ISSN 2355-3286

Convolutional Neural Network Roasted Coffee Bean Classification Based on Color

Natanael Michael Hanes[1], Endah Setyaningsih[2], Meirista Wulandari[3]
[1,2,3] Department of Electrical Engineering, Universitas Tarumanagara, Jakarta Barat, Indonesia
natanael.525210018@stu.untar.ac.id[1], endahs@ft.untar.ac.id[2], meiristaw@ft.untar.ac.id[3]

Accepted on August 04, 2025
Approved on November 18, 2025

***Abstract*— Coffee quality is significantly influenced by the roasting level, which is typically determined by observing the color of the beans. This color-based classification has traditionally been performed through manual sorting and human judgment. However, such methods are labor-intensive, time-consuming, subjective, and prone to inconsistencies or human error. These challenges highlight the need for a more reliable and automated solution to improve the accuracy and efficiency of coffee bean classification. In response, this project introduces an automated classification system based on deep learning, specifically utilizing Convolutional Neural Networks (CNNs). CNNs are particularly well-suited for image classification tasks, as they can automatically learn and extract relevant visual features such as color, texture, and shape. This allows the system to classify coffee beans into different roast levels with minimal human intervention. The proposed system is evaluated using several performance metrics, including accuracy, precision, recall, and F1 score. Experiments are conducted using various input image resolutions, and the results show that the CNN model achieves its best performance with 64×64 pixel images, obtaining an accuracy of 99%, which surpasses other tested resolutions such as 32×32 with an accuracy of 98%) and 128×128 with an accuracy of 98.6%). When compared to deeper architectures like ResNet50, which achieved 96.49% accuracy at 224×224 pixels. This demonstrates not only the model's effectiveness in correctly identifying roast levels but also its efficiency in working with lower-resolution images, making it practical for real-time applications or deployment in resource-constrained environments.**

***Index Terms*—Coffee Bean Classification; Convolutional Neural Network; LCD**

I. Introduction

Among the most popular beverages in the world, coffee's flavor and quality are significantly impacted by the caliber of the coffee beans used. Coffee's flavor is its most crucial component, and roasting has the biggest effect on it [1]. The most popular method of determining the roasting level is to look at the color of the coffee beans. Light color denotes a high acidity and low roasting level, medium color denotes a medium roasting level and balanced acidity, and dark color denotes a high roasting level and bitter flavor [2]. This process is laborious, capricious, and prone to human mistake because it always depends on human judgment and manual sorting [3]. The advent of modern technology, particularly computer vision and deep learning techniques [4], has raised interest in automating and increasing the accuracy of coffee bean classification. One area of machine learning called deep learning has demonstrated remarkable results in tasks involving picture recognition and classification.

Recent studies on classification tasks in food and agriculture have demonstrated advances with deep learning models. Convolutional Neural Networks (CNN) are a well-liked deep learning technique for classification jobs. A variety of fruits, cereals, and plants have been successfully categorized using CNN algorithms based on visual attributes[5]. The classification of coffee bean variations, which vary in size, shape, color, and texture, greatly benefits from these models' capacity to extract intricate patterns from photos. Large annotated datasets of coffee bean photos can be used to train deep learning models to accurately identify and categorize coffee beans. This lessens the need for manual effort and increases consistency in quality evaluation. CNN has an accuracy of more than 90% in classifying coffee beans according to their quality and place of origin [6].

The application of deep learning to the classification of coffee beans is consistent with the larger movement in the food and beverage sector to incorporate Artificial Intelligence (AI). AI-powered solutions are being utilized more and more to enhance supply chains, meet consumer needs for sustainability and transparency, and increase product quality. A coffee bean classification system that is automated and based on color may support importers and producers maintain a uniform product look, which is frequently linked to customer preference and quality. Coffee bean maturity, roasting levels, and processing methods are some of the elements that might create variations in coffee bean color profiles, which can be identified by deep learning models [7].

The application of deep learning for classifying coffee beans presents challenges. The requirement for large and high-quality datasets to properly train models is one of the main obstacles. Many resources are needed to collect and annotate such datasets, especially for small-scale coffee farmers. In addition, the classification procedure may become more difficult due to variations in coffee bean appearance caused by factors such as roasting levels and regional differences. Recent developments in transfer learning and data augmentation techniques have shown promising results in addressing these issues [8].

