Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 3, June 2026, Page. 2038-2050
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5537

A Novel Hybrid CNN Model Integrating Resnet and Inception for Precision Classification of Coffee Beans

Rahmat Zulpani[1], Agus Perdana Windarto*[2], Poningsih[3]

[1]Master's Student in Informatics, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia
[2,3]Department of Informatics, Master's Program, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

Email: [2]agus.perdana@amiktunasbanga.ac.id

Received : Dec 5, 2025; Revised : Dec 9, 2025; Accepted : Dec 13, 2025; Published : Jun 15, 2026

Abstract

Coffee is one of Indonesia's key strategic commodities with substantial economic value for farmers and exporters. However, inconsistencies in post-harvest coffee bean quality remain a major challenge due to manual, subjective, and expertise-dependent classification. This study addresses this issue by developing an automated and objective computer vision–based classification system using a hybrid deep learning architecture. The proposed model, named RI-Net, integrates the residual learning capability of ResNet with the multi-scale feature extraction of the Inception module to improve the precision and robustness of coffee bean classification across four roasting levels: Green, Light, Medium, and Dark. The model was trained and evaluated on a locally collected dataset and benchmarked against three standard architectures—ResNet50, InceptionV3, and a Fully Convolutional Neural Network (FCNN). Experimental results show that RI-Net outperforms all baseline models, achieving perfect scores of 100% in accuracy, precision, recall, and F1-score. These findings confirm the effectiveness of combining residual and multi-scale features in capturing subtle visual differences across roasting levels. The study demonstrates the potential of advanced hybrid CNN architectures to enhance post-harvest quality control, supporting faster, more consistent, and standardized classification processes that strengthen the competitiveness of Indonesia's coffee industry.

Keywords : *Coffee bean classification, ResNet, Inception, Image classification, RI-Net.*

1. INTRODUCTION

Coffee is one of Indonesia's strategic commodities [1], [2], [3], playing a crucial role in supporting the national economy both as an export product and as a livelihood source for millions of local farmers [4], [5], [6]. Indonesia ranks as the world's fourth largest coffee producer after Brazil, Vietnam, and Colombia, offering various local types and varieties with unique flavors and high market value [7]. However, a persistent challenge in the national coffee value chain is the inconsistent quality of post-harvest coffee beans [8]. A major contributor to this issue is the continued reliance on manual classification methods, which depend heavily on human expertise and visual perception, leading to subjectivity and lack of standardization [9], [10]. This adversely affects the market price, production efficiency, and distribution accuracy of coffee products. In today's digital transformation era, there is an urgent need for precise, fast, and consistent classification systems based on technology. Artificial intelligence, particularly Convolutional Neural Networks (CNN), has shown great promise in addressing these challenges by excelling in visual pattern recognition and image classification tasks [11], [12], [13], [14].

Popular CNN architectures such as ResNet [15], [16] and Inception [17], [18] have been widely applied in image processing across various domains. Nevertheless, their application specifically for classifying coffee bean types, especially within the Indonesian local context, remains limited. Moreover, existing standard models are often not optimally adapted to the unique visual characteristics of Indonesian coffee beans. The use of CNN in agricultural visual classification has rapidly evolved,

P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5537

demonstrating success in identifying complex visual patterns in fruits, leaves, and plant diseases [19] [20], [21], [22]. In the coffee domain, studies such as Anto et al. (2024) developed CNN-based systems for classifying roasting levels but did not address classification of green bean varieties [23]. Similarly, Arwatchananukul et al. (2024) and Auliya et al. (2024) applied CNN and FCNN models for defect and quality classification of green coffee beans, yet their approaches lacked architectural modifications aimed at improving efficiency and overall accuracy [24], [25]. Other researchers like Chang and Liu (2024) and Unal et al. (2022) demonstrated CNN's success in defect detection and species classification but relied on standard architectures without systematic benchmarking against alternative models [26], [27]. Gope and Fukai (2022) also reported limitations in CNN-based classification deployed on low-power hardware like Raspberry Pi, which affected accuracy [28]. In summary, although CNN utilization in coffee classification is common, research focused on designing, modifying, and comparatively evaluating novel CNN architectures tailored for coffee bean type classification remains scarce. The novelty of this study lies in developing a hybrid CNN architecture by modifying two popular models ResNet and Inception to better suit the visual traits of Indonesian coffee beans. Unlike merely applying pre-trained models, this approach constructs a new network that integrates residual learning from ResNet and efficient multi-scale feature extraction from Inception. The proposed model will be comparatively evaluated against its standard counterparts to validate its superior performance in coffee bean classification.

Therefore, this research focuses on designing and developing a CNN model based on a modified ResNet-Inception architecture that is more adaptive to local coffee bean characteristics. The study aims to deliver a more accurate and efficient visual classification solution applicable in precision agriculture, especially at the post-harvest stage. Based on this background, the research questions address how to design a CNN architecture that significantly improves classification accuracy of coffee bean types and how the modified model performs relative to existing standard models. The approach involves deep learning model development and optimization through architectural modifications of ResNet and Inception, employing transfer learning with pre-trained weights as a starting point and further adapting to the local coffee bean dataset.

