e-ISSN 2549-7472
Edumatic: Jurnal Pendidikan Informatika
Volume 10, Issue 1, **130-139**.
https://e-journal.hamzanwadi.ac.id/index.php/edumatic
DOI: 10.29408/edumatic.v10i1.34044

On the Effectiveness of Lightweight CNN Architectures for Fine-Grained Coffee Bean Classification

Akbar Muhamad Burhanudhin [1,*] , Usman Sudibyo [1] , Eka Putra Agus Meindiawan [1]

[1] Universitas Dian Nuswantoro, Indonesia
[*] Corresponding author: Akbar Muhamad Burhanudhin, Universitas Dian Nuswantoro, Indonesia
✉ 111202113731@mhs.dinus.ac.id

Received: 28 January 2025 | Revised: 9 February 2026 | Accepted: 30 March 2026 | Published: 11 April 2026

Abstract
Distinguishing coffee bean varieties remains a significant challenge in the agricultural industry due to high inter-class similarity and the subtle morphological differences between species. This study aims to conduct a comparative evaluation of MobileNetV2 and EfficientNetB0 for fine-grained coffee bean classification, specifically investigating how efficiency-oriented architectural mechanisms such as depthwise separable convolution and compound scaling influence feature extraction. The research employed a quantitative experimental method using a private dataset of 2,400 images comprising Arabica, Robusta, and Liberica varieties. Data preprocessing included resizing to 224×224 pixels and augmentation, followed by training the two architectures using transfer learning under a controlled experimental framework. The results showed that EfficientNetB0 achieved superior performance with a testing accuracy of 99.17%, while MobileNetV2 attained a competitive accuracy of 98.33% with lower computational complexity. These results demonstrate that while EfficientNetB0 is optimal for high-precision industrial sorting, MobileNetV2 offers a highly efficient alternative for resource-constrained mobile applications. This study provides a scalable framework for automating quality control, effectively balancing architectural efficiency with the sensitivity required for accurate coffee variety identification.

Keywords: cnn; coffee bean; deep learning; efficientnetb0; mobilenetv2

To cite this article: Burhanudhin, A. M., Sudibyo, U., & Meindiawan, E. P. A. (2026). On the Effectiveness of Lightweight CNN Architectures for Fine-Grained Coffee Bean Classification. *Edumatic: Jurnal Pendidikan Informatika*, *10*(1), 130–139. https://doi.org/10.29408/edumatic.v10i1.34044

INTRODUCTION

Fine-grained image classification presents a significant challenge in computer vision, particularly when objects share globally similar appearances but differ in subtle local visual characteristics (Liu et al., 2024; Pan et al., 2025). In agricultural imaging contexts, this problem becomes more complex due to the coexistence of high intra-class variability and strong inter-class similarity within relatively limited (Li et al., 2023; Lu et al., 2022). Coffee bean variety identification represents a practical manifestation of this challenge, as Arabica, Robusta, and Liberica beans exhibit highly similar shapes and color distributions, with their distinguishing characteristics primarily found in micro-textural patterns and surface morphology (Korkmaz et al., 2025; Shin et al., 2024). These subtle irregularities in surface structure require models with

high spatial sensitivity, as misclassification frequently occurs in agricultural datasets characterized by uneven feature distribution (Corthis et al., 2024).

The advancement of Convolutional Neural Networks (CNNs) has substantially improved automated image classification by enabling hierarchical feature learning directly from raw image data (Bhagat et al., 2024; Yaseliani et al., 2022). Within agricultural applications, CNN-based approaches have demonstrated strong capability in extracting discriminative visual features, including structural and textural patterns that are difficult to capture using handcrafted descriptors (Fareed et al., 2022; Lambert et al., 2024). However, the deployment of deep CNN architectures in practical agricultural environments often introduces computational challenges, particularly when implemented on mobile or embedded systems (González-Briones et al., 2025; Ni et al., 2024). Although deeper models such as ResNet and EfficientNet variants can achieve high accuracy, their computational complexity limits their applicability in resource-constrained scenarios (Kansal et al., 2024; Raja et al., 2022), and their high computational demands frequently result in unacceptable latency for real-time field monitoring (Aldea et al., 2023; Bian et al., 2022). Furthermore, previous studies on coffee bean classification, whether based on classical image processing or deep learning approaches, have reported promising performance (Hassan, 2024; Strelcenia & Prakoonwit, 2023), yet they predominantly treat CNN architectures as black-box systems, focusing on final accuracy without examining how architectural design influences feature representation and learning dynamics (Anto et al., 2025; Zhao et al., 2024).

