

Comparative analysis of deep learning models for post-roasting coffee bean classification

Faiza Osama Abdalla Hashim¹, Boon Chin Yeo^{1*}, Akaraphunt Vongkunghae², Way Soong Lim¹, Jiraporn Pooksook², Kia Wai Liew¹, Jakir Hossen¹

¹Faculty of Engineering and Technology, Multimedia University, Melaka, Malaysia; bcyeo@mmu.edu.my (B.C.Y.).

²Faculty of Engineering, Naresuan University, Phitsanulok, Thailand.

Abstract: The classification of coffee beans is crucial for maintaining quality and consistency within the coffee industry. Manual inspection, however, is labor-intensive, error-prone, and susceptible to human biases. To address these challenges, this study aims to automate coffee bean classification using deep learning models to improve accuracy and efficiency. Four pre-trained models—Xception, ResNet50V2, EfficientNetB0, and VGG16—were evaluated for predicting post-roasting coffee bean quality based on two datasets: a Kaggle dataset and a self-collected dataset with an image scanner. The datasets included images of coffee beans at four roast levels: dark, green, light, and medium. The models were trained and tested using standard deep learning techniques, with performance assessed through metrics such as accuracy, precision, recall, and F1-score. The results demonstrated that Xception and EfficientNetB0 achieved the highest classification performance. On the Kaggle dataset, both models achieved 100% accuracy, while on the self-collected dataset, Xception achieved 99.3%, and EfficientNetB0 achieved 99.07%. These findings underscore the robustness of applying deep learning models in automating coffee quality control, reducing human intervention, and enhancing classification reliability.

Keywords: Coffee bean classification, Deep learning, EfficientNet, ResNet, VGG16, Xception.

1. Introduction

Coffee is one of the most consumed beverages globally, with over 2.25 billion cups consumed daily [1]. Its production has seen significant growth, highlighting its economic importance in the agricultural sector. Beyond its economic value, coffee offers notable health benefits, such as reducing the risk of cirrhosis and improving heart health [2]. These findings emphasize the importance of maintaining consistent coffee bean quality, as it directly impacts consumer satisfaction and producer profitability.

Deep learning has gained significant attention in the agricultural sector, particularly for tasks like coffee bean quality prediction. Recent studies have explored the use of deep learning models to automate this process, achieving promising results. This section reviews the most relevant studies that have employed CNN, ResNet, EfficientNet, and VGG for coffee bean quality prediction.

CNNs have been the most widely used deep learning models for coffee bean quality prediction due to their ability to automatically extract features from images [3]. Several studies have demonstrated the effectiveness of CNNs in tasks such as defect detection, roast level classification, and quality categorization. For example, a study by Lee and Jeong [4] in 2022 used a CNN model to predict defects in coffee beans, achieving an accuracy of 93% for normal beans and 81% for defective beans, although the dataset was imbalanced. Another study by Naik and Sethy [5] applied a CNN with a Euclidean Distance Algorithm to classify roasted coffee beans into four roasting levels, achieving an impressive accuracy of 97.5%. CNNs have also been used for categorizing coffee beans into different quality

categories, such as sour, black, broken, moldy, and insect-damaged beans, with accuracies ranging from 88% to 98% [6].

Despite their high accuracy, most CNN-based studies focus on the pre-roasting phase, particularly on identifying defects in green coffee beans [7-23]. This indicates a research gap in applying CNNs to post-roasting quality analysis, where the quality of roasted beans could be further evaluated. Recently, Xception has gained researchers' attention in coffee bean species classification [24].

EfficientNet, a family of scalable and efficient convolutional neural networks, has demonstrated remarkable performance in coffee bean quality prediction tasks. The EfficientNetV2S variant, in particular, achieved extremely high accuracy in defect detection, with accuracies ranging from 96% to 99.77% for detecting insect-infested, broken, and moldy beans [12]. EfficientNet also performed well in roast level classification, achieving 95.79% accuracy in a 2024 study [11]. However, the model's performance in measuring acidity from roasted coffee beans has an F1 score of 0.71 only [16].

ResNets, known for their skip connections that enable the training of very deep networks, have also been applied to coffee bean quality prediction. ResNets have shown high accuracy in defect detection tasks. For instance, ResNet50V2 achieved 99.54% accuracy in detecting insect-infested beans, 98.86% for broken beans, and 98.07% for moldy beans [12].

However, ResNets have struggled with roast level classification tasks. A 2024 study by Hassan [11] using ResNet50 to classify coffee beans based on four degrees of roasting achieved only 58.05% accuracy, highlighting a significant research gap in this area. This suggests that while ResNets are highly effective for defect detection, further improvements are needed for roast level classification.

VGG models, known for their simplicity and effectiveness in image classification tasks, have also been applied to coffee bean quality prediction. VGG models have shown exceptional performance in roast level classification, with one study achieving 100% accuracy in classifying coffee beans based on four degrees of roasting [11]. VGG16 has also been used for defect detection in green coffee beans, achieving accuracies ranging from 81% to 97% [10-20]. However, most studies using VGG models have focused on pre-roasting applications, leaving a gap in research for post-roasting quality analysis.

The review of the different deep learning models demonstrates their effectiveness in various tasks related to coffee bean quality prediction. CNNs and ResNets have shown high accuracy in defect detection and quality categorization, while EfficientNet and VGG have excelled in roast level classification and defect detection.

However, there are still research gaps, particularly in the post-roasting phase, where further improvements and applications of these models could be explored. Future research could focus on enhancing the performance of these models in roast level classification and extending their application to post-roasting quality analysis.

2. Methodology

This study aims to evaluate and compare the performance of four deep learning models: Xception, EfficientNetB0, ResNet50V2, and VGG16, for predicting coffee bean quality based on roast levels. The methodology involves three key phases: dataset preparation, model implementation, and performance evaluation. Two datasets were utilized to train and test the models, focusing on four roast levels: green, light, medium, and dark.

2.1. Datasets

Besides the public available Kaggle dataset, a separate dataset has been developed and used in this study. Kaggle Dataset: This dataset consists of 1,600 images of coffee beans at four roast levels. The images were captured using an iPhone 12 with a resolution of 3024×3032 pixels and resized to 224×224 pixels for consistency. The dataset was divided into 80% for training and validation and 20% for testing [25].

Self-Collected Dataset: This dataset contains 2,140 images of coffee beans at four roast levels. The processes of collecting this dataset involve several steps as shown in Figure 1.

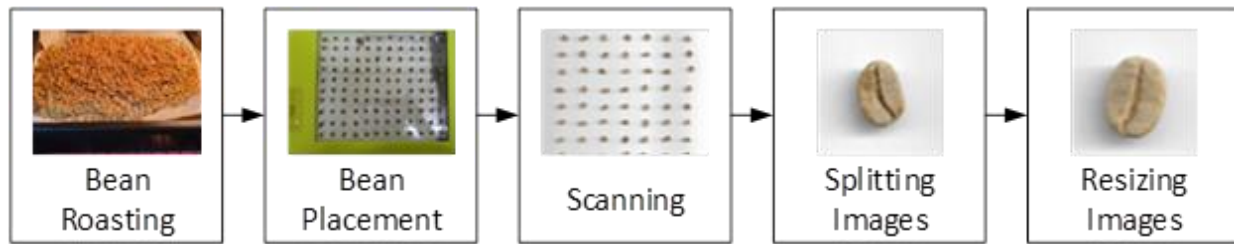





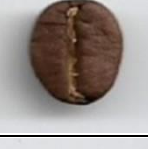

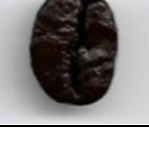


Figure 1.
Data Collection Process.