Multiple steps are involved in developing a deep learning model to categorize coffee beans according to color, where processing methods, bean origin, and roasting level all affect color variance [9],[10]. Four different roasting levels were used for the coffee beans used in this study at JJ Mall Jatujak's "Bona Coffee." The Laos Typica Bolaven (Coffea arabica) beans were first green or unroasted before being roasted at the light roast level, followed by Doi Chaang (Coffea arabica) at the medium roast level, and Brazil Cerrado (Coffea arabica) at the dark roast level [11]. When CNN processing techniques are used properly, they significantly improve coffee bean categorization accuracy and allow for a more automated and reliable procedure.

The objective of this design is to develop an image classification system for coffee beans using a (CNN) model as the main processing module. In order to recognize and categorize coffee beans into groups such as green, light, medium, and dark, the CNN processes the input photos to the system. CNN is capable of automatically and reliably classify coffee beans according to their roast levels when processing them using the right methods.

II. Literature Review

The effectiveness of the different approaches has been compared by reviewing a few papers on the use of deep learning in coffee bean classification. "Classification Model of 'Toraja' Arabica Coffee Fruit Ripeness Levels Using Convolution Neural Network Approach" is one paper on coffee bean classification [12]. Coffee beans in this paper are categorized as raw, overripe, fully ripe, and half-ripe. The CNN architecture for image analysis and a computer as the primary processor were used in its development. To increase system accuracy in differentiating each ripeness categorization, a dataset of coffee bean photos taken from different perspectives and lighting conditions was used to train the CNN model. Pre-processing steps such as colour normalization, contrast enhancement, and data augmentation are used to increase the diversity of training samples before the classification process starts. Once the photos of coffee beans have gone through a number of convolutional and pooling layers, CNN extracts important information and classifies the beans' maturity based on the visual patterns it has identified.

Coffee beans are categorized based on shape in according to their morphologies: premium, peaberry, long berry, and deformed shapes [13]. It was created with CNN for image analysis and a computer as the primary processor. To make sure the CNN model could accurately identify shape changes, it was trained on a dataset of coffee bean photos from multiple sources. The process begins with image pre-processing, such as normalization, contrast enhancement, and data augmentation to improve model generalization. The CNN model classified the coffee bean images into four predefined shape categories after extracting the key characteristics from the images after going through multiple convolutional and pooling layers. The classification results are displayed in the form of accuracy values, which serve as indicators of the system's reliability in recognizing coffee bean shapes automatically.

Indonesian coffee beans have been classified originating from Gayo Aceh, Kintamani Bali, and Toraja Tongkonan [14]. It was developed utilizing a computer as the main processor and implemented a CNN architecture based on AlexNet for image classification. The CNN model was trained using a dataset of coffee bean images from the three regions, covering various lighting variations and perspectives to improve model performance. The image pre-processing stage was performed before classification, including color normalization, contrast enhancement, and data augmentation to optimize input quality. After going through several convolutional and pooling layers, the AlexNet model automatically extracted unique visual features from the coffee beans and classified them into one of the three available categories. The final result was a classification accuracy value, which serves as the primary indicator for evaluating the model's reliability in identifying the origin of coffee beans with high precision.

III. Research Method

In order to identify coffee beans according to four roasting levels (light, medium, medium-dark, and dark), a CNN model that assesses various input image sizes was examined for this paper. The CNN model is developed, and the classification operation is obtained by simulation in Python programming language. An Intel(R) Core (TM) i7-8700 CPU running at 3.20 GHz was used for the simulations in order to execute programs and process images. This paper simulates varying input image sizes to determine the optimal configuration for achieving high accuracy in coffee bean categorization.

In order to create the CNN model, four distinct input image sizes consist of 32 × 32 pixels, 64 × 64 pixels, 128 × 128 pixels, and 224 × 224 pixels. The accuracy and loss metrics of the two models will be compared. The CNN model that has the highest accuracy and the lowest loss is selected for further evaluation. The block diagram of this research method is shown at Fig 1.