This leads to a clear limitation in the current literature, namely the lack of architecture-level analysis in CNN-based coffee bean classification. Most studies do not investigate how specific architectural choices affect the model's ability to capture fine-grained features, especially under conditions of extreme visual similarity. In addition, comparative evaluations conducted under consistent preprocessing pipelines and training configurations remain limited (Alimova et al., 2022; Issitt et al., 2022), making it difficult to draw reliable conclusions regarding model performance. The issue is further compounded by the limited exploration of model robustness against minor visual disturbances that frequently occur in real-world agricultural environments, such as variations in lighting, orientation, and surface contamination (Binning et al., 2025; Shao et al., 2024).

Beyond these limitations, a deeper scientific gap exists in understanding the relationship between efficiency-oriented CNN architectures and their ability to preserve high-frequency visual information. Architectural mechanisms such as depthwise separable convolutions in MobileNetV2 and compound scaling in EfficientNetB0 are specifically designed to optimize computational efficiency; however, their influence on the extraction of micro-textural features remains insufficiently explored. This issue is critical because precise botanical classification relies heavily on capturing subtle surface variations. Without a systematic investigation of these mechanisms, the selection of lightweight models for deployment in agricultural settings risks prioritizing computational efficiency at the expense of classification reliability.

Therefore, this study aims to conduct a comparative evaluation of MobileNetV2 and EfficientNetB0 within a controlled experimental framework for coffee bean classification. The objective is not only to measure classification performance but also to analyze how architectural design affects micro-texture feature extraction, inter-class confusion patterns, and convergence stability during training. The novelty of this research lies in its architecture-oriented approach, which goes beyond conventional accuracy-based evaluation by providing insight into the internal behavior of lightweight CNN models. The findings of this study are expected to contribute to a more informed selection of efficient and reliable deep learning models for real-world agricultural applications, particularly in resource-constrained environments.

METHOD

This study adopts a quantitative experimental approach to classify three coffee bean varieties: Arabica, Robusta, and Liberica. The primary dataset consists of 300 original images acquired through direct collection, which were subsequently expanded to 2,400 samples through data augmentation to enhance feature variation and prevent overfitting. This augmentation process which includes a transformation function applied to rotation (up to 90°), zooming, shearing, and brightness adjustments was applied exclusively to the training set to ensure the objectivity of the validation and testing phases. All images were resized to 224 × 224 pixels to comply with the input specifications of the selected convolutional neural network (CNN) architectures. The dataset was partitioned into training, validation, and testing subsets using an 85:10:5 ratio, resulting in 1,920 training images, 360 validation images, and 120 test images. This distribution prioritizes robust feature learning during the training phase while maintaining a separate, unbiased test set for performance evaluation. The overall experimental workflow, from data acquisition to performance evaluation, is systematically illustrated in Figure 1.

