TELKOMNIKA Telecommunication Computing Electronics and Control
Vol. 23, No. 6, December 2025, pp. 1590~1599
ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v23i6.27090

Comparative performance analysis of convolutional neural network-architectures on coffee-bean roast classification

Irfan Asfy Fakhry Anto[1], Jony Winaryo Wibowo[2], Aris Munandar[2], Taufik Ibnu Salim[2]
[1]Research Center for Data and Information Science, National Research and Innovation, Bandung, Indonesia
[2]Research Center for Smart Mechatronics, National Research and Innovation, Bandung, Indonesia

Article Info

Article history:

Received Mar 25, 2025
Revised Sep 12, 2025
Accepted Oct 19, 2025

Keywords:

Agtron level
Classification
Coffee-bean roast
Convolutional neural networks
Performance analysis

ABSTRACT

The classification of coffee bean roast levels using Agtron standards has evolved from traditional subjective methods to technology-driven approaches employing advanced artificial intelligence. Recent advancements in computer vision have demonstrated the capability of convolutional neural networks (CNNs) in providing objective and consistent roast level classification compared to human visual assessment, which is prone to variability and subjectivity. This research presents a performance analysis of five CNN architectures (AlexNet, ResNet, MobileNet, VGGNet, and DenseNet) for classifying coffee beans into eight distinct Agtron roast levels. The comprehensive methodology encompasses four phases: i) data acquisition, ii) image preprocessing, iii) model training and validation, and iv) evaluation metric. During training-validation, DenseNet outperformed other models, achieving 99.702% training accuracy and 77.68% validation accuracy. In the testing evaluation, DenseNet also led with an average testing accuracy of 93.8%, followed by ResNet at 92.6%, VGGNet and AlexNet both at 92.4%, and MobileNet at 89.7%. The results show that the DenseNet shows promise in classifying Agtron coffee-bean roast classification.

Corresponding Author:

Irfan Asfy Fakhry Anto
Research Center for Data and Information Science, National Research and Innovation
Bandung, 40135, Indonesia
Email: irfa013@brin.go.id

1. INTRODUCTION

Roasted coffee beans undergo a detailed classification process to maintain consistency and quality in the final product [1]. A well-known standard for this classification is the Agtron system [2]. The specialty coffee association (SCA), an international organization in the coffee industry, plays a key role in establishing methodologies for evaluating and classifying roasted coffee beans. For many years, the SCA has utilized tools like the Agtron system to assess coffee quality by assigning scores to samples [3], [4]. The Agtron classification provides an objective, standardized measure of roast levels, ensuring uniformity and quality in coffee production, which is crucial for both commercial and specialty coffee markets. Additionally, the Agtron system has been instrumental in coffee research, particularly in studies investigating how roast levels influence chemical composition [5]. Researchers have employed Agtron measurements to analyze the impact of different roast levels on compounds like chlorogenic acids and caffeine, which affect both flavor and health benefits [6]. These findings have guided commercial roasting practices and deepened the understanding of coffee science, emphasizing the importance of precise roast profiling in achieving desired flavors.

The Agtron classification relies on advanced technology to measure roast levels [7]. Recent technological developments have improved the accuracy and usability of the Agtron system. Modern spectrophotometers and software enhancements have simplified roast profile analysis, making it more accessible for both small-scale and large-scale coffee producers [8]. Convolutional neural network (CNN) has emerged as a powerful tool for automating the classification of coffee bean roast levels, offering a means to refine and standardize the process [9]. CNN's architecture, which draws inspiration from the human visual system, is particularly well-suited to the visual analysis required for classifying roasted coffee beans, and they have consequently shown considerable promise in this application [10]. By training on large datasets of labeled images, CNNs can learn to distinguish between different roast levels based on the visual characteristics of the beans, effectively automating a process that traditionally required manual input [11], [12]. This not only accelerates classification but also minimizes human error and variability, enhancing overall efficiency.

Recent research has increasingly focused on CNN-based methods for automated classification of coffee bean roast levels, offering more objective and consistent evaluation compared to traditional human visual assessment. In one notable study [13], proposed a custom CNN architecture demonstrated exceptional performance in categorizing coffee beans into four distinct roast levels (light, medium, medium-dark, and dark), achieving 97.5% classification accuracy through optimized hyperparameter tuning and data augmentation techniques. Another investigation on [14] used ResNet50 architecture implemented on an Android platform, which successfully identified different types of roasted coffee beans with 83.3% average accuracy, showcasing the potential for practical, on-the-go quality assessment in production facilities and coffee shops. The study specifically emphasized the model's proficient utilization of memory and its capacity for real-time analysis, attributes that render it a viable candidate for implementation on resource-constrained edge devices. Research on [15] conducted a systematic performance analysis including SqueezeNet, ShuffleNet, MobileNetV2, and NASNet Mobile in classifying four roast classes (green, light, medium, and dark). Similarly, work by Hassan [16] explored multiple CNN architectures for four-level roast classification from an online coffee-bean dataset, to determine the accuracy of pre-trained models in identifying four coffee-bean classes through advanced deep learning approaches.