Based on the literature gaps and practical challenges identified, this study addresses the following research problems: (1) How to design a CNN architecture that significantly improves the classification accuracy of coffee beans with different roasting levels? (2) How does the proposed architecture compare in performance against standard CNN models? The purpose of this research is to develop and evaluate a hybrid CNN model, RI-Net, that integrates the residual learning capabilities of ResNet and the multi-scale feature extraction of Inception. The main contribution of this study lies in the novel architectural design tailored to the unique visual features of Indonesian coffee beans, offering a more accurate and robust classification solution. The novelty of this work is in constructing a new CNN architecture not just applying pre-trained models that synergistically combines ResNet and Inception modules and is thoroughly benchmarked against other models using a locally sourced coffee bean dataset.

2. METHOD

2.1. Related Work

Recent advances in computer vision and deep learning have enabled a variety of applications in agricultural image classification, including in the domain of coffee bean processing. As shown in Table 1, previous research on coffee bean classification using CNNs often lacked either architectural innovation or comparative benchmarking. For instance, Anto et al. focused only on roasting levels using standard CNNs, while Auliya et al. used FCNN without comparing it to other models. This table highlights the research gaps that motivate the architectural development proposed in this study.

Vol. 7, No. 3, June 2026, Page. 2038-2050
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5537

Table 1. Related work on coffee bean classification using CNN

Study	Research Focus	Method	Limitations Identified
Anto et al. (2024) [23]	Classification of roasted coffee bean levels	Standard CNN	Focused solely on roasting levels; no differentiation by bean variety
Arwatchananukul et al. (2024) [24]	Visual defect classification in Thai Arabica green beans	CNN	Lacked architectural modifications or optimization
Auliya et al. (2024) [25]	Classification of green coffee beans using FCNN with Adam optimizer	Fully CNN + Adam	No comparative evaluation against standard models
Chang & Liu (2024) [26]	Multiscale defect detection in green coffee beans	Multiscale CNN	Did not benchmark against alternative architectures
Gope & Fukai (2022) [28]	Classification of peaberry and normal beans using embedded systems	CNN, SVM, KNN	Hardware limitations (Raspberry Pi) impacted accuracy
This Study (Proposed)	Development of a hybrid CNN architecture for classifying coffee bean types	Modified CNN (ResNet + Inception)	Proposes a new architecture, applies comparative evaluation, and adapts to local visual features

Table 1 presents a summary of recent studies that apply convolutional neural networks (CNNs) to coffee-related image classification. Anto et al. (2024) implemented a CNN-based approach to classify roasted coffee bean levels; however, their focus was limited to roast intensity without addressing varietal classification [23]. Arwatchananukul et al. (2024) developed a CNN model for visual defect detection in Thai Arabica green beans but did not explore architectural innovations or optimization strategies [24]. Similarly, Auliya et al. (2024) utilized a FCNN with the Adam optimizer to classify green coffee beans, yet their study lacked a comparative evaluation against standard baseline models, which limits the generalizability of the results [25]. Chang and Liu (2024) proposed a multiscale CNN for defect detection, offering improved spatial analysis but without benchmarking against alternative CNN architectures [26]. Gope and Fukai (2022) combined CNN, SVM, and KNN classifiers to differentiate between peaberry and normal beans using embedded devices (Raspberry Pi) [28]. While innovative in deployment, their model was constrained by hardware limitations that affected classification accuracy. In contrast, the present study proposes a novel hybrid CNN architecture that combines the strengths of ResNet (residual learning) and Inception (multi-scale feature extraction). Unlike prior works, this model is specifically adapted to the unique visual characteristics of local Indonesian coffee beans and is subjected to rigorous comparative evaluation against its baseline counterparts. This methodological advancement addresses multiple gaps in the literature, including architecture customization, dataset relevance, and performance benchmarking.

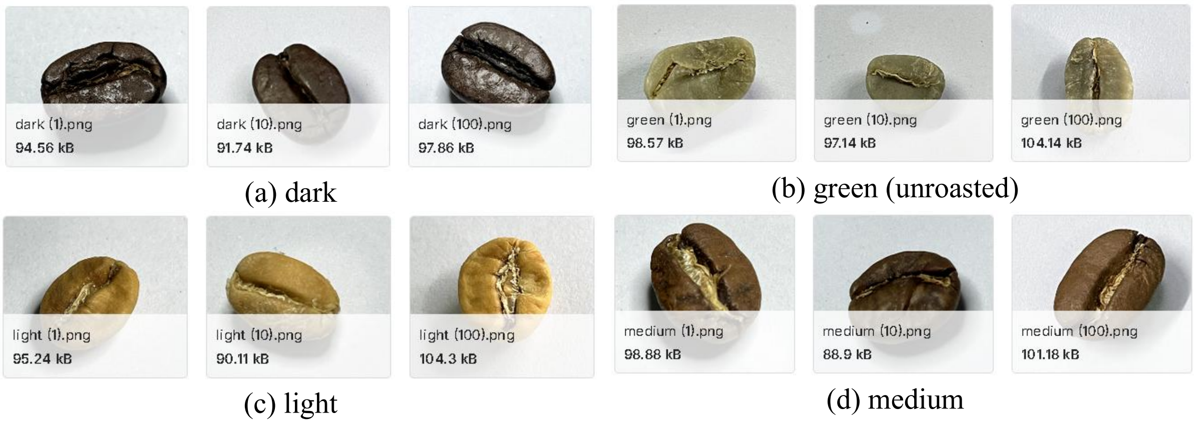
2.2. Dataset

The dataset used in this study consists of high-resolution images of Coffea arabica beans representing four distinct roasting levels: unroasted (green), light, medium, and dark. Each class corresponds to a specific variety and roast level: Laos Typica Bolaven for light roast, Doi Chaang for medium roast, and Brazil Cerrado for dark roast. The green beans are also derived from Laos Typica Bolaven. Images were captured using an iPhone 12 Mini equipped with a 12-megapixel ultra-wide and wide rear camera. The camera was mounted at a fixed distance, aligned parallel to the object surface to maintain a consistent imaging angle across all samples.