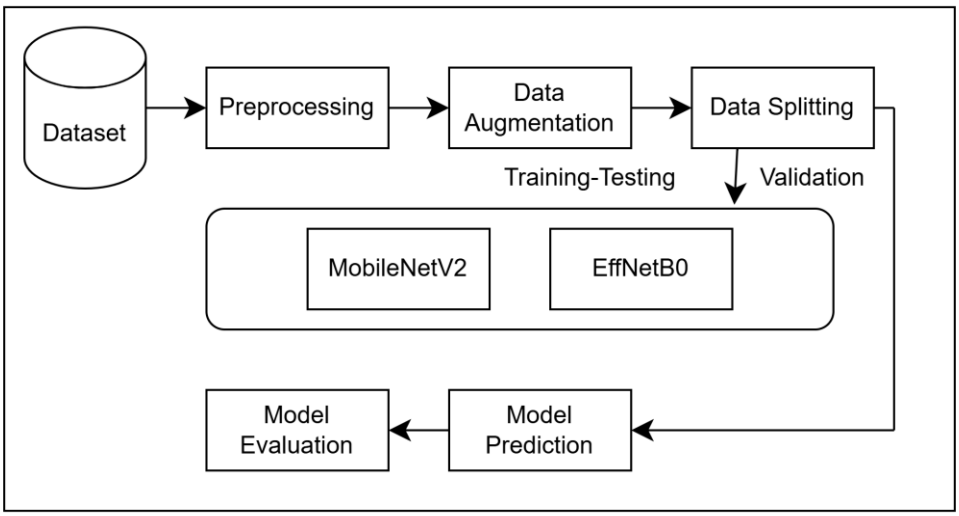

Figure 1. Research method

This study evaluates two lightweight CNN architectures: MobileNetV2 and EfficientNetB0. MobileNetV2 was selected for its depthwise separable convolutions and inverted residual blocks, which offer high computational efficiency for resource-constrained environments. In contrast, EfficientNetB0 was chosen for its compound scaling strategy, which systematically balances network depth, width, and resolution to capture fine-grained textures and morphological patterns essential for distinguishing visually similar coffee varieties. Experiments were conducted using the TensorFlow and Keras frameworks within a Google Colab environment, supported by an NVIDIA Tesla T4 GPU. A transfer learning strategy was implemented by utilizing pre-trained ImageNet weights. The convolutional backbones of both models remained frozen to preserve learned generic features, while only the custom classification heads consisting of Global Average Pooling, a dropout layer to mitigate overfitting, and a Dense layer with Softmax activation were trained. The optimization process employed the Adam algorithm with a learning rate of 0.0001, a batch size of 32, and a duration of 25 epochs. These hyperparameters were selected based on preliminary pilot experiments to achieve stable convergence. To further enhance training stability and efficiency, ReduceLROnPlateau and EarlyStopping mechanisms were integrated into the pipeline. The models were optimized by minimizing the Categorical Cross-Entropy loss function (L), which

measures the divergence between the predicted probability distribution and the ground-truth labels (see equation 1).

$$L_{CE} = -\sum_{i=1}^{C} y_i \log(\hat{y})_i \quad (1)$$

Where M denotes the total number of classes, yc represents the ground-truth label (one-hot encoded), and y-hat c is the predicted probability for class c. Finally, the model performance was evaluated using Accuracy, Precision, Recall, and F1-Score. These metrics provide a comprehensive assessment of the models' ability to minimize false predictions and distinguish between visually similar coffee bean classes under controlled experimental settings.

RESULTS AND DISCUSSION

Results

In alignment with the research workflow illustrated in Figure 1, the results of this study are presented in stages, beginning with the data preprocessing outcomes and concluding with a detailed analysis of architectural performance. The initial phase of the research successfully transformed the primary dataset of 300 original images into 2,400 training-ready samples through data augmentation techniques. This process, which involved geometric and photometric manipulations such as rotation, zooming, and brightness adjustments proved crucial in enriching feature variation without compromising the integrity of the test data. By partitioning the dataset using an 85:10:5 ratio, the models were able to learn the visual characteristics of coffee beans in depth during the training phase before being objectively evaluated against 120 independent images. The model training experiments, conducted in a GPU-based computing environment over 25 epochs, revealed distinct learning characteristics for each architecture.

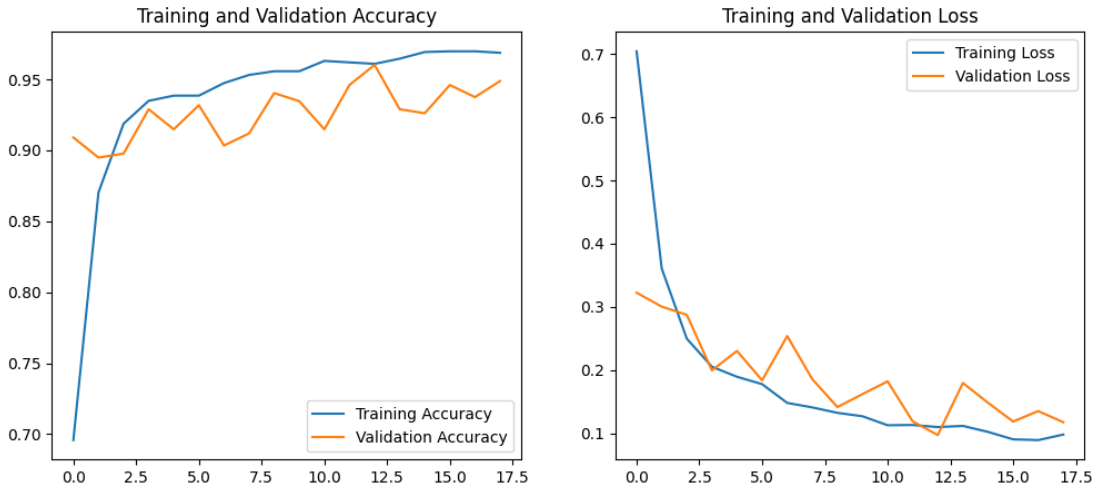

Figure 2. MobileNet-V2 performance

As monitored through the learning curves in Figure 2 and Figure 3, a consistent convergence between training and validation metrics was observed for both models. Figure 2 displays the performance profile of MobileNetV2, which reached peak validation accuracy stability at epoch 13 with a value of 96.02%. Meanwhile, Figure 3 illustrates the superiority of EfficientNetB0, which demonstrated more consistent feature learning behavior throughout the training process, achieving a validation accuracy of 99.15% with a significantly lower validation loss of 0.0718. The minimal gap between the training and validation lines in both