Current research trajectories position CNN architectures as a substantial advancement within coffee quality assessment. Ongoing research and development suggest these CNN-based methodologies are poised to assume a critical function in future quality control and roasting processes. However, an identified limitation in the extant literature is its constrained scope concerning the specific application of CNNs for classifying coffee bean roast levels. Unlike previous research, which typically categorizes roasting levels into only four classes (green, light, medium, and dark), this study conducts a comparative analysis of five CNN architectures to classify eight Agtron classes, such as very dark, dark, moderately dark, medium, medium light, moderately light, light, and very light. These offer a more detailed classification of coffee bean flavors. Furthermore, this research creates its dataset by employing images and a camera as sensors for experimentation, whereas previous studies often used datasets sourced from the internet. Thereby, this research contributes to creating a new dataset. Moreover, our prior research [17] developed a graphical user interface (GUI) within an embedded system for coffee bean classification, but it still utilized one CNN model, as the proposed system could replace the training output file. Building on this, this study aims to conduct a more in-depth analysis of multiple CNN-based models for classifying Agtron color levels in coffee beans. To achieve a highly accurate system, this paper presents a comparative performance analysis of five CNN architectures, such as AlexNet, ResNet, MobileNet, VGGNet, and DenseNet, to classify eight Agtron classes and identify the best-performing model for roasted coffee bean classification. The evaluation metrics include a confusion matrix used for each of the five architectures.

2. RELATED STUDIES

2.1. Agtron classification

The Agtron classification system has undergone significant advancements and found diverse applications in recent years, mirroring the evolving landscape of coffee technology [18]. This system constitutes a widely employed method for evaluating coffee bean roast levels. It employs a spectrophotometer to measure color and assign a numerical value on the Agtron scale, which ranges from light to dark roast [19]. The scale plays a crucial role in maintaining consistency in coffee roasting by offering an objective measure of roast degree, which directly influences flavor profiles and consumer preferences [20]. In this study, it is use the Agtron roast color standard is used for roasted coffee beans with eight color levels consisting of R-25, R-35, R-45, R-55, R-65, R-75, R-85, and R-95.

2.1. CNN architectures

CNNs have become a prevalent tool for classification, a utility largely attributable to their ability to perform autonomous feature extraction directly from raw data. This approach would address the growing demand for multimodal classification systems in real-world applications [18], [20]. Ongoing research in these areas will further enhance the potential of CNNs for classification tasks, such as determining coffee-bean Agtron classes. AlexNet was the first deep CNN to effectively utilize GPU acceleration for training. The accomplishments of this network were instrumental in establishing the viability of deep learning for complex image classification, thereby catalyzing the broad integration of CNN into computer vision [21]. For example, a study by [22] investigated combining AlexNet with ensemble genetic algorithms to boost image classification performance. This hybrid method sought to optimize feature selection and enhance classification accuracy, illustrating AlexNet's adaptability when integrated with other machine learning techniques. Collectively, this research highlights AlexNet's enduring influence and flexibility in image classification, while also acknowledging its pivotal role in the broader development of CNN architectures.

VGGNet is a CNN architecture that made significant strides in the part of computer vision [23]. Over the preceding five-year period, VGGNet has been extensively utilized and modified across diverse domains, with particular prominence in the fields of medical imaging and object detection [24]. This research utilized VGGNet, as it has not only driven advancements in image classification but also contributed significantly to progress in object detection and transfer learning [25]. ResNet has constituted a seminal contribution to the field of deep learning [26]. Its foundational element is the residual block [27], an architectural module comprising multiple convolutional and batch normalization layers, interspersed with activation functions. This architecture prioritizes a lightweight design and minimal computational overhead, thereby representing a significant advancement in the development of efficient deep learning models [28], [29]. Recent research efforts have focused on enhancing the operational efficiency and functional adaptability of MobileNet architectures. This has been pursued through the incorporation of optimization techniques such as numerical quantization and their integration with specialized edge artificial intelligence frameworks, including TensorFlow Lite and PyTorch Mobile [30], [31]. These advancements have established MobileNet as another choice for real-time applications, including object detection, facial recognition, and augmented reality on embedded or mobile devices [32]. DenseNet is a CNN architecture distinguished by its dense connectivity pattern, where each layer is connected to every other layer in a feed-forward manner. This design encourages feature reuse, reduces the number of parameters, and alleviates the vanishing gradient problem, making DenseNet highly efficient for deep learning tasks [33]. Research on DenseNet has centered on improving its scalability and adaptability for applications such as autonomous driving, medical image classification, and natural language processing [34], [35]. Furthermore, DenseNet remains a powerful and versatile architecture, fostering innovation in both academic research and industrial applications [36].

3. METHOD

The comparative performance analysis in this research was executed according to a structured, multi-stage methodology, outlined in Figure 1. The process began with the collection of a coffee-bean roast dataset, categorized into eight classes based on Agtron classes. The initial phase involved preprocessing the dataset to ensure the images were standardized and ready for analysis. Following this, the performance of several prominent CNN architectures namely AlexNet, ResNet, MobileNet, VGGNet, and DenseNet was evaluated. Each model underwent a comprehensive training and validation process to generate a classifier as the output. These classifiers were then rigorously tested using evaluation metrics consist of the confusion matrix and accuracy score, to assess their effectiveness in classifying the coffee-bean roast images. This systematic methodology facilitated a comprehensive evaluation of the models' capabilities, yielding critical insights regarding their performance and appropriateness for the intended application.

