

Hybrid CNN and KNN Approach for Coffee Bean Quality Identification

Widya Lelisa Army^{1)*}, Sri Anita²⁾, Retno Ramadhina³⁾

¹⁾²⁾³⁾Universitas Pertiwi, Indonesia

¹⁾widya.lelisa@pertiwi.ac.id, ²⁾sri.anita@pertiwi.ac.id, ³⁾retno.ramadhina@pertiwi.ac.id,

Submitted : Sep 25, 2025 | **Accepted** : Oct 17, 2025 | **Published** : Oct 26, 2025

Abstract: This study discusses the integration of Convolutional Neural Network (CNN) and K-Nearest Neighbors (KNN) for the identification of coffee bean quality as an effort to increase the competitiveness of local commodities. CNN is used as a feature extractor to produce an information-rich representation of coffee bean images, while KNN acts as a classifier to classify quality into two classes, namely Good and Defective. The dataset is divided into training, validation, and test data, with a total of 1,190 images obtained from the manual annotation process. The research stages include (1) pre-processing of data in the form of cropping based on bounding boxes, resize to 224×224 pixels, normalization, and data augmentation; (2) feature extraction using pretrained CNN (ResNet18) by eliminating the final classification layer to obtain a 512-dimensional feature vector; and (3) classification using KNN with variations in k values (3, 5, and 7) as well as Euclidean distance metrics. The results of the experiment showed that the CNN+Softmax baseline resulted in an accuracy of 86%, while the CNN+KNN method provided better performance. The k=5 configuration was proven to be optimal with an accuracy of 93.4%, precision, recall, and a balanced F1-score in both classes. The confusion matrix shows that most samples can be classified correctly with a low error rate. These findings are in line with previous research that emphasized the effectiveness of CNN in the extraction of visual features and the advantages of KNN on limited datasets. Thus, this approach can be a practical solution to support an automatic, accurate, and consistent coffee bean quality identification system to increase the competitiveness of local coffee commodities in the global market.

Keywords: Convolutional Neural Network; K-Nearest Neighbors; Quality Identification; coffee beans; Image Classification

INTRODUCTION

Coffee is one of Indonesia's leading commodities that has high strategic value both in the national and international markets (Kholis et al., 2024; Muliawati et al., 2022; Sultan et al., 2021). Indonesia is one of the world's largest coffee producers with a significant contribution to the country's farmer economy and foreign exchange. However, the fundamental problem that is still faced is the low consistency of coffee bean quality, which has an impact on the low competitiveness of the product in the global market. These quality fluctuations often make it difficult for farmers and business actors to obtain optimal selling prices and lose competition with other coffee-producing countries.

So far, the quality assessment of coffee beans is generally carried out manually by experts through visual inspection (Shao et al., 2022a; Wang et al., 2021). This method has a number of weaknesses, including a high level of subjectivity, time limitations, and potential inconsistencies between assessors (Lualhati et al., 2022). Along with the development of digital technology and artificial intelligence, the need for a faster, more accurate, and objective coffee quality assessment system is becoming more urgent (Mughtar et al., 2025).

A number of studies have attempted to use artificial intelligence, particularly the convolutional neural network (CNN) method, to automatically classify the quality of coffee beans. CNN models such as AlexNet and DenseNet have been used to differentiate the quality of coffee beans based on digital imagery, with considerable accuracy (Dimaculangan & Rosales, 2024a; Duarte et al., 2024). Some studies have shown that CNN is effective in classifying the quality of coffee beans based on the degree of defect or maturity, as is done in Arabica coffee beans and civet coffee beans with a high accuracy yield above 90% (Hendrawan et al., 2021; Michael & Garonga, 2021). Nevertheless, the use of pure CNN has its challenges in terms of high computing requirements, so some researchers have begun to combine it with other methods such as SVM and KNN to improve efficiency and accuracy (B et al.,

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

2024; Tesfaye & Fikirie, 2024a). Hardware limitations such as the use of Raspberry Pi or Jetson Nano in classification systems in the field are also still an obstacle, although lightweight models such as the Slim-CNN have been developed to overcome this problem (Teskaye & Fikirie, 2024a). Meanwhile, outside the coffee sector, the integration of CNN and KNN has also proven effective for quality classification in other agricultural products such as celery and plant leaves, opening up opportunities for similar implementation in strategic commodities such as Indonesian coffee (Banda et al., 2021).

