Leveraging deep learning for coffee bean grading: A comparative analysis of convolutional neural network models

Chris Chan Enriquez, Jinky Marcelo, Donah Rae Verula, Nathalie Joy Casildo
College of Information Sciences and Computing, Central Mindanao University, Maramag, Bukidnon, PHILIPPINES.
#Corresponding author. E-Mail: jinkymarcelo@cmu.edu.ph

ABSTRACT Accurate and efficient coffee bean grading is crucial for ensuring consistency, quality control, and standardization in the coffee industry. However, traditional manual methods are time-consuming, subjective, and costly. Deep learning approaches, particularly convolutional neural networks (CNNs), have shown remarkable performance in image classification tasks, offering a promising solution for automated coffee bean grading. However, these models encounter significant challenges due to the inherent characteristics of coffee beans, including their small size, limited visual features, and lack of texture. This study aims to address this challenge by comparatively analyzing various CNN models to identify the most effective architecture for automatic coffee bean grading. Specifically, we evaluate the performance of ten models: DenseNet, MobileNet, Inception, InceptionResNet, ResNet50, ResNet101, ResNet152, VGG16, VGG19, and Xception. Our experimental results demonstrate that DenseNet achieves the highest accuracy of 0.989, followed by MobileNet and ResNet152 with 0.982 and 0.980 accuracy, respectively. DenseNet has the highest precision and F1 score among all the models, with a precision of 0.996 and an F1 score of 0.992. VGG19 has the lowest accuracy of 0.902 and the lowest F1 score of 0.899. Overall, our analysis reveals that DenseNet, MobileNet and ResNet152 outperform other models for coffee bean grading accuracy. The findings of this study can contribute to the development of more accurate and efficient coffee bean grading systems that can benefit the coffee industry.

KEYWORDS: Coffee grading; Precision agriculture; Coffee beans; Computer vision; Deep learning.
Received 18 November 2023 **Revised** 24 February 2024 **Accepted** 29 February 2024 **Online** 5 April 2024

Original Article

TRANSACTIONS ON SCIENCE AND TECHNOLOGY

INTRODUCTION

Coffee quality (Samoggia *et al.*, 2019) is critical and depends on bean origin, processing, and roasting techniques (Amtate & Teferi, 2022; Chollet, 2017; Febriana *et al.*, 2022). Coffee bean grading, separating beans based on size, shape, and quality, is essential in determining bean value and final coffee taste (Garcia *et al.*, 2019). Traditional manual grading is time-consuming, subjective, and costly. Automatic coffee bean grading systems using deep learning techniques, particularly CNNs (He *et al.*, 2020), offer increased efficiency and consistency (Garcia *et al.*, 2019; Hsia *et al.*, 2016). However, CNN effectiveness in coffee bean grading remains a challenge (Chollet, 2017).

One of the problems which hinder the development CNN models for coffee bean grading is the limited availability of large-scale coffee bean datasets. For example, the dataset introduced by (Janandi & Cenggoro, 2020) only contains 617 images of four different types of coffee beans. Similarly, the datasets developed by (Kamilaris & Prenafeta-Boldú, 2018) and (Krizhevsky *et al.*, 2012) are also limited in terms of the types of coffee beans they include. Larger datasets can provide more representative and diverse samples for training and testing the models, and can also enable more sophisticated analyses of the features that affect coffee bean grading. In this study, we introduce an image dataset of coffee beans which are necessary to train and evaluate deep learning models for grading. This dataset offers a valuable resource for researchers in coffee bean classification and analysis.

Another major challenge of automatic coffee bean grading is the small size. The small size of coffee beans makes it hard to accurately classify them, which affects both traditional and modern grading methods. This challenge arises because convolutional neural networks (CNNs), commonly used for image classification, struggle to extract meaningful information from small objects with limited surface details and textures. Recent studies have explored different CNN architectures for this task, including SqueezeNet, InceptionV3, VGG16/19, MobileNet, NasNetMobile, and DenseNet achieving accuracy levels above from 87.3% to 97% (Hakim *et al*., 2020); Janandi & Cenggoro, 2020). Therefore, further research investigating more CNN architectures specifically designed for recognizing small objects is needed to improve the performance and reliability of automated coffee bean grading systems.