figures indicates that the applied regularization and augmentation strategies effectively mitigated overfitting, resulting in a stable convergence profile.

The superiority of EfficientNetB0 is further validated through the final evaluation using the test dataset. EfficientNetB0 achieved a testing accuracy of 99.17%, outperforming MobileNetV2, which reached 98.33%. To provide a clearer and more structured analysis, the evaluation results are separated into two tables. Table 1 presents the detailed performance of MobileNetV2. The model achieved strong classification results, with F1-Scores of 0.98 for Arabica, 0.97 for Liberica, and a perfect 1.00 for Robusta. These results indicate that MobileNetV2 is capable of capturing general discriminative features across classes. However, a slight imbalance between precision and recall is observed in the Arabica and Liberica classes, reflecting the model's limitation in distinguishing highly similar micro-textures. In terms of efficiency, MobileNetV2 recorded a faster inference time of 74.89 ms, highlighting its suitability for deployment in resource-constrained environments.

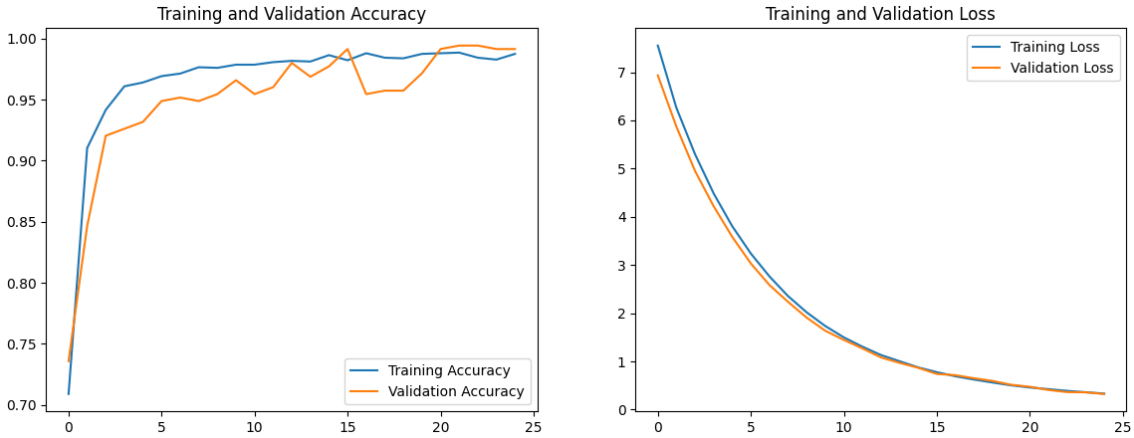

Figure 3. EfficientNetB0 perfomance

Table 1. Model evaluation mobilenetv2

Class	Precision	Recall	F1-Score	Accuracy	Inference Time
Arabica	0.95	1.00	0.98	0.98	74.89 ms
Liberica	1.00	0.95	0.97	0.98	74.89 ms
Robusta	1.00	1.00	1.00	0.98	74.89 ms

Table 2. Model evaluation efficientnetb0

Class	Precision	Recall	F1-Score	Accuracy	Inference Time
Arabica	0.98	1.00	0.99	0.99	83.74 ms
Liberica	1.00	0.97	0.99	0.99	83.74 ms
Robusta	1.00	1.00	1.00	0.99	83.74 ms

Table 2 shows the performance of EfficientNetB0, which consistently outperforms MobileNetV2 across all evaluation metrics. The model achieved near-perfect results, with average Precision, Recall, and F1-Score values reaching 0.99. Notably, EfficientNetB0 demonstrates a better balance between precision and recall in the Arabica and Liberica classes, which are known to exhibit high visual similarity. The inference time of 83.74 ms is slightly higher than that of MobileNetV2; however, this increase is justified by the significant improvement in classification accuracy and robustness.