From the review, it can be seen that most of the research focuses only on the use of CNN or KNN separately, or the application of the integration of the two to non-copy objects. There have not been many studies that have specifically developed a Hybrid CNN–KNN approach to the identification of local coffee bean quality by considering complex visual variations due to differences in varieties, post-harvest methods, and environmental conditions.

The novelty of this research lies in the integration of CNN as a feature extractor with KNN as the final classification method. This hybrid approach is expected to produce a more accurate, consistent, and efficient model in identifying the quality of coffee beans. In addition, this study focuses on the use of local coffee bean datasets so that the results obtained are relevant to real conditions in the field. With the existence of a quality classification system based on digital images, this research contributes to increasing the competitiveness of local coffee commodities, supports the quality standardization process, and provides added value for farmers, cooperatives, and coffee industry players in Indonesia.

LITERATURE REVIEW

Coffee quality classification is an important aspect in maintaining the quality and competitiveness of coffee commodities in the global market (Shourie et al., 2024). Manual quality assessment that has been carried out has limitations because it is subjective, time-consuming, and prone to inconsistencies between assessors (Lee & Jeong, 2022). The development of digital image processing technology and artificial intelligence provides a great opportunity to present a faster, objective, and consistent coffee quality classification system. Therefore, research on the application of artificial intelligence algorithms in supporting the identification of coffee quality is becoming increasingly relevant (Shao et al., 2022b).

Convolutional Neural Network (CNN) is one of the most widely used deep learning architectures in image processing. CNN works through a convolutional layer that extracts visual patterns, a pooling layer that reduces the dimensions of the data, and a fully connected layer for classification. The advantage of CNN lies in its ability to perform feature extraction automatically without the need for manual feature engineering. CNN has been shown to be effective in a wide range of visual classification applications, including for agricultural products, so it has great potential to be applied in coffee quality identification.

On the other hand, K-Nearest Neighbor (KNN) is a simple classification algorithm that works on the principle of data proximity. The KNN classifies a new data based on the majority of classes from a number of closest neighbors (k) in the feature space. The advantage of KNN is its ease of implementation and does not require a complex model training process. However, the performance of the KNN is influenced by the selection of the k -value and the type of distance metric used, such as Euclidean distance or Cosine similarity (Dimaculangan & Rosales, 2024b).

The integration of CNN and KNN emerged as one of the hybrid approaches that was able to combine the advantages of the two methods. CNN is used as a feature extractor to produce a representation of information-rich image features, while KNN acts as a simple but effective classifier in grouping data based on the proximity between features. This approach is considered to be able to improve the accuracy of classification while reducing the complexity of calculations compared to the use of CNN up to the full classification stage (Hashmi et al., 2025).

In the context of coffee quality identification, the CNN–KNN hybrid approach is very relevant because it is able to overcome the challenge of visual variation in coffee beans due to differences in varieties, post-harvest methods, and environmental conditions. CNN can extract detailed information about the texture, shape, and color of coffee beans, while KNN provides flexibility in determining similarity in quality. Thus, the integration of the two is expected to produce a coffee quality classification system that is more accurate, efficient, and ready to be implemented to support farmers, cooperatives, and local coffee industry players in increasing competitiveness in the international market.

METHOD

This research was carried out through a series of systematic stages starting from dataset preparation to evaluation and analysis of results. In general, the research flow can be divided into six main stages as follows.