Firstly, the coffee beans were roasted using an oven to achieve three desired roast levels: light, medium, and dark. The roasting duration was carefully controlled to ensure consistency across batches.

After roasting, the coffee beans were arranged systematically on a flatbed scanner to capture high-resolution images [24, 25]. The scanned images were processed using an online image editing tool, such as ImagesTool, to separate individual coffee beans efficiently. This tool significantly accelerated the image cropping process. The individual images were resized to 224×224 pixels using a Python script to ensure uniform input dimensions for deep learning models. Finally, the dataset was split into 80% for training and validation and 20% for testing. The sample images for each roast level in both datasets are shown in Table .

Table 1.
Samples of Datasets.

Roast Level	Kaggle Dataset		Self-Collected Dataset	
Green				
Light				
Medium				
Dark				

2.2. Deep Learning Models

The implementation of deep learning models was conducted using TensorFlow and Keras. Each model was trained and evaluated on both datasets, with hyperparameters selected based on experimental results to achieve the highest accuracy. Data augmentations such as horizontal flipping and rotation are

applied to the datasets. The following sections summarize the architecture and training parameters for the deep learning models.

2.2.1. Xception

The Xception model, which stands for Extreme Inception, represents an advanced deep convolutional neural network architecture that enhances the original Inception framework by substituting conventional Inception modules with depthwise separable convolutions. Introduced by François Chollet, the architect of Keras, the Xception model aims to optimize computational efficiency and enhance model performance by disentangling spatial and cross-channel dependencies. In contrast to traditional convolutional layers, depthwise separable convolutions execute filtering and feature integration in distinct phases, thereby substantially decreasing the parameter count while preserving a high level of accuracy. This architectural design enables the network to capture more intricate patterns with reduced computational expenditure. In the present study, the Xception model was utilized as the feature extractor within the customized CNN framework. A pre-trained Xception base, previously trained on the ImageNet dataset, was incorporated along with supplementary fully connected layers and dropout mechanisms for the classification of coffee bean roast levels. This integrative strategy harnesses the formidable feature extraction capabilities of Xception while tailoring the model to the specific objective of post-roasting coffee bean classification. The model's layer configuration is illustrated in Figure . The model was trained for 10 epochs using a batch size of 16 and a learning rate of 0.001. Early stopping with a patience of 3 was implemented to mitigate overfitting.

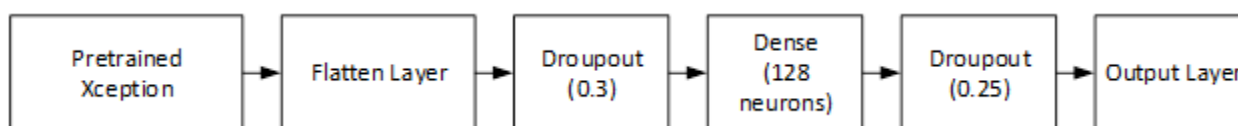


Figure 2.
Xception Layer Configuration.

2.2.2. EfficientNetB0

EfficientNetB0 serves as the foundational model within the EfficientNet architecture, a category of convolutional neural networks recognized for achieving elevated accuracy with a reduced number of parameters and computational demands. Originating from Google AI, EfficientNetB0 presents an innovative compound scaling methodology that systematically scales the network's depth, width, and input resolution through the application of a compound coefficient. This harmonized scaling approach enables the model to enhance performance optimization more proficiently compared to conventional scaling techniques. EfficientNetB0 is constructed utilizing mobile inverted bottleneck convolution (MBConv) blocks and incorporates squeeze-and-excitation optimization to recalibrate channel-wise feature responses, thereby augmenting its representational capability. In the context of this investigation, EfficientNetB0 was employed as a pre-trained foundational model, augmented with supplementary dense and dropout layers specifically designed for the four-class classification of coffee bean roast levels. The model's structure is depicted in Figure . The model was trained for 50 epochs with a batch size of 16 and an initial learning rate of 0.0001. A learning rate scheduler (ReduceLROnPlateau) was used to dynamically adjust the learning rate during training.

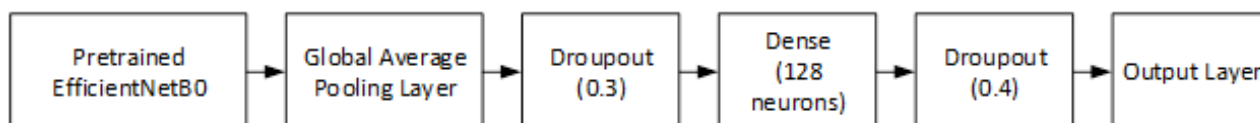


Figure 3.
EfficientNetB0 Layer Configuration.

2.2.3. ResNet50V2

ResNet50V2 represents a modification of the Residual Network (ResNet) framework, specifically developed to mitigate the difficulties associated with the training of extensive neural networks, most notably the vanishing gradient phenomenon. In contrast to conventional convolutional architectures, ResNet integrates residual connections, which permit the input data to circumvent one or more layers, thereby empowering the network to acquire identity mappings and sustain the gradient flow throughout the backpropagation process. This architectural configuration enables the training of significantly deeper models without incurring performance deterioration. ResNet50V2 enhances the foundational ResNet50 by reorganizing the elements within each residual block. It implements batch normalization and ReLU activation prior to each convolutional layer, a configuration referred to as pre-activation, which significantly bolsters training stability and generalization capabilities. In the present investigation, the ResNet50V2 architecture was employed as a pre-trained feature extractor, supplemented by custom classification layers to execute a four-class classification task pertaining to coffee bean roast levels. Its substantial depth and residual framework render it particularly adept at discerning intricate patterns and features within high-resolution imagery. The layer configuration is presented in Figure . The model was trained for 50 epochs with different batch sizes: 32 for the Kaggle dataset and 16 for the self-collected dataset. Early stopping was applied with a patience of 3 for the Kaggle dataset and 6 for the self-collected dataset. Early stopping was applied with a patience of 3 for the Kaggle dataset and 6 for the self-collected dataset.

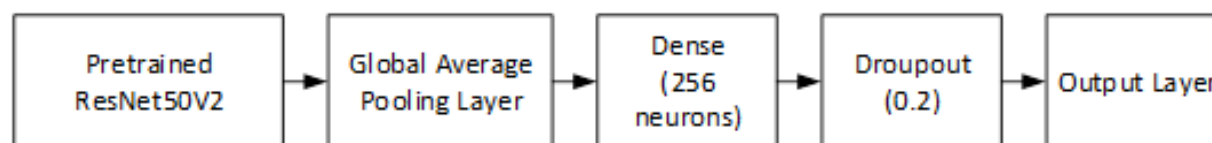


Figure 4.
ResNet50V2 Layer Configuration.

2.2.4. VGG16

VGG16 represents a sophisticated deep convolutional neural network architecture that was conceived by the Visual Geometry Group at the University of Oxford. Its acclaim stems from its methodological simplicity, uniform structural design, and proven efficacy in tasks pertaining to image classification. The architecture of VGG16 is comprised of 16 layers, which include 13 convolutional layers and 3 fully connected layers, employing diminutive 3×3 convolution filters uniformly across the network. This persistent application of small filters facilitates the model's capacity to capture intricate features while simultaneously preserving a manageable parameter count. Each convolutional block is succeeded by a max-pooling layer, which serves to diminish spatial dimensions and enhance computational efficiency. In the context of this study, VGG16 was utilized as a pre-trained model, with supplementary dense and dropout layers incorporated for the classification of coffee beans into four distinct roast levels. Although VGG16 is comparatively larger and less computationally efficient than more contemporary models such as EfficientNet, it continues to be extensively employed due to its consistent performance and straightforward architecture, thereby establishing it as a valuable benchmark for assessing model efficacy in the prediction of coffee bean quality. The VGG16 model was pre-trained and modified with global average pooling, dense, and dropout layers for classification. The model's architecture is outlined in Figure 2. The VGG16 model was trained for 100 epochs with a batch size of 16. Early stopping was applied with a patience of 5.