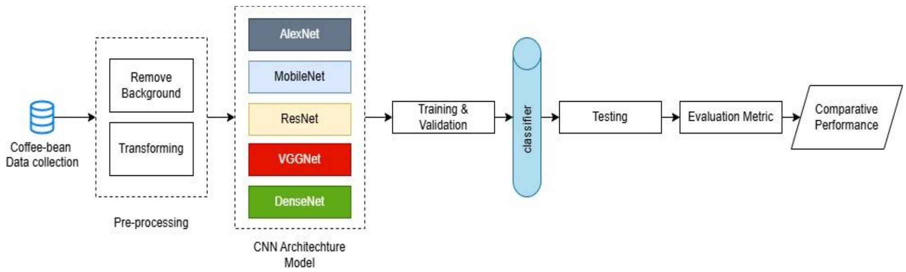
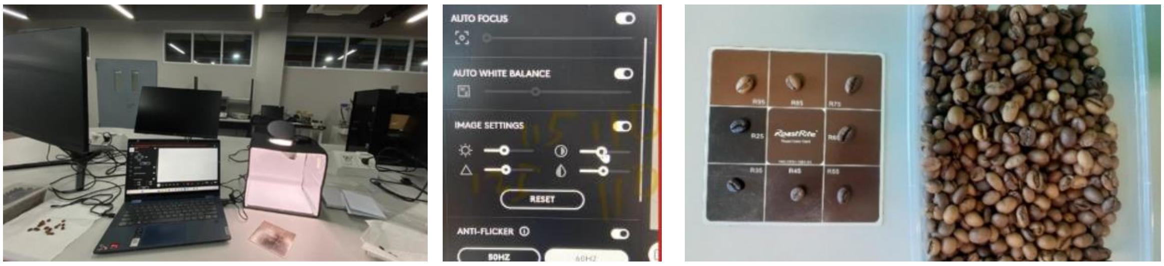


Figure 1. Research method

3.1. Data collection

This research is collected the dataset ourselves for the experiment. The materials used during the dataset collection are shown in Figure 2. A computer, a mini box photo studio, a Brio 4K Web Camera, and its software were used, as shown in Figure 2(a). To reduce the impact of environmental factors during image capture, the same parameters such as brightness, contrast, exposure, and other settings were maintained for all images, as seen in Figure 2(b). The materials consist of roasted coffee beans with various color levels and the Agtron roast color standard, as shown in Figure 2(c). The image was taken using lighting conditions using light from the LED on the miniphoto box and the following camera settings: i) brightness set to 115, ii) contrast set to 125, iii) sharpness or saturation set to 110, iv) noise reduction set to 110, and v) chroma key effect intensity set to 55. Coffee beans were chosen to precise Agtron specifications, and images were captured under controlled lighting conditions and settings to minimize variability. Each roasted coffee bean was matched with the Agtron roast color standard for the labeling process and then placed into the photo box studio for image capture. The captured images were organized into folders on the computer corresponding to the respective Agtron roast color classes. After all images were captured, data augmentation through image transformation was implemented to expand the dataset. By applying various image transformations, able to simulate different viewing conditions and perspectives, which helps improve the generalizability of the models trained on this dataset. Techniques such as vertical-horizontal flipping, rotation, zooming in and out, featurewise transformation, width-height shifting, and random cropping were applied to introduce greater variation in the images. The dataset laid the groundwork for training and evaluating for each AlexNet, ResNet, MobileNet, VGGNet, and DenseNet.



(a) (b) (c)

Figure 2. Materials for dataset collection; (a) tools for data collection, (b) software settings, and (c) Agtron roast color and coffee-bean roasted

3.2. Model training and validation

The training step was carefully monitored to ensure that the models were learning effectively. The training process for all five models AlexNet, ResNet, MobileNet, VGGNet, and DenseNet was conducted in parallel to ensure consistency and fairness in the comparison. Each model was trained for 50 epochs, a duration determined to be sufficient for the models to learn meaningful patterns from the dataset without excessive computational overhead. The environment for training was set up using the Anaconda distribution on the high-performance computing (HPC) Mahemeru BRIN facility, which provided the necessary computational resources for handling large-scale machine learning tasks. Hyperparameter tuning played a crucial role in optimizing the models' performance. This study employed the Adam optimizer on account of its adaptive learning rate capabilities, which are instrumental in attaining superior convergence speed over standard gradient descent. The learning rate of 0.005 was selected after several trials to balance the speed of convergence with the stability of training. Similarly, the batch size of 64 was chosen to ensure efficient use of computational resources while maintaining a sufficient number of samples for gradient updates. Weight momentum and decay were set to 0.9 and 0.005, to regularize the models and prevent overfitting while ensuring smooth updates to the model parameters.

3.3. Evaluation metric

Evaluation metrics constitute a fundamental component in the assessment of machine learning model performance, as they supply quantitative evidence of a model's capacity to generalize to unseen data. In the evaluation of the CNN architecture, the primary metrics employed were the confusion matrix and classification accuracy. The confusion matrix serves as a diagnostic tool that delineates model performance by tabulating the incidence of true and false classifications across all categories, thereby providing a detailed breakdown of predictive outcomes [37]. A confusion matrix provides a structured tabular overview for evaluating classification model performance, quantifying predictive outcomes through the categories of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Disparities in model performance can be further elucidated through the derivation of key metrics including accuracy, precision, recall, and F1-score—computed from the constituent values of the confusion matrix.