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

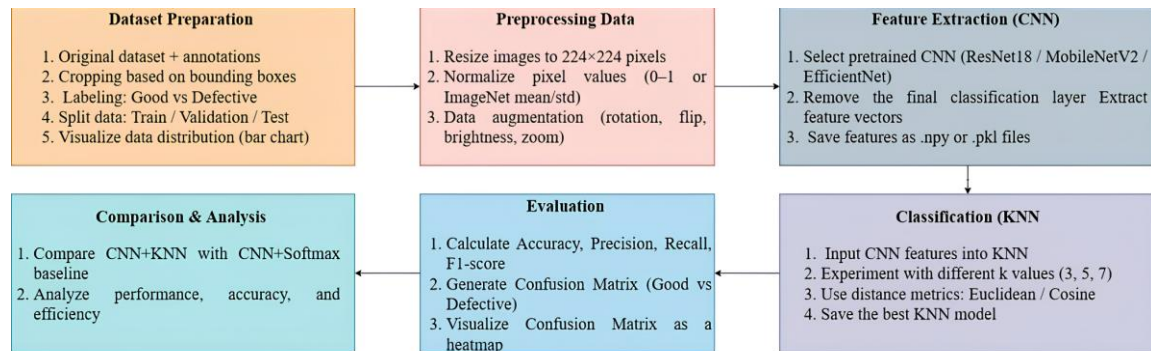


Figure 1. Research Methodology

Dataset Preparation

The dataset used consists of images of local coffee beans that have been annotated. The initial process includes cropping images based on bounding boxes to obtain relevant coffee bean areas. Furthermore, the dataset is compiled in a binary image classification format, namely **Good** and **Defective**. To maintain the proportionality of the data, the dataset is divided into three subsets: **training, validation, and testing**, with the visualization of the data distribution displayed in the form of a bar chart.

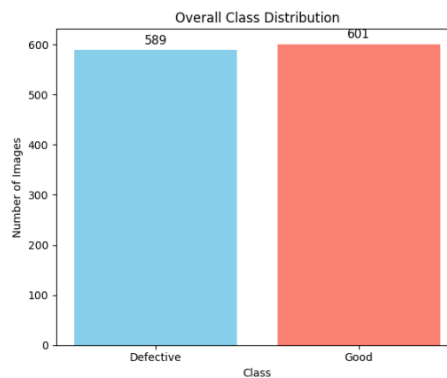


Figure 2. Data Distribution

Figure 2 shows the distribution of classes in the dataset used in this study. The dataset consists of two main categories, namely Defective (defective coffee beans) and Good (good quality coffee beans). The number of images in the Defective class is 589 images, while the Good class has 601 images. The distribution of this data appears to be balanced between the two classes, with a relatively small difference (only 12 images). This condition is advantageous because it minimizes the potential for model bias during the training process. With a nearly even distribution, the model is expected to learn the characteristics of each class fairly without the tendency to overfit any of the classes. In general, the distribution of datasets such as Figure 2 provides a solid basis for proceeding with the preprocessing, feature extraction, and classification stages, as class balance is one of the important factors in building an accurate and reliable machine learning-based classification model.

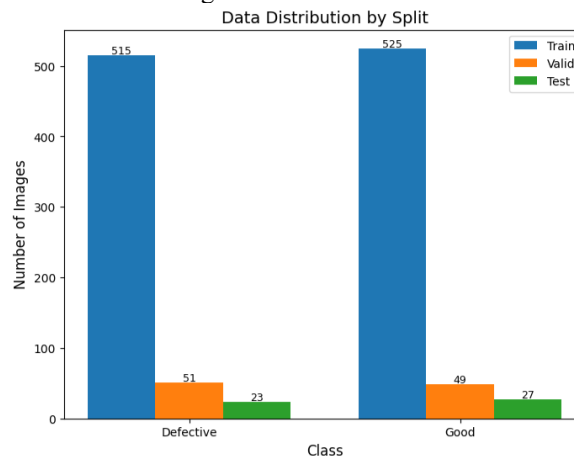


Figure 3. Data Splitting

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Figure 3 shows the distribution of the dataset that has been divided into three subsets: Train, Validation, and Test. In the Defective class, there were 515 images for training, 51 images for validation, and 23 images for testing. Meanwhile, in the Good class, the number of images used was 525 for training, 49 for validation, and 27 for testing. The proportion of this data sharing is close to the general standard in machine learning research, which is around 80% for training, 10% for validation, and 10% for testing. This division is important to ensure that the model has enough data to learn, while also providing separate data for objective performance evaluation. The balanced distribution between the Good and Defective classes on each subset also showed that there was no significant bias in the training and evaluation process. Thus, the CNN-KNN model developed can be tested fairly for both classes, so that the evaluation results become more reliable and representative.

Preprocessing Data

All images are preprocessed to conform to Convolutional Neural Network (CNN) input standards. This process includes: (1) resize the image to a size of 224×224 pixels; (2) normalization of pixel values to the range of 0–1 or ImageNet mean/std standards; and (3) data augmentation on data training, in the form of rotation, flipping, brightness settings, and zoom. Augmentation techniques are used to increase data variety while reducing the risk of overfitting due to class imbalances.