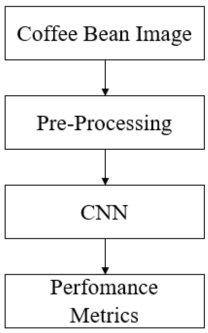

Fig. 1. Block Diagram

A. Coffee Bean Image

Four different roasting levels of coffee beans were used at "Bona Coffee," located in JJ Mall Jatujak. The coffee beans are Laos Typica Bolaven (Coffea arabica), which are mildly roasted and are known as green or unroasted coffee[11]. Brazil Cerrado (Coffea arabica) is dark roasted, and Doi Chaang (Coffea arabica) is medium roasted. The coffee bean images were taken using an iPhone 12 Mini with a 12-megapixel rear camera that has both Ultra-wide and WideCamera settings. During image collection, the camera was oriented parallel to the plane of the object. Images of roasted coffee beans were captured in a variety of settings to guarantee a range of image inputs. Each coffee bean variety was placed in a container to improve image noise, and the dataset was gathered using both natural light and LED light from a light box. Images were automatically taken and stored in PNG format. The coffee bean images were carefully labelled by professionals who were knowledgeable about coffee roasting stages. Light, green, medium, and dark were the four roast levels into which each photograph was assigned based on the beans' surface properties and visual hue. The size of each coffee bean image is 3024 × 3032 pixels. Each of the four roasting stages is represented by 1200 of the dataset's 4800 total images. The images used in this paper has been resized from 3024 × 3032 pixels into four different image sizes, such as, 32 × 32 pixels, 64 × 64 pixels, 128 × 128 pixels, and 224 × 224 pixels.

Fig 2 presents an unaltered coffee bean image from the Dark label in the dataset with a resolution of 3024 × 3032, alongside its resized version at 224 × 224 pixels.This transformation demonstrates how high-resolution images were standardized for model input while maintaining key visual features such as texture and color, which are essential for roast-level classification. Fig 3 shows an unaltered and resized coffee bean image from the Green label. The resizing process ensures consistency across all samples, allowing the model to effectively learn relevant color and surface characteristics without excessive computational cost. Fig 4 depicts an unaltered and resized coffee bean image from the Light label in the dataset. The comparison highlights that even after resizing, the key differences in hue and surface brightness, important indicators of roast intensity, remain visually distinguishable. Fig 5 displays an unaltered and resized coffee bean image from the Medium label. This figure illustrates that the preprocessing step successfully preserves visual cues necessary for accurate CNN-based classification, despite the significant reduction in image resolution.

Fig. 2. Unaltered Coffee Bean Image from the Dark Label in the Dataset in 3024x3032 and Resized Coffee Bean Image from the Dark Label in the Dataset in 224x224

Fig. 3. Unaltered Coffee Bean Image from the Green Label in the Dataset in 3024x3032 and Resized Coffee Bean Image from the Green Label in the Dataset in 224x224

Fig. 4. Unaltered Coffee Bean Image from the Light Label in the Dataset in 3024x3032 and Resized Coffee Bean Image from the Light Label in the Dataset in 224x224

Fig. 5. Unaltered Coffee Bean Image from the Medium Label in the Dataset in 3024x3032 and Resized Coffee Bean Image from the Medium Label in the Dataset in 224x224

B. Convlutional Neural Network (CNN)

CNN is a deep learning neural network that can resolve challenging visual issues. Numerous applications, including object detection, video processing, and picture classification, make extensive use of this technique [15]. A sequential architectural model with multiple convolutional, pooling, and fully connected layers is used.

The first layer is Conv2D, which generates an output of size (222, 222, 32) with 896 parameters and 32 3x3 filters with default padding set to "valid." After that, the output is down sampled to (111, 111, 32) using a MaxPooling2D layer with a 2×2 kernel size. With 64 filters, the second layer is another Conv2D, producing an output of (109, 109, 64) and 18,496 parameters. A second MaxPooling2D layer then decreases the dimensions to (54, 54, 64). Following the output of the third convolutional layer, which has 128 filters and 73,856 parameters, the feature map is reduced to (26, 26, 128) using MaxPooling2D. The CNN Architecture can be seen in Fig 6.