A deeper architectural analysis provides a scientific explanation for these performance differences, particularly regarding the challenges of similar surface textures and color distributions between Arabica and Liberica varieties. The use of Depthwise Separable Convolution in MobileNetV2 maximizes computational efficiency; however, this mechanism tends to lead to the loss of subtle spatial information required to distinguish micro-textures on coffee beans. Conversely, the Compound Scaling method in EfficientNetB0 allows the model to balance network depth, width, and resolution proportionally. This scaling capability enables EfficientNetB0 to extract micro-features, such as texture strokes, and macro-features, such as seed morphology, simultaneously with greater precision.

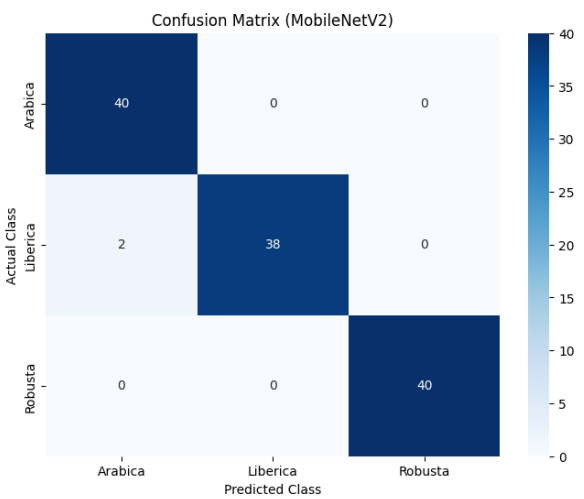

Figure 4. MobileNet-V2 prediction

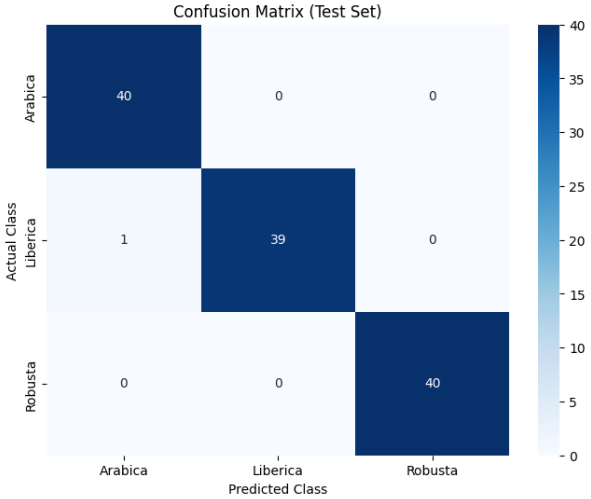

Figure 5. EfficientNetB0 prediction

The significance of this feature extraction capability is visualized in the model prediction results. The results in Figure 4 show that MobileNetV2 is capable of recognizing general patterns but still exhibits a small margin of uncertainty for varieties with complex textures. Meanwhile, the results in Figure 5 demonstrate that EfficientNetB0 provides higher and more stable prediction confidence, proving its robustness in handling inter-class similarity. Confusion matrix analysis reinforces these findings: MobileNetV2 showed minor errors in predicting Liberica varieties by misclassifying them as Arabica (recall of 0.95), whereas EfficientNetB0 successfully minimized this ambiguity with a higher recall value of 0.97. Overall, the choice of architecture should be tailored to implementation needs; EfficientNetB0 is recommended for tasks requiring high precision, while MobileNetV2 remains an efficient alternative for edge computing environments.

Discussion

The experimental results indicate that both MobileNetV2 and EfficientNetB0 achieve high classification performance under controlled experimental conditions; however, a deeper analytical perspective reveals substantial differences in learning behavior and representational capacity that are strongly influenced by architectural design. The convergence patterns observed in both models, characterized by minimal divergence between training and validation metrics, confirm that the applied augmentation and regularization strategies effectively promote generalization. EfficientNetB0 demonstrates superior convergence stability, as evidenced by lower validation loss and more consistent optimization trajectories across epochs. This behavior can be theoretically explained through the compound scaling principle, which systematically balances network depth, width, and input resolution, thereby improving gradient flow and enabling more stable hierarchical feature learning (Tan & Le, 2021).