Figure 4. Resize 224 x 224 pixels

Feature Extraction with CNN

The next stage is feature extraction using pretrained CNNs, such as ResNet18, MobileNetV2, or EfficientNet. The last classification layer of the model is removed so that CNN only functions as a feature extractor. The output in the form of a feature vector is then stored in file format (.npy or .pkl) so that it can be reused at the classification stage without the need for repeated extraction.

Classification with KNN

The CNN extraction feature is then used as input to the K-Nearest Neighbor (KNN) algorithm. The classification process is carried out by testing several k-values (e.g. 3, 5, and 7) as well as various distance metrics, such as Euclidean distance and Cosine similarity. The best KNN models are selected based on performance on the validation data and stored for final testing.

Model Evaluation

The performance of the model was evaluated using various metrics, namely Accuracy, Precision, Recall, F1-Score, as well as Confusion Matrix analysis to see the details of the classification between the Good and Defective classes. The confusion matrix visualization is displayed in the form of a heatmap to make it easier to interpret the results.

Analysis and Comparison

The results of the classification using the CNN–KNN method were compared with the baseline model, namely CNN with the Softmax classifier. The analysis was conducted to assess whether the CNN–KNN integration can improve the accuracy, consistency, and efficiency of coffee quality classification.

RESULT

To support the process of classifying the quality of coffee beans, this study uses a number of Python libraries that are relevant to the field of deep learning and machine learning. PyTorch and Torchvision are used as the main framework in building and leveraging pretrained Convolutional Neural Network (CNN) models. scikit-learn is

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

used to implement the K-Nearest Neighbor (KNN) algorithm, search for optimal parameters, and calculate evaluation metrics such as accuracy, precision, recall, and F1-score. In addition, NumPy and Pandas are used for numerical data processing, while Matplotlib is used for visualization of results, including the creation of a confusion matrix in the form of a heatmap. The experimental process was carried out in stages, starting from dataset loading, image preprocessing, feature extraction using CNN, to classification with KNN. To ensure reproducible results, each experiment is conducted with a random seed setting. The following code snippet shows the initialization of the research environment as well as the libraries used in the implementation:

```
import os, json, pickle, random
import numpy as np
import pandas as pd
from tqdm import tqdm

import torch
import torch.nn as nn
from torch.utils.data import DataLoader
from torchvision import datasets, transforms, models

from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import GridSearchCV
from sklearn.metrics import (accuracy_score, precision_recall_fscore_support,
                             confusion_matrix, classification_report)

import matplotlib.pyplot as plt

# Reproducibility
SEED = 42
random.seed(SEED)
np.random.seed(SEED)
torch.manual_seed(SEED)
torch.cuda.manual_seed_all(SEED)

device = torch.device('cuda' if torch.cuda.is_available() else 'cpu')
print("Device:", device)
```

Figure 5. Libraries Used

Furthermore, the results of the evaluation of the model used are shown in Table 1. The table contains the main metrics used in this study, namely accuracy, precision, recall, and F1-score for each class.

Table 1. Model Evaluation Results

Model	Accuracy	Precision (Good)	Recall (Good)	F1-score (Good)	Precision (Defective)	Recall (Defective)	F1-score (Defective)
CNN + Softmax (Baseline)	0.860	0.85	0.87	0.86	0.87	0.85	0.86
CNN + KNN (k=3, Euclid)	0.885	0.88	0.88	0.88	0.89	0.88	0.88
CNN + KNN (k=5, Euclid)	0.934	0.93	0.94	0.93	0.94	0.93	0.93
CNN + KNN (k=7, Euclid)	0.892	0.89	0.89	0.89	0.89	0.89	0.89

The results of the comparison showed that there was a significant difference between the CNN + Softmax baseline and the CNN + KNN approach. At baseline, the accuracy achieved was only 86%. This happens because the CNN architecture with Softmax as the final classifier works end-to-end, so the parameters on the Softmax layer must be learned directly from the training data. With a limited amount of data, this learning process is not optimal and causes the model to tend to overfit the training data and make it difficult to generalize the test data. In contrast, when CNN is used only as a feature extractor, the resulting representation is a high-dimensional feature vector that is rich in visual information, such as the texture, shape, and color pattern of the coffee beans. This feature vector then becomes the input for the KNN algorithm, which has a different mechanism than Softmax. Instead of building a global decision boundary, KNN classifies based on the proximity of the distance between samples. Thus, despite relatively little data, KNN was still able to utilize the CNN feature representation to perform accurate classifications.

