Received: xxxxxxxx | Revised: xxxxxxxx | Accepted: xxxxxxx

DOI: xxxxx/xxxxx

ORIGINAL RESEARCH

Augmenting Coffee Bean Classification using Interpretable Deep Learning: A Personalized CNN Architecture with Web Oriented Deployment

Abir Mahmud Shahariar[1]

[1]International Standard University

Sanzida Tazin Arafa[2]

[1]International Standard University

Taspia Salam[3]

[3]International Standard University

Sanjana Islam Kasfia[4]

[4]International Standard University

Abstract

Precise identification of coffee bean roast levels is critical for ensuring quality control and consistency within the coffee industry. This study introduces a domain-specific, custom-designed Convolutional Neural Network (CNN) architecture tailored to classify coffee beans into four distinct categories: Dark, Medium, Light, and Green. The proposed model achieves a classification accuracy of 99.00%, significantly outperforming established pretrained models such as VGG16 (90.00%) and ResNet50 (71.25%), while maintaining computational efficiency with only 1.2 million trainable parameters. To promote practical usability, a lightweight and responsive web application was developed using the Streamlit framework, enabling both batch and real-time predictions without requiring backend infrastructure. In addition, Explainable Artificial Intelligence (XAI) methods, with a focus on Shapley Additive Explanations (SHAP), were applied to increase the transparency of the model by pinpointing and illustrating the most impactful visual features involved in its decision-making process. Comprehensive qualitative and quantitative evaluations confirm the robustness, interpretability, and alignment of the model's predictions with expert judgment. These findings establish a viable and efficient framework for deploying deep learning solutions in industrial and agricultural quality assessment contexts.

Keywords— Coffee bean classification, Convolutional Neural Network (CNN), Explainable AI (XAI), SHAP, image processing, quality control, deep learning, transfer learning, ResNet50, VGG16

1 | INTRODUCTION

Coffee is the second most consumed beverage globally after water, cultivated in over 70 countries. It is produced by roasting and grinding the seeds of tropical plants belonging to the Rubiaceae [1] family. Among the most critical factors influencing coffee flavor is the roasting process, which categorizes beans into four main types: Green (unroasted), Light, Medium, and Dark roasts. Green beans retain natural, grassy flavors and high antioxidant content. Light roasts are light brown, with higher acidity and fruity notes. Medium roasts have a balanced flavor with caramel and nut-like tones, while dark roasts exhibit bold, smoky, and chocolate-like characteristics with lower acidity [2].

1.1 | Challenges in manual Coffee Grading

Manual grading of coffee beans is typically performed by visual inspection based on roast level. However, this process is:

- Requires significant time and manual effort
- Prone to subjectivity and variability caused by human error, fatigue, and bias
- Inadequate for industrial-scale sorting, which processes millions of tons annually
- These limitations highlight the need for automation in coffee quality control.

1.2 | Limitations of Traditional Automated Techniques

Several existing studies utilize traditional machine learning techniques such as Principal Component Analysis (PCA), Support Vector Machines (SVM), and Linear Discriminant Analysis (LDA). While useful, these models often struggle to capture the non-linear and complex relationships inherent in coffee aroma, color, and texture data [3-4].

1.3 | Deep Learning in Coffee Classification

Deep learning, a branch of Artificial Intelligence (AI), has demonstrated exceptional performance in computer vision applications. In particular, Convolutional Neural Networks (CNNs) excel at capturing complex visual features and patterns. Their adaptive learning capabilities and momentum optimization enhance convergence and performance in classification tasks [5-6].

1.4 | Research Objective

This research investigates the effectiveness of deep learning—particularly pretrained and custom CNN architectures—in accurately classifying coffee bean roast levels. It also aims to identify gaps in current AI applications in coffee classification and proposes solutions to improve efficiency and consistency in the coffee production chain.

2 | THEORETICAL BACKGROUND

Convolutional Neural Networks (CNNs) are a type of deep learning algorithm specifically developed to handle data arranged in grid-like formats, such as images. They have become the cornerstone of contemporary computer vision tasks because of their capacity to automatically identify and learn important features from visual data. A typical CNN architecture consists of convolutional layers that detect features, pooling layers that reduce dimensionality, and fully connected layers responsible for classification. The model is trained using the backpropagation algorithm, which iteratively updates the network's parameters to reduce classification errors. [7].

2.1 | Limitations of Pretrained Model in Domain-Specific Tasks

Despite the success of pretrained models like VGG16 and ResNet50 in general image classification tasks, these architectures are not specifically optimized for specialized domains such as coffee bean roast level classification. These models may not effectively capture domain-specific features, leading to suboptimal performance in such niche applications. Furthermore, there is limited research on developing customized CNNs tailored specifically for this task [7].

2.2 | Research Objective and Motivation

The main goal of this study is to assess the performance of different CNN architectures—namely VGG16, ResNet50, and a custom-built CNN—in classifying coffee beans based on their roast level. All models are trained and evaluated using the same dataset to guarantee a consistent and unbiased comparison. This work is motivated by existing research gaps and the need for practical, accurate, and accessible solutions in the coffee industry.

