysis, which is both expensive and time-consuming, making it impractical for providing rapid information for proper fertilizer management [1,2]. Furthermore, simple visual assessments using colorimetric test kits lack accuracy and are subjective, as results can vary depending on lighting conditions and the observer.

To address these limitations, various research studies have turned to the use of optical sensor technology in conjunction with machine learning for soil and fertilizer nutrient analysis [1,3–6]. Paper-based analytical devices combined with smartphones have also been developed to measure essential macronutrients [7,8]. For instance, a low-cost optical sensor was developed to measure NPK nutrients in soil [9]. Similarly, a spectrophotometric approach with optical fibers was explored to quantify nutrients [10]. The integration of machine learning has been shown to be effective in overcoming the limitations of spectral interference, as demonstrated by studies using UV absorption spectroscopy, with similar work achieving improved accuracy with hybrid machine learning [11].

In addition, systems based on the Internet of Things (IoT) have been used with sensors to analyze soil nutrients and provide crop recommendations [12]. Paper-based analytical devices combined with smartphones have also been developed to measure essential macronutrients [13]. Most recently, a 4-channel spectrophotometric sensor was developed to identify and quantify multiple fertilizers at once [14,15].

Current cost-effective nutrient analysis studies often focus on general soil or multiple nutrients, but they are limited by not reporting on ambient light, standardized calibration and by using only three coarse classification classes (Low, Medium, High). These issues necessitate using perceptually uniform color features and robust Machine Learning models to improve resolution and mitigate inaccuracies [16]. Our research directly addresses this gap by focusing on a specific and practical approach for analyzing nitrogen in vermicompost.

1.2. K-Nearest Neighbors Model and the TCS3200 Color Sensor

This work utilizes a low-cost TCS3200 color sensor that was systematically calibrated by carefully evaluating its raw frequency output against standard colorimetry values obtained using a commercial colorimeter. This calibrated sensor is then integrated with a K-Nearest Neighbors (KNN) machine learning model. KNN was selected strategically due to its computational simplicity and low memory efficiency [17,18], making it ideal for development on low power microcontroller in a portable a device [15,16]. The performance of four different color representations—RGB, Lab, LCh, and CMYK—is compared to identify the most suitable features for multi-class classification. The selection includes RGB and CMYK as common device-dependent standards, and Lab and LCh as device-independent models whose components theoretically offer greater separation between color and lightness, which is hypothesized to confer superior resilience against ambient light fluctuations [19,20]. The statistical robustness of the KNN model was validated through the

application of rigorous cross-validation to identify the optimal K hyperparameter selection and to effectively evaluate the model's stability [17].

In summary as shown in Table 1, our work provides a critical bridge between costly, lab-based analytical methods and unreliable on-field assessments. It demonstrates that with a targeted approach and careful data representation, an accessible and affordable system can deliver a level of performance that is not only competitive but also highly effective for empowering farmers to make data-driven decisions.

Table 1. A comparison of methods for soil/fertilizer nutrient analysis.

Author	Year	Subject of Study	Method	Key Findings/Performance	Reference
Masrie et al.	2017	Detection of Nitrogen, Phosphorus, and Potassium (NPK) nutrients in soil using an optical transducer	Optical transducer (an LED light source and a photodiode), with specific LED wavelengths to measure the absorbance of nutrient solutions	The sensor successfully assessed NPK levels in soil into three categories: High, Medium, and Low.	[9]
Monteiro-Silva et al.	2019	Optical Sensing of Nitrogen, Phosphorus and Potassium: A Spectrophotometrical Approach toward Smart Nutrient Deployment approach toward Smart Nutrient Deployment	UV-Vis spectroscopy combined with optical fibers to quantify N, P, and K in fertilizer solutions	Linear/log models were inaccurate due to spectral interference, but the issue was solved using AI algorithms.	[10]
Zhang et al.	2021	Simultaneous detection of nitrate and nitrite based on UV absorption spectroscopy and machine learning	Simultaneous detection of nitrate and nitrite using hybrid machine learning approach; UV absorption spectroscopy.	Improved accuracy (error < 1%) by overcoming spectral signal overlapping with AI.	[11]
Senapaty et al.	2023	IoT-Enabled soil nutrient analysis and crop recommendation model for precision agriculture.	IoT-based prototype for soil data collection (pH, N, P, K, moisture) using machine learning algorithms.	Accurately analyzes soil and recommends crops. Highlights importance of real-time monitoring.	[12]
Albuquerque et al.	2024	Low-cost precision agriculture for sustainable farming using paper-based analytical devices.	Paper-based platform for nutrient analysis; uses colorimetric methods and a smartphone.	Effectively detects typical soil nutrient concentrations (nitrate, calcium, magnesium, and ammonium) with strong linearity.	[13]
Li et al.	2024	Spectrophotometric-based sensor for the detection of multiple fertilizer solutions.	A 4-channel photoelectric sensor was developed to qualitatively and quantitatively identify four fertilizers (N, P, and K) using spectrophotometry.	The sensor achieved an 81.5% accuracy in fertilizer classification, with a relative error of less than ±15% for concentration measurements up to 80 mg/L.	[14]