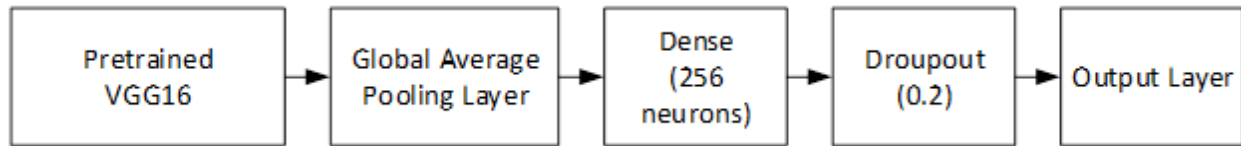


Figure 2.
VGG16 Layer Configuration.

3. Results

This section presents the results of the four deep learning models Xception, ResNet50V2, EfficientNetB0, and VGG16 for predicting coffee bean quality based on roast levels. The models were evaluated using two datasets: the Kaggle dataset and the self-collected dataset. Performance metrics such as accuracy, precision, recall, and F1-score were used to assess the models. Additionally, confusion matrices were generated to analyze classification behavior.

3.1. Xception

The Xception-based CNN model exhibited exceptional efficacy in the classification of post-roasting coffee beans, utilizing both the Kaggle dataset and a dataset collected independently. Within the Kaggle dataset, the model attained an impeccable classification accuracy of 100%, successfully predicting all images across the four defined roast levels: green, light, medium, and dark. The confusion matrix is shown in Figure 6. The assessment metrics, encompassing precision, recall, and F1-score, were uniformly recorded at 1.00 for each class, signifying an unblemished performance as shown in Table . The analysis of the confusion matrix further substantiated the absence of misclassifications within the test dataset.

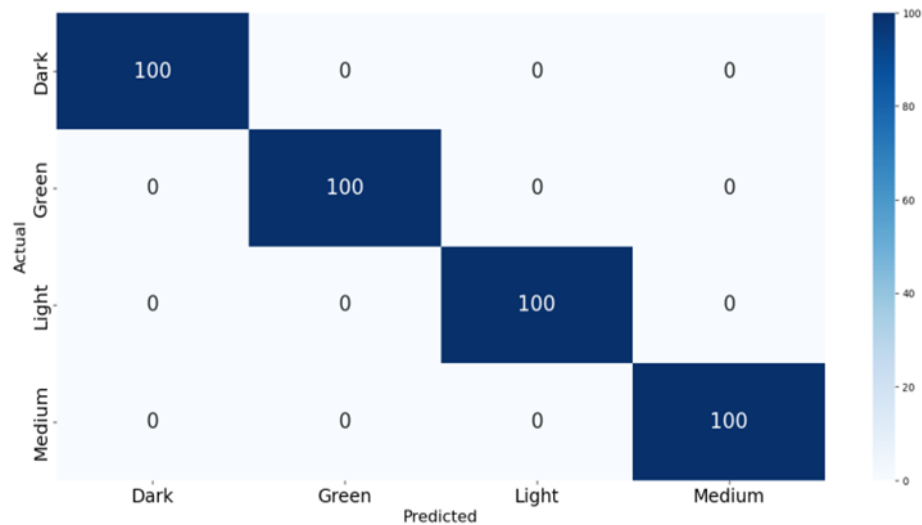


Figure 6.
Confusion Matrix for Xception with Kaggle Dataset.

Table 2.
Classification Report for Xception with Kaggle Dataset.

Class	Precision	Recall	F1-Score	Support
Dark	1.00	1.00	1.00	100
Green	1.00	1.00	1.00	100
Light	1.00	1.00	1.00	100
Medium	1.00	1.00	1.00	100
Accuracy	-	-	1.00	400
Macro Avg	1.00	1.00	1.00	400
Weighted Avg	1.00	1.00	1.00	400

When evaluated on the self-collected dataset, which encompassed increased variability in terms of image acquisition and background conditions, the model continued to exhibit commendable performance. It secured an accuracy rate of 99.3%, with precision and recall values ranging between 0.97 and 1.00 across the various roast classifications. The classification report in Table shows the scores for precision, recall, and F1-score across all roast levels. The confusion matrix in Figure shows that a limited number of images were misclassified, notably within the medium roast category, where three instances were inaccurately predicted as either light or dark. Notwithstanding this slight diminution in performance, the model illustrated robust generalization capabilities across datasets with distinct characteristics. These findings underscore the Xception-based CNN's efficacy and resilience in the automation of post-roasting coffee bean classification.

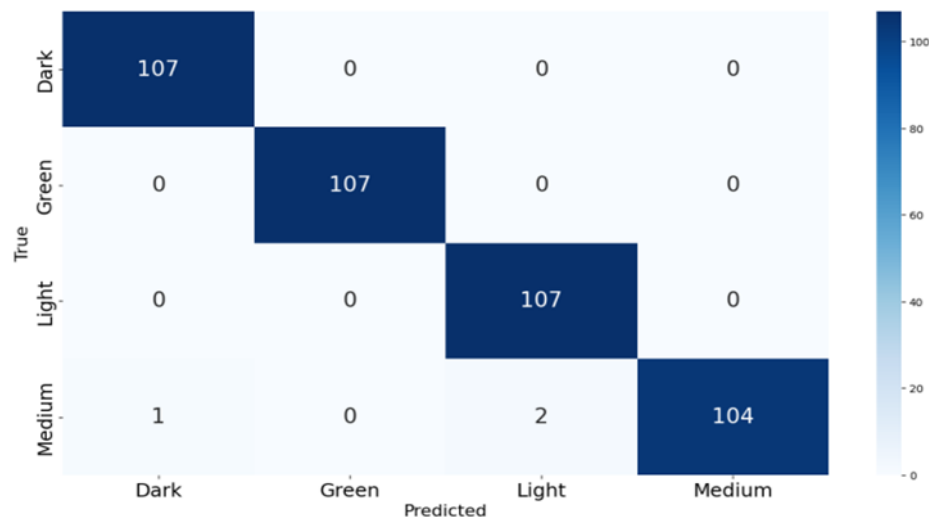


Figure 7.
Confusion Matrix for Xception with Self-Collected Dataset.

Table 3.
Classification Report for Xception with Self-Collected Dataset.

Class	Precision	Recall	F1-Score	Support
Dark	0.99	1.00	1.00	107
Green	1.00	1.00	1.00	107
Light	0.98	1.00	0.99	107
Medium	1.00	0.97	0.99	107
Accuracy	-	-	0.99	428
Macro Avg	0.99	0.99	0.99	428
Weighted Avg	0.99	0.99	0.99	428

3.2. EfficientNetB0

The EfficientNetB0 model exhibited exceptional efficacy in the classification of post-roasting coffee bean imagery. Within the Kaggle dataset, the model secured a flawless accuracy rate of 100%, with every test instance accurately categorized into the four distinct roast classifications: green, light, medium, and dark. The confusion matrix is shown in Figure . From the classification report shown in Table , the precision, recall, and F1-score for each category were uniformly recorded at 1.00, thereby signifying that the model has impeccably assimilated the distributional and visual attributes of the dataset. The confusion matrix indicated an absence of misclassifications, while the accuracy and loss trajectories illustrated a consistent and stable learning trajectory throughout the training phase.

