J-KOMA Journal of Computer Science and Applications

Universitas Negeri Jakarta
Phone: (021) 4890856
East Jakarta City
Jakarta, 13220

Open Access

E-ISSN: *2620 - 4872*
Vol.*08*, ***No***.*02*
Doi: *10.21009/JKOMA.082.05*

Recivied: *16 Dec 2025*
Accepted: *22 Dec 2025*
Published: *26 Dec 2025*

Keywords:
KNN;
PCA;
Color image;
Rice bran;
Rice husk.
****Correspondence Email:***
azizku@apps.ipb.ac.id

Utilizing KNN for Estimating Lignin in Rice Bran through Color Imagery with PCA Preprocessing

Rijal Triadi Sutrisno[1], Aziz Kustiyo[2*]

[1,2]Computer Science Study Program, IPB University, Indonesia.

Abstract
Feed is essential for enhancing livestock production, particularly in maintaining animal health and stamina. Rice bran is commonly used as animal feed; however, its quality can decline when mixed with other ingredients, such as rice husks. The addition of rice husks to rice bran increases the levels of crude fiber and lignin, which are difficult for livestock to digest and can lead to health issues. This mixing can be assessed by estimating the lignin content through the phloroglucinol dye reaction. This study aimed to estimate the lignin content in a mixture of rice bran and rice husks using the dye reaction and the resulting color images. The images were captured using the red-green-blue (RGB) color model. A feature extraction technique called principal component analysis (PCA) was employed on each RGB component. The results from the PCA were subsequently classified using the k-Nearest Neighbor (KNN) algorithm. The findings indicated that the red (R) color component yielded the highest classification accuracy of 77.27%.

INTRODUCTION

Feed refers to various food ingredients provided to livestock, including corn, rice bran, pollard, coconut meal, soybean meal, and fish meal [1]. It plays a crucial role in the success of a farm, as it is essential for enhancing livestock production by maintaining health and endurance [2].

To improve the quality and safety of feed distributed for animals, humans, and the environment, the Indonesian government, through the Directorate General of Animal Husbandry and Animal Health of the Ministry of Agriculture, issued the Regulation of the Minister of Agriculture of the Republic of Indonesia Number 22/Permentan/PK.110/6/2017 concerning the Registration and Distribution of Feed [3]. This regulation stipulates that all feed produced for distribution, whether sold or not, must possess a Feed Registration Number (NPP) and a Feed Quality and Safety Certificate. The quality and safety of feed are assessed based on the nutritional and anti-nutritional content of the ingredients, which must comply with the Indonesian National Standard or the Minimum Technical Requirements for feed without an SNI [4].

Despite these regulations, several factors can lead to a decline in feed quality, resulting in either damage to the feed or contamination. A prevalent issue is feed adulteration, where inferior ingredients are mixed in by suppliers or traders seeking higher profits. This practice involves adding other ingredients that closely resemble the original ones in terms of criteria and physical properties ([5] ; [6]).

Feed quality, especially that of rice bran, can decline when it is mixed with other ingredients like ground rice husks. The addition of ground rice husks to rice bran can increase the levels of crude fiber and lignin, which are challenging for poultry to digest . Lignin is a significant factor in preventing the enzymes produced by microbes from effectively digesting feed ingredients. This is because lignin binds with cellulose to form a robust lignocellulose bond, making it very difficult for rumen microbes to break down [7].

Universitas Negeri Jakarta
Phone: (021) 4890856
East Jakarta City
Jakarta, 13220

The quality of rice bran can be assessed through various tests, including organoleptic testing, which relies on direct observations using the five senses: sight, smell, and touch. The phloroglucinol test involves dissolving phloroglucinol in rice bran and using the resulting color to determine the quality of the rice bran [7]. While physical testing is quick and relatively inexpensive, the accuracy of the data obtained may be less reliable [8]. Estimating lignin content through a dyeing reaction takes a longer time, requiring at least two days.