Table 1. *Cont.*

Author	Year	Subject of Study	Method	Key Findings/Performance	Reference
Our study	2025	Optimal Color Space Selection for Vermicompost Nitrogen Classification: A Comparative Study Using the KNN Model	Low-cost TCS3200 sensor using KNN model based on optical color space	High accuracy (97.08%) for classifying specific N level in vermicompost, using a dedicated, low-cost sensor.	

2. Materials and Methods

2.1. Colorimetric Set Up for Nitrogen Assessment in Vermicompost

An infographic illustrating the overall system for nitrogen level classification in vermicompost fertilizer is shown in Figure 1.

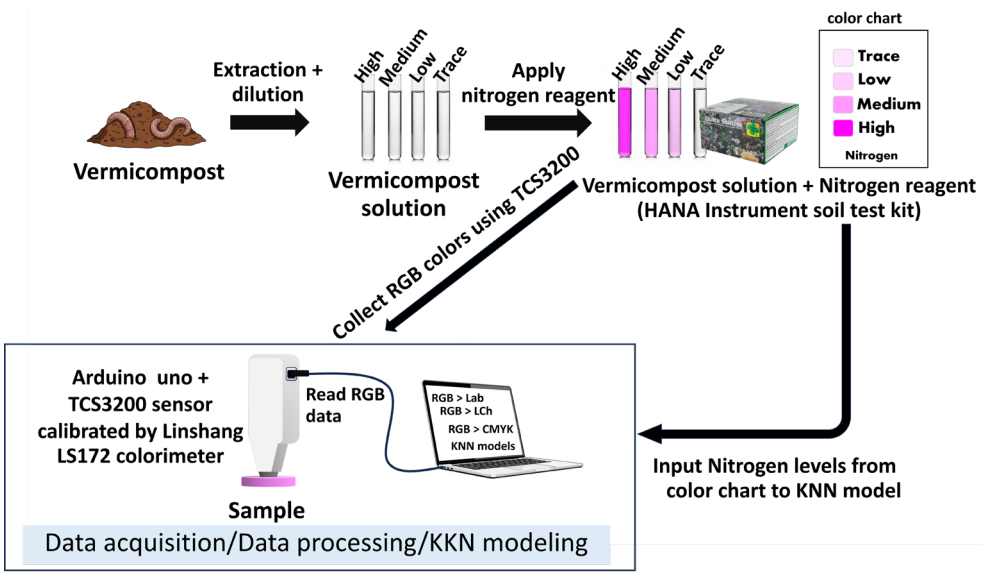

Figure 1. Colorimetric Detection System for Assessing Nitrogen Levels in Vermicompost Samples Using a Nitrogen test kit.

The system comprises three main stages: (1) Data Acquisition, where a TCS3200 sensor (OSRAM Opto Semiconductors (China) Co., Ltd., Wuxi, China) and Arduino Uno (R3, Shenzhen, China) are used to measure the color of prepared samples; (2) Data Processing, where raw RGB values are calibrated using a transformation matrix derived from a commercial Colorimeter (Shenzhen Linshang Technology Co., Ltd., Shenzhen, China), and then converted to various color spaces (Lab, LCh, CMYK); and (3) Machine Learning Model Development, where the processed data is used to train and validate a K-Nearest Neighbors (KNN) classification model using a 60-fold cross-validation approach to ensure robust performance [21].