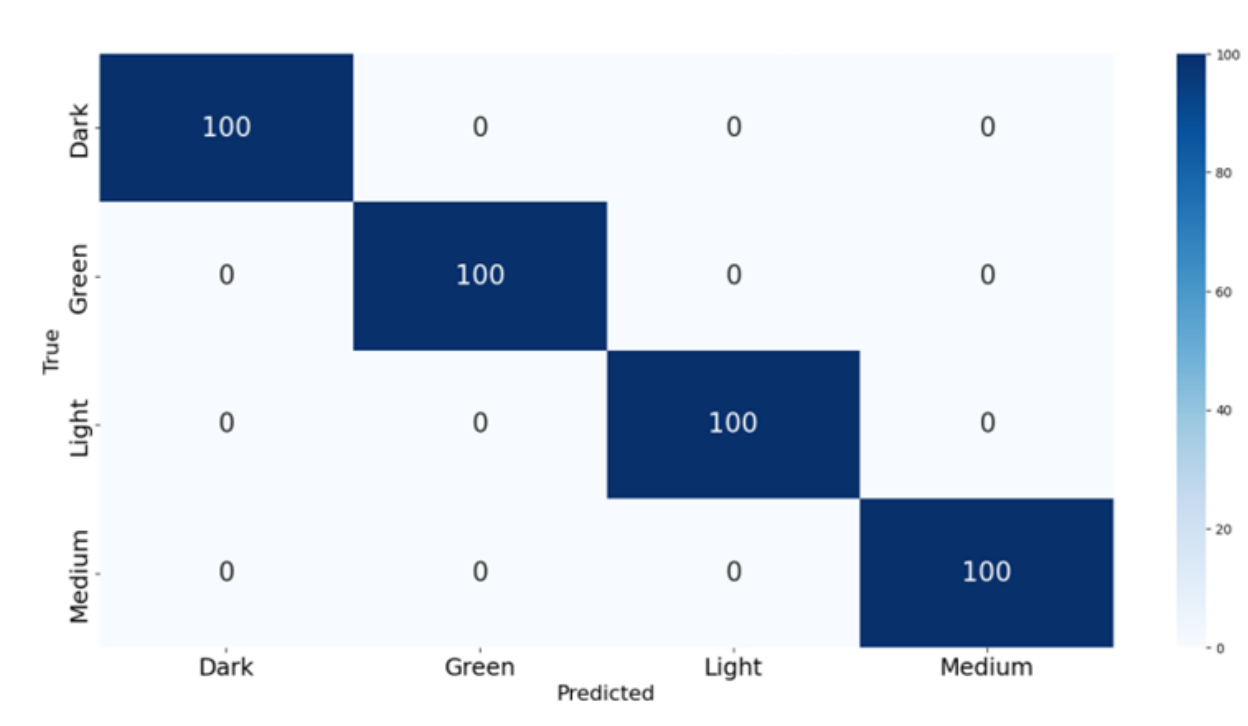


Figure 8.
Confusion Matrix for EfficientNetB0 with Kaggle Dataset.

Table 4.
Classification Report for EfficientNetB0 with Kaggle Dataset.

Class	Precision	Recall	F1-Score	Support
Dark	1.00	1.00	1.00	100
Green	1.00	1.00	1.00	100
Light	1.00	1.00	1.00	100
Medium	1.00	1.00	1.00	100
Accuracy	-	-	1.00	400
Macro Avg	1.00	1.00	1.00	400
Weighted Avg	1.00	1.00	1.00	400

Upon assessment with the self-collected dataset, which comprised images procured via flatbed scanning under meticulously controlled lighting conditions, the EfficientNetB0 model sustained its robust performance. It achieved a classification accuracy rate of 99.07%, with the majority of classes displaying near-optimal precision and recall metrics. Confusion matrix in Figure shows that a minimal number of misclassifications were noted, primarily involving medium roast beans being erroneously

categorized alongside adjacent roast classifications. Based on the classification report in Table , the model demonstrated impeccable precision, recall, and F1-score of 1.00 for the green roast category, whereas the light and dark roast classifications also exhibited commendable outcomes with F1-scores of 0.99. The medium roast classification, however, exhibited a marginally diminished performance, characterized by a precision of 0.98, a recall of 0.98, and an F1-score of 0.98, attributable to minor misclassifications involving adjacent roast categories. Notwithstanding the slight decline in accuracy relative to the Kaggle dataset, the model preserved a commendable degree of generalization, thus exhibiting its resilience to variations in image quality and acquisition methodologies. These findings substantiate that EfficientNetB0 constitutes a highly dependable and efficient framework for the automation of roast-level classification of coffee beans across varying contexts.

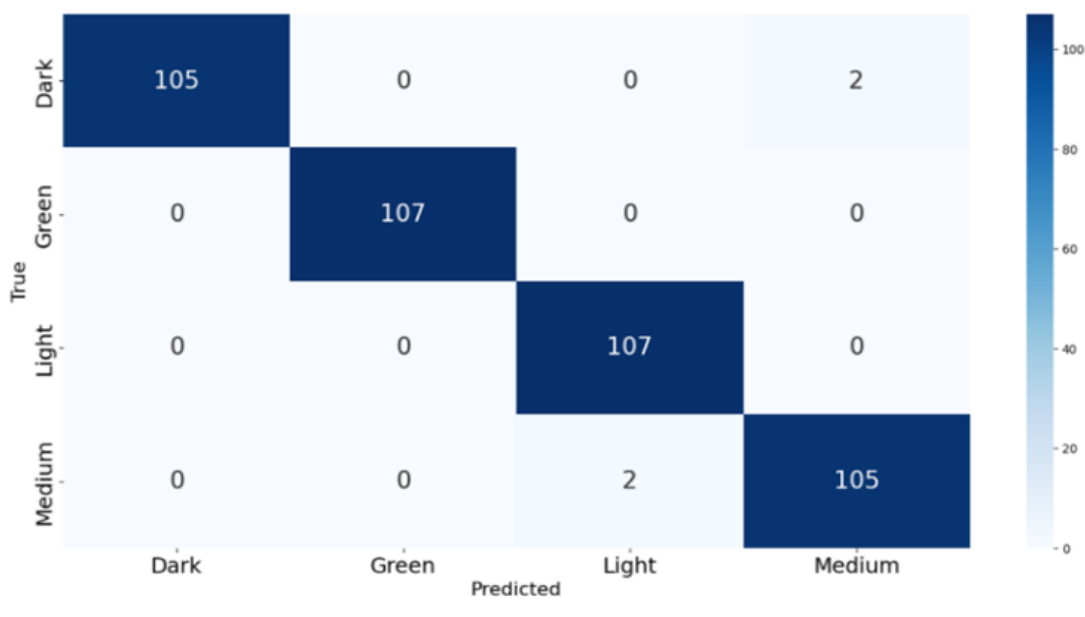


Figure 9.
Confusion Matrix for EfficientNetB0 with Self-Collected Dataset.

Table 5.
Classification Report for EfficientNetB0 with Self-Collected Dataset.