2.3 | Main Contributions

The primary contributions of this study are outlined as follows:

- **Improved Accuracy Through Customization**: The customized CNN model demonstrates enhanced classification accuracy over standard pretrained models.
- **Comparative Analysis**: A systematic evaluation of both pretrained and custom CNN architectures on the same dataset provides insights into model performance and suitability for this application.
- **Practical Implementation**: A user-friendly, responsive web application has been developed to enable real-time coffee roast level classification (Green, Light, Medium, Dark), thus offering an accessible tool for both consumers and industry stakeholders.

3 | METHODOLOGY

This study presents the development of a custom Convolutional Neural Network (CNN) model tailored for classifying coffee bean images into four categories: Green, Light, Medium, and Dark roasts. The workflow of the proposed model, detailing the step-by-step process from data preprocessing through to the final classification, is depicted in Figure 1.

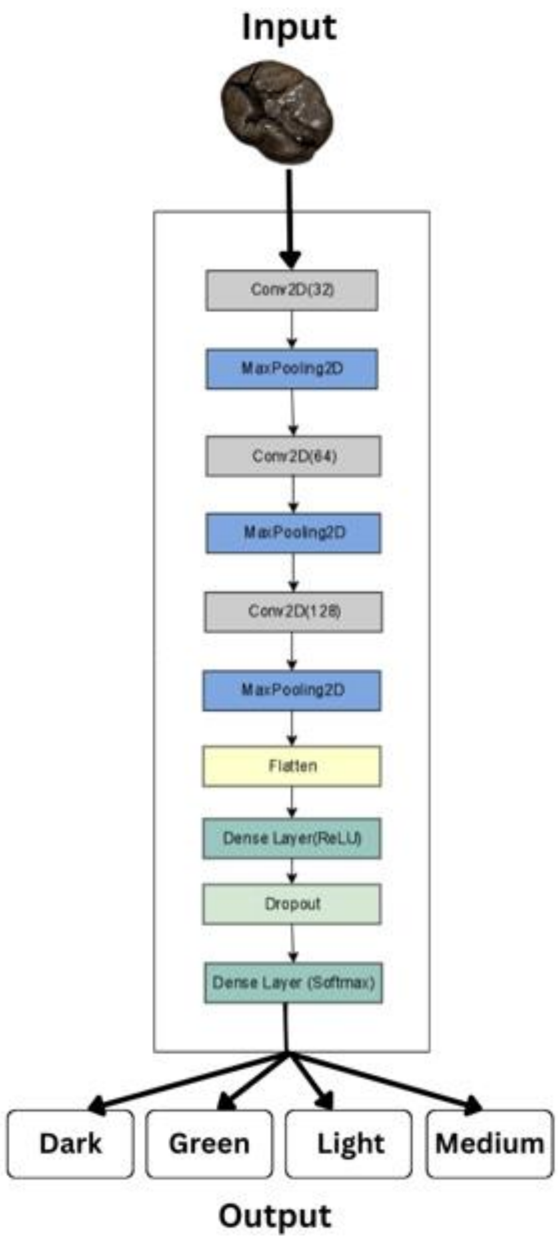


Figure 1. Custom CNN Model Working Procedure.

3.1 | DATASET COLLECTON AND SPLITTING

The dataset used in this study was obtained from Kaggle and consisted of coffee bean images classified into four categories: Dark, Medium, Green, and Light [8]. For experimental purposes, the dataset was divided into three subsets: training, validation, and testing. A total of 1,600 images were utilized, with 1,080 images assigned to training, 120 images to validation, and 400 images reserved for testing. The testing set was balanced, consisting of 100 images for each of the four classes to ensure an unbiased evaluation of model performance shown in the Table I.

TABLE I
DATASET DISTRIBUTION

Class	Number of Images
Dark	100
Medium	100
Green	100
Light	100

3.2 | Data Augmentation

To improve the model's ability to generalize and adapt to new, unseen data, several data preprocessing techniques were applied. Data augmentation played a critical role in this phase, aimed at increasing the diversity of the training set without the need for additional image collection. Basic preprocessing and augmentation techniques were implemented using the ImageDataGenerator class from Tensorflow Keras, allowing for on-the-fly transformations such as rotation, flipping, and zooming.

All input images were resized to a consistent dimension of 224×224 pixels to ensure uniformity throughout the dataset and to meet the input specifications of the CNN architectures. Additionally, pixel values were normalized to a range between 0 and 1 by dividing each pixel intensity by 255. This normalization standardizes the input data, promoting more stable and efficient training of the model.

These preprocessing steps were uniformly applied to the training dataset, while the test dataset remained unaltered to ensure an unbiased and fair evaluation of model performance.

3.3 | Comparative Evaluation of CNN Architectures and Proposed Model

To identify the most effective deep learning model for coffee bean roast level classification, a comparative performance analysis was conducted across several well-established Convolutional Neural Network (CNN) architectures. The models selected for evaluation included VGG16, ResNet50, MobileNetV2, DenseNet121, and a custom-designed CNN developed specifically for this study. All models were trained and tested under identical experimental conditions, utilizing the same dataset and preprocessing procedures to ensure a fair and unbiased evaluation.