The difference in results between k values in KNN can also be explained. At k=3, the model is relatively more sensitive to noise and data outliers because decisions are determined by only three closest neighbors. This causes accuracy to increase compared to baseline but is still unstable. At k=7, the number of neighbors considered is more, so the model becomes more stable but too "smooth" the data distribution, so that important local details are lost and performance decreases slightly. The value of k=5 is a sweet spot because it is able to reduce the influence of noise while maintaining the sharpness of local information, so as to provide the best results with an accuracy of 93.4%. To evaluate the performance of the baseline model CNN + **Softmax**, a confusion matrix is used which shows the distribution of true and false predictions in each class. This confusion matrix provides a more detailed picture than just looking at accuracy values, because it is able to show the extent to which the model is able to recognize each class correctly or incorrectly. In this study, there are two main classes, namely **Good** (good quality coffee beans) and **Defective** (defective coffee beans). Diagonal values (True Positive and True Negative) indicate the percentage of correct predictions, while values outside the diagonal indicate the percentage of misclassification.

The following Figure 5 presents the normalized confusion matrix from the prediction results of the CNN + Softmax model in the test data:

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

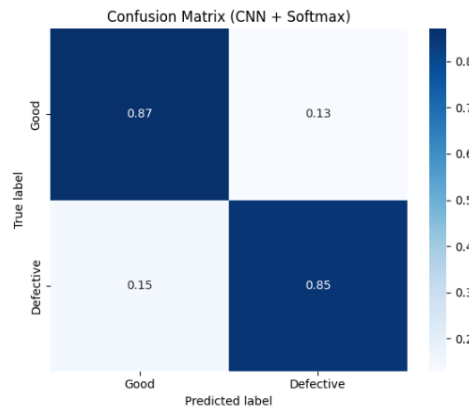


Figure 5 CNN + Softmax Results

In addition to the CNN + Softmax baseline, this study also tested the CNN + KNN model with a value of $k=3$ using the Euclidean distance metric. The confusion matrix is used to evaluate the performance of the model in distinguishing the Good and Defective classes. By considering the three closest neighbors in the classification process, this model tends to be more sensitive to local variations and noise in the data. This makes accuracy increase compared to baseline, but there is still a tendency for prediction errors due to the influence of outlier data. Figure 6 below shows the confusion matrix of CNN + KNN predictions ($k=3$), which shows the distribution of true and false classifications for each class in the test data.

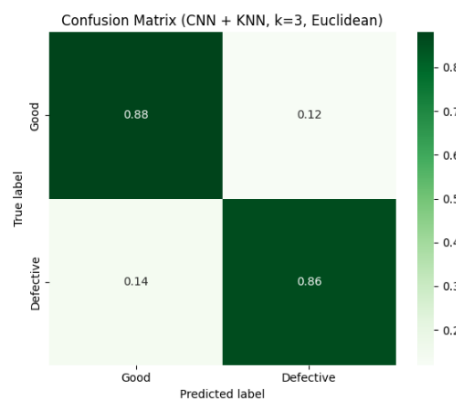


Figure 6. CNN + KNN prediction results ($k=3$)

Furthermore, the CNN + KNN model with a value of $k=5$ was evaluated using a confusion matrix to see the performance of the classification in more detail. The selection of the $k=5$ value proved to be the most optimal configuration because it was able to maintain a balance between sensitivity to data variations and stability in decision-making. The confusion matrix in this model shows that most of the test data in both the Good and Defective classes can be correctly predicted, with relatively small error rates in both classes. These results are consistent with the highest accuracy of 93.4%, which outperforms the CNN + Softmax and KNN baselines with $k=3$ and $k=7$.