Class	Precision	Recall	F1-Score	Support
Dark	1.00	0.98	0.99	107
Green	1.00	1.00	1.00	107
Light	0.98	1.00	0.99	107
Medium	0.98	0.98	0.98	107
Accuracy	-	-	0.99	428
Macro Avg	0.99	0.99	0.99	428
Weighted Avg	0.99	0.99	0.99	428

3.3. Residual Networks (ResNet)

In the Kaggle dataset, the ResNet50V2 architecture exhibited remarkable efficacy, attaining an overall classification accuracy of 99.25%, which is closely aligned with the performance exhibited by the Xception-based CNN and the EfficientNetB0 models, both of which reached a perfect accuracy of 100%. Figure shows the confusion matrix, and Table presents the classification report. The ResNet50V2 model achieved near-optimal precision, recall, and F1-scores across all four levels of roast, with only 3 instances of misclassification recorded. The confusion matrix corroborated that this model was

exceptionally adept at discerning the nuanced color and texture distinctions between green, light, medium, and dark roasted coffee beans. Although ResNet50V2 fell slightly short of the impeccable predictions demonstrated by Xception and EfficientNetB0, the disparity observed was minimal and statistically negligible for practical applications. Nevertheless, the marginally lower accuracy implies that ResNet50V2 may exhibit heightened sensitivity to finely delineated class boundaries, particularly in contrast to EfficientNetB0, which incorporates compound scaling to achieve a harmonious balance of depth and resolution, or Xception, which employs depthwise separable convolutions to facilitate efficient feature extraction. In general, ResNet50V2 remains a robust candidate for tasks related to roast-level classification, particularly when utilized with structured datasets characterized by minimal variability.

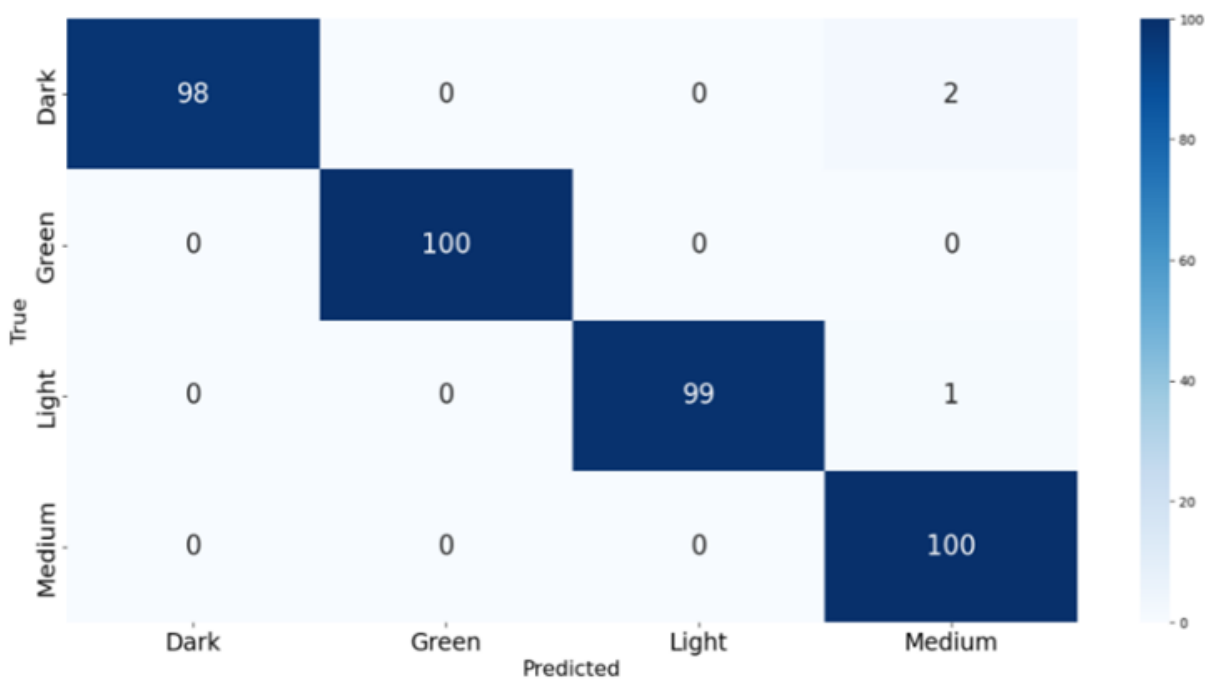


Figure 10.
Confusion Matrix for ResNet50V2 with Kaggle Dataset.

Table 6.
Classification Report for ResNet50V2 with Kaggle Dataset.

Class	Precision	Recall	F1-Score	Support
Dark	1.00	0.98	0.99	100
Green	1.00	1.00	1.00	100
Light	1.00	0.99	0.99	100
Medium	0.97	1.00	0.99	100
Accuracy	-	-	0.99	400
Macro Avg.	0.99	0.99	0.99	400
Weighted Avg.	0.99	0.99	0.99	400

When evaluated utilizing the self-collected dataset, the ResNet50V2 model attained a diminished classification accuracy of 95.09%, signifying a substantial decrease relative to its 99.25% accuracy on the Kaggle dataset. Table 7 shows the classification reports. Although the model preserved a commendable overall performance, the precision, recall, and F1-scores for specific classes, particularly the medium and light roast category, experienced a decline. The confusion matrix in Figure 11 disclosed that the

majority of misclassifications involved medium roast beans being erroneously categorized as either light or dark roasts, likely attributable to the visual resemblances between neighboring roast categories. Meanwhile, a number of misclassifications also involve light roast beans being erroneously classified as medium and green classes. This reduction accentuates ResNet50V2's susceptibility to data variability, as the self-collected dataset encompassed diverse imaging conditions, including scanning as opposed to conventional photography and a broader range of bean appearances. In contrast to its performance on the well-structured and uniformly illuminated Kaggle dataset, ResNet50V2 exhibited challenges in sustaining an equivalent level of generalization with less standardized inputs. This disparity indicates that, as opposed to EfficientNetB0 and Xception, which maintained high accuracy across both datasets, ResNet50V2 may necessitate supplementary preprocessing or fine-tuning to adapt proficiently to real-world variations in image data.

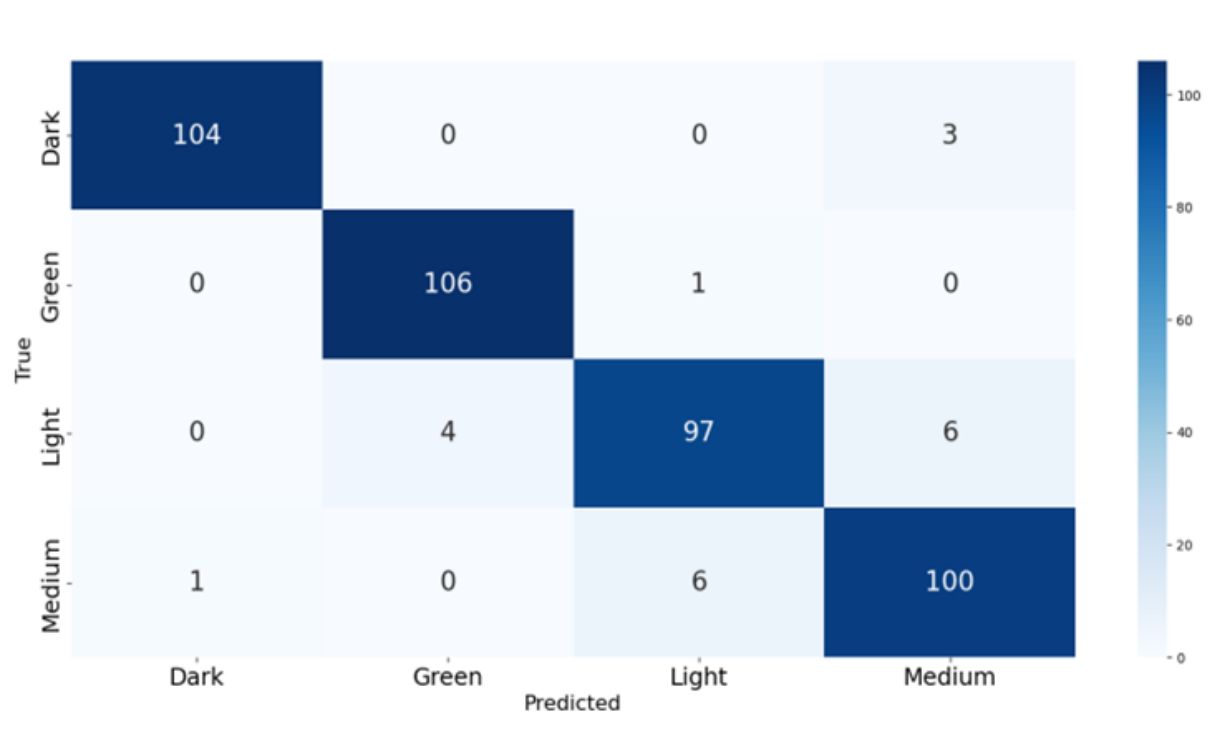


Figure 11.
Confusion Matrix of ResNet50V2 with Self-Collected Dataset.