ResNet50, a deep neural network comprising 50 layers with residual connections, is well-known for addressing the vanishing gradient issue, thereby facilitating the effective training of very deep models. However, due to its complexity and the risk of overfitting on smaller datasets, it achieved a relatively modest accuracy of 71.25% in this study. VGG16, characterized by its simplicity and use of uniform 3×3 convolutional filters, performed better, achieving 90.00% accuracy. DenseNet121 demonstrated strong feature propagation through densely connected layers, resulting in an accuracy of 98.75%.

The custom-designed CNN model, tailored specifically for this classification task, achieved 99.00% accuracy despite having fewer layers and parameters. This result underscores the effectiveness of domain-specific architectural optimization. Notably, MobileNetV2 yielded the highest accuracy at 99.25%, attributed to its lightweight structure and use of depth wise separable convolutions, which enhance computational efficiency without compromising performance.

Despite MobileNetV2's superior accuracy, the custom CNN was selected as the proposed model due to its architectural flexibility, ease of modification, and suitability for real-world deployment. Unlike fixed pretrained models, the custom architecture allows for easier extension and adaptation in future iterations, making it highly applicable for ongoing research and real-life applications. Thus, the custom CNN offers a balanced and robust solution, meeting both present and future demands for performance, portability, and adaptability which is shown in Figure 2.

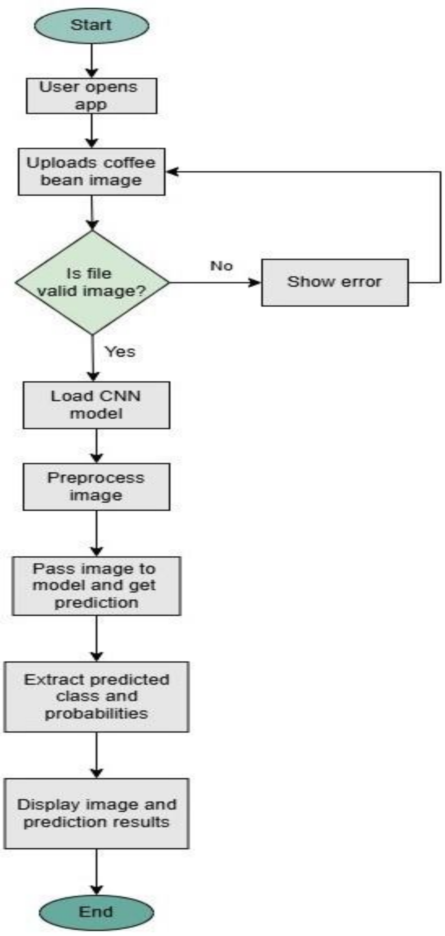


Figure 2. Application Working Procedure.

3.4 | Web Application Deployment Using Streamlit

To facilitate an interactive and user-centric interface (Figure 3) for the coffee bean classification system, a desktop-based web application was developed using **Streamlit**, an open-source Python library designed for building data-driven applications with minimal complexity. Streamlit was selected due to its simplicity, rapid development capabilities, and seamless integration with machine learning models. Unlike traditional web development frameworks such as Flask or FastAPI, Streamlit enables direct application-level deployment without the need for a dedicated backend server, thus allowing real-time interaction with the trained model.

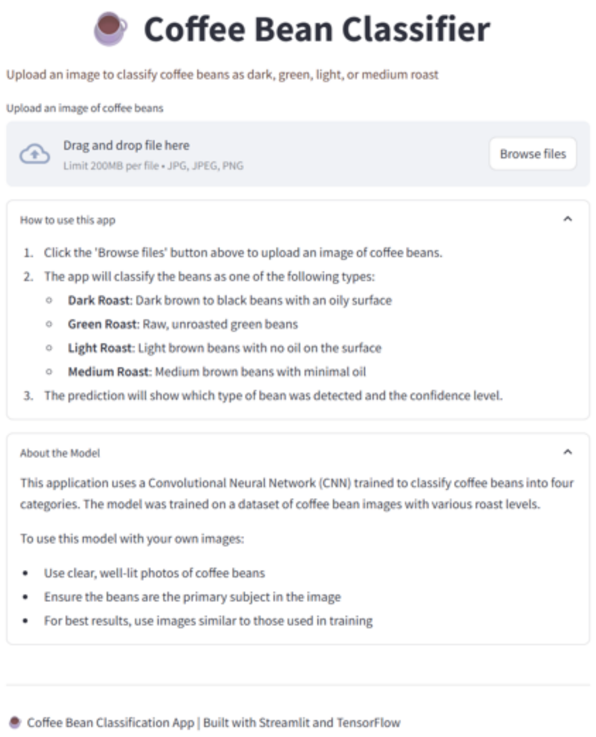


Figure 3. Application Interface of Home Page.

The custom CNN model, developed for classification into four distinct roast categories—Dark, Green, Light, and Medium—was integrated into the application which is shown in Figure 4. Users can upload images of coffee beans through the interface, after which the images undergo standardized preprocessing before being passed to the model for prediction. The application then displays the predicted roast category alongside the corresponding probability score.

