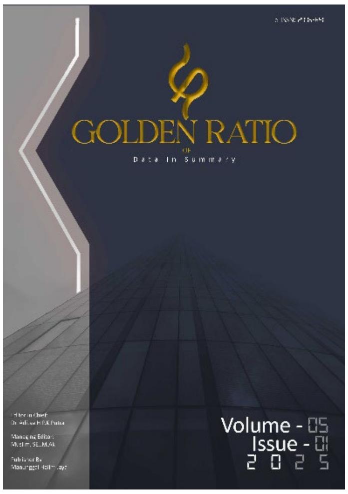
Rezkya Nadilla Putri *et al.*, *Golden Ratio of Data in Summary*, Vol.5, Issue.1 (2025)

https://doi.org/10.52970/grdis.v5i1.910

Website: https://goldenratio.id/index.php/grdis

OPEN ACCESS

ISSN [Online]: 2776-6411




Received: December 17, 2024
Revised: January 01, 2025
Accepted: February 01, 2025

*Corresponding author: Rezkya Nadilla Putri1, Department of Mathematics, Universitas Negeri Medan, Medan, Indonesia.

E-mail: rezkyanadillap@gmail.com

DATA IN SUMMARY | ENGINEERING, MATHEMATICS AND STATISTICS

# Classification of Pear Varieties Using the K-Nearest Neighbor Algorithm and Extraction of Shape, Color, Texture, and Size Features

**Rezkya Nadilla Putri[1]*, Dedy Kiswanto[2], Keysa Shifa Adwitia Sitepu[3]**

[1,2,3] Department of Mathematics, Universitas Negeri Medan, Medan, Indonesia.
Email: rezkyanadillap@gmail.com[1*], dedykiswanto@unimed.ac.id[2], chacasitepuu@gmail.com[3]

**Abstract:** This study develops a pear variety classification system based on digital images using the K-Nearest Neighbor (KNN) algorithm. The data used included 195 images from three pear varieties, namely Century, Forel Afrika, and Singo, which were analyzed by utilizing various features such as color (RGB), texture (Local Binary Pattern), shape (area, circumference, length-width ratio), and size (bounding box dimensions). The preprocessing process removes the image's background to increase focus on the main object, thus allowing for more optimal feature extraction. The dataset is divided into 80% for training and 20% for model testing. The evaluation results show that the KNN model can achieve an accuracy of 85%, with an average precision value of 0.85, recall of 0.89, and F1-score of 0.85. These results prove that the KNN algorithm is effective in accurately classifying pear varieties, which can significantly contribute to applying digital image-based technology for automatic classification needs in the agricultural sector.

Keywords: Pear Classification, K-Nearest Neighbor Algorithm, Digital Imagery, Feature Extraction.

## 1. INTRODUCTION

Pears are an agricultural commodity with high economic value in many countries, including Indonesia. Pear (Pyrus bretschneideri) is one of the most popular fruits in Indonesia, and the public often consumes it. Its popularity can be seen from the high consumption of the public, which is reflected in the import volume of 69,000 tons from countries such as China, Australia, South Korea, and the United States in 2012 (Setyawan et al., 2023). Pears also have considerable water content in fruit and peel (Arthania et al., 2021). Pear (Pyrus) is one of the fruits rich in nutrients, containing various nutrients such as vitamins, niacin, pantothenic acid, and folacine. In Indonesia, pears are many people's favorite because of their unique taste, juicy characteristics, master texture, and sweet taste. Each pear variety has its characteristics that distinguish it in terms of taste and specialties. This makes each type of pear priced differently by consumers according to their respective preferences (Juliansyah & Laksito, 2021). Pears have several different types and different shapes for each type. This can also be done by recognizing pear types using artificial intelligence (Paniza et al., 2021). The diversity of pear varieties, spread across different parts of the world, reflects genetic adaptation to different geographical environments, making it an interesting subject for further research.

The advancement of the times that continues to develop rapidly has brought significant improvements in the technology field. This innovation penetrates various sectors, including agriculture and food. Regarding fruit identification and classification, digital image-based methods are beginning to replace manual approaches. This technology allows the process to be faster, more efficient, and more accurate so that human time and workforce can be allocated to other activities that are more valuable and productive (Inzaghi et al., 2024). Using image processing techniques,

https://doi.org/10.52970/grdis.v5i1.910

Website: https://goldenratio.id/index.php/grdis

ISSN [Online]: 2776-6411

automated systems can recognize and classify fruits based on visual features such as color, texture, and shape. In addition, machine learning-based algorithms can improve accuracy in data grouping (Sya'ban et al., 2022). This makes this method a potential solution in the classification of pear varieties.