Table 7.
Classification Report for ResNet50V2 with Self-Collected Dataset.

Class	Precision	Recall	F1-Score	Support
Dark	0.99	0.97	0.98	107
Green	0.96	0.99	0.98	107
Light	0.93	0.91	0.92	107
Medium	0.92	0.93	0.93	107
Accuracy	-	-	0.95	428
Macro Avg	0.95	0.95	0.95	428
Weighted Avg	0.95	0.95	0.95	428

3.4. Visual Geometry Group (VGG16)

The VGG16 architecture attained an admirable accuracy rate of 95.5% on the Kaggle dataset, effectively categorizing the predominant portion of coffee bean images across the four distinct roasting

levels. The confusion matrix and classification report for VGG16 are shown in Figure 12 and Table 8, respectively. Although this level of performance suggests that VGG16 possesses the capability to undertake the classification task, it fell short compared to the other models analyzed in this research. In contrast, both Xception and EfficientNetB0 achieved an impeccable accuracy of 100%, whereas ResNet50V2 recorded an accuracy of 99.25%, thereby significantly surpassing VGG16 in terms of overall metrics such as precision, recall, and the F1-score. The analysis of the confusion matrix for VGG16 uncovered numerous misclassifications, especially within the medium and dark roast categories, where the visual distinctions tend to be nuanced and necessitate a more sophisticated feature extraction process. This disparity in performance can be ascribed to VGG16's comparatively antiquated architecture, which does not incorporate the depth and efficiency provided by the residual connections found in ResNet, nor the compound scaling and separable convolutions utilized by EfficientNetB0 and Xception, respectively. Nonetheless, while VGG16 has demonstrated itself to be a dependable and straightforward model for image classification tasks, its shortcomings become evident in nuanced applications such as roast-level classification, wherein more contemporary architectures exhibit enhanced accuracy and generalization capabilities.

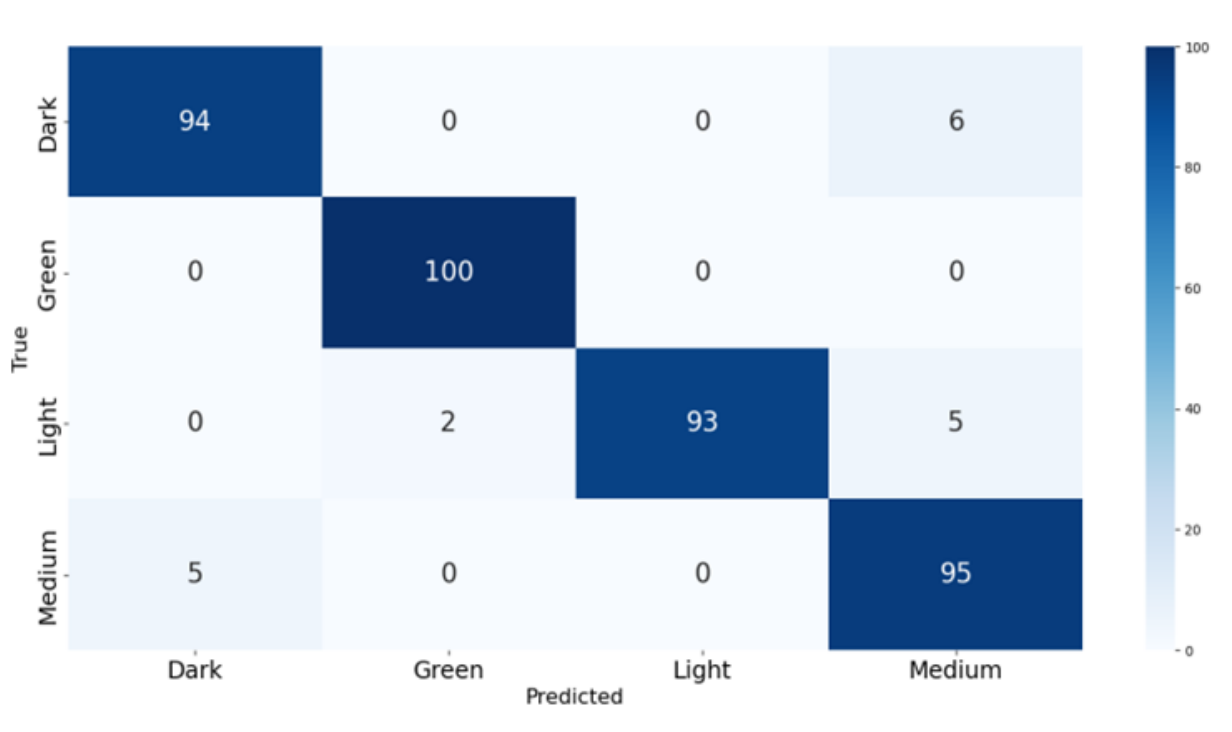


Figure 12.
Confusion Matrix for VGG16 with Kaggle Dataset.

Table 8.
Classification Report for VGG16 with Kaggle Dataset.

Class	Precision	Recall	F1-Score	Support
Dark	0.95	0.94	0.94	100
Green	0.98	1.00	0.99	100
Light	1.00	0.93	0.96	100
Medium	0.90	0.95	0.92	100
Accuracy	-	-	0.95	400
Macro Avg	0.96	0.96	0.96	400
Weighted Avg	0.96	0.95	0.96	400

Upon assessment utilizing the self-compiled dataset, the VGG16 model attained a classification accuracy of 92.53%, which is marginally inferior to its efficacy on the Kaggle dataset, where it achieved an accuracy of 95.5%. This decline in accuracy underscores the model's constrained capacity to generalize across datasets that exhibit disparate characteristics. The self-compiled dataset introduced increased variability in the appearance of beans, background texture, and the method of image acquisition (scanner-based imaging), which likely posed challenges to VGG16's relatively shallow and static architecture. The confusion matrix in Figure 13 indicated a heightened incidence of misclassifications, particularly among medium roast beans, which were frequently misidentified as light or dark roasts due to their visual similarities. In comparison to its performance on the more uniform Kaggle dataset, VGG16 demonstrated heightened sensitivity to variations in imagery and encountered difficulties in sustaining consistent precision and recall across all classifications. These results substantiate the premise that while VGG16 may exhibit commendable performance under controlled circumstances, it lacks the adaptability and resilience of more sophisticated models such as Xception and EfficientNetB0, both of which have exhibited superior generalization capabilities and elevated accuracy on the self-compiled dataset. The classification report in Table 9 shows lower precision, recall, and F1-score values for VGG16, when comparing to the performance results of other models.