2.2. Sample Preparation and Colorimetric Analysis

This research utilized vermicompost samples sourced from local farmers who practice organic agriculture. The vermicompost was stored under controlled conditions to maintain stability, with its moisture content at approximately 35 ± 5% relative humidity (RH). Importantly, to maintain exacting sample homogeneity and improve the repeatability of the system testing, the same batch of manufactured vermicompost solution was continuously

employed as the control sample over the whole experimental run. The samples were prepared as a solution for optical analysis by first weighing 1 g of vermicompost and adding it to a container with 15 g of deionized water. The mixture was then thoroughly dispersed to ensure all compost particles were broken down in the water. To separate the solid particles, the solution was centrifuged at a speed of 4000 rpm for 10 min. The supernatant was then carefully extracted with a pipette and used as the primary solution for subsequent analysis. The primary solution was serially diluted with DI water to create various concentrations, with the goal of producing a range of samples that corresponded to all four-color levels specified by the HANA Instrument soil test kit (Hana Instruments SRL, Salaj, Romania). The four-color levels (High, Medium, Low, and Trace), representing a predefined range of nitrogen content relevant to general plant growth, served as the ground truth for a classification model. Ten vermicompost solutions, each at a different concentration, were prepared for each of the four-color levels to accurately represent every nitrogen classification. This resulted in 40 distinct samples used for color data collection, covering the full range indicated by the color chart. To achieve statistically reliable experimental data, three replicates were prepared at the same concentration. To create a dataset, parallel data collections were performed. First, the color chart from the HANA test kit was used to assign the nitrogen level to each sample, with these values serving as the "true labels" for the machine learning model. Simultaneously, an affordable TCS3200 color sensor, controlled by a microcontroller, was used to measure and record the raw RGB color values of each sample solution. The raw RGB data was then converted into other color spaces, namely Lab, LCh, and CMYK, to be used as features for training and comparing the performance of the KNN model.

2.3. Color Data Acquisition and Calibration

A TCS3200 color sensor, integrated with an Arduino Uno microcontroller, was used to capture the color of the prepared solutions. The sensor was set up within a custom light-shielding enclosure (dark box). The distance between the sample surface and the sensor lens was fixed at 5 mm, and all measurements were achieved at ambient room temperature (~25 °C ± 2 °C).

The calibration data was obtained using a colorimeter model Linshang LS172. The calibration was performed using a linear transformation based on the following transformation matrix (M) and bias vector (b):

$$\mathrm{M} = \begin{bmatrix} 0.5687 & -0.0586 & 0.0741 \\ -0.0897 & 0.4839 & 0.0687 \\ -0.1375 & -0.2196 & 0.4786 \end{bmatrix} \text{ and } \mathrm{b} = \begin{bmatrix} 15.5770 \\ 42.5167 \\ 76.5055 \end{bmatrix}$$

The transformation matrix and bias vector were derived using ordinary least square linear regression. The model achieved high coefficient of determination (R^2) of 0.9973 and a root mean square error (RMSE) of 1.4314.

A statistical technique based on the Z-score ($|Z| \geq 2$) was used to determine and eliminate significant outliers to confirm the dataset's quality and robustness before analysis and model training. Following the Z-score filtering, the 20 data points were randomly sampled. Therefore, a total of 60 raw data points were initially collected per sample (20 data points × 3 repetitions), resulting there were 600 RGB data points were drawn from multiple samples which represent entire concentration range of each color level. The collected 2400 calibrated RGB data points were then converted into several other color spaces, including Lab, LCh, and CMYK, to explore which feature representation would yield the best model performance.

2.4. Data Processing and Machine Learning Model Development

The classification model was developed using the K-Nearest Neighbors (KNN) algorithm, implemented with the Python programming language (Python version 3.13.5) and the Scikit-learn library (Version 1.7.1). To guarantee consistent representation of all four nitrogen classes in both subsets, the data was divided using stratified random sampling into a training set (80%) and a testing set (20%). This stratification method enables a comprehensive and objective evaluation of the model's performance. Before training, the features of the color spaces (RGB, Lab, LCh, and CMYK) were standardized using the Standard Scaler method to ensure all features contributed equally to the distance calculation. Critically, the scaler was fitted only on the training set (80%) and then applied to transform both the training and testing sets. This rigorous approach was implemented to prevent Data Leakage from the test set into the training process, thereby ensuring an unbiased model evaluation.

For all feature sets, the Euclidean distance was selected as the distance metric, and the model used uniform weights, where each of the K nearest neighbors contributed equally to the classification decision. To ensure the model's reliability and to prevent overfitting, a 60-fold cross-validation approach was implemented during the training process. A systematic search was performed to identify the best K-value by evaluating the average classification accuracy across the 60 folds over a range from 1 to 50. This process helped to select the optimal model hyperparameters and ensure a robust performance estimate. The final performance of each model was then validated using its unseen testing data. The performance of the models will be evaluated and presented using standard classification metrics, including Accuracy, Precision, Recall, F1-Score, and the Confusion Matrix. This comprehensive evaluation will allow for a detailed comparison of the model's predictive capability across all tested color spaces.