Until now, there are various methods or algorithms used in the classification process, one of which is k-Nearest Neighbor (k-NN) (Pawening et al., 2020). The K-Nearest Neighbors (KNN) algorithm is a classification method that determines the class of an object based on the value (K) of its closest neighbors (Reswan et al., 2024). The KNN algorithm is known for its simplicity in processing training and test data, even in large quantities (Cholil et al., 2021). In addition, K-NN also has the advantage of quickly classifying data, resisting noise in training data, and producing accurate classification (Nuraeni et al., 2023). With these various advantages, K-NN can effectively classify pear varieties, utilizing image data to recognize and classify pear types more efficiently and accurately. With the rapid development of technology, digital imagery has become an effective tool for identifying and classifying various objects with precision (Irwan et al., 2022). Research shows that digital imagery can be analyzed through three main features: color, shape, and texture, each of which makes a significant contribution to describing the characteristics of objects (Fahrurozi et al., 2023). Analysis of these parameters allows the system to recognize subtle differences between varieties, which is often tricky for human observation to capture. For example, the texture of a pear peel can be deciphered using the Gray Level Co-occurrence Matrix (GLCM) method, which detects pixel distribution patterns in detail. On the other hand, color analysis uses color models such as RGB or HSV, which can identify shades of color more accurately, supporting the classification process more efficiently and reliably.

Although research in digital image-based fruit classification offers a wide range of opportunities, it still faces significant challenges. One of the main challenges is the variation in lighting and shooting angles, which can affect the accuracy of the analysis. Background complexity, lighting fluctuations, and different viewing angles require an in-depth analysis approach to thoroughly evaluate the model's performance (Tikasni et al., 2024). In addition, multi-parameter data, such as color and texture combinations, presents new challenges in data processing. This requires algorithms that are efficient and capable of managing data on a scale. Therefore, an adaptive, fast, and accurate approach is indispensable for complex data dynamics. This study aims to overcome these obstacles by developing a more robust method and optimizing multi-parameter analysis to improve the accuracy of pear variety classification.

## 2. RESEARCH DESIGN AND METHOD

### *2.1. Data Collection*

This study utilizes images from three pear varieties: Century Pear, African Forel Pear, and Singo Pear. From each array, 65 photos were taken, so a total of 195 images were used in this study. The shooting was done using the iPhone 11 rear camera with a 1:1 ratio. Additionally, a white background is used consistently to ensure that the focus is on the outer skin of the pear without distraction from other visual elements. This approach was chosen to maximize the quality of the resulting images so that the extraction process of color, texture, shape, and size features can be carried out with higher precision. This strategy aims to support optimal data analysis, ensuring that every critical characteristic of the pear variety can be accurately identified (Wijaya & Ridwan, 2019).

### *2.2. Preprocessing Data*

At this stage, the researcher preprocessed the pear image (Saputro & Sumantri, 2022). The preprocessing process includes removing the image's background using a web-based platform. This background removal ensures that the pear object, specifically the outer shell, is more precise and focused. Thus, essential features such as color, texture, shape, and size can be extracted more accurately, resulting in more optimal data for the classification process. After the background is

https://doi.org/10.52970/grdis.v5i1.910
Website: https://goldenratio.id/index.php/grdis
ISSN [Online]: 2776-6411

removed, the feature extraction is carried out. In the feature extraction stage, important information from the pear image is captured using several methods. First, the RGB feature is calculated by measuring the color intensity histogram on three component channels, Red, Green, and Blue (Batubara et al., 2020), to describe the color of the pear skin. Second, the Local Binary Pattern (LBP) feature is calculated to analyze the surface texture of the pear by converting the image into a frequency-calculated local binary pattern, providing a stable texture representation. Third, the size features are extracted by calculating the height and width of the image, as well as the length-width ratio of the image, and fourth, the shape features are extracted by calculating the contour area and contour circumference, which describes the physical dimensions and shape of the pear based on the detected contours.