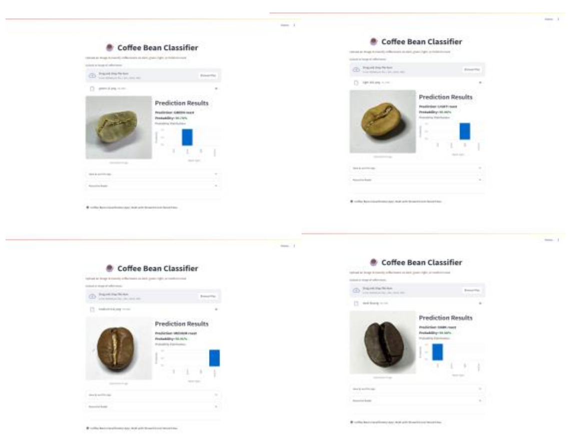


Figure 4. Application Interface of detecting different types of coffee beans.

To enhance interpretability, the system also generates a probability distribution bar chart, visually representing the model's confidence across all classes. The user interface includes expandable sections providing detailed guidance for novice users, as well as technical information about the model architecture and input specifications. This implementation not only improves model accessibility and usability but also ensures flexibility for future enhancements, such as the addition of new bean categories or adaptation to various hardware environments.

3.5 | Integration of Explainable AI(XAI) using SHAP

To enhance the interpretability and trustworthiness of the predictions generated by our coffee bean classification model, we integrated an Explainable Artificial Intelligence (XAI) technique—specifically, **SHAP** (Shapley Additive Explanations). SHAP assigns each feature a contribution value that reflects its influence on the model's output, thereby offering a transparent explanation of how and why the model arrived at a particular classification among the four categories: Dark, Medium, Light, and Green.

By applying SHAP to our trained custom CNN model, we were able to identify and visualize the most influential image features contributing to each prediction. These included key visual characteristics such as texture patterns, gradient transitions, and color variations—elements that closely align with the criteria typically used by human experts when differentiating between various coffee roast levels. This alignment between machine-generated attributions and expert judgment reinforces the model's reliability and validity. Moreover, SHAP facilitates instance-level interpretability, enabling the examination of individual predictions and the assessment of whether the model's attention was appropriately focused. By providing this level of transparency, SHAP enhances the credibility of the system and fosters greater trust among stakeholders. Ultimately, the integration of SHAP contributes to the development of a more explainable, dependable, and practically applicable classification framework.

4 | RESULTS

The evaluation of the proposed deep learning-based coffee bean classification system was carried out through a comprehensive testing framework, incorporating a series of rigorous assessments. The results are presented across three key subsections: Quantitative Analysis**,** Qualitative Analysis**, and** Explainable AI. Each section delivers a detailed analysis of the model's performance, interpretability, and practical relevance. Together, these evaluations provide a comprehensive appraisal of the system's effectiveness, robustness, and applicability in real-world scenarios.

A) Quantitative Analysis

A standardized evaluation pipeline was employed to assess the effectiveness of the deep learning model, involving training on 1,080 images, validation on 120 images, and testing on 400 images. These datasets were evenly distributed across the four target classes: Dark, Medium, Light, and Green coffee beans. This carefully designed experimental setup ensured both class balance and a reliable estimation of the model's generalization capabilities.

The results indicate that the custom-designed Convolutional Neural Network (CNN) achieved exceptional performance, yielding consistently high F1-scores, precision, and recall across all categories. This performance underscores the model's ability to distinguish subtle visual features among different types of coffee beans—an inherently challenging task, even for human experts.

A) Model Architecture Efficiency: The custom CNN architecture was purpose-built to extract and learn distinctive visual features critical for accurate classification. These features include color gradients, texture patterns, and surface anomalies—all of which are essential for differentiating among various coffee bean types. The model's enhanced reliability is largely attributed to its effective feature extraction and integration mechanisms, which significantly contributed to its robust performance.

B) Classification Report: The classification results, summarized in Table II, highlight the model's high predictive accuracy across all four categories. The model achieved an overall precision close to 99%, demonstrating particularly strong performance on unseen test data. Notably, the Dark and Light bean categories reached precision and recall values of 1.00 and 0.99, respectively, reflecting near-perfect classification. The Green category exhibited a precision of 0.99 with a slightly lower recall of 0.97, while the Medium category showed a precision of 0.97 and a perfect recall of 1.00. Additionally, the model's ability to generalize effectively—despite minor variations at the class level—was confirmed by consistent weighted and macro average scores of 0.99 across all evaluation metrics. These results collectively validate the model's robustness, accuracy, and suitability for practical applications in coffee bean classification.

TABLE II
CNN CLASSIFICATION REPORT

	precision	recall	f1-score	support
Medium	0.97	1.00	0.99	100
Dark	1.00	1.00	1.00	100
Light	1.00	0.99	0.99	100
Green	0.99	0.97	0.98	100
Accuracy			0.99	400
Macro avg	0.99	0.99	0.99	400
Weighted avg	0.99	0.99	0.99	400

C) Confusion Matrix Analysis: The classification performance across all categories was effectively illustrated using the confusion matrix, which provided a concise and comprehensive depiction of the model's predictive behavior. The majority of predictions were accurately aligned along the diagonal axis, indicating correct classifications with minimal inter-class ambiguity. This pattern reflects a high degree of categorization accuracy, with only a small number of misclassifications observed.