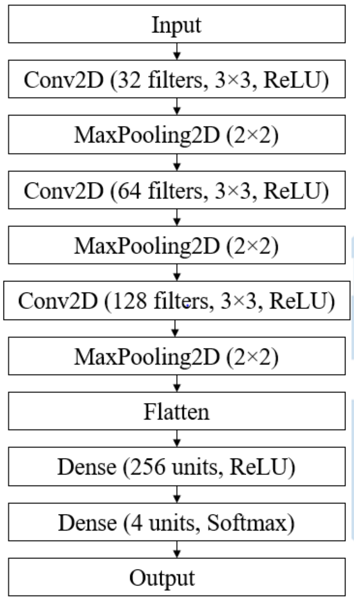

Fig. 6. Proposed CNN Architecture

C. *Performance Metrics*

The parameters used to evaluate a deep learning algorithm's performance are called performance metrics. These measurements include precision, recall, and F1 score. The ratio of correctly predicted positive observations to all expected positive observations is known as precision. The ratio of correctly predicted positive observations to all actual positive observations is known as recall [16]. The F1 score, defined as a calculated average of precision and recall, is a balanced measure of a model's accuracy, especially when dealing with unbalanced datasets.

A contingency table called a confusion matrix is used to show how well a classification model performs in both binary and multi-class scenarios. The true positives (TP), true negatives (TN), false negatives (FN), and false positives (FP) for each class are displayed in this matrix by comparing the actual labels (rows) with the predicted labels (columns). Researchers can compute other evaluation metrics like accuracy, precision, recall, and F1-score using this visual representation, in addition to directly observing which classes the model frequently misclassifies [17].

IV. Results

Analysis of the proposed CNN performance compared to ResNet 50 architecture showing the CNN accuracy value can be seen in TABLE I, confusion matrix results of input size 224×224 can be seen at Fig 8; confusion matrix results of input size 128×124 can be seen at Fig 9; confusion matrix results of input size 64×64 can be seen at Fig 10; and confusion matrix results of input size 32×32 can be seen at Fig 11; The suggested CNN model and the ResNet-50 architecture perform noticeably differently in terms of classification accuracy when compared across a range of input image sizes.

The suggested CNN outperformed ResNet-50 with an accuracy of 98.87% at the standard input size of 224×224 pixels, surpassing its 96.49% accuracy. The accuracy of the suggested model remained high at 98.62% when the input size was decreased to 128×128 pixels, whereas ResNet-50's accuracy decreased to 94.12%. Interestingly, ResNet-50 scored 92.87%, whereas the proposed CNN's best accuracy of 99% was observed at 64×64 pixels, demonstrating its resilience to lower-resolution inputs. ResNet-50's accuracy dropped significantly to 78.62% at the smallest input size of 32×32 pixels, but the suggested CNN exhibited a modest decline to 98%. The accuracy graph as ashown in Fig 7 makes it evident that the proposed CNN model does not overfit. By the tenth epoch, the training accuracy (blue line) and validation accuracy (orange line) have both converged close to 1.0 (99–100%) after increasing gradually throughout the first few epochs. The training procedure shows no discernible difference between the two curves, which follow a similar trajectory. This close alignment suggests that the model has acquired generalizable characteristics rather than simply memorizing the training data, as it shows consistent performance on both training and validation datasets. Further supporting the lack of overfitting is the validation accuracy, which shows no signs of declining after peaking.

TABLE I. Accuracy Results

Input Size	*ResNet 50[18]*	*Proposed CNN*
224×224	96,49%	98,87%
128×128	94,12%	98,62%
64×64	92,87%	99%
32×32	78,62%	98%