To enhance dataset variability and generalization, each sample was photographed under two lighting conditions: natural daylight and controlled LED illumination from a light box. The coffee beans were placed in shallow containers to introduce controlled background noise and increase intra-class

Vol. 7, No. 3, June 2026, Page. 2038-2050
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5537

variation. All images were saved in PNG format with a resolution of 3024 × 3024 pixels. In total, 4,800 images were collected, evenly distributed across the four roasting levels, with 1,200 images per class. Figure 1 presents representative images of Arabica coffee beans at four distinct roasting levels, illustrating the visual differences associated with each stage of the roasting process.



(a) dark (b) green (unroasted) (c) light (d) medium

Figure 1. The example image of Coffea arabica beans representing four different roasting levels

Figure 1 displays representative images of Coffea arabica beans at four roasting levels: Dark, Green (unroasted), Light, and Medium. These images illustrate the visual characteristics such as color, texture, and surface pattern that form the basis for model classification. By presenting samples from each class, this figure helps clarify the diversity and intra-class variation that the model needs to distinguish.

2.3. The Research Framework

In order to systematically address the research objectives and ensure a comprehensive understanding of the proposed approach, this section presents the overall research framework. Figure 2 illustrates the overall research framework applied in this study, consisting of stages from problem identification, data acquisition, preprocessing, to model development and evaluation. This diagram is important for understanding the flow and logic of the research methodology. It helps the reader visualize how each component particularly the integration of ResNet and Inception into the hybrid model is situated within the broader process.

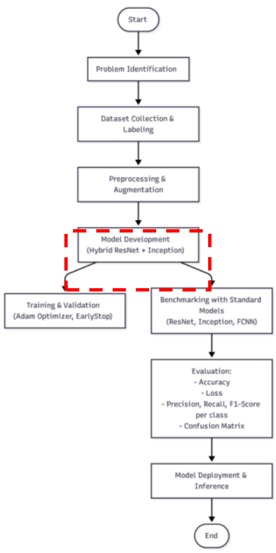


Figure 2. The research framework

Vol. 7, No. 3, June 2026, Page. 2038-2050
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5537

Figure 2 illustrates the sequential stages starting from problem identification to the systematic collection and labeling of coffee bean image datasets. Preprocessing techniques, including normalization and augmentation, are applied to enhance model generalization. Central to this study is the development of a novel hybrid convolutional neural network that integrates the residual learning capabilities of ResNet with the multi-scale feature extraction strength of the Inception module. The model is trained and validated using the Adam optimizer alongside early stopping to prevent overfitting. A pivotal component of the framework involves benchmarking the proposed hybrid model against established standard architectures ResNet50, InceptionV3 as well as a custom FCNN baseline. Evaluation metrics such as accuracy, precision, recall, and F1-score are employed to provide a detailed assessment of classification performance, complemented by confusion matrix analyses to identify class-specific prediction errors. Additionally, training and validation loss curves are examined to ensure robust model fitting and generalization.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

$$F1 - Score = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

2.4. Proposed Model

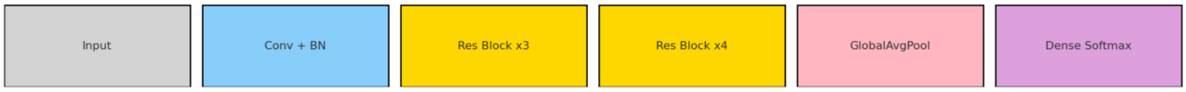
This study introduces a novel hybrid convolutional neural network architecture that integrates the strengths of Residual Networks (ResNet) and Inception modules to enhance the classification accuracy of coffee bean images. Residual learning, as proposed by He et al. (2016), effectively addresses the degradation problem in deep neural networks by allowing identity mappings through shortcut connections, enabling the training of substantially deeper models without performance loss . ResNet's ability to learn residual functions facilitates efficient gradient flow and stabilizes the training of deep architectures, which is critical for extracting robust features from complex image data [29]. In contrast, as introduced by Szegedy et al. (2015) the Inception architecture captures multi-scale spatial features by employing parallel convolutional filters of varying sizes within its modules [30]. This design allows the network to learn features at different granularities simultaneously, which is beneficial for texture-rich and heterogeneous data such as coffee bean images. Combining these two architectures leverages the complementary advantages of residual learning and multi-scale feature extraction. The proposed hybrid model (RI-Net) as a novelty can be seen in Table 2.