3. Results and Discussion

3.1. Model Performance Comparison

The classification performance of the K-Nearest Neighbors (KNN) model was evaluated across four different color spaces: RGB, Lab, LCh, and CMYK. The optimal number of neighbors (K) for each feature set was determined through a systematic search using 60-fold cross-validation, a more robust approach that yielded slightly improved results. The relationship between the mean cross-validation accuracy and the number of neighbors (K-value) for each color space is illustrated in Figure 2.

The performance results for the optimal K-value of each color space are summarized in Table 2.

Table 2. Comparative Classification Report of KNN Model Performance on Different Color Spaces.

Feature Set	Best k-Value	Precision	Recall	F1-Score	Accuracy
RGB	9	0.9627 (0.9527 ± 0.0292)	0.9625 (0.9464 ± 0.0328)	0.9625 (0.9460 ± 0.0332)	0.9625 (0.9464 ± 0.0328)
Lab	10	0.9690 (0.9506 ± 0.0310)	0.9688 (0.9437 ± 0.0351)	0.9686 (0.9428 ± 0.0364)	0.9688 (0.9437 ± 0.0351)
LCh	6	0.9707 (0.9526 ± 0.0297)	0.9708 (0.9458 ± 0.0332)	0.9707 (0.9451 ± 0.0340)	0.9708 (0.9458 ± 0.0332)
CMYK	7	0.9585 (0.9518 ± 0.0321)	0.9583 (0.9458 ± 0.0351)	0.9584 (0.9455 ± 0.0355)	0.9583 (0.9458 ± 0.0351)

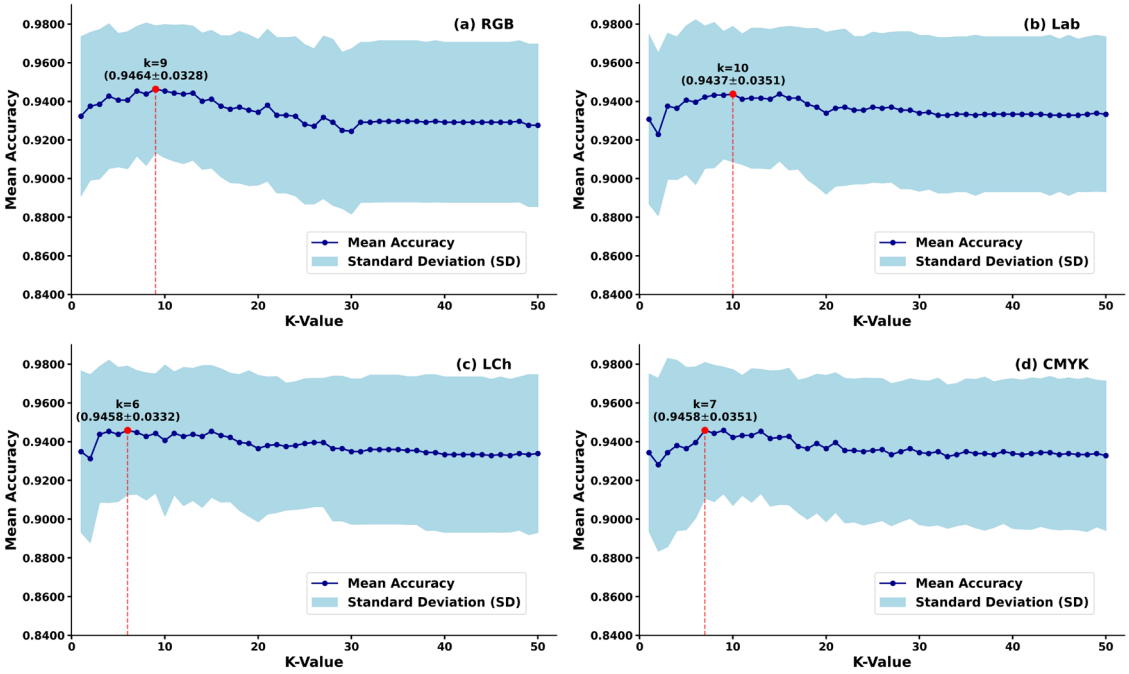

Figure 2. Mean cross-validation accuracy for the KNN model as a function of the number of neighbors (K) for each color space: (**a**) RGB, (**b**) Lab, (**c**) LCh and (**d**) CMYK. The red dotted line in each subplot indicates the optimal K-Value that yields the highest mean accuracy.