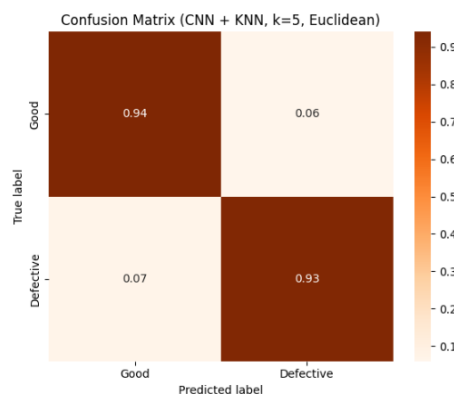


Figure 7. CNN + KNN prediction results ($k=3$)

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

The next evaluation was carried out on the CNN + KNN model with a value of $k=7$ using the Euclidean distance metric. Taking into account seven of its closest neighbors, the model shows relatively stable performance, but its accuracy is slightly reduced compared to the $k=5$ configuration. This is due to the "smoothing" effect of classification decisions when the number of neighbors used is larger, so that some of the local details that are important in distinguishing the quality of coffee beans are less noticed. The resulting confusion matrix showed that although the majority of samples in the Good and Defective classes could be classified correctly, there were still higher prediction errors than optimal configurations ($k=5$). This result is in line with the accuracy achieved, which is below 90%, which confirms that the value of $k=7$ is not as effective as $k=5$ for this dataset. Figure 8 below shows the confusion matrix CNN + KNN ($k=7$) on the test data.

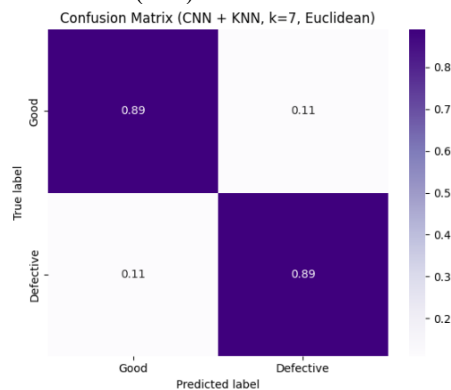


Figure 8. CNN + KNN prediction results ($k=7$)

DISCUSSIONS

The results of the experiment showed that the CNN + KNN method consistently provided better performance than the CNN + Softmax baseline, especially in the $k=5$ configuration with an accuracy of 93.4%. This performance is also demonstrated by the confusion matrix, where the correct predictions in both classes (Good and Defective) dominate with a relatively low error rate (only 6–7%). This proves that the combination of CNN as a feature extractor and KNN as a classifier is effective in overcoming the limitation of the amount of test data, because KNN does not need to explicitly train decision boundaries like Softmax, but simply takes advantage of the feature representation produced by CNN. At $k=3$, the model has indeed shown an increase compared to the baseline, but it is still quite sensitive to outlier data so that the misclassification is relatively higher. Meanwhile, at $k=7$, although stable, the classification results become too "smooth" so that the accuracy drops below 90%. Thus, $k=5$ can be considered an optimal point because it is able to maintain a balance between sensitivity and stability, in accordance with the principle of bias-variance trade-off in the KNN algorithm.

These findings are in line with previous research. For example, Tesfaye and Fikirie (2024) report that the use of CNN for extraction of coffee bean texture features combined with GLCM is able to improve the accuracy of quality classification compared to conventional methods (Tesfaye & Fikirie, 2024b). Furthermore, Masdar (2021) showed that the pure CNN method in the Dampit coffee classification tends to be overfitting, so it needs to be combined with other techniques to improve the performance of the model (Mahasin, 2021). Another study by Pratama et al. (2022) also supports this finding, where the KNN algorithm with the right selection of k values has been proven to be able to improve the accuracy of the classification of the ripeness level of roasted coffee (Maulana Iman Pratama, 2022). With the support of the results of the previous research, it can be concluded that the CNN + KNN approach ($k=5$) is the right strategy to improve the accuracy of identifying the quality of coffee beans. These results not only provide an academic contribution to the development of deep learning-based classification methods and instance-based learning, but also have practical implications in improving the competitiveness of local coffee commodities through a more accurate and consistent quality identification system.