Table 2. Proposed architecture of the hybrid model (RI-Net)

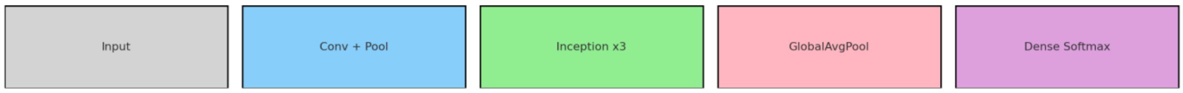
Layer	Output Size	Detail
Input	224x224x3	Input image
Conv2D + BatchNorm	112x112x64	7x7 kernel, stride 2
MaxPooling2D	56x56x64	3x3 pool, stride 2
Residual Block x3	56x56x64	3x3 conv, stride 1, residual connections
Residual Block x3	28x28x128	Downsampling (stride 2)
Inception Module	28x28x256	1x1, 3x3, 5x5 conv + pooling concatenation
Conv2D + BatchNorm	28x28x256	1x1 conv to reduce dimensionality
GlobalAvgPooling	1x1x256	Global average pooling
Dense + Dropout	256	Fully connected layer, ReLU + Dropout (0.5)
Output Dense	num_classes=4	Softmax classifier

Vol. 7, No. 3, June 2026, Page. 2038-2050
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5537

Table 2 outlines the detailed architecture of the proposed hybrid CNN model (RI-Net), integrating residual blocks from ResNet and inception modules from the Inception architecture. Each layer is listed with its output size and key configuration. The model starts with an input size of 224×224×3, progresses through convolutional and residual layers, and includes an inception module followed by dimensionality reduction and classification layers. This structured layout provides a layer-by-layer understanding of how RI-Net processes input images and extracts multi-scale visual features for classification. Overall, the hybrid design aspires to provide a more accurate and robust classification framework tailored specifically for the unique visual features of local coffee bean varieties. Here is a comparison between the proposed model (RI-Net) and the standard models (ResNet and Inception), as illustrated in Figure 3 below.



(a) Standard ResNet Architecture



(b) Standard Inception Architecture

(c) Proposed Hybrid Architecture (RI-NET)

: The orange color indicates new components (novelty) that are not present in the standard ResNet or Inception models.

Figure 3. Comparison of Model Architectures

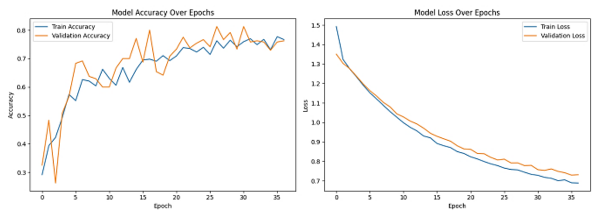
Figure 3 compares the standard architectures (a) ResNet, (b) Inception with the proposed hybrid model (c) RI-Net. The orange components in Figure 3(c) indicate novel elements not found in the original models, such as modified inception modules integrated into a residual learning backbone. This comparison highlights the architectural novelty and justifies the hybrid approach by visually contrasting structural enhancements.

3. RESULT

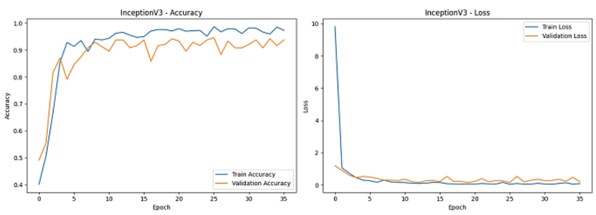
This section presents the empirical results obtained from testing four deep learning models, followed by a comprehensive discussion of the findings. The models evaluated include ResNet50, InceptionV3, FCNN, and the proposed hybrid RI-Net.

3.1. Training Performance Analysis

This study involved training four key models: ResNet50, InceptionV3, FCNN, and the proposed hybrid ResNet-Inception model (RI-Net). The training and validation accuracy and loss curves provide comprehensive insights into each model's performance and characteristics as shown in Figure 4.

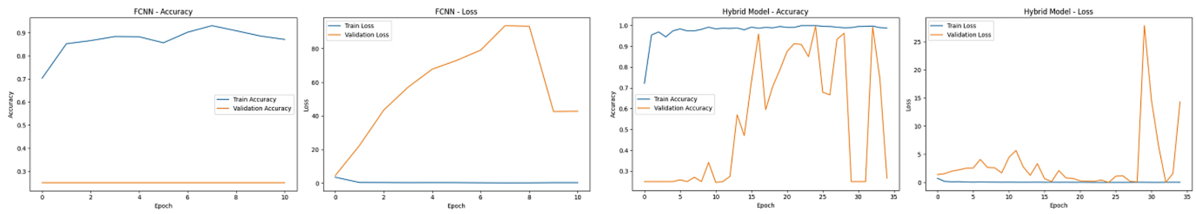


(a) ResNet50



(b) InceptionV3

Jurnal Teknik Informatika (JUTIF) Vol. 7, No. 3, June 2026, Page. 2038-2050
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5537



(c) FCNN (d) Hybrid Model (RI-Net) - Proposed

Figure 4. Comparison of training results for all models (accuracy and loss)

Figure 4 presents the training and validation curves (accuracy and loss) for the four models evaluated. ResNet50 (Figure 4.a) demonstrated a steady and consistent increase in both training and validation accuracy, reaching approximately 80% training accuracy with validation closely following and minimal fluctuation, indicative of good model generalization and stable training. Similarly, InceptionV3 (Figure 4.b) achieved higher training accuracy, approaching 98%, with relatively stable and high validation accuracy, and both training and validation losses dropped rapidly before stabilizing at low values. In contrast, the FCNN model (Figure 4.c) exhibited a different pattern: despite attaining around 90% training accuracy, its validation accuracy stagnated near 25%, accompanied by highly variable and elevated validation loss, suggesting significant overfitting and poor generalization to unseen data. The hybrid RI-Net model (Figure 4.d) showed the highest training accuracy, nearing 100%, underscoring its superior capacity for feature learning by combining residual connections and multi-scale feature extraction. However, its validation accuracy fluctuated noticeably across epochs, and the validation loss displayed sharp spikes despite a consistent decrease in training loss, indicating potential overfitting and instability during validation phases. Overall, ResNet50 and InceptionV3 models showed more stable and reliable generalization on the validation set, while the hybrid model demonstrated greater learning potential but with challenges related to validation stability. The FCNN's limited generalization reflects its simpler architecture, making it more suitable as a baseline rather than a final solution. These findings highlight that the hybrid approach effectively captures the complex visual features characteristic of coffee beans but requires further refinement, such as enhanced regularization or data augmentation strategies, to improve its validation robustness and practical applicability.