As shown in Table 2, while the mean values and Standard Deviation (SD) for all four-color spaces exhibited comparable results, the LCh color space demonstrated the highest test set accuracy overall. LCh achieved an accuracy of 0.9708 at an optimal K-value of 6. This result was closely followed by the Lab color space, which achieved an accuracy of 0.9688, while RGB and CMYK provided strong, though slightly lower, accuracies of 0.9625 and 0.9583, respectively. The high F1-Scores across all models further confirm a strong balance between precision and recall, suggesting that the models are highly effective at correctly identifying positive instances for each of the four nitrogen levels.

The inclusion of SD further confirms the statistical reliability and robustness of these findings. The SD values for all four models are comparably low, ranging narrowly from 0.0328 (RGB) to 0.0351 (CMYK). This low overall variance across the 60 cross-validation folds indicates that all color spaces provide a highly stable and reliable feature representation. Therefore, the direct comparison of the Mean Classification Accuracy is statistically reliable, leading us to conclude that LCh—with its best predictive performance (0.9708) supported by high stability—is the optimal choice for this classification task.

This suggests that representing color data in terms of Luminosity (L), Chroma (C), and Hue (h) is more effective for this specific application than the additive RGB or subtractive CMYK models [19,22]. The superior performance of the LCh color space, derived from CIELab, over the device-dependent RGB and CMYK spaces serves as strong empirical evidence supporting its theoretical robustness against minor fluctuations in ambient lighting [20,22]. Since LCh separates the intensity information (Lightness) from the chromatic information (Chroma and Hue), the model built on these features is inherently less sensitive to the noise that may arise from the low-cost sensor under non-perfectly controlled conditions. This design principle allows the KNN model to focus on the more stable Chroma and Hue features for classification, a finding that is critical for future field applications of colorimetric sensors, where full external lighting control is often impractical.

In a practical context, these results have strong implications for precision agriculture. The proposed system, utilizing readily available and inexpensive hardware (TCS3200

sensor and Arduino Uno) and a robust machine learning model, provides a cost-effective alternative to traditional, expensive laboratory tests. Farmers could potentially use a similar portable device to quickly and accurately assess the nitrogen level of their compost or soil samples, enabling them to make timely and informed decisions regarding fertilization. The use of a rigorous 60-fold cross-validation approach further confirms the reliability of these findings.

3.2. Confusion Matrix Analysis and Discussion

To provide a detailed insight into the classification performance, the confusion matrices for each feature set were analyzed and are presented in Table 3.

Table 3. Confusion matrices for the KNN model trained on each color space: (**a**) RGB, (**b**) Lab, (**c**) LCh, and (**d**) CMYK.

(**a**)					
RGB (k = 9)		Predicted			
		High	Medium	Low	Trace
Real	High	119	1	0	0
	Medium	0	117	3	0
	Low	0	2	110	8
	Trace	0	0	4	116
(**b**)					
Lab (k = 10)		Predicted			
		High	Medium	Low	Trace
Real	High	119	1	0	0
	Medium	0	120	0	0
	Low	0	2	110	8
	Trace	0	0	4	116
(**c**)					
LCh (k = 6)		Predicted			
		High	Medium	Low	Trace
Real	High	120	0	0	0
	Medium	1	119	0	0
	Low	0	2	113	5
	Trace	0	0	6	114
(**d**)					
CMYK (k = 7)		Predicted			
		High	Medium	Low	Trace
Real	High	119	1	0	0
	Medium	0	117	3	0
	Low	0	2	111	7
	Trace	0	0	7	113

The analysis of these matrices reveals both the strengths and weaknesses of each model. While all models performed exceptionally well, the LCh model stands out with a near-perfect classification for the ‘High’ category, correctly identifying all 120 instances.

To provide a quantitative basis for the misclassifications observed between the 'Low' and 'Trace' categories, a perceptual color difference (ΔE_{00}) analysis was performed using the LCh color space data [23]. Analysis of the misclassified samples (6 Trace -> Low and 5 Low -> Trace) yielded a low mean ΔE_{00} value of 0.8206 (Median: 0.5752). As a ΔE_{00} value below 1.0 is generally considered to be perceptually indistinguishable by the human eye, this quantitative finding strongly confirms that the color distinction between the 'Low' and 'Trace' classes is minimal. This validates our assertion that the primary limitation of model stems from the inherent, subtle color overlaps on the physical test kit itself.