### 2.3. Modelling

The classification model is built and trained using datasets extracted from its features in the modeling stage. First, the K-Nearest Neighbors (KNN) model is constructed using features generated from RGB, Local Binary Patterns (LBP), Size, and Shape. The extracted data is used as input for the KNN model. The model determines the class of the new data based on most of its closest "neighbors." The process involves calculating the distance between data unknown to its class and previously classified data, usually using the Euclidean distance. The parameter K, a positive integer, determines the number of neighbors considered in this process. Choosing the right K-value is essential for improving the model's accuracy, usually by trying different K-values during the testing phase. Once the KNN model is prepared, the dataset is divided into 80% for training and 20% for testing (Fadli et al., 2024). This division aims to train the model with most of the data, while the rest is used to evaluate its ability to classify new data. This approach is essential to prevent overfitting when the model is too adaptable to the training data and cannot handle the new data well.

### 2.4. Model Evaluation

The model was evaluated by measuring several metrics such as accuracy, precision, recall, and F1-score. The confusion matrix is also calculated to give an idea of the model's performance and identify the distribution of predictions and proper labels.

## 3. RESULT AND DISCUSSION

### 3.1. Data Collection

This study used data from three pear varieties: Century Pear, African Forel Pear, and Singo Pear. From each array, 65 pictures were taken. The entire image was taken using the iPhone 11's rear camera with a 1:1 ratio, with an adjusted white background to ensure focus on the outer skin of the pear. The following will display the results of taking pictures of each pear variety.

**Figure 2. African Forel Pear**

https://doi.org/10.52970/grdis.v5i1.910

ISSN [Online]: 2776-6411

Figure 3. Pear Singo

Figure 4. Pear Century

### *3.2. Preprocessing Data*

In data preprocessing, background removal and feature extraction are performed. Background removal is done manually using a website-based platform. The following will display the results of the background removal from the image image.

Figure 5. Results of Background Removal of African Forel Pears

Figure 6. Results of Pear Singo Background Removal

Figure 7. Results of Pear Century Background Removal

https://doi.org/10.52970/grdis.v5i1.910

Website: https://goldenratio.id/index.php/grdis

ISSN [Online]: 2776-6411

Once the background removal is complete, the next step is the feature extraction process, which includes the image's color, texture, shape, and size features. This extraction is performed directly on the original image to ensure accuracy than previous extraction results. All the extracted features are then combined into a single vector or array as a representation of the data to be used in the training and testing stages of the model. This process is crucial to generating optimal data, which will be explained later.

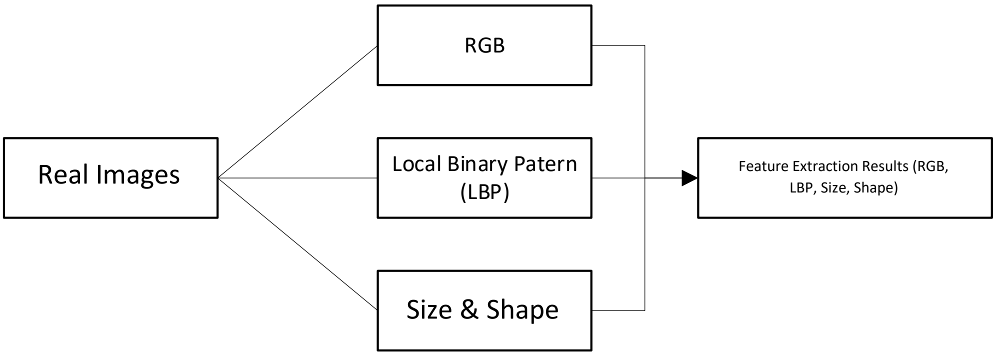



**Figure 8. Feature Extraction Process**

### *3.3. Modelling*

In the modeling stage, the KNN algorithm is implemented using the data from feature extraction that has been carried out previously. The data is divided into training data and test data with a ratio of 80:20. The selected K value is 3; the optimal selection of the K value is essential to improve the model's accuracy, and it is usually achieved by trying multiple K values during the testing process. After the dataset is divided, the next thing is to divide the data using the training data that has been shared and then continue to test using the test data.

### *3.4. Evaluation*

Once the test is conducted using test data, the next stage is to evaluate how well the model can perform. The evaluation metrics used in this study include accuracy, precision, recall, and F1-score. The results of these evaluation metrics will provide an overview of the model's performance in classifying data appropriately. The following are the full results of the applied evaluation metrics.