3.2. Evaluation Metrics and Comparative Analysis

The evaluation metrics encompass accuracy, precision, recall, and F1-score, further supported by comprehensive visual analyses, including the confusion matrix (Figure 5), ROC curve (Figure 6), precision-recall curve (Figure 7), and class-wise accuracy comparison (Figure 8).

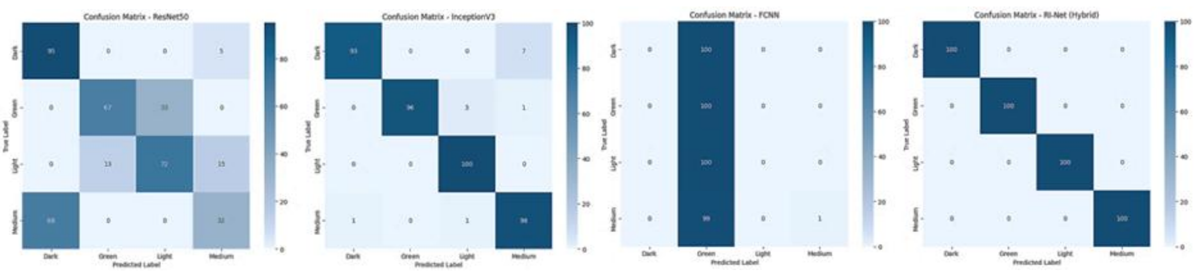


Figure 5. Confusion Matrices models: ResNet50, InceptionV3, FCNN, and the proposed model (RI-Net)

Figure 5 displays the confusion matrices for all four models, showing the distribution of predictions across the four classes. The proposed hybrid RI-Net model achieved perfect classification with no misclassifications in any class, demonstrating its strong ability to distinguish the subtle visual

Vol. 7, No. 3, June 2026, Page. 2038-2050
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5537

differences among coffee bean types. While ResNet50 and InceptionV3 also performed well, they exhibited some misclassifications, particularly in the Medium and Dark classes. ResNet50 showed notable confusion between Medium and Dark beans, whereas InceptionV3 made minor errors in the Green and Medium classes. In contrast, the FCNN model performed poorly, with significant misclassification of Dark beans as Green, reflecting its limitations in handling the dataset's visual complexity. The next, Figure 6 presents the Receiver Operating Characteristic (ROC) curves for each model and corresponding class, providing a visual representation of their classification performance.

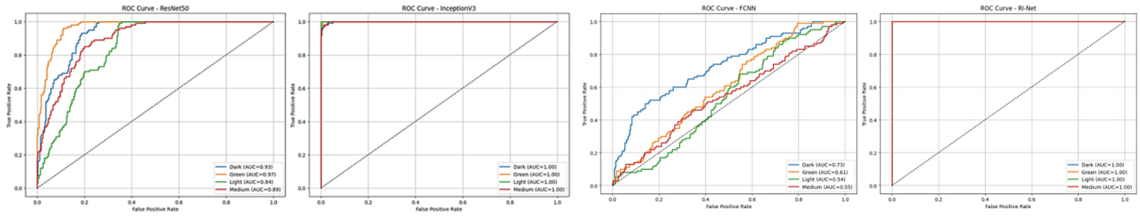


Figure 6. ROC curves models: ResNet50, InceptionV3, FCNN, and the proposed model (RI-Net)

Figure 6 shows the Receiver Operating Characteristic (ROC) curves for each model and class. These curves reflect each model's ability to distinguish between classes, with the area under the curve (AUC) representing the overall performance. The RI-Net model showed perfect Area Under the Curve (AUC) scores of 1.0 for all classes, indicating flawless class separation. InceptionV3 closely matched this performance with AUC values near 1.0, confirming its robust discriminative ability. ResNet50 achieved good AUC scores ranging from 0.84 to 0.97, depending on the class, while FCNN lagged with AUCs between 0.54 and 0.73, reinforcing its relatively poor classification performance. The next, Figure 7 displays the precision-recall (PR) curves across all classes, offering insights into the trade-off between precision and recall for each classification outcome.