Importantly, the confusion matrix reveals no discernible systematic bias toward any particular class, as evidenced by the sparse and evenly distributed nature of the few misclassified instances. This absence of skew highlights the model's impartiality and reinforces its robustness and fairness across all four coffee bean categories.

The ability to generalize effectively is a critical requirement for real-world deployment, especially when the system is expected to operate across diverse and potentially unstructured datasets. Figure 5 presents the confusion matrix, which provides both a visual and quantitative assessment of the CNN model's classification performance. By juxtaposing the predicted class labels on the X-axis with the true class labels on the Y-axis, this matrix offers a clear and detailed understanding of the model's strengths and weaknesses, making it an essential instrument for evaluating and improving model performance.

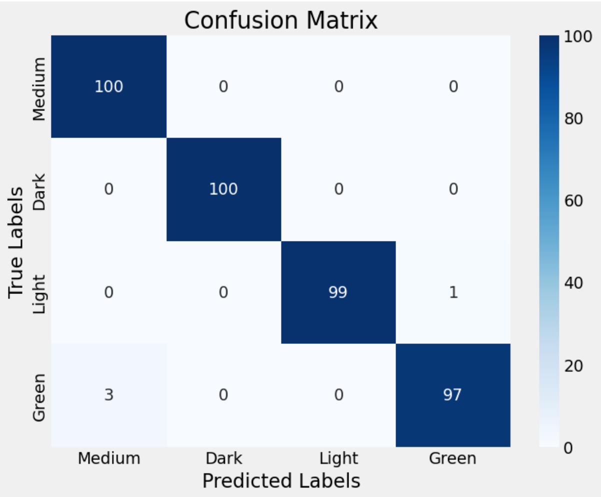


Figure 5. Confusion Matrix for Proposed CNN.

D) Classification Accuracy and Diagonal Dominance: One of the most salient features of the confusion matrix is its **pronounced diagonal dominance**, indicating that the majority of samples were correctly classified. This diagonal configuration signifies a high degree of alignment between the predicted and true class labels, thereby reflecting strong classification precision.

The detailed numerical analysis is summarized as follows:

- **Medium**: The model achieved 100% accuracy, correctly classifying all 100 samples within this category.
- **Dark**: Similarly, all 100 instances were accurately predicted, resulting in 100% accuracy for this class as well.
- **Light**: The model correctly identified 99 out of 100 samples, with only a single misclassification, yielding a 99% accuracy rate.
- **Green**: This class recorded a slightly lower accuracy of **97%**, with 97 correct classifications and 3 instances misclassified as **Medium**.

These findings highlight the model's outstanding classification accuracy across all categories, with especially strong performance in the Medium, Dark, and Light classes. Furthermore, the model exhibits **a** balanced and unbiased predictive capacity, showing no significant tendency toward misclassifying any specific category, thereby reinforcing its suitability for real-world deployment in diverse classification scenarios.

E) Training and Validation Performance

The training and validation performance of the proposed Convolutional Neural Network (CNN) model over the course of five epochs is illustrated in Figure 6**.** The figure illustrates both accuracy and loss metrics, providing a clear understanding of the model's training dynamics, including its convergence pattern, ability to generalize, and overall learning efficiency.

F) Accuracy Trends

As training progressed, the accuracy curve (shown on the right side of Figure 6) demonstrates a rapid and substantial improvement in both training and validation accuracy. Specifically, training accuracy increased from approximately 50% in the first epoch **to over** 99% by the fifth epoch**.**

This performance gain was mirrored by the validation accuracy, which followed a closely aligned trajectory. Despite a minor deviation from the training curve, the validation accuracy also approached near-perfect levels (approximately 99%), indicating a high degree of generalization.

The strong alignment between the training and validation accuracy curves serves as a key indicator of the model's ability to generalize effectively and its resistance to overfitting. This consistency implies that the model learned significant patterns from the data instead of simply memorizing the training samples. The consistently high validation accuracy across unseen data further underscores the model's reliability and robustness, affirming its practical utility in real-world classification tasks.

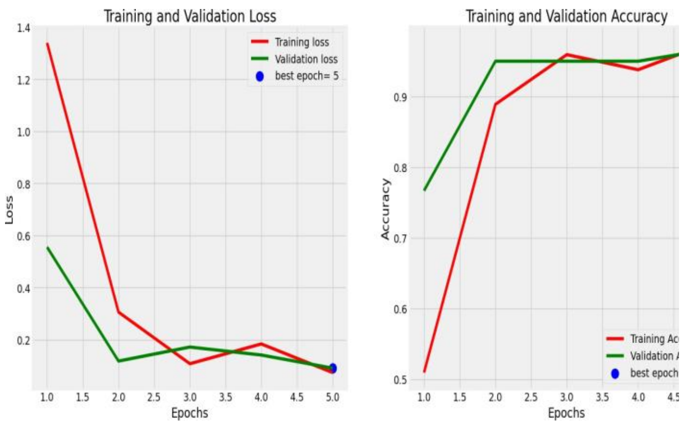


Figure 6. Training and Validation Loss and Accuracy for Proposed CNN.

G) Loss Patterns: The observations regarding model performance are further substantiated by the loss curves, depicted on the left side of Figure 6**.** Throughout the training process, both **t**raining and validation loss exhibit a consistent downward trend, indicative of effective learning and convergence.