The Lab model also performed very well, achieving perfect classification for the 'High' and 'Medium' classes, with 120 samples correctly identified in each. The RGB and CMYK models showed comparable performance, with the majority of misclassifications also concentrated between the 'Low' and 'Trace' classes, though with slightly more errors scattered across other categories.

The confusion matrix in Table 3c highlights the model's accuracy and its specific limitations. The model correctly identified every sample in the 'High' category and only misclassified one sample from the 'Medium' category. The most notable misclassifications occurred between the 'Low' and 'Trace' categories. Specifically, 5 samples from the 'Low' class were incorrectly predicted as 'Trace', and 6 samples from the 'Trace' class were incorrectly predicted as 'Low'. This suggests that the subtle color differences captured by the sensor between these two closely related categories are the most challenging for the model to distinguish.

3.3. Detailed Performance Analysis of the Best Model (LCh)

The overall performance of the best-performing model, trained on the LCh color space, is presented in a detailed classification report and a confusion matrix to provide a comprehensive analysis of its predictive capability.

The detailed classification report in Table 4 reveals that the model performs exceptionally well across all four nitrogen levels. It achieves near-perfect classification for the 'High' and 'Medium' classes, with F1-Scores of 1.00 and 0.99, respectively. The performance for the 'Low' and 'Trace' classes is also very strong, with F1-Scores of 0.95. This indicates a high level of balance between the model's precision (correctly identified positive cases) and recall (correctly identified all positive cases) for each category.

Table 4. Detailed Classification Report for the Best Model (LCh).

Feature Set	Precision	Recall	F1-Score	Support
High	0.99	1.00	1.00	120
Medium	0.98	0.99	0.99	120
Low	0.95	0.94	0.95	120
Trace	0.96	0.95	0.95	120

In comparison to existing literature, our low-cost system (<$20) achieves a high level of accuracy via cross-validation that is competitive with and, in some cases, surpasses more expensive methods (>$2000) or less-controlled approaches [4,23,24]. While high-end spectrophotometers offer near-perfect accuracy, their cost and lack of portability limit practical application [23]. Conversely, accessible smartphone-based systems often face performance challenges due to lighting variability. With an analysis time of less than a minute and a reliability of 97.08% employing a dedicated [5], budget-friendly sensor, our research successfully overcomes the gap between expensive laboratory techniques and erratic on-field evaluations [6,25–27].

Finally, it is crucial to position the accuracy of this low-cost method against comparable studies that leverage optical detection and machine learning for nutrient analysis [7,28,29]. Prior work on similar optical detection approaches reports varied performance: while high accuracy has been achieved with expensive spectroscopic instrumentation [11], a comparable low-cost colorimetric system [30] reported a significantly lower accuracy of 80% and classified soil fertility into only three coarse levels (low, medium, high). In comparison, our proposed system, utilizing the LCh color space with a KNN model, achieved a high multi-class classification accuracy of 97.08% for vermicompost nitrogen levels. This superior performance confirms the efficacy of combining the TCS3200 low-cost sensor with the perceptually uniform LCh features, providing an exceptionally accurate and cost-effective alternative for on-site fertilizer quality assessment.

4. Conclusions

In conclusion, this study successfully demonstrated the effectiveness of the LCh color space as the optimal feature representation for the multi-class nitrogen classification of vermicompost using a low-cost TCS3200 sensor and a KNN model. The LCh model achieved the highest mean accuracy of 0.9708, robustly validating its theoretical advantage in minimizing the influence of varying light conditions.

The combination of high accuracy, minimal hardware cost, and rapid analysis time (requiring less than 1 min per sample) positions this system as a highly scalable and practical solution for agricultural monitoring. The compact and low-power nature of this design makes it an ideal candidate for direct integration with Internet of Things (IoT) platforms or mobile interfaces. This integration would enable real-time, on-field analysis, instant data logging, and immediate fertilization recommendations via a user's smartphone, effectively transforming a high-accuracy laboratory concept into an accessible, field-ready digital tool for precision agriculture. Future work should therefore explore ensemble learning or deep learning (e.g., CNN-based color analysis) for higher generalizability, as well as testing the system under field lighting conditions and expanding the detection capability to include Phosphorus (P) and Potassium (K) to create a comprehensive nutrient monitoring solution.