Research on estimating lignin content using the K-nearest neighbors (KNN) approach was conducted by [9]. In this study, the KNN classification method utilized color components and histogram intervals to estimate the lignin content in rice bran. The most effective results were achieved with 64 histogram intervals and by focusing on the green color component (G). The highest prediction accuracy recorded using the KNN method was 72.12% [9].

Research conducted by [10] develop an automatic classification system for determining the ripeness of Robusta coffee cherries using digital images. This system employed the K-Nearest Neighbor (KNN) algorithm and was optimized through Principal Component Analysis (PCA). The study examined three levels of ripeness for Robusta coffee cherries and achieved a peak accuracy of 93.33%. Similarly, research by [11] developed a PCA-KNN method for classifying medical images to detect brain tumors, lung cancer, and kidney lesions, achieving an accuracy of 90%. Additionally, researh by [12] focused on classifying ripeness levels in matoa fruit based on RGB color features. By using PCA and KNN methods, this study reached an accuracy of 92%.

This study aims to estimate lignin content using the K-Nearest Neighbors (KNN) algorithm, with Principal Component Analysis (PCA) applied for data preprocessing. The analysis will utilize RGB color images captured from a mixture of rice bran and rice husks, which has been treated with a phloroglucinol solution.

METHOD

In general, the research began with rice bran images obtained from a study by [9]. The rice bran images were initially in color, using the RGB model. Each image was then separated into its red (R), green (G), and blue (B) components. Feature extraction was performed on each component, followed by dimensionality reduction using principal component analysis. Finally, classification was conducted using the K-Nearest Neighbors (KNN) method. The stages of the research are illustrated in Figure 1.

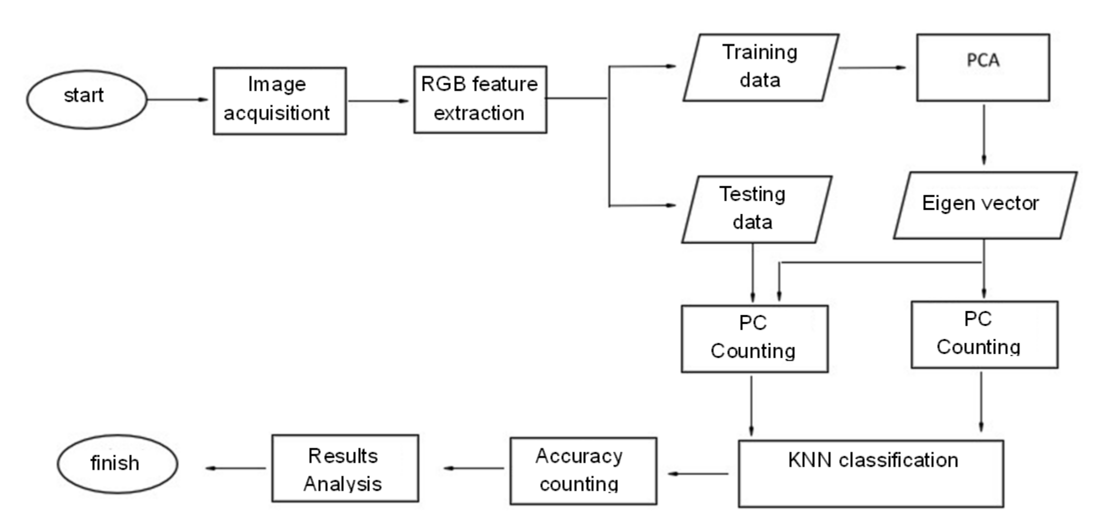


FIGURE 1. Stages of the Research Process

Image Data Acquisition

The data utilized in this study are identical to those employed by [9], specifically 11 images captured with a 5184×3456 pixel digital camera. The image acquisition took place in the Feed Science and Technology Laboratory, which is part of the Department of Nutrition Science and Feed Technology at the Faculty of Animal Husbandry, IPB University. A white paper was used as the background during the image

Universitas Negeri Jakarta
Phone: (021) 4890856
East Jakarta City
Jakarta, 13220

acquisition process. The rice husks were ground before being mixed with rice bran. A mixture containing 10 grams of rice bran and rice husks of different compositions was prepared, using 1 gram of this mixture for the dyeing process, along with 3.5 ml of phloroglucinol solution. The composition of rice bran mixed with rice husk content was shown in Table 1.