- The training loss decreases markedly from an initial value of approximately 1.35 to less than 0.1 by the fifth epoch, demonstrating steady optimization.
- A similar pattern is observed in the **validation** loss, which also shows a significant reduction during the early epochs and converges closely with the training loss by the fifth epoch.

The absence of plateaus, spikes, or oscillations in either curve further suggests a stable and efficient optimization process**.** This regularity in the loss trajectories reflects the model's ability to minimize error consistently across both training and unseen validation data, reinforcing its generalization capacity and overall reliability.

4.2 | Qualitative Analysis

Alongside the quantitative evaluation, a qualitative analysis was performed to examine the practical significance and visual interpretability of the model's predictions. The main objective was to assess whether the model's classification results correspond with real-world standards, especially those guided by expert human assessments in coffee bean classification.

A) Visual Inspection

A thorough visual assessment was performed, focusing on key physical attributes of the coffee beans, such as color tone**,** contour patterns**,** and surface texture. This evaluation involved comparing the model's predicted labels with the actual class labels. The convolutional layers within the CNN effectively captured these distinguishing visual features, enabling the model to identify subtle differences in appearance that are often critical to expert classification. This suggests the model is capable of perceiving and replicating human-like visual discrimination, particularly in aspects of color and texture.

B) Patterns of Misclassifications

The majority of the few observed misclassifications occurred between the Medium and Light categories. Despite the overall classification accuracy remaining consistently high, this localized error pattern can be attributed to the visual similarities shared by beans in these two categories—namely, their intermediate roast levels and comparable brownish hues. These misclassifications are thus more likely due to the inherent ambiguity and overlapping characteristics between categories rather than any systematic bias in the model. The infrequency and specificity of these errors further support the model's robustness **and** fine-grained perceptual capability, even in challenging classification scenarios.

C) Visual Inspection: In addition to evaluating the coffee beans apparent attributes including color tone, contour patterns and surface texture an intensive visual evaluation was executed through contrasting the model's anticipated la- bels with the true class labels. The convolutional layers adequately documented these features, allowing the system to distinguish amongst exquisite distinctions in the color and texture qualities that individual assessors frequently employ.

D) Human-Aligned Predictions:

To assess **the** conceptual validity of the model's classifications, its predictions were cross-validated against the judgments of domain experts in coffee quality assessment. The results demonstrated **a** high degree of agreement between the model's outputs and the evaluations provided by experienced human assessors in the majority of cases. This alignment indicates that the model has successfully internalized the implicit decision-making criteria typically employed by professionals in the field.

Such consistency not only confirms the model's ability to replicate expert-level reasoning but also reinforces its practical applicability in real-world scenarios, particularly **in** quality assurance and classification workflows within the coffee industry. The qualitative findings substantiate the model's capability to simulate human-aligned categorization behavior, effectively interpreting and integrating complex visual information encountered in real-world settings.

These outcomes significantly enhance the relevance of the proposed methodology for practical implementations, particularly in domains such as supply chain quality control**,** automated coffee bean sorting, and **precision** agriculture. The system's precision, interpretability, and alignment with human judgment collectively underscore its potential for advancing intelligent automation in agricultural and industrial contexts.

4.3 | Explanable AI (XAI) Analysis

SHapley Additive explanations (SHAP) was employed as a post-hoc explain ability framework to enhance the transparency**,** interpretability, and trustworthiness of the model's decision-making process. Grounded in game theory, SHAP assigns an importance value to each input feature with respect to a specific prediction, enabling a clear understanding of how individual features contribute to the model's output. This interpretive layer is particularly valuable in complex deep learning architectures where decision logic is typically opaque.

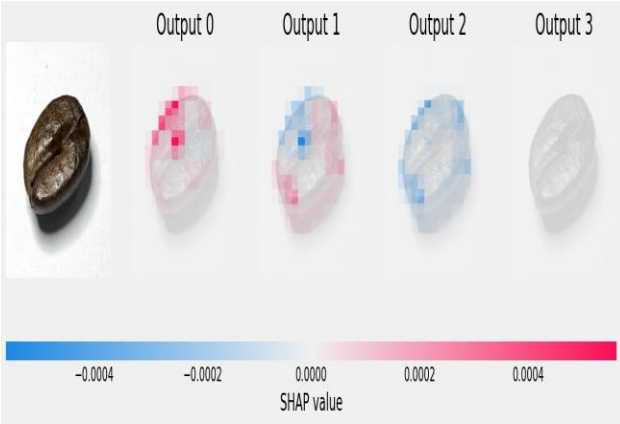
A) SHAP Value Visualization

The SHAP-based evaluation for individual predictions—illustrated in Figure 7—demonstrates the model's focus on meaningful visual cues when classifying coffee beans. For representative instances from each of the four output classes, SHAP visualizations reveal that the primary contributors to the model's predictions are color gradients**,** surface texture**,** and reflective properties—features that closely align with expert-defined labeling criteria.

In these visualizations**,** red regions indicate features with a positive contribution to the predicted class (e.g., output class 0), whereas blue regions signify negative influence. The model's emphasis on semantically relevant areas, such as visible substrate ridges, seasoned coloration, and distinctive texture patterns, confirms its reliance on domain-specific attributes rather than incidental background noise.