In this paper, we aim to develop and compare the performance of various CNN models for grading coffee beans using our own dataset. Our contributions are as follows: [1] We present the Coffee Bean Dataset, an image dataset containing 5,500 images of coffee beans obtained from a coffee farm, comprising various quality levels. We demonstrate the usefulness of the Coffee Bean Dataset for coffee bean grading tasks. The dataset can be used for future related research; [2] We exhaustively trained and evaluated our coffee bean dataset using transfer learning. We demonstrate that among the ten convolutional neural networks used for training, DenseNet, MobileNet and ResNet152 are the most effective models for coffee bean grading. This result contributes to the literature by demonstrating that exploring different CNN architectures can lead to more accurate and efficient coffee bean grading models. Our study could provide valuable insights into the potential of transfer learning for this task and could lead to the development of more accurate and efficient automatic grading systems for the coffee industry.

METHODOLOGY

Dataset Collection

The images were collected from a coffee farm in Maramag, Bukidnon, Philippines, with two quality levels, including good quality and bad quality. The image dataset of coffee beans was obtained through a systematic process in which each coffee bean was photographed individually using a smartphone camera. The images were taken in a single location with controlled lighting conditions during daytime to maintain consistency in the background and lighting. Figure 1 shows how a cluster of coffee beans was individually photographed into single images for training dataset.

Figure 1. Samples of coffee beans manually sorted and labeled for dataset collection

Data Preprocessing and annotation

We also applied data augmentation techniques such as rescaling, shear range, zoom range, horizontal flip, and rotation range to increase the size of the training dataset and reduce overfitting. The dataset was annotated with two classes: good quality and bad quality. Each class contains 2,500 images. The coffee bean dataset was split into three sub-datasets: 3,500 images for training, 1,000 images for validation, and 1,000 images for testing.

Classifier Training and Evaluation

As shown in Figure 2, the training consists of two main phases: transfer learning and classification. In the transfer learning phase, a pre-trained neural network model is imported and fine-tuned on the coffee bean dataset. The pre-trained model serves as a starting point and has already learned to extract useful features from images, which saves time and computational resources. Fine-tuning the final layers of the model on the new data makes it more suitable for the coffee bean grading task. Using transfer learning, we trained several CNN models (Howard *et al*., 2017; Huang *et al*., 2017; Simonyan *et al*., 2015; Szegedy *et al*., 2017) on our own coffee bean dataset. The pre-trained CNN models included DenseNet, Inception, InceptionResNet, MobileNet, ResNet50, ResNet101, ResNet152, VGG16, VGG19 and Xception. These network architectures are considered as state-of-the-art models for image classification. In the classification phase, the input image is fed into the trained model, which extracts features and predicts its class label. The trained model is evaluated on a validation dataset to determine its performance, which involves calculating various performance metrics such as accuracy, precision, recall, and F1 score. Finally, the model can be used to classify new images based on their features, which enables automated coffee beans grading.

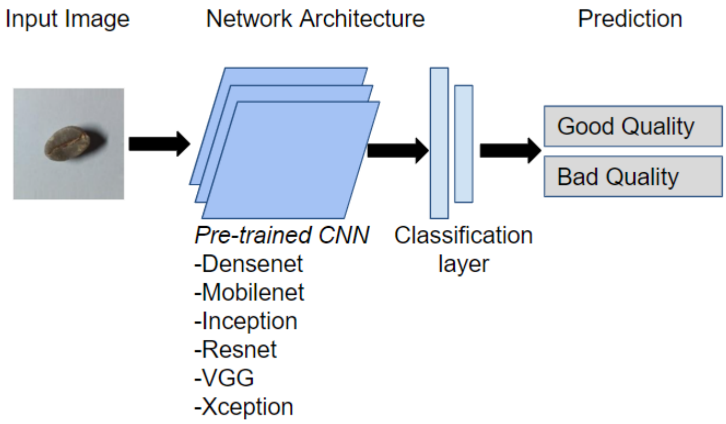


Figure 2. Research framework leveraging transfer learning and classification phase

RESULT AND DISCUSSION

We evaluated the performance of several deep learning models on the Coffee Bean Dataset for coffee bean grading tasks. Table 1 shows the performance comparison of various pre-trained CNN models on the coffee bean dataset. The results are presented in terms of the accuracy, precision, recall, F1 score, inference time and the number of parameters in the model. The results show that most of the models performed well, with accuracy ranging from 0.902 to 0.989.