TABLE 1. The composition of rice bran mixed with rice husk content

No	Sekam (gram)	Dedak (gram)	Lignin (%)
1.	0	10	10.19
2.	1	9	10.39
3.	2	8	10.59
4.	3	7	12.31
5.	4	6	12.81
6.	5	5	13.44
7.	6	4	13.50
8.	7	3	14.31
9.	8	2	16.00
10.	9	1	16.07
11.	10	0	16.16

Source: Mutya et al. (2022)

Before conducting further analysis, we collected 10 samples, each measuring 100x100 pixels, from each of the 11 acquired images. This process resulted in a total of 110 image samples.

The lignin content data presented in Table 1 indicates that several compositions of the bran and husk mixtures had similar lignin values. Therefore, for classification purposes, we combined the compositions with closely related lignin content values into one class. This resulted in four distinct classes, as illustrated in Table 2.

TABLE 2. The lignin content of 4 classes

No	Class	Lignin (%)
1.	1	10.19 – 10.59
2.	2	12.31 – 12.81
3.	3	13.44 – 14.31
4.	4	16.00 – 16.16

RGB Feature Extraction

The feature extraction stage in this study involves analyzing the RGB color components of each image and calculating the values for each. The RGB color model divides an image into three components: Red (R), Green (G), and Blue (B) [13]. The calculation process happens sequentially—first for the Red component, then the Green, and finally the Blue. This process results in a 100x100 matrix being generated for each color component, containing specific values.

Data Splitting

The image data of rice bran was divided into two parts: training data and testing data. The training data was utilized for data reduction using Principal Component Analysis (PCA) [14], followed by classification using the k-nearest neighbor (K-NN) algorithm. The testing data was then employed to evaluate the K-NN classification results. In this study, three testing processes were carried out, involving 88 training images and 22 testing images. Each of the three testing processes used different compositions of training and testing data.

Counting Eigenvectors and Principal Components(PCs)

Each RGB color component of the training data, represented as a 100x100 image, was converted into a 1x10,000 row vector. This transformed data was then analyzed using Principal Component Analysis (PCA), a technique that reduces a high-dimensional set of variables to a smaller set while preserving as much information as possible from the original data. The output of PCA is a new set of variables known as principal components (PCs), which replace the original variables (X) in subsequent analyses. PCA was conducted on the training data, and the test data utilized the eigenvectors generated from this training data. As a result, new variables called principal components were produced, significantly decreasing the number of variables from the initial 10,000. According to [15], the minimum number of PCs to retain is the number of PCs that represents cumulative contribution of 70. In this study, the number of PCs retained represented the maximum cumulative contribution of approximately 90% for each RGB color component.

K-Nearest Neighbors (KNN) Classification

After completing the principal component (PC) calculation, the results will be classified using the K-Nearest Neighbors (KNN) algorithm. The fundamental principle of KNN is to identify the shortest distance between the data being evaluated and its k nearest neighbors in the training dataset, where k represents the number of neighbors considered. The distance measure used in this study is the Euclidean distance. The Euclidean distance between two points, $X_1 = (x_{11}, x_{12},..., x_{1n})$ and $X_2 = (x_{21}, x_{22},..., x_{2n})$, is shown in equation 1 [16].

$$dist(\boldsymbol{X}_1, \boldsymbol{X}_2) = \sqrt{\sum_{i=1}^{n} (x_{1i} - x_{2i})^2} \tag{1}$$

The application of KNN requires determining the K value, which is the number of neighbors needed to determine the class of the test data. According to [16], the K value can be determined through experimentation. In this study, the K value was determined simultaneously with the number of PCs that produced the highest KNN accuracy, with K values ranging from 1 to 30.