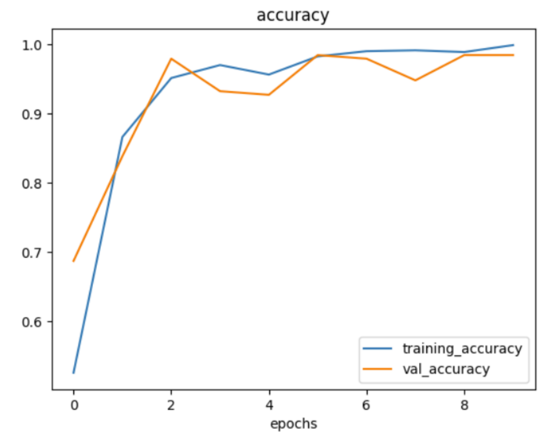

Fig. 7. Training and Validation Accuracy of the Proposed Model

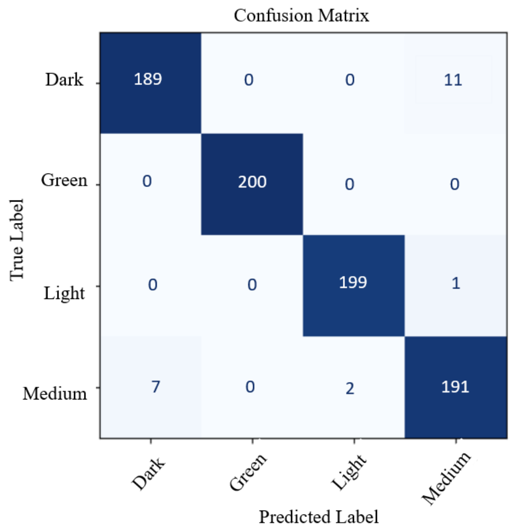

Fig. 8. Confusion Matrix Results of Proposed CNN using 224x224 input

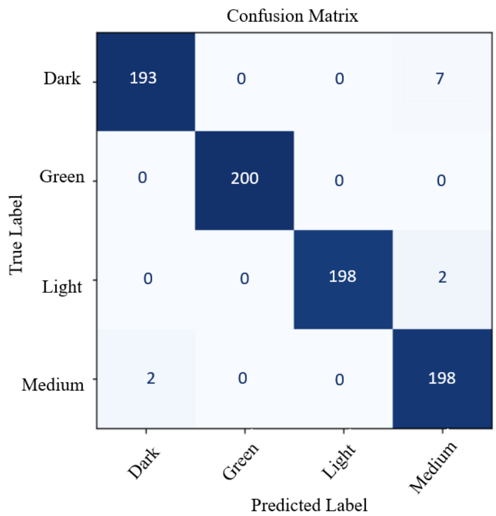

Fig. 9. Confusion Matrix Results of Proposed CNN using 128x128 input

As shown in Fig 8, the model performs accurately but shows some misclassification, particularly between dark and medium roasts (11 instances), indicating that increased resolution does not always improve class separation. In contrast, the model achieves better results with 128×128 input as shown with Fig 9, where misclassifications decrease as dark vs. medium errors drop to 7, and other classes remain highly accurate.

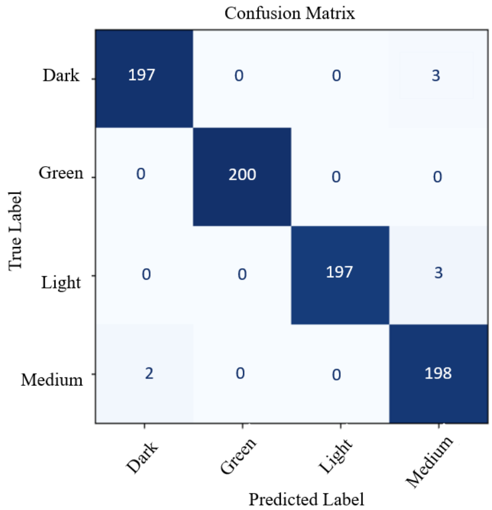

Fig. 10. Confusion Matrix Results of Proposed CNN using 64x64 input

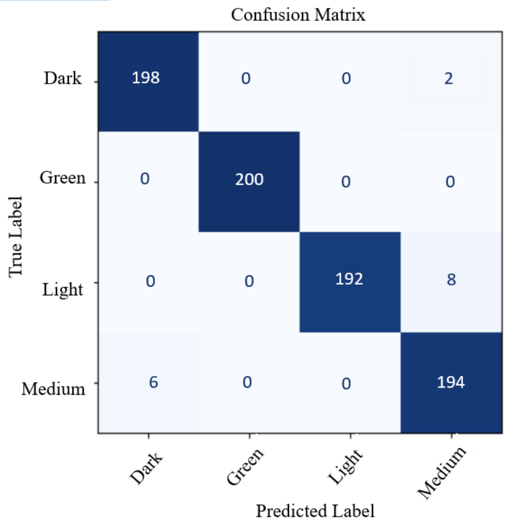

Fig. 11. Confusion Matrix Results of Proposed CNN using 32x32 input

The best performance is observed at 64×64 input as shown with Fig 10, where the model reaches 99% accuracy with only a few errors with just 3 dark roast beans misclassified as medium, and 3 light roasts also predicted as medium. At the lowest input size of 32×32 as shown with Fig 11, the model still performs strongly but shows a slight increase in confusion between visually similar classes such as light and medium, with 8 misclassified samples, and 6 medium roasts mislabeled as dark. These results collectively highlight that the 64×64 input offers the most balanced performance, with high accuracy and minimal misclassification, making it an ideal resolution for efficient and reliable coffee bean classification.