Author Contributions: Conceptualization, P.L. and S.L.; methodology, P.L. and S.L.; software, P.L.; validation, P.L. and S.L.; formal analysis, P.L. and S.L.; investigation, P.L. and S.L.; resources, P.L. and S.L.; data curation, P.L. and S.L.; writing—original draft preparation, P.L. and S.L.; writing—review and editing, P.L. and S.L.; visualization, P.L.; supervision, S.L.; project administration, S.L.; funding acquisition, P.L. and S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Rajamangala University of Technology Suvarnabhumi (Fundamental Fund: fiscal year 2023 by National Science Research and Innovation Fund (NSRF)) grant number FRB66005/0173-2. And The APC was funded by Rajamangala University of Technology Suvarnabhumi.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to ethical considerations regarding the privacy and consent required for sharing data collected directly from local farmer groups.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Najdenko, E.; Lorenz, F.; Dittert, K.; Olfs, H.W. Rapid in-field soil analysis of plant-available nutrients and pH for precision agriculture—A review. *Precis. Agric.* **2024**, *25*, 3189–3218. [CrossRef]
2. Kim, J.S.; Kim, A.H.; Oh, H.B.; Goh, B.J.; Lee, E.S.; Kim, J.S.; Jung, G.I.; Baek, J.Y.; Jun, J.H. Simple LED spectrophotometer for analysis of color information. *Bio-Med. Mater. Eng.* **2025**, *26*, S1773–S1780. [CrossRef] [PubMed]

3. Abu Hassan, S.N.S.; Masrie, M.; Burham, N. Optical Sensor Analysis of NPK Nutrients Using Visible and Near-Infrared Light. In Proceedings of the 2024 IEEE 22nd Student Conference on Research and Development (SCOReD), Shah Alam, Malaysia, 19–20 December 2024.
4. Caya, M.V.; Arcagua, A.J.; Cleofas, D.R.; Hornada, R.N.; Ballado, A.H., Jr. Soil Survey Implementation Using Rasberry PI Through Colorimetric Determination of Nitrogen, Phosphorus, Potassium, and pH Level. In Proceedings of the 40th Asian Conference on Remote Sensing (ACRS 2019), Daejeon, Republic of Korea, 14–18 October 2019.
5. Kumar, B.V.P.; Tirupathi, P.; Avaniketh, P.; Akhila, M.; Dorthi, K. AI-Powered Adaptive Fertilizer Recommendation System Using Soil And Weather Data. *Int. J. Environ. Sci.* **2025**, *11*, 386–393. [CrossRef]
6. Stevens, J.D.; Murray, D.; Diepeveen, D.; Toohey, D. A low-cost spectroscopic nutrient management system for Microscale Smart Hydroponic system. *PLoS ONE* **2024**, *19*, e0302638. [CrossRef] [PubMed]
7. Han, F.; Huang, X.; Aheto, J.H.; Zhang, D.; Feng, F. Detection of Beef Adulterated with Pork Using a Low-Cost Electronic Nose Based on Colorimetric Sensors. *Foods* **2020**, *9*, 193. [CrossRef]
8. Oliveira, G.C.; Machado, C.C.S.; Inácio, D.K.; Petruci, J.F.S.; Silva, S.G. RGB color sensor for colorimetric determinations: Evaluation and quantitative analysis of colored liquid samples. *Talanta* **2022**, *241*, 123244. [CrossRef]
9. Masrie, M.; Rosman, M.S.A.; Sam, R.; Janin, Z. Detection of nitrogen, phosphorus, and potassium (NPK) nutrients of soil using optical transducer. In Proceedings of the 2017 IEEE 4th International Conference on Smart Instrumentation, Measurement and Application (ICSIMA), Putrajaya, Malaysia, 28–30 November 2017.
10. Silva, F.M.; Jorge, P.A.; Martins, R.C. Optical sensing of nitrogen, phosphorus and potassium: A spectrophotometrical approach toward smart nutrient deployment. *Chemosensors* **2019**, *7*, 51. [CrossRef]
11. Zhang, H.; Wu, Q.; Li, Y.; Xiong, S. Simultaneous detection of nitrate and nitrite based on UV absorption spectroscopy and machine learning. *Adv. UV-Vis-NIR Spectrosc.* **2021**, *36*, 38–44.
12. Senapaty, M.K.; Ray, A.; Padhy, N. IoT-enabled soil nutrient analysis and crop recommendation model for precision agriculture. *Computers* **2023**, *12*, 61. [CrossRef]
13. Albuquerque, J.R.P.; Makara, C.N.; Ferreira, V.G.; Brazaca, L.C.; Carrilho, E. Low-cost precision agriculture for sustainable farming using paper-based analytical devices. *RSC Adv.* **2024**, *14*, 23392–23403. [CrossRef]
14. Li, J.; Wu, Z.; Liang, J.; Gao, Y.; Wang, C. Spectrophotometric-Based sensor for the detection of multiple fertilizer solutions. *Agriculture* **2024**, *14*, 1291. [CrossRef]
15. Wang, P.; Zhang, Y.; Jiang, W. Application of K-Nearest Neighbor (KNN) algorithm for human action recognition. In Proceedings of the 2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference, Chongqing, China, 18–20 June 2021.
16. Alstrøm, T.S.; Larsen, J.; Kostesha, N.V.; Jakobsen, M.H.; Boisen, A. Data representation and feature selection for colorimetric sensor arrays used as explosives detectors. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Beijing, China, 18–21 September 2011.
17. Syriopoulos, P.K.; Kalampalikis, N.G.; Kotsiantis, S.B.; Vrahatis, M.N. *k*NN Classification: A review. *Ann. Math. Artif. Intell.* **2025**, *93*, 43–75. [CrossRef]
18. Uddin, S.; Haque, I.; Lu, H.; Moni, M.A.; Gide, E. Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. *Sci. Rep.* **2022**, *12*, 6256. [CrossRef]
19. Fay, C.D.; Wu, L. Critical importance of RGB color space specificity for colorimetric bio/chemical sensing: A comprehensive study. *Talanta* **2024**, *266*, 124957. [CrossRef]
20. Yang, R.; Cheng, W.; Chen, X.; Qian, Q.; Zhang, Q.; Pan, Y.; Duan, P.; Miao, P. Color Space Transformation-Based Smartphone Algorithm for Colorimetric Urinalysis. *ACS Omega* **2018**, *3*, 12141–12146. [CrossRef]
21. Tania, M.H.; Shabut, A.M.; Lwin, K.T.; Hossain, M.A. Clustering and classification of a qualitative colorimetric test. In Proceedings of the 2018 International Conference on Computing, Electronics & Communications Engineering, Southend, UK, 16–17 August 2018.
22. Velastegui, R.; Pedersen, M. CMYK-CIELAB Color Space Transformation Using Machine Learning Techniques. *Lond. Imaging Meet.* **2021**, *2021*, 73–77. [CrossRef]
23. Samec, J.; Štruc, E.; Berzina, I.; Naglič, P.; Cugmas, B. Comparative Analysis of Low-Cost Portable Spectrophotometers for Colorimetric Accuracy on the RAL Design System Plus Color Calibration Target. *Sensors* **2024**, *24*, 8208. [CrossRef]
24. Trinh, V.Q.; Babilon, S.; Myland, P.; Khanh, T.Q. Processing RGB color sensors for measuring the circadian stimulus of artificial and daylight light sources. *Appl. Sci.* **2022**, *12*, 1132. [CrossRef]
25. Alberti, G.; Zanoni, C.; Magnaghi, L.R.; Biesuz, R. Disposable and Low-Cost Colorimetric Sensors for Environmental Analysis. *Int. J. Environ. Res. Public Health* **2020**, *17*, 8331. [CrossRef]
26. Diaz, F.J.; Ahmad, A.; Parra, L.; Sendra, S.; Lloret, J. Low-Cost Optical Sensors for Soil Composition Monitoring. *Sensors* **2024**, *24*, 1140. [CrossRef]