A more nuanced interpretation emerges when examining class-level performance in the context of fine-grained visual similarity. MobileNetV2 achieves strong overall accuracy; however, the imbalance between precision and recall in the Arabica and Liberica classes indicates a limitation in capturing subtle micro-textural variations. This observation is consistent with prior findings that depthwise separable convolution, although computationally efficient, constrains inter-channel feature interaction and reduces the richness of spatial representations (Zhao et al., 2024). In contrast, EfficientNetB0 consistently achieves a more balanced precision–recall profile, indicating improved sensitivity to high-frequency visual patterns. Such performance aligns with studies demonstrating that deeper and well-scaled architectures enhance discriminative feature extraction in fine-grained classification tasks by enabling multi-level feature abstraction (Bera et al., 2023; Valarmathi et al., 2023).

The confusion matrix analysis further strengthens this interpretation by revealing differences in class separability. The misclassification of Liberica as Arabica in MobileNetV2 reflects the inherent difficulty of distinguishing categories with overlapping morphological and textural characteristics. EfficientNetB0 reduces this ambiguity and achieves higher recall, suggesting that the architecture is more effective in learning compact and separable feature embeddings. This finding is consistent with theoretical perspectives in deep learning, where increased network capacity and balanced scaling contribute to improved intra-class compactness and inter-class separability (Li et al., 2023; Lu et al., 2022). Consequently, EfficientNetB0 demonstrates a stronger ability to model complex visual distributions associated with fine-grained agricultural datasets.

The analysis of computational efficiency reveals a clear trade-off between inference speed and classification performance. MobileNetV2 achieves lower inference time, confirming its suitability for deployment in edge computing and resource-constrained environments. This efficiency is primarily attributed to the use of lightweight convolutional operations designed to minimize computational overhead (Kansal et al., 2024). EfficientNetB0, on the other hand, requires slightly higher computational resources; however, the improvement in accuracy and robustness justifies this increase, particularly in applications where precision is critical. This trade-off reflects a well-established principle in deep learning, where enhanced representational capacity is associated with increased computational demand (González-Briones et al., 2025).

In relation to prior studies on coffee bean classification, existing research predominantly emphasizes accuracy as the primary evaluation metric while treating CNN architectures as black-box systems (Hassan, 2024; Korkmaz et al., 2025). The present study advances this line of research by providing an architecture-oriented analysis that explicitly links design mechanisms to feature extraction capability and classification behavior. The consistent classification of the Robusta class by both models indicates that categories with distinct morphological characteristics can be effectively learned even by lightweight architectures. Conversely, the reduced performance observed in Arabica and Liberica highlights the necessity of architectures capable of capturing fine-grained visual cues. This observation reinforces the argument that architectural configuration plays a critical role in determining model effectiveness in complex classification scenarios.

The controlled experimental design employed in this study further enhances the reliability of the findings. By maintaining identical preprocessing procedures, training configurations, and evaluation protocols, the observed performance differences can be directly attributed to architectural characteristics rather than external experimental variations. Such methodological consistency addresses limitations reported in previous works, where heterogeneous experimental setups often obscure the true impact of model design (Alimova et al., 2022; Issitt et al., 2022).

Several limitations should be acknowledged to provide a balanced interpretation of the results. The dataset was collected under controlled environmental conditions, which may not

fully represent the variability encountered in real-world agricultural settings, including illumination changes, occlusions, and surface contamination. Consequently, the reported performance represents an upper-bound estimation of model capability. Future investigations should incorporate more diverse and heterogeneous datasets to evaluate generalization performance more rigorously. In addition, the integration of explainability techniques, such as attention visualization methods, would provide deeper insight into the spatial regions utilized by each model during classification, thereby enhancing interpretability and transparency (Lambert et al., 2024).

These findings demonstrate that the effectiveness of lightweight CNN architectures in fine-grained classification is fundamentally determined by their architectural design. EfficientNetB0 exhibits superior capability in capturing subtle visual patterns and maintaining classification stability through balanced scaling mechanisms, whereas MobileNetV2 offers a computationally efficient alternative suitable for real-time applications.

CONCLUSION

This study evaluated the performance of lightweight convolutional neural network architectures, namely MobileNetV2 and EfficientNetB0, for fine-grained classification of coffee bean varieties under identical experimental conditions. The results indicate that EfficientNetB0 achieves higher classification accuracy, reflecting its stronger capability in capturing subtle visual features of visually similar coffee beans, while MobileNetV2 provides competitive performance with lower computational requirements, making it suitable for implementation on resource-limited systems. These findings confirm that architectural selection should consider both accuracy and computational efficiency according to application needs. The experiments were conducted using a limited dataset collected in controlled conditions and focused on a three-class classification task; therefore, the obtained results represent indicative performance. Future research may involve larger and more diverse datasets, repeated evaluations, and real-world testing scenarios to further assess model generalization and robustness. This research contributes to the application of lightweight deep learning models in agricultural image classification, particularly in supporting efficient and practical decision-making systems.