TABLE II. PRECISION, RECALL, F1-SCORE, SUPPORT RESULTS OF 64X64 INPUT

Label	*Precicion*	*Recall*	*F1-score*	*Support*
Dark	0.99	0.98	0.99	200
Green	1.00	1.00	1.00	200
Light	1.00	0.98	0.99	200
Medium	0.97	0.99	0.98	200

Table II presents the precision, recall, F1-score, and support results for the CNN model using a 64×64 input image size. Precision, recall, and F1-scores range from 0.97 to 1.00, indicating consistently high performance across all coffee roast levels. The model can reliably and consistently identify unroasted beans without misclassification, as shown by the Green class's accurate outcomes (precision, recall, and F1-score = 1.00). With F1-scores of 0.99, the Light and Dark roast levels also performed almost flawlessly, demonstrating the model's excellent capacity to discern minute variations in roast color and texture. The F1-score (0.98) and precision (0.97) of the Medium class were marginally lower but still very high, indicating slight feature overlaps between the medium and adjacent roast levels.

V. CONCLUSION

The proposed CNN model demonstrates its best performance at an input size of 64×64 pixels, achieving the highest accuracy of 99%. At this resolution, the confusion matrix indicates minimal misclassifications across all coffee roasting levels, suggesting the model's strong ability to distinguish between different categories even with relatively low image quality. This result highlights the model's efficiency, as it maintains excellent classification performance without the need for high-resolution inputs.

Compared to larger input sizes like 224×224, where accuracy slightly decreases and misclassifications increase especially between similar roast levels the 64×64 input provides an optimal balance between accuracy and computational efficiency, making it highly suitable for practical applications in coffee bean classification.

The paper makes a significant contribution in comparison to previous studies, in addition to these numerical results. Many earlier works on coffee bean classification used complicated designs like ResNet50 or VGG16, which demand greater processing power and longer training times, as well as high-resolution photos. Alternatively, the suggested CNN model achieves similar or even better accuracy using a simpler architecture and inputs with much lower resolution. The key markers of roast intensity, color and texture, can be effectively extracted using feature extraction without the need for deep or computationally costly networks. In just 10 epochs, the model achieves the required accuracy, demonstrating its quick convergence and effectiveness during training. Additionally, it is evident that the model does not display overfitting because the training and validation accuracy curves closely overlap.

Beyond technical performance, both coffee farmers and consumers can benefit from these results in the real world. The use of a CNN-based classification system by manufacturers can drastically cut down on the need for manual sorting, lowering labor expenses and human error while ensuring constant product quality throughout large-scale manufacturing. In addition to enabling faster turnaround times and more impartial grading standards, automated classification can help speed up the quality control procedure. As roast levels have a direct impact on flavor, aroma, and the overall coffee experience, this translates into increased consumer confidence in product consistency. The suggested approach promotes a more effective and transparent coffee supply chain, which is advantageous to all parties involved, from growers to consumers.

The study does, however, have a number of shortcomings that may be resolved in subsequent research. The model performs well under controlled imaging conditions, which include constant background, camera distance, and lighting. Such variances could impact the model's accuracy in actual industrial settings. It's also probable that not all potential coffee bean varietals, flaws, or environmental factors were covered by the dataset. Future research should broaden the dataset to encompass a variety of coffee varieties and take pictures in various settings to increase the model's resilience. Future studies could also look into deploying the model on embedded or mobile systems for real-time, on-site coffee quality assessment and using explainable AI (XAI) tools to interpret classification results.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to LPPM Universitas Tarumanegara for the support provided through the completion of this paper.