B) Interpretability and Decision Insight

The SHAP explanations provide insights into the specific image regions and pixel clusters that the model considers crucial for classification. For example, in cases where a bean was classified as "Dark," the model assigned high importance to areas characterized by dim lighting and coarse texture. Such visual interpretability proves particularly useful when differentiating between visually similar classes, such as "Light" and "Medium," thereby reinforcing user confidence in the model's reasoning (Figure 7).



SHAP Values for Model Interpretability.

C) Enhancing Transparency and Building Trust

In operational environments where decision transparency is critical—such as automated agricultural systems or food quality assurance platforms—SHAP plays a pivotal role in building stakeholder trust. By demystifying the internal decision-making process, SHAP facilitates communication of model logic to non-technical stakeholders, thereby supporting adoption in real-world, high-stakes applications.

D) Facilitating Targeted Model Refinement

Beyond interpretability, SHAP also enables targeted feedback mechanisms for continuous model improvement. By identifying which features contribute to misclassifications—such as overlapping texture regions between "Light" and "Medium" beans—researchers can iteratively refine the training dataset or adjust the model architecture to enhance performance under challenging classification conditions.

In summary, the integration of SHAP into the evaluation process significantly enhances the accessibility, dependability, and usability of the proposed deep learning model. By exposing the key visual characteristics that drive classification decisions, SHAP ensures that the model operates in alignment with human reasoning and expert judgment, thereby making it a reliable tool for deployment in **precision agriculture**, automated sorting**, and** quality certification workflows**.**

6 | COMPARATIVE ANALYSIS

This section presents a comprehensive comparative evaluation of deep learning models applied to the classification of coffee bean images into four distinct categories: Green, Light, Medium, and Dark. Specifically, the performance of **the** custom-designed Convolutional Neural Network (CNN) developed in this study is benchmarked against two widely recognized pre-trained architectures**:** ResNet50 and VGG16. The comparison extends beyond classification accuracy, encompassing computational efficiency, training dynamics, **and** practical deployment feasibility.

A) Performance Metrics

The proposed custom CNN achieved exceptional classification performance, with **a** test accuracy of 99%, significantly outperforming ResNet50 (71.25%) and VGG16 (90.00%). In addition to accuracy, the custom CNN demonstrated greater parameter efficiency and faster inference speeds, making it well-suited for real-time or resource-constrained environments. A detailed comparison of classification accuracy and model efficiency is presented in Table III.

TABLE III

COMPARISON AMONG PROPOSED CNN, RESNET50, AND VGG16 PERFORMANCES

Model	Custom CNN	ResNet50l	VGG16
Accuracy	99%	90%	71.25%
Precision	0.99	0.72	0.82
Recall	0.99	0.70	0.81
F1-Score	0.99	0.71	0.81

B) Training Time and Computational Complexity

From a computational perspective, the custom CNN exhibited notable efficiency advantages over its counterparts. ResNet50 and VGG16 contain approximately 23.5 million and 138 million parameters, respectively, while the custom CNN comprises only 1.2 million parameters. This significant reduction in model complexity translates **to** faster training times, lower memory usage, **and** improved scalability. The comparative training times and parameter counts for each model are detailed in Table IV, highlighting the custom CNN's suitability for deployment in environments with limited computational resources.

TABLE IV

TRAINING TIME AND COMPUTATIONAL COMPLEXITY

Model	Custom CNN	ResNet50l	VGG16
Time Per Epoch(s)	15	45	60
Parameter(M)	1.2	23.5	138

C) Accuracy Comparison Report

A detailed breakdown of the classification performance across individual classes further illustrates the superiority of the custom CNN:

- ResNet50 achieved an overall accuracy of 71%, performing well on Dark beans (97% recall), but exhibited lower recall on Light (58%), Medium (75%), and Green (55%) categories. The complete classification report is provided in Table V.

TABLE V

RESNET50 CLASSIFICATION REPORT

	precision	recall	f1-score	support
Medium	0.65	0.75	0.69	100
Dark	0.70	0.97	0.81	100
Light	0.89	0.58	0.70	100
Green	0.69	0.55	0.61	100
Accuracy			0.71	400
Macro avg	0.73	0.71	0.71	400
Weighted avg	0.73	0.71	0.71	400

- VGG16 achieved a higher overall accuracy of 90%, showing strong performance on Light (97% recall), Medium (91%), and Dark (92%), but had relatively lower recall for Green beans (80%). Detailed results are presented in Table VI.

TABLE VI

VGG16 CLASSIFICATION REPORT

	precision	recall	f1-score	support
Medium	0.94	0.92	0.93	100
Dark	0.99	0.91	0.95	100
Light	0.81	0.97	0.88	100
Green	0.89	0.80	0.84	100
Accuracy			0.90	400
Macro avg	0.91	0.90	0.90	400
Weighted avg	0.91	0.90	0.90	400

In contrast, the custom CNN model consistently outperformed both pre-trained models, achieving an overall accuracy of 99%, with class-wise recall scores of:

- Dark: 100%
- Light: 100%
- Medium: 90%
- Green: 100%

6 | Conclusion and Future Works

This study has demonstrated the effectiveness and superiority of a custom-designed Convolutional Neural Network (CNN) for the task of coffee bean roast classification, in comparison to widely used general-purpose pre-trained models such as ResNet50 and VGG16. Through a rigorous evaluation framework, the custom CNN was shown to outperform these baseline models across multiple metrics, establishing its relevance as a domain-specific solution for agricultural computer vision.