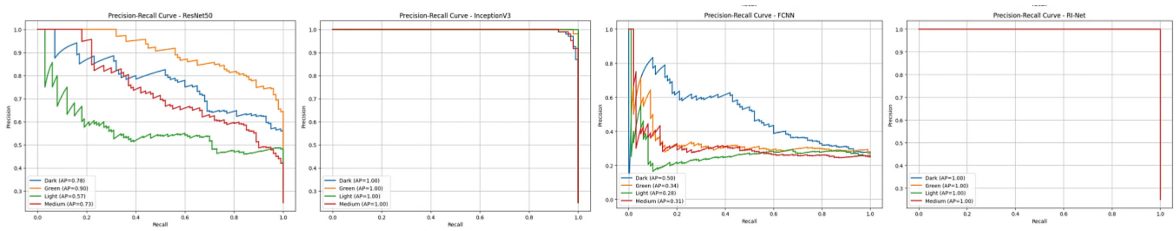


Figure 7. Precision-Recall Curves models: ResNet50, InceptionV3, FCNN, and the proposed model (RI-Net)

Figure 7 presents the precision-recall (PR) curves across all classes. These curves help evaluate model performance when dealing with imbalanced datasets or unequal class distributions. The RI-Net model consistently attained an average precision (AP) of 1.0 across all classes, demonstrating high accuracy and reliability in identifying true positives. InceptionV3 also delivered excellent AP values close to 1.0, whereas ResNet50 showed more variation with AP scores between 0.57 and 0.90, reflecting class-dependent performance differences. The FCNN's low AP values (0.28–0.50) further emphasized its limitations in precise and consistent classification. The next, Figure 8 provides a comparative analysis of class-wise accuracy for each model, highlighting performance variations across individual classes.

Figure 8 compares accuracy across individual classes for each model. This view highlights class-level weaknesses in standard models. For example, ResNet50 underperforms in the Green class, while InceptionV3 misclassifies some Medium beans. RI-Net shows uniform 100% accuracy across all classes, reaffirming its robustness and consistent performance.

Vol. 7, No. 3, June 2026, Page. 2038-2050
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5537

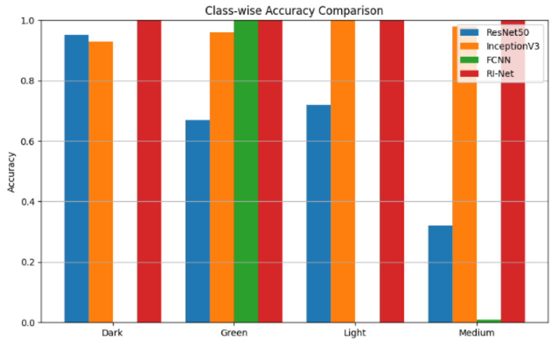


Figure 8. Class-Wise Accuracy Comparisons four key models: ResNet50, InceptionV3, FCNN, and the proposed model (RI-Net)

Overall, The comprehensive evaluation confirms that the hybrid RI-Net model significantly outperforms the standard ResNet50, InceptionV3, and FCNN models in classifying local coffee bean varieties. Its consistent superiority in both accuracy and stability across classes demonstrates the effectiveness of integrating residual and inception modules. While InceptionV3 presents a strong competitive baseline, ResNet50 and FCNN lag behind, highlighting the advantage of the hybrid approach in capturing complex visual features for coffee bean classification.

3.3. Discussion of Testing Outcomes

The testing results, as summarized in Table 3, offer valuable insights into the practical performance of each model in classifying different types of coffee beans.

Table 3. Performance comparison of models based on accuracy, precision, recall, and f1-score

Model	Accuracy	Precision	Recall	F1-score
ResNet50	0.6650	0.680355	0.6650	0.647592
InceptionV3	0.9675	0.968857	0.9675	0.967551
FCNN [25]	0.2525	0.312657	0.2525	0.105151
RI-Net	1.0000	1.000000	1.0000	1.000000

Table 3 compares the classification performance of four CNN models ResNet50, InceptionV3, FCNN, and the proposed RI-Net based on accuracy, precision, recall, and F1-score. The RI-Net model achieved perfect scores (1.000) across all metrics, indicating flawless classification of the test dataset. InceptionV3 followed with strong performance (96.75% accuracy), while ResNet50 and FCNN lagged significantly. This table provides quantitative evidence that RI-Net substantially outperforms both standard and baseline models, reinforcing its effectiveness in coffee bean classification tasks. The following are the results of the hybrid model prediction (RI-Net) on 20 sample images as shown in Figure 9.

P-ISSN: 2723-3863
https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5537

Figure 9. The results of the hybrid model prediction (RI-Net)

Figure 9 presents the prediction outcomes of the proposed RI-Net model on a set of 20 coffee bean images, representing all four roasting levels: Green, Light, Medium, and Dark. Each image is annotated with the actual class label (Ground Truth/GT), the predicted label (Pred), and the prediction outcome (Hasil), which is marked as "Benar" (Correct) if the classification matches.

This figure serves to visually demonstrate the real-world applicability and robustness of RI-Net beyond numerical metrics. Despite natural variations in texture, color gradient, and surface features within each class, RI-Net accurately identifies all samples. For example, Dark roast beans typically exhibit deeper brown tones and shiny surfaces, while Green beans maintain a pale, matte appearance. These subtle visual cues are effectively captured by RI-Net’s hybrid architecture, highlighting its strong feature extraction capabilities.

By displaying the predicted versus actual labels alongside correctness validation, this figure allows the reader to visually verify the model's classification reliability. The perfect agreement between predictions and ground truth across all samples underscores the model's generalization power, especially when dealing with high intra-class variation. This visual evidence further validates the RI-Net model’s suitability for automated coffee quality control systems, contributing to improved consistency and objectivity in post-harvest classification tasks.