4. RESULTS AND DISCUSSION

4.1. Dataset collection

The results of the dataset collection process are summarized in Figure 3. Also, Figure 4 provides visual examples of images from each class, offering a clear representation of the dataset's diversity and quality. This dataset consists of images of coffee beans roasted at eight distinct Agtron roast classes, representing a range of roast profiles from very dark to very light. In total, 2,522 images of coffee beans were gathered and categorized into eight classes, each corresponding to a specific Agtron class. These classes are labeled as R25 as shown as Figure 4(a), R35, as shown in Figure 4(b), R45, as shown in Figure 4(c), R55, as shown in Figure 4(d), R65 as shown in Figure 4(e), R75 as shown in Figure 4(f), R85, as shown in Figure 4(g), and R95, as shown in Figure 4(h). Each class is stored in a separate folder named for its respective Agtron class. Furthermore, the dataset was partitioned into training, validation, and testing subsets in accordance with established machine learning protocols, utilizing an 80 (training), 10 (validation), and 10 (testing) ratio. Adherence to this conventional framework ensures the dataset provides a robust foundation for developing models proficient in the accurate classification of coffee bean roast levels.

Agtron Class	Number Files	Train	Valid	Test
R25 Very Dark	313	251	31	31
R35 Dark	321	257	32	32
R45 Moderately Dark	310	248	31	31
R55 Medium	314	252	31	31
R65 Light Medium	310	248	31	31
R75 Moderately Light	349	279	35	35
R85 Light	331	265	33	33
R95 Very Light	274	220	27	27
Total	2522	2020	251	251

Figure 3. Dataset composition

Figure 4. Sample images from the dataset for each class of Agtron class; (a) R25, (b) R35, (c) R45, (d) R55, (e) R65, (f) R75, (g) R85, and (h) R95

4.2. Training and validating

After the training step on HPC Mahameru, the training output files for each model were obtained in the form of ".pth" and ".pkl" formats. These files contained the trained weights, configurations, and other relevant data necessary for further analysis. The training result files were then analyzed using Jupyter Notebook for evaluation, where performance metrics, including accuracy, loss, and confusion matrices, were computed and visualized. Metrics such as loss and accuracy were logged, providing insights into the models' behavior over epochs. Figure 5 depict the training performance metrics, including accuracy and loss, for both training and validation phases across all epochs for the five CNN models. The results demonstrate a positive correlation between epoch count and model performance, characterized by increasing accuracy and a concomitant reduction in loss for both training and validation datasets. DenseNet achieved the highest training accuracy at 99.702% and the highest validation accuracy at 77.68%. The remaining models' training accuracies were as follows: ResNet at 97.07%, MobileNet at 95.544%, VGGNet at 89.851% and AlexNet at 85.346%, which was the lowest. The best-performing model displayed exceptional convergence and reached peak accuracy with minimal training and validation losses. It achieved faster convergence, kept the disparity between training and validation loss to a minimum, and consistently delivered superior accuracy compared to the other models. ResNet also showed strong performance, achieving near-perfect training accuracy, indicating its ability to learn complex patterns effectively. MobileNet, while slightly less accurate, demonstrated its efficiency and suitability for resource-constrained environments due to its lightweight architecture. AlexNet, despite being the oldest architecture among the tested models, still achieved reasonable accuracy, showcasing its foundational role in the evolution of CNNs. Upon completing the training process, the models were evaluated using a separate test dataset to measure their generalization capabilities. Performance metrics described in section 4.3, to provide a comprehensive assessment of each model's effectiveness. The results were compared across the five CNN architectures to identify the best-performing model for the coffee bean roast classification task. This systematic approach to training and evaluation ensured that the models were rigorously tested and optimized, providing reliable insights for the study.

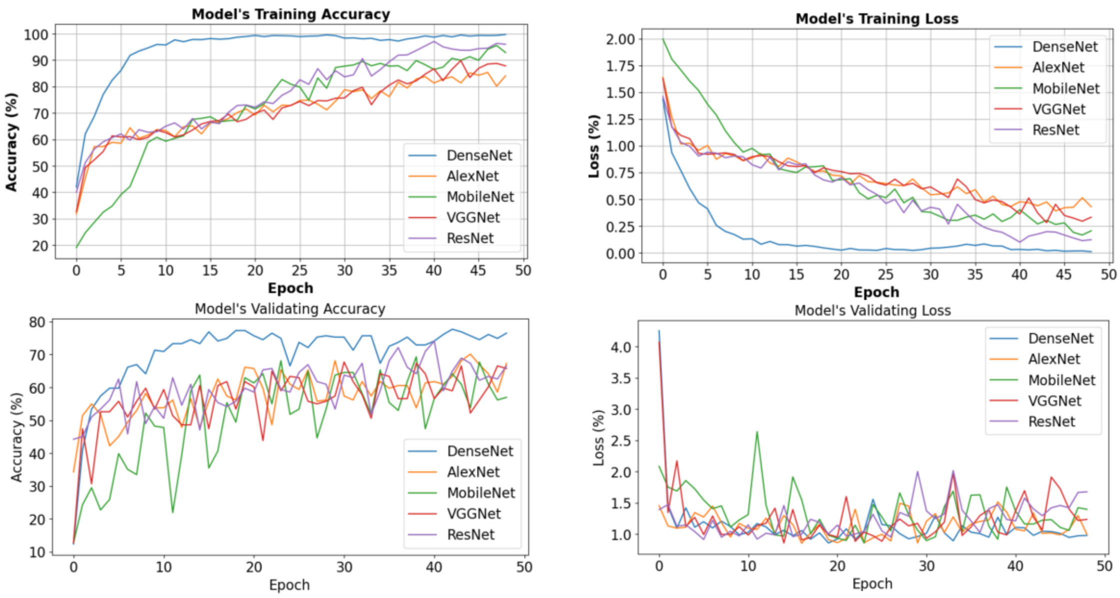
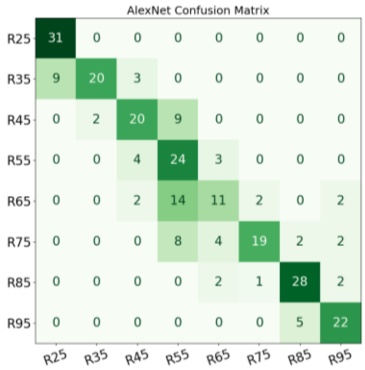