Accuracy Calculation

After classification is carried out using KNN, the next step is to carry out accuracy calculation using the equation 2.

$$accuracy = \frac{number\ of\ data\ correctly\ classified}{number\ of\ testing\ data}\ x\ 100\% \tag{2}$$

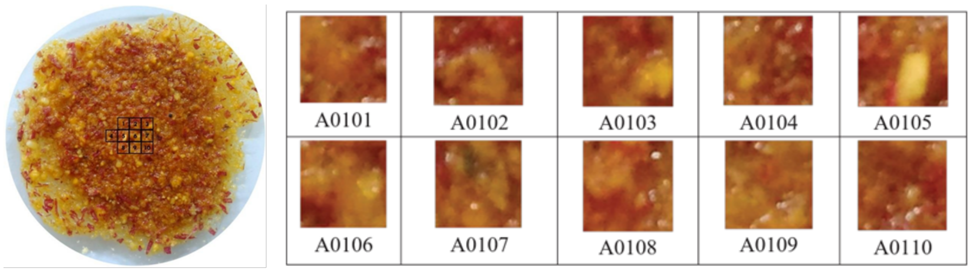
RESULTS AND DISCUSSION

Sample Data

This study utilized images sourced from research conducted by [9]. A total of 11 images of rice bran, each measuring 5184 × 3456 pixels and representing different compositions of rice bran and husk, were included. From each image, 10 samples measuring 100 × 100 pixels were extracted, resulting in a total of 110 images. Figure 2 shows original image and sample image of rice bran. The distribution of the number of images for each class is shown in Table 3.

RGB Color Extraction and PCA

In this section, RGB color feature extraction was performed on the 100x100 pixels images. The image, represented in the RGB color model, is divided into its three color components: Red (R), Green (G), and Blue (B). After extraction, each image is transformed into a 100 x 100 matrix.



(a) (b)

FIGURE 2. (a) Original image of 5184 × 3456 pixels (b) Sample image of 100x100 pixels

TABLE 3. Number of images for each class

No	Class	Number of images
1.	1	30
2.	2	20
3.	3	30
4.	4	30

Next, these matrices are reshaped into row vectors. Subsequently, the eigenvalues and eigenvectors were calculated, resulting in a total of 87 eigenvalues, along with their cumulative contributions. For the K-nearest neighbors (KNN) process, the Principal Components (PC) selected range from 1 to 50, as the maximum cumulative contribution in this interval approximately 90%. Table 4 presents the eigenvalues and their contributions for the first 10 eigenvalues.

TABLE 4. The eigenvalues and their contributions for the first 10 eigenvalues for each RGB components

No	R		G		B	
	Eigen value	Contribution* (%)	Eigen value	Contribution* (%)	Eigen value	Contribution* (%)
1.	38.77	38.77	40.29	40.29	47.56	47.56
2.	5.53	44.31	5.18	45.47	4.37	51.94
3.	4.25	48.57	4.72	50.19	3.57	55.51
4.	3.74	52.31	4.18	54.38	3.07	58.58
5.	3.64	55.96	3.23	57.61	2.44	61.03
6.	2.86	58.83	2.90	60.52	2.33	63.36
7.	2.55	61.39	2.72	63.25	2.09	65.46
8.	2.32	63.72	2.38	65.64	1.86	67.32
9.	1.88	65.61	1.94	67.59	1.66	68.98
10.	1.79	67.40	1.92	69.51	1.48	70.47

*cumulative contribution

KNN Test Results

The experiment to determine the K value in KNN and the number of PCs that produce the highest KNN accuracy per RGB color component consists of 1500 combinations (30 K values and 50 PC numbers). The combination of K values from KNN and the number of PCs for the highest accuracy per RGB color component is presented in Table 5 while the accuracy value is presented in Table 6. From Table 5 it can be seen that the highest KNN accuracy mostly occurs at the value of K = 1. Meanwhile, for the number of PCs, it has various values with a minimum value of 3 and a maximum value of 37.