27. Liu, N.; Wei, Z.; Wei, H. Colorimetric detection of nitrogen, phosphorus, and potassium contents and integration into field irrigation decision technology. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *651*, 042044. [CrossRef]
28. Deshpande, S.; Nidoni, U.; Hiregoudar, S.; Ramappa, K.T.; Maski, D.; Naik, N. Performance of advanced machine learning models in the prediction of amylose content in rice using internet of things-based colorimetric sensor. *Curr. Sci.* **2023**, *124*, 722–730. [CrossRef]
29. Yamin, M.; Ismail, W.I.B.W.; Kassim, M.S.B.M.; Aziz, S.B.A.; Akbar, F.N.; Shamshiri, R.R.; Ibrahim, M.; Mahns, B. Modification of colorimetric method based digital soil test kit for determination of macronutrients in oil palm plantation. *Int. J. Agric. Biol. Eng.* **2020**, *3*, 188–197. [CrossRef]
30. Agarwal, S.; Bhangale, N.; Dhanure, K.; Gavhane, S.; Chakkarwar, V.A.; Nagori, M.B. Application of colorimetry to determine soil fertility through naive bayes classification algorithm. In Proceedings of the 9th International Conference on Computing, Communication and Networking Technologies, Bengaluru, India, 10–12 July 2018.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.