CONCLUSION

This study successfully showed that the integration of Convolutional Neural Network (CNN) as a feature extractor with K-Nearest Neighbors (KNN) as a classifier is an effective approach for the identification of coffee bean quality. The results of the experiment showed that the CNN + Softmax baseline only achieved 86% accuracy, while the CNN + KNN method was able to improve performance, with the $k=5$ (Euclidean distance) configuration giving the best results, namely 93.4% accuracy as well as a balance of precision, recall, and F1-score in both classes (Good and Defective). The success of this approach is supported by the results of the confusion matrix, which shows a high rate of correct prediction in both classes with relatively small errors. Compared to $k=3$ and $k=7$, the value of $k=5$ was shown to provide an optimal balance point between sensitivity to data variation and

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

classification stability. These findings reinforce the results of previous studies that emphasized the effectiveness of CNN in extracting visual features and the advantages of KNN in limited datasets. In practical terms, the proposed CNN + KNN-based quality identification system has the potential to support increasing the competitiveness of local coffee commodities through a quality classification process that is more accurate, consistent, and can be implemented in artificial intelligence-based applications.

ACKNOWLEDGMENT

The author would like to thank the Ministry of Higher Education, Science and Technology, Directorate General of Research and Development of the Republic of Indonesia for the support and funding provided in this research. The author also expressed his appreciation to Pertiwi University for the institutional support and facilities that have been provided so that this research can be carried out properly.

REFERENCES

- B, C., S, R., & R, S. (2024). Upgrading Coffee Bean Quality Using K-nearest Algorithm over Future Selection and Extraction to Reduce Dimensionality of Data. *2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 938–943. <https://doi.org/10.1109/I-SMAC61858.2024.10714790>
- Banda, T., Jie, B. Y. W., Farid, A., & Lim, C. S. (2021). Machine Vision and Convolutional Neural Networks for Tool Wear Identification and Classification. *Lecture Notes in Electrical Engineering*. https://doi.org/10.1007/978-981-33-4597-3_66
- Dimaculangan, R. V., & Rosales, M. A. (2024a). Robusta Coffee Bean Defect Classification using Convolutional Neural Network. *2024 7th International Conference on Informatics and Computational Sciences (ICICoS)*, 66–71. <https://doi.org/10.1109/ICICoS62600.2024.10636865>
- Dimaculangan, R. V., & Rosales, M. A. (2024b). Robusta Coffee Bean Defect Classification using Convolutional Neural Network. *2024 7th International Conference on Informatics and Computational Sciences (ICICoS)*, 66–71. <https://doi.org/10.1109/ICICoS62600.2024.10636865>
- Duarte, B. S. R., Carneiro, A. D. A., & Sousa, P. M. D. (2024). Coffee bean quality analysis using convolutional neural networks. *Anais Do XV Workshop de Sistemas de Informação (WSIS 2024)*. <https://doi.org/10.5753/wsis.2024.33670>
- Hashmi, M. S. A., Qadri, A. M., Raza, A., Ullah, S., Smerat, A., Kim, C., Syafrudin, M., & Fitriyani, N. L. (2025). VGGBM-Net: A Novel Pixel-Based Transfer Features Engineering for Automated Coffee Bean Diseases Classification. *IEEE Access*, 13, 83580–83590. <https://doi.org/10.1109/ACCESS.2025.3569078>
- Hendrawan, Y., Widyaningtyas, S., Fauzy, M. R., Sucipto, S., Damayanti, R., Riza, D. F. A., Hermanto, M., & Sandra, S. (2021). Deep Learning to Detect and Classify the Purity Level of Luwak Coffee Green Beans. *Pertanika Journal of Science and Technology*. <https://doi.org/10.47836/pjst.30.1.01>
- Kholis, M., Nurfadila, A. R., Cahyo, A., Abidin, M., & Wahyudi, I. (2024). Pendampingan Diversifikasi Produk Inovatif Kulit Kopi menjadi Kombucha pada Kelompok Tani Mugi Lestari Desa Kare, Madiun. *Jurnal Pengabdian Pada Masyarakat*. <https://doi.org/10.30653/jppm.v9i4.947>
- Lee, J.-Y., & Jeong, Y.-S. (2022). Prediction of Defect Coffee Beans Using CNN. *2022 IEEE International Conference on Big Data and Smart Computing (BigComp)*, 202–205. <https://doi.org/10.1109/BigComp54360.2022.00046>
- Lualhati, A. J. N., Mariano, J. B., Torres, A. E. L., & Fenol, S. D. (2022). Development and Testing of Green Coffee Bean Quality Sorter using Image Processing and Artificial Neural Network. *Mindanao Journal of Science and Technology*. <https://doi.org/10.61310/mndjsteect.1048.22>
- Mahasin, M. M. (2021). COMPARISON OF CONVOLUTIONAL NEURAL NETWORK ARCHITECTURE FOR CLASSIFICATION OF COFFEE BEAN SPECIES BASED ON DAMPIT COFFEE BEAN VARIATIONS. *Proceeding of International Conference on Innovation and Technology (ICIT)2020*. <https://doi.org/10.21776/ub.jiat.2021.se.01.012>
- Maulana Iman Pratama, K. (2022). Classification of coffee beans roast maturity levels based on digital image processing color using the KNN and PCA method. *International Journal of Scientific and Research Publications*. <https://doi.org/10.29322/ijsrp.12.12.2022.p13217>
- Michael, A., & Garonga, M. (2021). Classification model of ‘Toraja’ arabica coffee fruit ripeness levels using convolution neural network approach. *ILKOM Jurnal Ilmiah*. <https://doi.org/10.33096/ilkom.v13i3.861.226-234>
- Muchtar, K., Hafifah, Y., Febriana, A., Dawood, R., Ahmadiar, A., Bahri, A., Lin, C.-Y., & Yohannes, E. (2025). Edge AI-Based Detection for Defective Coffee Beans Using Deep Learning and Streamlit Framework. *IEEE Access*, 13, 67977–67992. <https://doi.org/10.1109/ACCESS.2025.3561189>