REFERENCES

Aldea, C. L., Bocu, R., & Solca, R. N. (2023). Real-Time Monitoring and Management of Hardware and Software Resources in Heterogeneous Computer Networks through an Integrated System Architecture. *Symmetry*, *15*(6). https://doi.org/10.3390/sym15061134

Alimova, I., Tutubalina, E., & Nikolenko, S. I. (2022). Cross-Domain Limitations of Neural Models on Biomedical Relation Classification. *IEEE Access*, *10*, 1432–1439. https://doi.org/10.1109/ACCESS.2021.3135381

Anto, I. A. F., Wibowo, J. W., Munandar, A., & Salim, T. I. (2025). Comparative performance analysis of convolutional neural network-architectures on coffee-bean roast classification. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, *23*(6), 1590-1599. https://doi.org/10.12928/telkomnika.v23i6.27090

Bera, A., Nasipuri, M., Krejcar, O., & Bhattacharjee, D. (2023). Fine-Grained Sports, Yoga, and Dance Postures Recognition: A Benchmark Analysis. *IEEE Transactions on Instrumentation and Measurement*, *72*(0), 1–12. https://doi.org/10.1109/TIM.2023.3293564

Bhagat, D., Vakil, A., Gupta, R. K., & Kumar, A. (2024). Facial Emotion Recognition (FER) using Convolutional Neural Network (CNN). *Procedia Computer Science*, *235*(2023), 2079–2089. https://doi.org/10.1016/j.procs.2024.04.197

Bian, J., Arafat, A. Al, Xiong, H., Li, J., Li, L., Chen, H., Wang, J., Dou, D., & Guo, Z. (2022).

Machine Learning in Real-Time Internet of Things (IoT) Systems: A Survey. *IEEE Internet of Things Journal*, *9*(11), 8364–8386. https://doi.org/10.1109/JIOT.2022.3161050

Binning, S. A., Ackerly, K. L., Cooke, S. J., Fusi, M., Gomez Isaza, D. F., Hardison, E. A., Martin, S., Munson, A., Pineda, M., Schwieterman, G. D., Reichard, M., Rummel, A., & Blewett, T. A. (2025). The lab-field continuum in conservation physiology research: leveraging multiple approaches to inform policy and practice. *Conservation Physiology*, *13*(1), 1–14. https://doi.org/10.1093/conphys/coaf063

Corthis, P. B., Ramesh, G. P., García-Torres, M., & Ruíz, R. (2024). Effective Identification and Authentication of Healthcare IoT Using Fog Computing with Hybrid Cryptographic Algorithm. *Symmetry*, *16*(6). https://doi.org/10.3390/sym16060726

Fareed, M. M. S., Zikria, S., Ahmed, G., Mui-Zzud-Din, Mahmood, S., Aslam, M., Jillani, S. F., Moustafa, A., & Asad, M. (2022). ADD-Net: An Effective Deep Learning Model for Early Detection of Alzheimer Disease in MRI Scans. *IEEE Access*, *10*, 96930–96951. https://doi.org/10.1109/ACCESS.2022.3204395

González-Briones, A., Florez, S. L., Chamoso, P., Castillo-Ossa, L. F., & Corchado, E. S. (2025). Enhancing Plant Disease Detection: Incorporating Advanced CNN Architectures for Better Accuracy and Interpretability. *International Journal of Computational Intelligence Systems*, *18*(1). https://doi.org/10.1007/s44196-025-00835-2

Hassan, E. (2024). Enhancing coffee bean classification: a comparative analysis of pre-trained deep learning models. *Neural Computing and Applications*, *36*(16), 9023–9052. https://doi.org/10.1007/s00521-024-09623-z

Issitt, R. W., Cortina-Borja, M., Bryant, W., Bowyer, S., Taylor, A. M., & Sebire, N. (2022). Classification Performance of Neural Networks Versus Logistic Regression Models: Evidence From Healthcare Practice. *Cureus*, *14*(2), 1–8. https://doi.org/10.7759/cureus.22443

Kansal, K., Chandra, T. B., & Singh, A. (2024). ResNet-50 vs. EfficientNet-B0: Multi-Centric Classification of Various Lung Abnormalities Using Deep Learning "session id: ICMLDsE.004." *Procedia Computer Science*, *235*, 70–80. https://doi.org/10.1016/j.procs.2024.04.007