Among the evaluated models, DenseNet achieved the highest accuracy of 0.989, with precision, recall, and F1 scores of 0.996, 0.988, and 0.992, respectively. This indicates that DenseNet is the best performing CNN model for coffee bean grading tasks, and it was able to correctly classify 98.9% of the image samples in the testing dataset. The relatively high precision and recall scores of 0.996 and 0.988, respectively, indicate that DenseNet was able to accurately classify both the positive and negative examples, and that it had relatively low false positive and false negative rates. The high F1 score of 0.992 also indicates that DenseNet achieved a good balance between precision and recall.

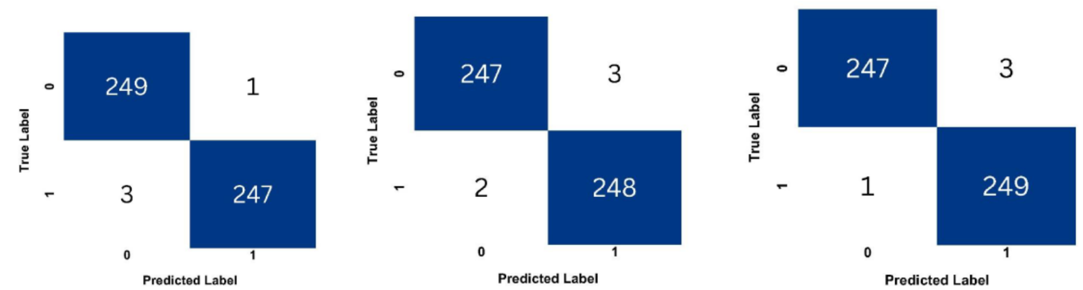
The MobileNet architecture also performed very well, achieving an accuracy score of 0.982, with a precision score of 0.988 and a recall score of 0.992. This suggests that the model has a low rate of false positives and false negatives, respectively, and its F1 score of 0.990 suggests that it has a high level of overall performance. On the other hand, the ResNet architecture, with variations of ResNet50, ResNet101, and ResNet152, also performed well, with accuracy scores ranging from 0.970

to 0.980. However, they have longer inference times compared to other models. The Inception and Inception ResNet architectures achieved accuracy scores of 0.961 and 0.962, respectively, with F1 scores of 0.972 and 0.980. Conversely, the VGG16 and VGG19 architectures had lower accuracy scores, with 0.938 and 0.902, respectively, but still achieved high precision and F1 scores. However, their recall scores were lower than other models, potentially indicating a higher rate of false negatives. Finally, the Xception architecture achieved an accuracy score of 0.970, with a precision score of 0.968, recall score of 0.972, and F1 score of 0.970.

Based on the accuracy metric, DenseNet, MobileNet, and ResNet152 are the top three models for automated coffee bean grading system with accuracy scores of 0.989, 0.982, and 0.980, respectively. These models have the highest accuracy scores among all the models listed in Table 1. As shown in the confusion matrix in Figure 3, the top three models have very few misclassified images as evident in its false positive and false negative rates. The grading system of DenseNet predicted 249 true positives and 247 true negatives while MobileNet predicted 247 true positives and 248 true negatives. Finally, Resnet152 correctly classified 247 true positives and 249 true negatives. Hence, the low number of misclassifications demonstrates the robustness and reliability of the models in distinguishing between good and bad quality coffee beans.

Table 1. Performance comparison of various pretrained models on coffee bean grading task.