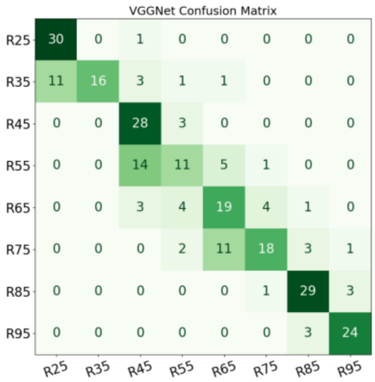
Figure 5. Models and validating accuracy and loss

4.3. Evaluation matrix

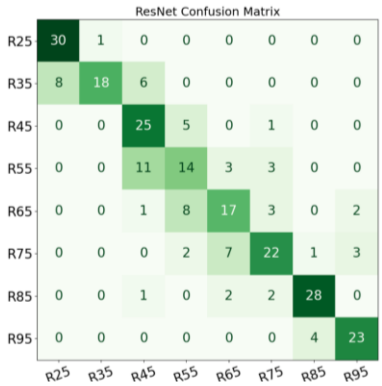
Within the domain of image classification, several evaluation metrics are extensively employed, notably the confusion matrix (CM) and classification accuracy. The evaluation metrics and comparative performance of this study are illustrated in Figure 6. Specifically, the CM corresponds to AlexNet Figure 6(a), VGGNet Figure 6(b), ResNet Figure 6(c), MobileNet Figure 6(d), and DenseNet Figure 6(e). Based on the empirical results, the testing accuracy, precision, recall, and F1-score were quantified for each model and individual class, as comprehensively detailed in Figure 6(f). The DenseNet model emerged as the top-performing architecture after averaging the experimental results, achieving an impressive average for accuracy of 93.8%, precision of 76%, recall of 75,4% and f1-score of 0.755. ResNet followed closely with an accuracy of 92.6%, while VGGNet and AlexNet both achieved identical accuracies of 92.4%. MobileNet recorded the lowest accuracy at 89.7%. When evaluating the averaged accuracy scores across all locations, DenseNet consistently outperformed the other models, whereas MobileNet struggled to deliver competitive results in this classification task. Notably, DenseNet's training and testing accuracy metrics aligned perfectly, reinforcing DenseNet's superior performance across the board. DenseNet encourages feature reuse and substantially improves parameter efficiency, making it more effective in tasks requiring fine-grained visual distinctions. In other cases, while DenseNet's architecture enhances feature reuse and improves parameter efficiency, it is also susceptible to overfitting. Traditional dropout methods can interfere with the dense feature propagation in DenseNet, reducing their regularization effectiveness and unintentionally harming generalization [33], [38]. A closer examination of class-specific performance revealed that AlexNet achieved perfect accuracy 100% in the R25 class, making it the best performer for that particular class. VGGNet and ResNet followed with 99.2% accuracy, while DenseNet achieved 96.8%. MobileNet lagged significantly behind, recording only 88.048% accuracy in the same class. The observed variation in performance across classes indicates that certain architectures may be more effective for particular data types, underscoring the necessity of aligning model selection with the distinct characteristics of the dataset. Additionally, combining multiple preprocessing methods or concatenating feature vectors derived from different approaches could enhance the overall quality of the input data. Exploring alternative neural network architectures, such as transformers or hybrid models, could also yield better results. Furthermore, the low performance of certain models in specific classes may indicate a need for methodological adjustments. Additionally, exploring the integration of unsupervised or self-supervised learning techniques might reduce the reliance on large labeled datasets. Future developments in deep learning techniques, including refined attention mechanisms and adaptive learning rates, are anticipated to considerably expand the frontiers of achievable accuracy in coffee-bean roast classification tasks.



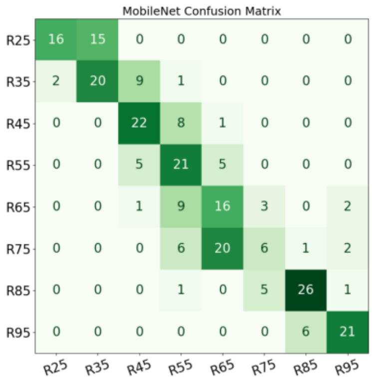
(a)



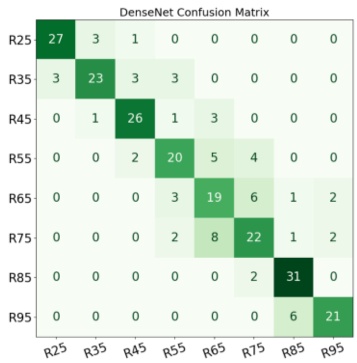
(b)



(c)



(d)



(e)