**Table 1. Results of Evaluation Metrics**

| Classification Report | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| 1 | 0.75 | 1.00 | 0.86 | 9 |
| 2 | 1.00 | 0.68 | 0.81 | 19 |
| 3 | 0.79 | 1.00 | 0.88 | 11 |
| | | | | |
| Accuracy | | | 0.85 | 39 |
| Macro avg | 0.85 | 0.89 | 0.85 | 39 |
| Weighted avg | 0.88 | 0.85 | 0.84 | 39 |

In addition to the evaluation metrics, the results of the confusion matrix calculation will also be displayed as follows.

https://doi.org/10.52970/grdis.v5i1.910

Website: https://goldenratio.id/index.php/grdis ISSN [Online]: 2776-6411

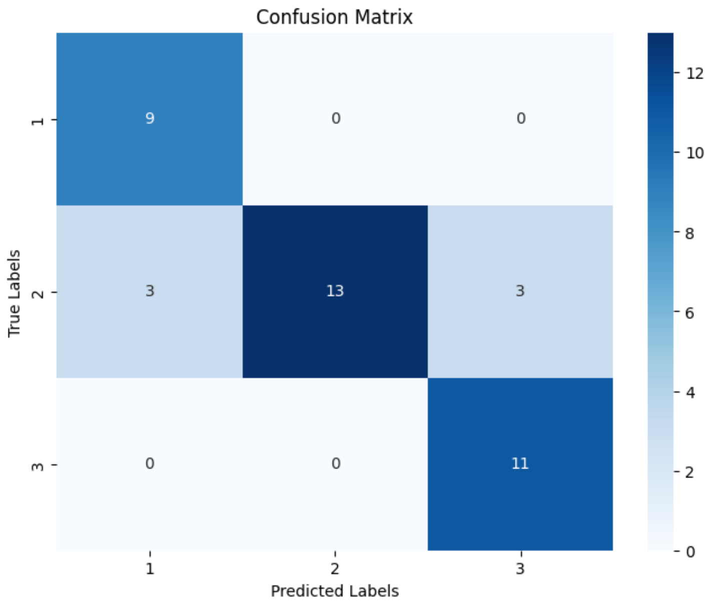



**Figure 8 Convincial Matrix**

The evaluation results show that the model has achieved a pretty good accuracy, around 85%. Although there are some classification errors in some classes, the model has proven effective in classifying images with high accuracy. With accuracy that has exceeded the 80% threshold, this model can be considered optimal and does not require further development according to the original plan.

## 4. CONCLUSION

This study shows that the pear variety classification system based on digital images with the K-Nearest Neighbor (KNN) algorithm has achieved satisfactory accuracy, which is 85%. The system utilizes key features such as color, texture, shape, and size, which have proven effective in accurately recognizing and grouping pear varieties. The model evaluation yielded high-performance metrics, with precision, recall, and F1-scores of 0.85, 0.89, and 0.85, respectively, reflecting the reliability of the KNN algorithm in digital image-based classification. The results of this study confirm the great potential of digital image-based technology as an innovative solution for classification automation in the agricultural sector, which improves efficiency and accuracy. In addition, the KNN-based approach is essential for developing a more adaptive and resilient system facing challenges such as lighting variations and shooting angles. This research also opens up development opportunities through the expansion of data coverage or the exploration of other algorithms, such as Support Vector Machines (SVMs) or Convolutional Neural Networks (CNNs), to improve the flexibility, generalization, and performance of the system, thereby contributing to technological innovation in the agriculture and food sectors.

Further research can be conducted by expanding the amount of data, including adding more varieties of pears or other types of fruit, to improve the model's accuracy, generalization, and flexibility. In addition, exploration of other algorithms, such as Support Vector Machines (SVMs) or Convolutional Neural Networks (CNNs), can be carried out to evaluate the performance and advantages of each method. To ensure a more resilient and adaptive model, model trials are also recommended under more diverse conditions, such as different lighting, complex backgrounds, or varying shooting angles. As a final step, the development of this system-based application is expected

https://doi.org/10.52970/grdis.v5i1.910

Website: https://goldenratio.id/index.php/grdis ISSN [Online]: 2776-6411

to be applied practically in various sectors, such as agriculture and the food industry, to accelerate the automatic fruit classification process, increase efficiency, and support technological innovation.

## REFERENCES

Arthania, T., Purwati, E., Puspadina, V., & Safitri, C. I. N. H. (2021, October). Formulas dan uji mutu fisik body lotion ekstrak kulit buah pir (Pyrusbretschneideri). In Prosiding SNPBS Seminar Nasional Pendidikan Biologi dan Saintek (pp. 416-422).

Batubara, N. P., Widiyanto, D., & Chamidah, N. (2020). Klasifikasi rempah rimpang berdasarkan ciri warna rgb dan tekstur glcm menggunakan algoritma naive bayes. Informatik: Jurnal Ilmu Komputer, 16(3), 156-163.