Universitas Negeri Jakarta
Phone: (021) 4890856
East Jakarta City
Jakarta, 13220

TABLE 5. The Combinations of K value for KNN and number of PCs for the highest accuracy of KNN for each RGB color components

RGB component	First testing		Second testing		Third testing	
	K*	PC**	K*	PC**	K**	PC**
R	22	6	1	37	1	9
G	1	23	1	23	1	3
B	3	4	1	4	4	3

K* = K value of KNN
PC** = number of principal components

The average accuracy of the K-Nearest Neighbors (KNN) algorithm, analyzed by color component, shows that the red color component (R) achieves the highest average accuracy at 74.24%. In contrast, the accuracies for the other color components fall below 70%, as shown in Table 6. The table also indicates that the maximum KNN accuracy for the red component (R) is 77.27%, for the green component (G) it is 72.73%, and for the blue component (B) it is 73.00%. These results confirm that the red color component (R) yields the highest KNN accuracy among the three components. This is likely attributed to the characteristics of the image data used, where higher lignin content correlates with a more pronounced red hue in the resulting images.

TABLE 6. The average accuracy of the K-nearest neighbors (KNN) algorithm using RGB color components

RGB component	Accuracy (%)			
	First testing	Second testing	Third testing	Average accuracy(%)
R	77.27	72.73	72.73	74.24
G	72.73	72.73	63.64	69.70
B	63.64	63.64	72.73	66.67
Average accuracy(%)	71.21	69.70	69.70	

Table 7 presents the confusion matrix for the highest accuracy achieved, which is 77.27%. This accuracy was obtained from the first test concerning the red color component (R). Overall, prediction errors predominantly occurred between adjacent classes. For instance, class 2 was mistakenly predicted as class 1 and class 3, while class 4 was incorrectly classified as class 3. These results indicate that PCA and KNN performed well in classifying images of rice bran mixed with rice husks.

TABLE 7. The confusion matrix of the 77.27% accuracy of the first KNN testing

Actual class	Predicted class			
	1	2	3	4
1	6	0	0	0
2	2	1	1	0
3	0	0	6	0
4	0	1	1	4

Regarding lignin content per class, Table 2 reveals that class 1 has a lignin content ranging from 10.19% to 10.59%, whereas the other three classes have lignin contents exceeding 12%. When comparing this information to the confusion matrix results, it is evident that class 1 was correctly predicted 100% of the time. In contrast, only two data points, or approximately 9%, were misclassified as class 1 among the other three classes. This suggests that KNN can effectively distinguish between images of rice bran with low lignin content (around 10%) and those with higher lignin content (more than 12%).

CONCLUSIONS AND SUGGESTIONS

Based on the conducted research, it can be concluded that utilizing the PCA method for feature extraction, combined with the KNN classification technique, is effective for identifying colored images and estimating the lignin content in rice bran. The best accuracy achieved was 77.27% for the red (R) color component. Furthermore, the KNN method successfully distinguishes class 1 (around 10% of low lignin content) from other classes with higher lignin content, particularly classes 3 and 4 (more than 13% of lignin content).

This study has certain limitations that could be addressed in future research. Future studies may explore additional classification techniques, such as Probabilistic Neural Networks (PNN) or Artificial Neural Networks (ANN). Additionally, image acquisition methods could be enhanced through the use of more advanced cameras. It is also recommended that the images considered extend beyond rice bran mixed with husks to include other mixtures, such as sawdust.

REFERENCES

[1] Ramayulis, Salvia, and M. Dewi, *Ransum Ruminansia*. Lima Puluh Kota: Politeknik Pertanian Negeri Payakumbuh, 2023.

[2] Dwitara Wulandari Sihombing, Muhammad Aidil Ghifari, Fitra Haikal, Said M. Sajahtra, Kiagus Muhammad Zain Basriwijaya, and Faoeza Hafiz Saragih, “Pengaruh Suplemen Pakan Terhadap Kesehatan dan Pertumbuhan Sapi Pedaging,” *Botani : Publikasi Ilmu Tanaman dan Agribisnis*, vol. 2, no. 1, pp. 290–295, Dec. 2024, doi: 10.62951/botani.v2i1.205.