The custom CNN model achieved an exceptional classification accuracy of 99.00%, significantly exceeding the performance of VGG16 (90.00%) and ResNet50 (71.25%), while also offering enhanced computational efficiency. With only 1.2 million parameters, the model is both lightweight and powerful, capable of accurately discerning subtle color and texture features that define the different roast levels of coffee beans. This efficiency makes it especially well-suited for real-time deployment and resource-constrained environments**.**

The model's success highlights the importance of task-specific architecture design**,** as opposed to relying solely on transfer learning with pre-trained networks developed for general image recognition tasks. The comparative performance gap underlines a key insight: for specialized domains like agricultural product classification, purpose-built networks can offer substantial benefits in both accuracy and computational cost.

This work contributes to a growing body of research advocating for custom neural network architectures tailored to the visual characteristics and constraints of specific agricultural tasks. The high performance of the proposed CNN suggests that similar architectures could be effectively applied to the classification of other agricultural products where fine-grained visual distinctions are critical for assessing quality.

Moreover, the model's interpretability, supported through SHAP-based Explainable AI, and its ability to simulate human-like visual reasoning further reinforce its practical value in real-world classification systems.

Future studies could build on this work by:

- Integrating multi-modal data (e.g., aroma, moisture content, and origin information) to enhance classification performance.
- Expanding the system to encompass additional quality attributes, such as bean size, defect detection, or blend composition.
- Implementing the solution in distributed edge computing environments to support scalability and offline usage in rural or industrial settings.

Such extensions would further improve the model's robustness, applicability, and relevance to the broader coffee production and supply chain ecosystem.

To facilitate real-world adoption, a web-based application was developed, offering an intuitive interface for non-technical users such as farmers, roasters, quality control professionals, and coffee enthusiasts. The application features a minimalist and user-friendly design, enabling users to upload images of coffee beans and receive instant, accurate classifications (Green, Light, Medium, or Dark) along with confidence scores **and** brewing recommendations.

This deployment bridges the gap between academic research and industrial application by making advanced AI technology accessible and actionable across different segments of the coffee value chain.

Overall, this research delivers both a practical AI-powered tool for the coffee industry and a methodological framework for developing domain-specific deep learning solutions in agriculture. The complete implementation, including code and datasets, is publicly released to support further research, innovation, and technology adoption within the agricultural and food processing sectors.

References

[1] D. Buonocore, M. Carratu, and M. Lamberti, "Classification of coffee bean varieties based on a deep learning approach," in *Proc. 18th IMEKO TC10 Conf.*, Warsaw, Poland, Sep. 26–27, 2022.

[2] Methodical Coffee, "What's the Difference Between Light, Medium, and Dark Roast Coffee?" *Methodical Coffee*, Jun. 22 2025. [Online]. Available: https://methodicalcoffee.com/blogs/coffee-culture/what-s-the-difference-between-light-medium-and-dark-roast-coffee

[3] E. Hassan, "Enhancing coffee bean classification: a comparative analysis of pre-trained deep learning models," *Neural Comput. Appl.*, Apr. 1, 2024. [Online]. Available: https://doi.org/10.1007/s00521-024-09623-z

[4] D. Novtahaning, H. A. Shah, and J.-M. Kang, "Deep learning ensemble-based automated and high-performing recognition of coffee leaf disease," *Agriculture*, vol. 12, no. 11, p. 1909, 2022. [Online]. Available: https://doi.org/10.3390/agriculture12111909

[5] D. Velásquez, A. Sánchez, S. Sarmiento, M. Toro, M. Maiza, and B. Sierra, "A method for detecting coffee leaf rust through wireless sensor networks, remote sensing, and deep learning: case study of the Caturra variety in Colombia," *Appl. Sci.*, vol. 10, no. 2, p. 697, 2020. [Online]. Available: https://doi.org/10.3390/app10020697

[6] A. Barantsov, A. B. Pnev, K. I. Koshelev, V. S. Tynchenko, V. A. Nelyub, and A. S. Borodulin, "Classification of acoustic influences registered with phase-sensitive OTDR using pattern recognition methods," *Sensors*, vol. 23, no. 2, p. 582, 2023. [Online]. Available: https://doi.org/10.3390/s23020582

[7] H. S. Metha, "Classification of types of roasted coffee beans using convolutional neural network method," *Sinkron: J. dan Penelit. Tek. Inform.*, vol. 8, no. 2, Apr. 2024. [Online]. Available: https://doi.org/10.33395/v8i2.13517

[8] G. Sakdipat Ontoum, T. K. Thitaree, and P. Sroison, "Coffee Bean Dataset Resized (224 × 224)," *Kaggle*, [Online]. Available: https://www.kaggle.com/datasets/gpiosenka/coffee-bean-dataset-resized-224-x-224