REFERENCES

[1] D. Buonocore, M. Carratu, and M. Lamberti, “Classification of coffee bean varieties based on a deep learning approach,” in *18th IMEKO TC10 Conference “Measurement for Diagnostics, Optimization and Control to Support Sustainability and Resilience” Warsaw, Poland*, 2022.
[2] P. Bahrumi, R. Ratna, and R. Fadhil, “Levelisasi penyangraian kopi: suatu kajian,” *Jurnal Ilmiah Mahasiswa Pertanian*, vol. 7, no. 1, pp. 522–525, 2022.
[3] N. O. Adiwijaya, H. I. Romadhon, J. A. Putra, and D. P. Kuswanto, “The quality of coffee bean classification system based on color by using k-nearest neighbor method,” in *Journal of Physics: Conference Series*, IOP Publishing, 2022, p. 012034.
[4] D. S. Leme, S. A. da Silva, B. H. G. Barbosa, F. M. Borém, and R. G. F. A. Pereira, “Recognition of coffee roasting degree using a computer vision system,” *Comput Electron Agric*, vol. 156, pp. 312–317, 2019.

[5] F. Yu *et al.*, “Progress in the application of cnn-based image classification and recognition in whole crop growth cycles,” *Remote Sens (Basel)*, vol. 15, no. 12, p. 2988, 2023.

[6] E. Hassan, “Enhancing coffee bean classification: a comparative analysis of pre-trained deep learning models,” *Neural Comput Appl*, vol. 36, no. 16, pp. 9023–9052, 2024.

[7] D. H. Suryana and W. K. Raharja, “Applying artificial intelligence to classify the maturity level of coffee beans during roasting,” *Int J Eng Sci Inf Technol*, vol. 3, no. 2, pp. 97–105, 2023.

[8] C. C. Enriquez, J. Marcelo, D. R. Verula, and N. J. Casildo, “Leveraging deep learning for coffee bean grading: A comparative analysis of convolutional neural network models,” *2024. Transactions on Science and Technology*, vol. 11, no. 1, pp. 1–6, 2024.

[9] J. N. C. Sarino, M. M. Bayas, E. R. Arboleda, E. C. Guevarra, and R. M. Dellosa, “Classification of coffee bean degree of roast using image processing and neural network,” *Int. J. Sci. Technol. Res*, vol. 8, no. 10, pp. 3231–3233, 2019.

[10] W. Manosroi, A. Phromfaiy, and W. Janthong, “Classification of Coffee Beans Color Level during Roasting by Feed Forward Back Propagation Neural Network,” *The Journal of King Mongkut’s University of Technology North Bangkok*, vol. 34, Sep. 2022, doi: 10.14416/j.kmutnb.2022.09.006.

[11] S. Ontoum, T. Khemanantakul, P. Sroison, T. Triyason, and B. Watanapa, “Coffee roast intelligence,” *arXiv preprint arXiv:2206.01841*, 2022.

[12] A. Michael and M. Garonga, “Classification model of ‘Toraja’arabica coffee fruit ripeness levels using convolution neural network approach,” *ILKOM J. Ilm*, vol. 13, pp. 226–234, 2021.

[13] P. F. I. R. H. Zer and R. Rosnelly, “Classification of Shape Bean Coffee Using Convolutional Neural Network,” in *Proceeding International Conference on Information Science and Technology Innovation*, 2023, pp. 6–10.

[14] Y. Hendrawan *et al.*, “AlexNet convolutional neural network to classify the types of Indonesian coffee beans,” in *IOP Conference Series: Earth and Environmental Science*, IOP Publishing, 2021, p. 012059.

[15] P. Purwono, A. Ma’arif, W. Rahmaniar, H. I. K. Fathurrahman, A. Z. K. Frisky, and Q. M. ul Haq, “Understanding of convolutional neural network (cnn): A review,” *International Journal of Robotics and Control Systems*, vol. 2, no. 4, pp. 739–748, 2022.

[16] Z. Haimeed, M. Amir, and M. J. Applied Sciences, “Comprehensive Image Classification using Hybrid CNN-LSTM Model with Advanced Feature Extraction on Coco Dataset,” Oct. 2024.

[17] J. Görtler *et al.*, “Neo: Generalizing Confusion Matrix Visualization to Hierarchical and Multi-Output Labels,” in *Conference on Human Factors in Computing Systems - Proceedings*, Association for Computing Machinery, Apr. 2022. doi: 10.1145/3491102.3501823.

[18] B. Koonce, “ResNet 50,” 2021, pp. 63–72. doi: 10.1007/978-1-4842-6168-2_6.