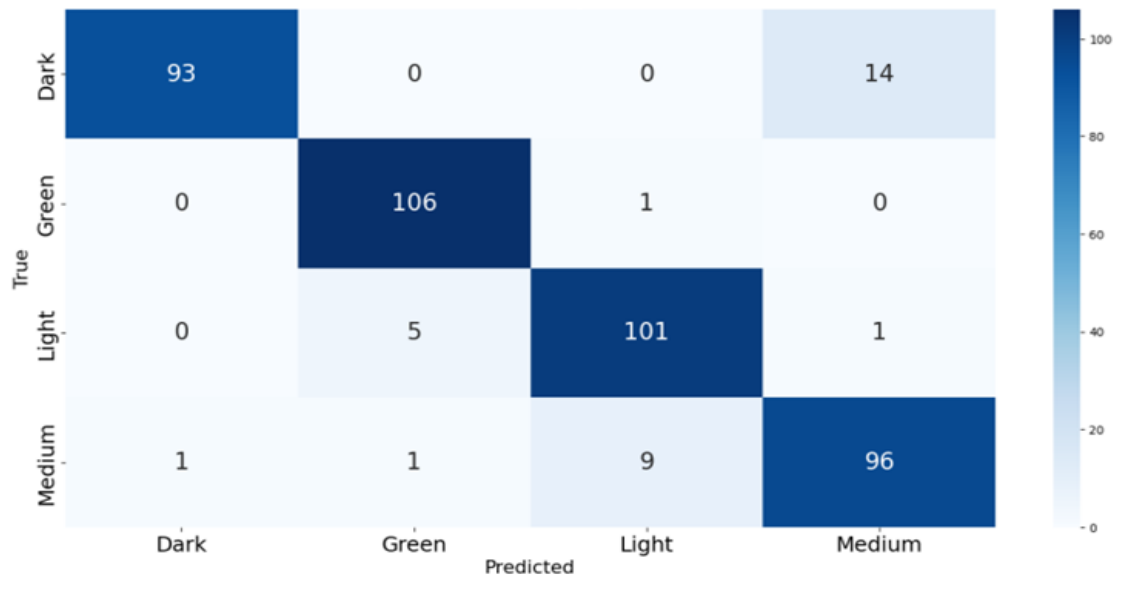


Figure 13.
Confusion Matrix for VGG16 with Self-Collected Dataset.

Table 9.
Classification Report for VGG16 with Self-Collected Dataset.

Class	Precision	Recall	F1-Score	Support
Dark	0.99	0.87	0.93	107
Green	0.95	0.99	0.97	107
Light	0.91	0.94	0.93	107
Medium	0.86	0.90	0.88	107
Accuracy	-	-	0.93	428
Macro Avg	0.93	0.93	0.93	428
Weighted Avg	0.93	0.93	0.93	428

4. Comparative Analysis

The comparative analysis of the performance of four deep learning architectures, particularly Xception, EfficientNetB0, ResNet50V2, and VGG16, demonstrates significant variations in their classification efficacy when assessed using both the Kaggle dataset and the self-acquired dataset. A comparative analysis of the four models is presented in Table 10, summarizing their performance on both datasets.

Table 10.
Performance Summary for Different Models.

Model	Dataset	Accuracy	Precision	Recall	F1-Score	Misclassification
Xception	Kaggle	100.00%	1.00	1.00	1.00	0
	Self-Collected	99.30%	0.99	0.99	0.99	3
EfficientNetB0	Kaggle	100.00%	1.00	1.00	1.00	0
	Self-Collected	99.07%	0.99	0.99	0.99	4
ResNet50V2	Kaggle	99.25%	0.99	0.99	0.99	3
	Self-Collected	95.09%	0.95	0.95	0.95	21
VGG16	Kaggle	95.50%	0.96	0.96	0.96	18
	Self-Collected	92.53%	0.93	0.93	0.93	33

In the context of the Kaggle dataset, characterized by high-resolution, uniformly illuminated images captured via smartphone under controlled conditions, both Xception and EfficientNetB0 attained a flawless accuracy of 100%, exhibiting perfect precision, recall, and F1-scores across all four roast classifications: green, light, medium, and dark. These findings signify impeccable classification performance, with no misclassifications discerned within the confusion matrices.

ResNet50V2 demonstrated a commendable performance, achieving a marginally lower accuracy of 99.25%. The model misclassified merely a single instance, reflecting commendable, albeit not flawless, performance. It sustained high precision and recall metrics across all categories; however, it exhibited a slight deficiency in capturing the nuanced distinctions between visually similar roast levels when juxtaposed with Xception and EfficientNetB0. This minor disparity suggests that while ResNet50V2 is proficient, its architectural design may not be as finely tuned for this specific application as the compound scaling of EfficientNetB0 or the depthwise separable convolutions utilized by Xception.

Conversely, VGG16 exhibited a markedly reduced accuracy of 95.5%. A number of misclassifications were noted, particularly within the medium and dark roast classifications, where the differences in color and texture are subtle. This diminished performance underscores the inherent limitations associated with VGG16's antiquated architecture, which is deficient in depth, residual connections, and parameter efficiency when compared to the other models. Although still functional, VGG16 did not perform at the same caliber as the more sophisticated networks on this rigorously structured dataset.

When evaluated on the self-collected dataset, which introduced increased variability regarding image acquisition (utilizing a flatbed scanner), background uniformity, and real-world noise, the models exhibited significantly more marked discrepancies in their generalization capabilities.

Xception and EfficientNetB0 consistently demonstrated robust performance, with Xception attaining an accuracy of 99.3% and EfficientNetB0 achieving 99.07%. Both models sustained elevated levels of precision, recall, and F1-scores across all roast classifications, although a limited number of misclassifications were noted, predominantly within the medium roast category, which was occasionally misidentified as light or dark roasts. Notwithstanding this minor deterioration from the flawless outcomes observed on the Kaggle dataset, the performance continued to be remarkably dependable, underscoring the models' resilience to novel data conditions.

Conversely, ResNet50V2 encountered a more substantial decline in accuracy, reducing to 95.09%. While it still demonstrated satisfactory performance, the frequency of misclassifications rose, particularly within the medium roast category. This reduction indicates that ResNet50V2 exhibits heightened sensitivity to fluctuations in imaging conditions and may necessitate additional data

augmentation or fine-tuning to adapt proficiently. The disparity between its performance on the Kaggle dataset and that on the self-collected dataset highlights its dependence on consistent feature representations, which were more readily captured in the standardized Kaggle dataset.

VGG16 exhibited the most pronounced decline, with its accuracy diminishing to 92.53% on the self-collected dataset. The confusion matrix illustrated a more extensive distribution of errors, with misclassification of roast levels occurring with greater frequency than observed in the Kaggle evaluation. This finding further accentuates VGG16's deficiencies in generalization, particularly in real-world contexts where image quality and conditions exhibit greater variability. Its comparatively shallow architecture and dependence on fixed convolutional blocks render it less adaptable in comparison to other more contemporary and deeper neural networks.

In general, both Xception and EfficientNetB0 emerged as the foremost performers across the two datasets, exhibiting not only remarkable accuracy but also significant generalization capabilities across varied image sources. ResNet50V2 demonstrated commendable performance on the Kaggle dataset; however, it exhibited greater difficulty when confronted with variability in the self-collected dataset, thereby suggesting certain limitations in its robustness. VGG16, although providing essential functionality, consistently attained the lowest rankings in both contexts, thereby underscoring the performance disparity between traditional architectures and contemporary deep learning models. These observations imply that for practical implementation in the classification of post-roasting coffee beans, models such as EfficientNetB0 and Xception are preferable due to their adaptability, precision, and efficiency.

5. Conclusion

This research investigated the efficacy of four advanced deep learning architectures, namely Xception, EfficientNetB0, ResNet50V2, and VGG16, in the classification of post-roasting coffee beans into four discrete roast categories: green, light, medium, and dark. The evaluation of the models was conducted using two datasets: a publicly accessible Kaggle dataset and a self-collected dataset reflecting real-world variabilities. The findings revealed that both Xception and EfficientNetB0 attained superior performance across both datasets, achieving perfect accuracy on the Kaggle dataset and exceeding 99 percent accuracy on the self-compiled dataset. Furthermore, these models exhibited robust generalization capabilities, rendering them exceptionally suitable for practical implementation in automated systems for coffee quality assessment.