<table>
<tr><th>Agtron Class</th><th>AlexNet (accuracy|precision|recall|F1-score)</th><th>VGGNet (accuracy|precision|recall|F1-score)</th><th>ResNet (accuracy|precision|recall|F1-score)</th><th>MobileNet (accuracy|precision|recall|F1-score)</th><th>DenseNet (accuracy|precision|recall|F1-score)</th></tr>
<tr><td>R25</td><td>1.0|
0.775|
1.0|
0.873</td><td>0.992|
0.732|
0.968|
0.833</td><td>0.992|
0.789|
0.968|
0.87</td><td>0.88|
0.889|
0.516|
0.653</td><td>0.968|
0.9|
0.871|
0.885</td></tr>
<tr><td>R35</td><td>0.904|
0.909|
0.625|
0.741</td><td>0.873|
1.0|
0.5|
0.667</td><td>0.888|
0.947|
0.562|
0.706</td><td>0.904|
0.571|
0.625|
0.597</td><td>0.928|
0.852|
0.719|
0.78</td></tr>
<tr><td>R45</td><td>0.912|
0.69|
0.645|
0.667</td><td>0.976|
0.571|
0.903|
0.7</td><td>0.952|
0.568|
0.806|
0.667</td><td>0.928|
0.595|
0.71|
0.647</td><td>0.96|
0.812|
0.839|
0.825</td></tr>
<tr><td>R55</td><td>0.944|
0.436|
0.774|
0.558</td><td>0.841|
0.524|
0.355|
0.423</td><td>0.865|
0.483|
0.452|
0.467</td><td>0.92|
0.457|
0.677|
0.545</td><td>0.912|
0.69|
0.645|
0.667</td></tr>
<tr><td>R65</td><td>0.841|
0.55|
0.355|
0.431</td><td>0.904|
0.528|
0.613|
0.567</td><td>0.888|
0.586|
0.548|
0.567</td><td>0.88|
0.381|
0.516|
0.438</td><td>0.904|
0.543|
0.613|
0.576</td></tr>
<tr><td>R75</td><td>0.873|
0.864|
0.543|
0.667</td><td>0.865|
0.75|
0.514|
0.61</td><td>0.896|
0.71|
0.629|
0.667</td><td>0.769|
0.429|
0.171|
0.245</td><td>0.896|
0.647|
0.629|
0.638</td></tr>
<tr><td>R85</td><td>0.96|
0.8|
0.848|
0.824</td><td>0.968|
0.806|
0.879|
0.841</td><td>0.96|
0.848|
0.848|
0.848</td><td>0.944|
0.788|
0.788|
0.788</td><td>0.984|
0.795|
0.939|
0.861</td></tr>
<tr><td>R95</td><td>0.96|
0.786|
0.815|
0.8</td><td>0.976|
0.857|
0.889|
0.873</td><td>0.968|
0.821|
0.852|
0.836</td><td>0.952|
0.808|
0.778|
0.792</td><td>0.952|
0.84|
0.778|
0.808</td></tr>
<tr><td>Average</td><td>0.924|
0.726|
0.701|
0.695</td><td>0.924|
0.721|
0.703|
0.689</td><td>0.926|
0.719|
0.708|
0.704</td><td>0.897|
0.615|
0.598|
0.588</td><td>0.938|
0.76|
0.754|
0.755</td></tr>
</table>

(f)

Figure 6. Confusion matrix and comparative performance; (a) AlexNet, (b) VGGNet, (c) RestNet, (d) MobileNet, (e) DenseNet, and (f) testing accuracy, precision, recall, and F1-score

5. CONCLUSION

Recent advancements in CNN architectures have significantly enhanced coffee-bean roasting level classification, particularly in predicting Agtron classes. This study provides a comprehensive performance comparison of several CNN models AlexNet, ResNet, MobileNet, VGGNet, and DenseNet through a structured evaluation process involving data collection, preprocessing, training, validation, and evaluation matrix. This research utilized cost-effective image-based data collection with a camera. During training and validation, DenseNet outperformed other models, achieving 99.702% training accuracy and 77.68% validation accuracy. In the final evaluation, DenseNet also led with an average testing accuracy of 93.8%, followed by ResNet at 92.6%, VGGNet and AlexNet both at 92.4%, and MobileNet at 89.7%. The findings reveal critical trade-offs between computational efficiency and accuracy, which must be considered when deploying coffee-bean classification systems in real-world applications. While DenseNet delivered the highest accuracy, lightweight models like MobileNet could be more practical. Future research could explore hybrid architectures that combine the strengths of different CNNs while optimizing computational efficiency. Next, unsupervised learning techniques could reduce reliance on large labeled datasets, making the system more scalable. Emerging advancements in attention mechanisms and adaptive learning rate optimization could further refine Agtron-level classification, paving the way for more precise and automated coffee roasting processes.

ACKNOWLEDGMENTS

The authors acknowledge the support of the National Research and Innovation Agency (BRIN). Also, the computations performed in this work utilized the MAHAMERU HPC facility, managed by BRIN.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Irfan Asfy Fakhry Anto	✓	✓	✓		✓	✓			✓	✓	✓			
Jony Winaryo Wibowo	✓	✓		✓				✓	✓					
Aris Munandar		✓		✓	✓				✓	✓				
Taufik Ibnu Salim	✓		✓			✓			✓					