The RI-Net model achieved a perfect classification accuracy of 100%, significantly outperforming the baseline models ResNet50 and InceptionV3, which obtained accuracies of 66.5% and 96.75%, respectively. When compared with previous studies, such as Anto et al. (2024), who reported approximately 92% accuracy using a standard CNN for roasted coffee bean classification, our hybrid model demonstrates a marked improvement. Similarly, Chang and Liu (2024) applied a multiscale CNN for defect detection in coffee beans and reported accuracy levels between 85% and 90%, which remain below the performance achieved by RI-Net. These comparisons underscore the effectiveness of integrating residual learning and multi-scale feature extraction in addressing the complex visual characteristics of local coffee beans, resulting in superior classification performance

4. CONCLUSION

This study successfully developed and evaluated a novel hybrid convolutional neural network model (RI-Net) that integrates residual learning from ResNet and multi-scale feature extraction from Inception architectures to classify local Indonesian coffee bean types with high accuracy. The hybrid model consistently outperformed established baseline models, including ResNet50, InceptionV3, and a FCNN, by achieving perfect classification metrics 100% accuracy, precision, recall, and F1-score demonstrating its superior capability in capturing the complex visual characteristics inherent in different coffee roasting levels. Experimental results showed that while traditional models such as ResNet50 and InceptionV3 delivered strong and stable performance, they were less effective than the hybrid model in distinguishing subtle differences among coffee bean classes, particularly in challenging cases such as Medium and Dark roasts. The FCNN model, though simple and computationally efficient, exhibited significant limitations in generalization and classification accuracy. The comprehensive evaluation including confusion matrices, ROC and precision-recall curves, and detailed class-wise analyses confirmed the robustness and practical applicability of the proposed RI-Net architecture. Moreover, qualitative assessment of sample predictions further validated its effectiveness in real-world scenarios. This research provides a significant contribution to precision agriculture by delivering a robust, high-performance CNN-based classification system tailored to the specific visual features of Indonesian coffee beans. The findings not only demonstrate improved classification accuracy but also offer a viable framework for automated coffee quality control systems, which can enhance post-harvest processing and market competitiveness. Future work should focus on extending the model to larger and more diverse datasets and incorporating explainability techniques to further enhance its practical deployment.

CONFLICT OF INTEREST

The authors declares that there is no conflict of interest between the authors or with research object in this paper.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the financial support provided by the Ministry of Education, Culture, Research, and Technology through the Directorate General of Research and Development for the Research and Community Service Program, Fiscal Year 2025. This research was conducted under the postgraduate research scheme (Master's Thesis Research), and the funding has been instrumental in enabling the successful completion of this study. We sincerely thank the Ministry and Directorate General for their commitment to advancing research and higher education in Indonesia.

REFERENCES

[1] D. Apriani, A. Bashir, F. Marissa, and Mukhlis, "The Structure-Conduct-Performance of Indonesian Coffee Processing Industry," *KnE Social Sciences*, vol. 2024, pp. 100–120, 2024, doi: 10.18502/kss.v9i14.16096.

[2] A. Wahyudi, S. Wulandari, A. Aunillah, and J. C. Alouw, "Sustainability certification as a pillar to promote Indonesian coffee competitiveness," *IOP Conference Series: Earth and Environmental Science*, vol. 418, no. 1, 2020, doi: 10.1088/1755-1315/418/1/012009.

[3] E. Putri, S. Astuti, and C. Safitri, "Analysis of Elements Business Model in Coffee Shop," *Economic Education Analysis Journal*, vol. 12, no. 3, pp. 255–265, 2023, doi: 10.15294/eeaj.v12i3.70595.

[4] N. Gizaw, J. Abafita, and T. M. Merra, "Impact of coffee exports on economic growth in Ethiopia; An empirical investigation," *Cogent Economics \& Finance*, vol. 10, no. 1, p. 2041260, 2022, doi: 10.1080/23322039.2022.2041260.

[5] S. H. Muhie, "Strategies to improve the quantity and quality of export coffee in Ethiopia, a look at multiple opportunities," *Journal of Agriculture and Food Research*, vol. 10, p. 100372, 2022, doi: https://doi.org/10.1016/j.jafr.2022.100372.

[6] V. Poncet, P. van Asten, C. P. Millet, P. Vaast, and C. Allinne, "Which diversification trajectories make coffee farming more sustainable?," *Current Opinion in Environmental Sustainability*, vol. 68, p. 101432, 2024, doi: https://doi.org/10.1016/j.cosust.2024.101432.

[7] G. M. Ngure and K. N. Watanabe, "Coffee sustainability: leveraging collaborative breeding for variety improvement," *Frontiers in Sustainable Food Systems*, vol. 8, 2024, doi: 10.3389/fsufs.2024.1431849.

[8] D. Fauzi, U. Purnamasari, S. A. Wicaksono, and C. Maharani, "When local customs meet the market: an analysis of coffee value chain in Tebat Benawa customary community, Indonesia," *International Journal of Agricultural Sustainability*, vol. 21, no. 1, p. 2231769, 2023, doi: 10.1080/14735903.2023.2231769.

[9] I. M. Pakaya, R. Radi, and B. Purwantana, "Classification of Roasting Level of Coffee Beans Using Convolutional Neural Network with MobileNet Architecture for Android Implementation," *Jurnal Teknik Pertanian Lampung (Journal of Agricultural Engineering)*, vol. 13, no. 3, p. 924, 2024, doi: 10.23960/jtep-l.v13i3.924-932.

[10] M. García, J. E. Candelo-Becerra, and F. E. Hoyos, "Quality and Defect Inspection of Green

Coffee Beans Using a Computer Vision System," *Applied Sciences*, vol. 9, no. 19, 2019, doi: 10.3390/app9194195.