[3] Ministry of Agriculture, “PermentanNo 22 Tahun 2017,” 2017. Accessed: Dec. 20, 2025. [Online]. Available: https://peraturan.bpk.go.id/Details/161900/permentan-no-22permentanpk11062017-tahun-2017

[4] Ministry of Agriculture, “Regulasi Untuk Tingkatkan Penjaminan Mutu dan Keamanan Pakan Ternak yang Beredar,” https://www.pertanian.go.id/home/?show=news&act=view&id=2228.

[5] A. M. Rusyidi, “Evaluasi kualitas dedak padi sebagai bahan pakan yang terkontaminasi sekam padi,” Universitas Hasanuddin, Makasar, 2022.

[6] C. S. Utama and B. Sulistiyanto, “Kajian pemalsuan bekatul dan tepung di wilayah Jawa Tengah,” *Livestock and Animal Research*, vol. 19, no. 1, pp. 32–39, 2021.

[7] U. Rosani, I. Hernaman, R. Hidayat, and D. Hidayat, “Karakterisasi Dedak Padi dan Campuran Sekam Padi berdasarkan Sifat Fisik dan Kimia,” *Jurnal Agripet*, vol. 24, no. 1, pp. 14–22, Apr. 2024, doi: 10.17969/agripet.v24i1.33257.

[8] M. Ridla *et al.*, “Korelasi Sifat Fisik dan Kandungan Nutrien Dedak Padi Correlation of Physical Characteristics and Nutrient Content of Rice Bran,” *Jurnal Peternakan*, vol. 20, no. 1, pp. 1–8, 2023, doi: 10.24014/jupet.v20i1:18374.

[9] H. M. H. Mutya, A. Kustiyo, and A. Jayanegara, “Estimasi kandungan lignin pada dedak padi yang bercampur sekam mengggunakan KNN berbasis warna citra,” in *Prosiding Seminar Nasional MIPA UNIPA*, 2022, pp. 95–101.

[10] Ayu Wulandari, Rudi Heriansyah, and Lastri Widya Astuti, “Implementasi Algoritma K-nearest Neighbor dan Principal Component Analysis untuk Klasifikasi Tingkat Kematangan Ceri Kopi Robusta Berdasarkan Warna,” *Jurnal Komputer, Informasi dan Teknologi*, vol. 5, no. 2, p. 9, Oct. 2025, doi: 10.53697/jkomitek.v5i2.3048.

[11] N. T. Ujianto, Gunawan, H. Fadillah, A. P. Fanti, A. D. Saputra, and I. G. Ramadhan, “Penerapan algoritma K-Nearest Neighbors (KNN) untuk klasifikasi citra medis,” *IT-Explore: Jurnal Penerapan Teknologi Informasi dan Komunikasi*, vol. 4, no. 1, pp. 33–43, Feb. 2025, doi: 10.24246/itexplore.v4i1.2025.pp33-43.

[12] R. Efando and A. Sonita, “Klasifikasi Tingkat Kematangan Buah Matoa Menggunakan Metode PCA dan KNN Berdasarkan Warna RGB,” *Journal Scientific and Applied Informatics* , vol. 8, no. 2, pp. 337–346, 2025.

[13] R. C. . Gonzalez and R. E. . Woods, *Digital image processing*, 4th ed. London: Pearson, 2018.

[14] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, “MULTIVARIATE DATA ANALYSIS EIGHTH EDITION,” 2019. [Online]. Available: www.cengage.com/highered

[15] I. T. Jollife and J. Cadima, “Principal component analysis: A review and recent developments,” Apr. 13, 2016, *Royal Society of London*. doi: 10.1098/rsta.2015.0202.

Universitas Negeri Jakarta
Phone: (021) 4890856
East Jakarta City
Jakarta, 13220

[16] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Burlington, MA, USA: Morgan Kaufmann Publishers, 2011.