Cholil, S. R., Handayani, T., Prathivi, R., & Ardianita, T. (2021). Implementasi algoritma klasifikasi k-nearest neighbor (knn) untuk klasifikasi seleksi penerima beasiswa. IJCIT (Indonesian Journal on Computer and Information Technology), 6(2), 118-127.

Fadli, F., Van FC, L. L., & Lisnawita, L. (2024, November). Klasifikasi Buah Pinang Berdasarkan Tingkat Kematangan Menggunakan Metode Naïve Bayes. In SEMESTER: Seminar Nasional Teknologi Informasi & Ilmu Komputer (Vol. 3, No. 1, pp. 423-433).

Fahrurozi, A., Nugraha, N., & Riminarsih, D. (2023). Sistem Klasifikasi Kayu Berbasis Citra Tekstur Menggunakan Machine Learning. Uwais Inspirasi Indonesia.

Inzaghi, R., Cahyani, A. D., Valenda, V., Ananda, M. A., & Pratama, D. (2024). Analisis Penerapan Artificial Intelligence (AI) di Berbagai Bidang. Jurnal Rekayasa Informatika, 1(1), 36-45.

Irwan, A. S. M., Arifien, Y., Jannah, A., & Masnang, A. (2022). Identifikasi Benih Beberapa Varietas Melon (Cucumis Melo L.) Menggunakan Pengolahan Citra Digital. Agrisintech (Journal of Agribusiness and Agrotechnology), 3(2), 79-84.

Juliansyah, S., & Laksito, A. D. (2021). Klasifikasi Citra Buah Pir Menggunakan Convolutional Neural Networks. InComTech: Jurnal Telekomunikasi dan Komputer, 11(1), 65-72.

Nuraeni, S., Syam, S. P. A., Wajdi, M. F., Firmansyah, B., & Malkan, M. (2023). Implementasi Metode K-NN Untuk Menentukan Jurusan Siswa di SMAN 02 Manokwari. G-Tech: Jurnal Teknologi Terapan, 7(1), 89-95.

Paniza, M., Rusbandi, R., & Alamsyah, D. (2021). Identifikasi Jenis Buah Pir Berdasarkan Bentuk Menggunakan Metode HOG dan JST. Jurnal Algoritme, 2(1), 62-72.

Pawening, R. E., Shudiq, W. J. F., & Wahyuni, W. (2020). Klasifikasi Kualitas Jeruk Lokal Berdasarkan Tekstur dan Bentuk Menggunakan Metode k-Nearest Neighbor (k-NN). COREAI: Jurnal Kecerdasan Buatan, Komputasi dan Teknologi Informasi, 1(1), 10-17.

Reswan, Y., Toyib, R., Witriyono, H., & Anggraini, A. (2024). Klasifikasi Tingkat Kematangan Buah Nanas Berdasarkan Fitur Warna Menggunakan Metode K–Nearest Neighbor (KNN). Jurnal Media Infotama, 20(1), 280-287.

Saputro, W., & Sumantri, D. B. (2022). Implementasi Citra Digital Dalam Klasifikasi Jenis Buah Anggur Dengan Algoritma K-Nearest Neighbors (KNN) Dan Data Augmentasi. INTECOMS: Journal of Information Technology and Computer Science, 5(2), 248-253.

Setyawan, H. F., Nugroho, B., & Via, Y. V. (2023). Analisis Perbandingan Performansi Algoritma Cnn Dan Nn Pada Buah Pir. JATI. Jurnal Mahasiswa Teknik Informatika, 7(6), 3114-3120.

Sya'ban, D. R., Hamzah, A., & Susanti, E. (2022). Klasifikasi Buah Segar dan Busuk Menggunakan Algoritma Convolutional Neural Network dengan TFLite sebagai Media Penerapan Model Machine Learning. Prosiding Snast, F7-16.

Tikasni, E., Utami, E., & Ariatmanto, D. (2024). Analisis Akurasi Object Detection Menggunakan Tensorflow Untuk Pengenalan Bahasa Isyarat Tangan Menggunakan Metode SSD. Jurnal Fasilkom, 14(2), 385-393.

Wijaya, N., & Ridwan, A. (2019). Klasifikasi Jenis Buah Apel Dengan Metode K-Nearest Neighbors Dengan Ekstraksi Fitur HSV dan LBP. Jurnal Sisfokom (Sistem Informasi dan Komputer), 8(1), 74-78.