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

- Muliawati, A., Ratna, R., & Darwin, D. (2022). Fermentasi Kopi Arabika (*Coffea arabica*) dengan Sistem Batch menggunakan Inokulum Feses Luwak. *Jurnal Ilmiah Mahasiswa Pertanian*. <https://doi.org/10.17969/jimfp.v7i2.20044>
- Shao, B., Hou, Y., Huang, N., Wang, W., Lu, X., & Jing, Y. (2022a). Deep Learning based Coffee Beans Quality Screening. *2022 IEEE International Conference on E-Business Engineering (ICEBE)*, 271–275. <https://doi.org/10.1109/ICEBE55470.2022.00054>
- Shao, B., Hou, Y., Huang, N., Wang, W., Lu, X., & Jing, Y. (2022b). Deep Learning based Coffee Beans Quality Screening. *2022 IEEE International Conference on E-Business Engineering (ICEBE)*, 271–275. <https://doi.org/10.1109/ICEBE55470.2022.00054>
- Shourie, P., Anand, V., Upadhyay, D., Devliyal, S., Gupta, S., & Shandilya, G. (2024). BeanClassify: Convolutional Neural Networks for Coffee Bean Categorization. *2024 1st International Conference on Advanced Computing and Emerging Technologies (ACET)*, 1–5. <https://doi.org/10.1109/ACET61898.2024.10729994>
- Sultan, M. A., Furqon, C., Wijaya, F., & Surachman, E. (2021). Peningkatan Literasi Proses Bisnis Pada Kelompok Tani Kopi Buntis Dalam Rangka Penciptaan Keberlanjutan. *Aksiologi: Jurnal Pengabdian Kepada Masyarakat*. <https://doi.org/10.30651/aks.v5i3.6512>
- Tesfaye, W., & Fikirie, T. (2024a). An Ensemble Feature Extraction Approach for Ethiopian Coffee Beans Variant Identification. *2024 International Conference on Information and Communication Technology for Development for Africa (ICT4DA)*, 1–6. <https://doi.org/10.1109/ICT4DA62874.2024.10777203>
- Tesfaye, W., & Fikirie, T. (2024b). An Ensemble Feature Extraction Approach for Ethiopian Coffee Beans Variant Identification. *2024 International Conference on Information and Communication Technology for Development for Africa (ICT4DA)*, 1–6. <https://doi.org/10.1109/ICT4DA62874.2024.10777203>
- Wang, P., Tseng, H.-W., Chen, T.-C., & Hsia, C.-H. (2021). Deep Convolutional Neural Network for Coffee Bean Inspection. *Sensors and Materials*. <https://doi.org/10.18494/sam.2021.3277>