[11] L. Chen, S. Li, Q. Bai, J. Yang, S. Jiang, and Y. Miao, "Review of Image Classification Algorithms Based on Convolutional Neural Networks," *Remote Sensing*, vol. 13, no. 22, 2021, doi: 10.3390/rs13224712.

[12] B. S. Rao, "INTELLIGENT SYSTEMS AND APPLICATIONS IN Advancements in Image Classification and Object Detection : Leveraging Deep Learning for Enhanced Performance," vol. 12, pp. 466–473, 2024.

[13] N. Sharma, V. Jain, and A. Mishra, "An Analysis Of Convolutional Neural Networks For Image Classification," *Procedia Computer Science*, vol. 132, pp. 377–384, 2018, doi: https://doi.org/10.1016/j.procs.2018.05.198.

[14] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar, "A review of convolutional neural networks in computer vision," *Artificial Intelligence Review*, vol. 57, no. 4, 2024, doi: 10.1007/s10462-024-10721-6.

[15] K. Kanwal *et al.*, "Efficient CNN architecture with image sensing and algorithmic channeling for dataset harmonization," *Scientific Reports*, vol. 15, no. 1, p. 7552, 2025, doi: 10.1038/s41598-025-90616-w.

[16] W. Xu, Y.-L. Fu, and D. Zhu, "ResNet and its application to medical image processing: Research progress and challenges," *Computer Methods and Programs in Biomedicine*, vol. 240, p. 107660, 2023, doi: https://doi.org/10.1016/j.cmpb.2023.107660.

[17] G. Meena, "Sentiment analysis on images using convolutional neural networks based Inception-V3 transfer learning approach," *International Journal of Information Management Data Insights*, vol. 3, no. 1, 2023, doi: 10.1016/j.jjimei.2023.100174.

[18] C. Mawardi, A. Buono, K. Priandana, and Herianto, "Performance Analysis of ResNet50 and Inception-V3 Image Classification for Defect Detection in 3D Food Printing," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 14, no. 2, pp. 798–804, 2024, doi: 10.18517/ijaseit.14.2.19863.

[19] S. Castillo-Girones, S. Munera, M. Martínez-Sober, J. Blasco, S. Cubero, and J. Gómez-Sanchis, "Artificial Neural Networks in Agriculture, the core of artificial intelligence: What, When, and Why," *Computers and Electronics in Agriculture*, vol. 230, p. 109938, 2025, doi: https://doi.org/10.1016/j.compag.2025.109938.

[20] S. Thaseentaj and S. S. Ilango, "Deep Convolutional Neural Networks for South Indian Mango Leaf Disease Detection and Classification," *Computers, Materials and Continua*, vol. 77, no. 3, pp. 3593–3618, 2023, doi: https://doi.org/10.32604/cmc.2023.042496.

[21] L. C. Ngugi, M. Abdelwahab, and M. Abo-Zahhad, "A new approach to learning and recognizing leaf diseases from individual lesions using convolutional neural networks," *Information Processing in Agriculture*, vol. 10, no. 1, pp. 11–27, 2023, doi: https://doi.org/10.1016/j.inpa.2021.10.004.

[22] B. Tugrul, E. Elfatimi, and R. Eryigit, "Convolutional Neural Networks in Detection of Plant Leaf Diseases: A Review," *Agriculture*, vol. 12, no. 8, 2022, doi: 10.3390/agriculture12081192.

[23] 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*, vol. 12, no. 4, pp. 1005 – 1018, 2024, doi: 10.52549/ijeei.v12i4.5531.

[24] Y. A. Auliya, I. Fadah, Y. Baihaqi, and I. N. Awwaliyah, "Green Bean Classification: Fully Convolutional Neural Network with Adam Optimization," *Mathematical Modelling of Engineering Problems*, vol. 11, no. 6, pp. 1641 – 1648, 2024, doi: 10.18280/mmep.110626.

[25] S. Arwatchananukul, D. Xu, P. Charoenkwan, S. Aung Moon, and R. Saengrayap, "Implementing a deep learning model for defect classification in Thai Arabica green coffee beans," *Smart Agricultural Technology*, vol. 9, 2024, doi: 10.1016/j.atech.2024.100680.

[26] S.-J. Chang and K.-H. Liu, "Multiscale Defect Extraction Neural Network for Green Coffee Bean Defects Detection," *IEEE Access*, vol. 12, pp. 15856 – 15866, 2024, doi: 10.1109/ACCESS.2024.3356596.

[27] Y. Liu, S. Zhou, W. Han, W. Liu, Z. Qiu, and C. Li, "Convolutional neural network for

Vol. 7, No. 3, June 2026, Page. 2038-2050
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5537

hyperspectral data analysis and effective wavelengths selection," *Analytica Chimica Acta*, vol. 1086, pp. 46 – 54, 2019, doi: 10.1016/j.aca.2019.08.026.

[28] H. L. Gope and H. Fukai, "Peaberry and normal coffee bean classification using CNN, SVM, and KNN: Their implementation in and the limitations of Raspberry Pi 3," *AIMS Agriculture and Food*, vol. 7, no. 1, pp. 149 – 167, 2022, doi: 10.3934/AGRFOOD.2022010.

[29] 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)*, 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[30] C. Szegedy, S. Reed, P. Sermanet, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," pp. 1–12.