Model	Performance Metrics				
	Precision	Recall	F1	Accuracy	Inference Time
Densenet	***0.996***	***0.988***	***0.992***	***0.989***	***6.732***
Inception	0.980	0.964	0.972	0.961	4.278
Inception ResNet	0.984	0.976	0.980	0.962	4.374
MobileNet	***0.988***	***0.992***	***0.990***	***0.982***	***1.561***
ResNet50	0.996	0.976	0.986	0.972	2.932
ResNet101	0.976	0.964	0.970	0.970	5.293
ResNet152	***0.988***	***0.996***	***0.992***	***0.980***	***8.025***
VGG16	0.983	0.900	0.939	0.938	1.025
VGG19	0.928	0.872	0.899	0.902	1.049
Xception	0.968	0.972	0.970	0.970	2.664



(a) (b) (c)

Figure 3. Confusion matrix of DenseNet (a), MobileNet (b) and ResNet152 (c) architecture on coffee bean dataset

Our results are consistent with previous studies (Unal *et al.*, 2022; Wallelign *et al.*, 2019; Rivalto *et al.*, 2020) that have shown that deep learning models, especially CNNs, are effective in coffee bean grading. A study by (Unal *et al.*, 2022) found that Lightweight Deep CNN and MobileNet achieved the highest classification accuracy and outperformed other models. Relatedly, MobileNet demonstrated super performance in classifying roasted coffee beans embedded into android smartphones (Najafabadi *et al.*, 2015). Furthermore, DenseNet outperformed other CNN models in coffee fruit classification (Vilcamiza *et al.*, 2022). However, the study was focused on different stages of ripening. To the best of our knowledge, our study is the first that demonstrates the superiority of Densenet for coffee bean grading, indicating its potential for practical applications in the coffee industry. It is worth noting that the dataset used in the present study is significantly larger than the datasets used in the recent studies of (Vilcamiza *et al.*, 2022; Walleligh *et al.*, 2019), which only used 600 images and 1266 images, respectively.

Overall, the results of the present study support the potential of CNN models for accurate coffee bean grading, and suggest that using larger datasets and evaluating multiple models can improve classification accuracy. However, it is important to note that the results of this study are based on a single dataset, and may not generalize to other datasets. More research is needed to further explore the optimal models and datasets for coffee bean classification in different contexts to improve the inference speed. Moving forward, the next step in our research is to integrate the model into an embedded system using low-cost hardware materials such as Raspberry Pi that can be used in real-world settings.

CONCLUSION

In this paper, we developed and evaluated the performance of various CNN models for coffee bean grading tasks using the Coffee Bean Dataset, an image dataset containing 5,500 images of coffee beans obtained from a coffee farm, comprising various quality levels. We demonstrated the utility of the dataset for coffee bean classification and analysis using deep learning models. Our results demonstrate that transfer learning with DenseNet, MobileNet, and ResNet152 models are outperform other models for coffee bean grading tasks, with DenseNet achieving the highest accuracy of 0.989. Future work in this field should focus on expanding the Coffee Bean Dataset to include more diverse coffee beans from different regions, processing techniques, and roasting levels. This would enable the development of more robust and generalizable models that can handle variations in coffee bean characteristics. Finally, the developed models could be integrated into a real-world coffee bean grading system, and their effectiveness and efficiency could be evaluated in practical applications.

ACKNOWLEDGEMENTS

We extend our deepest gratitude to the coffee grower and coffee sorting experts who generously provided us with the data used in this study. We are also grateful for the resources and opportunities provided by the College of Information Sciences and Computing of Central Mindanao University, which allowed us to undertake this research study.

REFERENCES

[1] Amtate, G. & Teferi, D. 2022. Multiclass classification of Ethiopian coffee bean using deep learning. *Ethiopian Journal of Science*, 45(3), 309-321.

[2] Chollet, F. 2017. Xception: Deep learning with depthwise separable convolutions. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 21 - 26 July 2017, Honolulu, HI, USA. pp. 1251-1258.