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

DATA AVAILABILITY

The datasets used and/or analyzed in this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] A. A. Catão *et al.*, "Coffee-roasting variables associated with volatile organic profiles and sensory evaluation using multivariate analysis," *Applied Food Research*, vol. 2, no. 2, Dec. 2022, doi: 10.1016/j.afres.2022.100223.
[2] M. Lin, C. Chen, and J. Lu, "Intelligent classifier for various degrees of coffee roasts using smart multispectral vision system," *Journal of Field Robotics*, vol. 41, no. 3, pp. 639–653, May 2024, doi: 10.1002/rob.22285.
[3] L. Anokye-Bempah, I. R. Donis-González, and W. D. Ristenpart, "What color is your coffee?" Accessed: Oct. 19, 2024. [Online]. Available: https://sca.coffee/sca-news/25/issue-21/what-color-is-your-coffee
[4] F. de Carvalho Pires, R. G. F. A. Pereira, M. R. Baqueta, P. Valderrama, and R. A. da Rocha, "Comparison between binary and multiclass partial least square with discriminant analysis: A case study on coffee classification based on near-infrared spectra," in *Advances in Spectroscopic Analysis of Food and Drink*, IOP Publishing, 2024, doi: 10.1088/978-0-7503-5573-5ch4.

[5] I. Santoso, S. A. Mustaniroh, and A. Choirun, "Methods for quality coffee roasting degree evaluation: A literature review on risk perspective," *IOP Conference Series: Earth and Environmental Science*, vol. 924, no. 1, Nov. 2021, doi: 10.1088/1755-1315/924/1/012058.
[6] S. E. Yeager *et al.*, "Roast level and brew temperature significantly affect the color of brewed coffee," *Journal of Food Science*, vol. 87, no. 4, pp. 1837–1850, Apr. 2022, doi: 10.1111/1750-3841.16089.
[7] R. E. Bumbaugh *et al.*, "An electrochemical appraisal of coffee qualities," *Prepr. arXiv.2501.14950*, Oct. 2025.
[8] J. Alstrup, M. A. Petersen, F. H. Larsen, and M. Münchow, "The effect of roast development time modulations on the sensory profile and chemical composition of the coffee brew as measured by NMR and DHS-GC–MS," *Beverages*, vol. 6, no. 4, Dec. 2020, doi: 10.3390/beverages6040070.
[9] V. Somaudon and T. Kerdcharoen, "Unlocking the aroma profiles of coffee roasting levels with an electronic nose coupled with machine learning," in *2024 21st International Joint Conference on Computer Science and Software Engineering (JCSSE)*, IEEE, Jun. 2024, pp. 678–681, doi: 10.1109/JCSSE61278.2024.10613715.
[10] R. G. Pereira Neto, P. M. de Sousa, L. F. R. Moreira, P. I. V. G. God, and J. F. Mari, "Enhancing green coffee quality assessment through deep learning," in *Anais do XVIII Workshop de Visão Computacional (WVC 2023)*, Sociedade Brasileira de Computação - SBC, Nov. 2023, pp. 84–89, doi: 10.5753/wvc.2023.27537.
[11] Q.-H. Wu and D.-F. Shen, "Color video and convolutional neural networks deep learning based real-time agtron baking level estimation method," in *New Trends in Computer Technologies and Applications*, 2019, pp. 197–206, doi: 10.1007/978-981-13-9190-3_21.
[12] P. Wang, H.-W. Tseng, T.-C. Chen, and C.-H. Hsia, "Deep convolutional neural network for coffee bean inspection," *Sensors and Materials*, vol. 33, no. 7, Jul. 2021, doi: 10.18494/SAM.2021.3277.
[13] N. K. Naik and P. K. Sethy, "Roasted coffee beans classification based on convolutional neural network," in *2022 International Conference on Futuristic Technologies (INCOFT)*, IEEE, Nov. 2022, pp. 1–3, doi: 10.1109/INCOFT55651.2022.10094378.
[14] E. I. Muhlisin, R. R. Nurmalasari, L. Kamelia, and R. W. Sururie, "Implementation of convolutional neural network (CNN) in The Android-based application for detecting coffee bean maturity," in *2024 10th International Conference on Wireless and Telematics (ICWT)*, IEEE, Jul. 2024, pp. 1–5, doi: 10.1109/ICWT62080.2024.10674676.
[15] K. Saddami, N. Aulia, V. Maulidia, and Andriansyah, "Comparative analysis of lightweight pre-trained CNN models for coffee bean roasting level identification," in *2023 2nd International Conference on Computer System, Information Technology, and Electrical Engineering (COSITE)*, IEEE, Aug. 2023, pp. 13–18, doi: 10.1109/COSITE60233.2023.10249546.
[16] E. Hassan, "Enhancing coffee bean classification: a comparative analysis of pre-trained deep learning models," *Neural Computing and Applications*, vol. 36, no. 16, pp. 9023–9052, Jun. 2024, doi: 10.1007/s00521-024-09623-z.
[17] I. A. F. Anto, J. W. Wibowo, T. I. Salim, and A. Munandar, "Implementation of image processing and CNN for roasted-coffee level classification," *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, vol. 12, no. 4, pp. 1005–1018, Dec. 2024, doi: 10.52549/ijeei.v12i4.5531.
[18] M. Bolka and S. Emire, "Effects of coffee roasting technologies on cup quality and bioactive compounds of specialty coffee beans," *Food Sci Nutr*, vol. 8, no. 11, pp. 6120–6130, Nov. 2020, doi: 10.1002/fsn3.1904.
[19] F. de Carvalho Pires *et al.*, "Feasibility of using colorimetric devices for whole and ground coffee roasting degrees prediction," *Journal of the Science of Food and Agriculture*, vol. 104, no. 9, pp. 5435–5441, Jul. 2024, doi: 10.1002/jsfa.13376.
[20] M. Münchow, J. Alstrup, I. Steen, and D. Giacalone, "Roasting conditions and coffee flavor: A multi-study empirical investigation," *Beverages*, vol. 6, no. 2, May 2020, doi: 10.3390/beverages6020029.
[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.
[22] W. Hussain *et al.*, "Ensemble genetic and CNN model-based image classification by enhancing hyperparameter tuning," *Scientific Reports*, vol. 15, no. 1, Jan. 2025, doi: 10.1038/s41598-024-76178-3.
[23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Prepr. arXiv.1409.1556*, Apr. 2015.
[24] S. Ghosh, A. Chaki, and K. Santosh, "Improved U-Net architecture with VGG-16 for brain tumor segmentation," *Physical and Engineering Sciences in Medicine*, vol. 44, no. 3, pp. 703–712, Sep. 2021, doi: 10.1007/s13246-021-01019-w.
[25] P. Udayaraju, P. Jeyanthi, and B. V. D. S. Sekhar, "A hybrid multilayered classification model with VGG-19 net for retinal diseases using optical coherence tomography images," *Soft Computing*, vol. 27, no. 17, pp. 12559–12570, Sep. 2023, doi: 10.1007/s00500-023-08928-w.
[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[27] J. Guo, H. Wang, X. Xue, M. Li, and Z. Ma, "Real-time classification on oral ulcer images with residual network and image enhancement," *IET Image Processing*, vol. 16, no. 3, pp. 641–646, Feb. 2022, doi: 10.1049/ipr2.12144.
[28] A. G. Howard *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision applications," *Prepr. arXiv.1704.04861*, Apr. 2017.
[29] E. Elfatimi, R. Eryigit, and L. Elfatimi, "Beans leaf diseases classification using MobileNet models," *IEEE Access*, vol. 10, pp. 9471–9482, 2022, doi: 10.1109/ACCESS.2022.3142817.
[30] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," *Prepr. arXiv.1801.04381*, Mar. 2019.
[31] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *Prepr. arXiv.1905.11946*, Sep. 2020.
[32] K. Sun, X. Wang, X. Miao, and Q. Zhao, "A review of AI edge devices and lightweight CNN and LLM deployment," *Neurocomputing*, vol. 614, Jan. 2025, doi: 10.1016/j.neucom.2024.128791.
[33] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jul. 2017, pp. 2261–2269, doi: 10.1109/CVPR.2017.243.
[34] T. Zhou, X. Ye, H. Lu, X. Zheng, S. Qiu, and Y. Liu, "Dense convolutional network and its application in medical image analysis," *BioMed Research International*, vol. 2022, no. 1, Jan. 2022, doi: 10.1155/2022/2384830.
[35] P. K. Mall *et al.*, "A comprehensive review of deep neural networks for medical image processing: Recent developments and future opportunities," *Healthcare Analytics*, vol. 4, Dec. 2023, doi: 10.1016/j.health.2023.100216.
[36] X. Wang and W. Jia, "Optimizing edge AI: A comprehensive survey on data, model, and system strategies," *Prepr. arXiv.2501.03265*, Jan. 2025.
[37] K. M. Ting, "Confusion matrix," in *Encyclopedia of Machine Learning*, Boston, MA: Springer US, 2011, pp. 209–209, doi: 10.1007/978-0-387-30164-8_157.