Korkmaz, A., Talan, T., Koşunalp, S., & Iliev, T. (2025). Comparison of deep learning models in automatic classification of coffee bean species. *PeerJ Computer Science*, *11*, 1–29. https://doi.org/10.7717/peerj-cs.2759

Lambert, B., Forbes, F., Doyle, S., Dehaene, H., & Dojat, M. (2024). Trustworthy clinical AI solutions: A unified review of uncertainty quantification in Deep Learning models for medical image analysis. *Artificial Intelligence in Medicine*, *150*, 102830. https://doi.org/10.1016/j.artmed.2024.102830

Li, W., Fan, Z., Huo, J., & Gao, Y. (2023). Modeling Inter-Class and Intra-Class Constraints in Novel Class Discovery. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2023-June*, 3449–3458. https://doi.org/10.1109/CVPR52729.2023.00336

Liu, S., Zhong, W., Guo, F., Cong, J., & Gu, B. (2024). Fine-Grained Few-Shot Image Classification Based on Feature Dual Reconstruction. *Electronics (Switzerland)*, *13*(14). https://doi.org/10.3390/electronics13142751

Lu, J., Zhang, W., Zhao, Y., & Sun, C. (2022). Image local structure information learning for fine-grained visual classification. *Scientific Reports*, *12*(1), 1–10. https://doi.org/10.1038/s41598-022-23835-0

Ni, X., Wang, F., Huang, H., Wang, L., Wen, C., & Chen, D. (2024). A CNN- and Self-Attention-Based Maize Growth Stage Recognition Method and Platform from UAV Orthophoto Images. *Remote Sensing*, *16*(14). https://doi.org/10.3390/rs16142672

Pan, Q., Liu, K., Zheng, S., & Wang, G. (2025). A Fine-Grained Image Classification Method Based on ConvNeXt Heatmap Localization and Contrastive Learning. *IEEE Access*, 80123–80132. https://doi.org/10.1109/ACCESS.2025.3567488

Raja, S. P., Sawicka, B., Stamenkovic, Z., & Mariammal, G. (2022). Crop Prediction Based on Characteristics of the Agricultural Environment Using Various Feature Selection Techniques and Classifiers. *IEEE Access*, *10*, 23625–23641. https://doi.org/10.1109/ACCESS.2022.3154350

Shao, Y., Li, L., Li, J., Li, Q., An, S., & Hao, H. (2024). Out-of-plane full-field vibration displacement measurement with monocular computer vision. *Automation in Construction*, *165*, 105507. https://doi.org/10.1016/j.autcon.2024.105507

Shin, J., Kaneko, Y., Miah, A. S. M., Hassan, N., & Nishimura, S. (2024). Anomaly Detection in Weakly Supervised Videos Using Multistage Graphs and General Deep Learning Based Spatial-Temporal Feature Enhancement. *IEEE Access*, *12*(March), 65213–65227. https://doi.org/10.1109/ACCESS.2024.3395329

Strelcenia, E., & Prakoonwit, S. (2023). Improving Cancer Detection Classification Performance Using GANs in Breast Cancer Data. *IEEE Access*, *11*, 71594–71615. https://doi.org/10.1109/ACCESS.2023.3291336

Tan, M., & Le, Q. V. (2021). EfficientNetV2: Smaller Models and Faster Training. *Proceedings of Machine Learning Research*, *139*, 10096–10106.

Valarmathi, B., Srinivasa Gupta, N., Prakash, G., Hemadri Reddy, R., Saravanan, S., & Shanmugasundaram, P. (2023). Hybrid Deep Learning Algorithms for Dog Breed Identification - A Comparative Analysis. *IEEE Access*, 77228–77239. https://doi.org/10.1109/ACCESS.2023.3297440

Yaseliani, M., Hamadani, A. Z., Maghsoodi, A. I., & Mosavi, A. (2022). Pneumonia Detection Proposing a Hybrid Deep Convolutional Neural Network Based on Two Parallel Visual Geometry Group Architectures and Machine Learning Classifiers. *IEEE Access*, *10*, 62110–62128. https://doi.org/10.1109/ACCESS.2022.3182498

Zhao, X., Wang, L., Zhang, Y., Han, X., Deveci, M., & Parmar, M. (2024). A review of convolutional neural networks in computer vision. *Artificial Intelligence Review*, *57*(4). Springer Netherlands. https://doi.org/10.1007/s10462-024-10721-6