ResNet50V2 also demonstrated commendable performance, particularly on the Kaggle dataset; however, it displayed a more pronounced decline in accuracy when assessed on the self-collected dataset, thereby indicating a necessity for enhanced optimization in variable environments. VGG16, although operationally adequate under controlled settings, consistently yielded lower accuracy and reflected limited robustness, thereby underscoring the architectural constraints of older convolutional networks in fine-grained classification endeavors.

In summary, this research substantiates that contemporary deep learning architectures, particularly Xception and EfficientNetB0, possess the potential to significantly augment both the accuracy and efficiency of coffee bean roast-level classification. Their application could diminish dependence on manual inspection, enhance quality control methodologies, and foster greater consistency in coffee production. Future research may entail the expansion of the dataset, the inclusion of supplementary roast attributes such as texture and surface imperfections, and the deployment of these models within real-time quality monitoring frameworks.

Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Acknowledgment:

The researchers at Multimedia University (MMU) would like to express their deepest gratitude for the invaluable support provided through the Matching Grant Collaboration between MMU and Naresuan University (NU). This generous initiative has played a crucial role in facilitating the research and development efforts that have led to the successful completion of this publication.

Copyright:

© 2025 by the authors. This open-access article is distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

References

- [1] C. Hou *et al.*, "Medical conditions associated with coffee consumption: Disease-trajectory and comorbidity network analyses of a prospective cohort study in UK Biobank," *The American Journal of Clinical Nutrition*, vol. 116, no. 3, pp. 730-740, 2022. <https://doi.org/10.1093/ajcn/nqac148>
- [2] C.-H. Hsia, Y.-H. Lee, and C.-F. Lai, "An explainable and lightweight deep convolutional neural network for quality detection of green coffee beans," *Applied Sciences*, vol. 12, no. 21, p. 10966, 2022. <https://doi.org/10.3390/app122110966>
- [3] G. Vilcamiza, N. Trelles, L. Vincés, and J. Oliden, "A coffee bean classifier system by roast quality using convolutional neural networks and computer vision implemented in an NVIDIA Jetson Nano," in *2022 Congreso Internacional de Innovación y Tendencias en Ingeniería (CONITI)*, 2022.
- [4] J. Y. Lee and Y. S. Jeong, "Prediction of defect coffee beans using CNN," presented at the 2022 IEEE International Conference on Big Data and Smart Computing (BigComp) (pp. 202–205), 2022.
- [5] N. K. Naik and P. K. Sethy, "Roasted coffee beans classification based on convolutional neural network," in *2022 International Conference on Futuristic Technologies (INCOFT)* (pp. 1-3). IEEE, 2022.
- [6] B. Shao, Y. Hou, N. Huang, W. Wang, X. Lu, and Y. Jing, "Deep learning based coffee beans quality screening," in *2022 IEEE International Conference on e-Business Engineering (ICEBE)*, pp. 271-275. IEEE, 2022.
- [7] Y.-F. Wang, C.-C. Cheng, and J.-K. Tsai, "Implementation of green coffee bean quality classification using slim-cnn in edge computing," in *2022 IEEE 5th International Conference on Knowledge Innovation and Invention (ICKII)* (pp. 133-135). IEEE, 2022.
- [8] R. E. Angelia, K. C. R. Villaverde, K. E. Recto, and R. B. Bactat, "Dried Robusta coffee bean quality classification using convolutional neural network algorithm," in *Proceedings of the 2021 7th International Conference on Computing and Artificial Intelligence* (pp. 57-61), 2021.
- [9] N.-F. Huang, D.-L. Chou, and C.-A. Lee, "Real-time classification of green coffee beans by using a convolutional neural network," in *2019 3rd International Conference on Imaging, Signal Processing and Communication (ICISPC)*, pp. 107-111. IEEE, 2019.
- [10] 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. <https://doi.org/10.1109/ACCESS.2024.3356596>
- [11] 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, 2024. <https://doi.org/10.1007/s00521-024-09623-z>
- [12] C.-S. Liang, Z.-Y. Xu, J.-Y. Zhou, C.-M. Yang, and J.-Y. Chen, "Automated detection of coffee bean defects using multi-deep learning models," in *2023 VTS Asia Pacific Wireless Communications Symposium (APWCS)* (pp. 1-5). IEEE, 2023.
- [13] T. Micaraseth, K. Pornpipatsakul, R. Chancharoen, and G. Phanomchoeng, "Coffee bean inspection machine with deep learning classification," in *2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)* (pp. 1-5). IEEE, 2022.
- [14] A. Febriana, K. Muchtar, R. Dawood, and C.-Y. Lin, "USK-COFFEE dataset: A multi-class green arabica coffee bean dataset for deep learning," in *2022 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom)*, pp. 469-473. IEEE, 2022.
- [15] Y. Hendrawan *et al.*, "Deep learning to detect and classify the purity level of luwak coffee green beans," *Pertanika Journal of Science & Technology*, vol. 30, no. 1, pp. 1-18, 2022. <https://doi.org/10.47836/pjst.30.1.01>
- [16] P. Sajjacholapunt, A. Supratak, and S. Tuarob, "Automatic measurement of acidity from roasted coffee beans images using efficient deep learning," *Journal of Food Process Engineering*, vol. 45, no. 11, p. e14147, 2022. <https://doi.org/10.1111/jfpe.14147>
- [17] K. Saddami, N. Aulia, and V. Maulidia, "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)* (pp. 13-18). IEEE, 2023.

- [18] M. Hakim, T. Djatna, and I. Yuliasih, "Deep learning for roasting coffee bean quality assessment using computer vision in mobile environment," in *2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS)* (pp. 363-370). IEEE, 2020.
- [19] A. J. Manansala and E. C. C. Paglinawan, "Classification of Coffea Liberica quality using convolution neural networks (Slim-CNN, YOLOv5, and VGG-16)," in *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-6). IEEE, 2024.
- [20] A. Pratondo, T. Zani, A. Novianty, and B. Pudjoatmodjo, "Raw coffee bean classification for roasting suitability assessment using transfer learning," in *2023 IEEE 11th Conference on Systems, Process & Control (ICSPC)* (pp. 1-6). IEEE, 2023.
- [21] L.-Y. Ke, P.-H. Chiang, C.-H. Hsia, and S.-L. Chen, "Lightweight deep convolution neural network for green coffee bean defects detection," in *2023 IEEE 6th international conference on knowledge innovation and invention (ICKII)*, pp. 461-463. IEEE, 2023.
- [22] D. Buonocore, M. Carratu, and M. Lamberti, "Classification of coffee beans varieties based on deep learning approach," in *18th IMEKO TC10 Conference "Measurement for Diagnostics, Optimization and Control to Support Sustainability and Resilience" Warsaw, Poland, 2022*.
- [23] R. V. Dimaculangan and M. A. Rosales, "Robusta coffee bean defect classification using convolutional neural network," in *2024 7th International Conference on Informatics and Computational Sciences (ICICoS)* (pp. 66-71). IEEE, 2024.
- [24] A. Korkmaz, T. Talan, S. Koşunalp, and T. Iliev, "Comparison of deep learning models in automatic classification of coffee bean species," *PeerJ Computer Science*, vol. 11, p. e2759, 2025. <https://doi.org/10.7717/peerj-cs.2759>
- [25] Coffee Bean Dataset Resized (224×224), "Kaggle," 2025. <https://www.kaggle.com/datasets/gpiosenka/coffee-bean-dataset-resized-224-x-224>