[38] K. Wan, B. Feng, L. Xie, and Y. Ding, "Reconciling feature-reuse and overfitting in DenseNet with specialized dropout," *Prepr. arXiv.1810.00091*, Sep. 2018.

BIOGRAPHIES OF AUTHORS

Irfan Asfy Fakhry Anto was received his Bachelor of Computer Science from the University of Brawijaya in 2009, and his Master's degree in School of Electrical Engineering and Informatics from the Bandung Institute of Technology (STEI-ITB), Bandung, Indonesia, in 2019. He is currently a Researcher at the Research Center for Data and Information Science, National Research and Innovation, Indonesia. His current research interests include data science, artificial intelligence, information retrieval, and internet of things. He can be contacted at email: irfa013@brin.go.id.

Jony Winaryo Wibowo was completed his Bachelor of Science in Computer Science from Padjajaran University of Bandung, Indonesia, in 2007. His Master of Informatics Engineering degree from Institut Teknologi Bandung (ITB), Indonesia, in 2010. He is a Researcher at the Research Center for Smart Mechatronics, National Research and Innovation Agency of the Republic of Indonesia. He is an active member of the intelligent instrumentation research group, focusing his research interests on technology instrumentation and control systems. He can be contacted at email: jony001@brin.go.id.

Aris Munandar earned his bachelor's degree in 2009 from the Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, majoring in Computer Science. He is currently a Junior Researcher at the Smart Mechatronics Research Center, Smart Instrumentation Research Group, National Research and Innovation Agency (BRIN). His research interests include instrumentation, internet of things (IoT), computer vision, and machine learning. He can be contacted at email: aris013@brin.go.id.

Taufik Ibnu Salim recieved his Bachelor's degree in Electronics and Instrumentation from Gadjah Mada University in 2010, and his Master's degree in Instrumentation and Control from the Bandung Institute of Technology, Bandung, Indonesia in 2021. He is currently a researcher at the Research Center for Smart Mechatronics, National Research and Innovation Agency, Indonesia. His research interests are related to instrumentation, IoT, embedded machine learning, and autonomous systems. He can be contacted at email: tauf021@brin.go.id.