[3] Febriana, A., Muchtar, K., Dawood, R. & Lin, C. Y. 2022. USK-COFFEE Dataset: A Multi-class Green Arabica Coffee Bean Dataset for Deep Learning. *2022 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom)*. 16-18 June 2022, Malang, Indonesia. pp. 469-473.

[4] García, M., Candelo-Becerra, J. E. & Hoyos, F. E. 2019. Quality and defect inspection of green coffee beans using a computer vision system. *Applied Sciences*, 9(19), 4195.

[5] Hakim, M., Djatna, T. & Yuliasih, I. 2020. Deep learning for roasting coffee bean quality assessment using computer vision in mobile environment. *2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*. 17-18 October, 2020, Depok, Indonesia, pp 363-370.

[6] He, K., Zhang, X., Ren, S. & Sun, J. 2016. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* 27-30 June, 2016, Las Vegas, USA. pp 770-778.

[7] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T. & Adam, H. 2017. MobileNets: efficient convolutional neural networks for mobile vision applications. *ArXiv, abs/1704.04861.*

[8] Hsia, C. H., Lee, Y. H. & Lai, C. F. 2021. An explainable and lightweight deep convolutional neural network for quality detection of green coffee beans. *Sensors*, 21(6), 2046.

[9] Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. 2017. Densely connected convolutional networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 21-26 July, 2017, Honolulu, Hawaii, USA. pp 4700-4708.

[10] Janandi, R. & Cenggoro, T. W. 2020. An implementation of convolutional neural network for coffee beans quality classification in a mobile information system. *2020 International Conference on Information Management and Technology (ICIMTech)*. 13-14 August, 2020, Bandung, Indonesia. pp 218-222.

[11] Kamilaris, A. & Prenafeta-Boldú, F. 2018. A review of the use of convolutional neural networks in agriculture. *The Journal of Agricultural Science*, 156(3), 312-322.

[12] Krizhevsky, A., Sutskever, I. & Hinton, G. E. 2012. ImageNet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 1097-1105.

[13] Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R. & Muharemagic, E. 2015. Deep learning applications and challenges in big data analytics. *Journal of Big Data,* 2, Article number 1.

[14] Rivalto, A., Pranowo, & Santoso, A. J. 2020. Classification of indonesian coffee types with deep learning. *AIP Conference Procceedings,* 2217(1), 030014.

[15] Samoggia, A. & Riedel, B. 2019. Consumers' Perceptions of Coffee Health Benefits and Motives for Coffee Consumption and Purchasing. *Nutrients* 18;11(3):653.

[16] Simonyan, K. & Zisserman, A. 2015. Very deep convolutional networks for large-scale image recognition. *arXiv*, 1409.1556.

[17] Szegedy, C., Ioffe, S., Vanhoucke, V. & Alemi, A. A. 2017. Inception-v4, inception-resnet and the impact of residual connections on learning. *AAAI'17: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*. 4-9 February, 2017, San Francisco, California, USA. pp 4278–4284.

[18] Tamayo-Monsalve, M. A., Mercado-Ruiz, E., Villa-Pulgarin, J. P. & Bravo-Ortiz, M. A. 2022. Coffee Maturity Classification Using Convolutional Neural Networks and Transfer Learning. *IEEE Access*, 10, 42971-42982.

[19] Unal, Y., Taspinar, Y. S., Cinar, I. & Kursun, R. 2022. Application of pre-trained deep convolutional neural networks for coffee beans species detection. *Food Analytical Methods*, 15(1), 3232-3243.

[20] Vilcamiza, G., Trelles, N., Vinces, L. & Oliden, J. 2022. A coffee bean classifier system by roast quality using convolutional neural networks and computer vision implemented in an NVIDIA Jetson Nano. *Congreso Internacional de Innovación y Tendencias en Ingeniería (CONIITI)*. 5-7 October, 2022, Bogota, Colombia, pp 1-6.

[21] Wallelign, S., Polceanu, M., Jemal, T. & Buche, C. 2019. Coffee grading with convolutional neural networks using small datasets with high variance. *Journal of WSCG*, 1-2, 113-120.

TRANSACTIONS ON SCIENCE AND TECHNOLOGY