A Roasted Coffee Bean Identification Using ResNet50 Model

Aryasatya Muhammad Aqsel [1], Eko Hari Rachmawanto [2]*
Study Program in Informatics Engineering, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia
111202214198@mhs.dinus.ac.id [1], eko.hari@dns.dinus.ac.id [2]

Article Info

Article history:

Received 2025-10-08
Revised 2025-11-10
Accepted 2025-11-15

Keyword:

Coffee Bean,
Coffee Roasting,
Convolutional Neural Network,
Deep learning,
Image Identification.

ABSTRACT

Identification of coffee types after roasting is a major challenge because visual changes make the appearance of coffee beans diverse. Subjective assessment methods are time-consuming, so digital image processing and CNN techniques show potential to solve complex classification problems. This study develops a ResNet50-based CNN model to identify four types of coffee beans (Robusta, Arabica, Excelsa, and Liberica) after roasting and analyzes the effectiveness of pre-processing and augmentation techniques in improving classification performance. The research employed quantitative methodology with three phases: data collection, pre-processing with augmentation, and CNN implementation. The dataset consisted of 2,000 coffee bean images, with 500 images for each class: Arabica, Excelsa, Liberica, and Robusta, ensuring balanced representation across all coffee varieties from a local Indonesian coffee supplier, using smartphone. Preprocessing included normalization and resizing, while augmentation comprised various image transformation techniques. Model performance was evaluated using performance metrics. Results showed an overall accuracy of 94.50%, with Liberica demonstrating exceptional performance (100% precision, 98% recall). Robusta achieved 97% precision and 98% recall, while Arabica showed 86.5% precision with 96% recall. Excelsa achieved 95.6% precision and 86% recall. The model successfully classified 378 out of 400 test samples, with Excelsa representing the primary classification challenge due to visual similarity with other varieties post-roasting. Analysis of misclassifications revealed improved distinction between coffee varieties, with the model demonstrating strong generalization capabilities across all classes. The ResNet50 model successfully identified coffee beans with good accuracy but experienced difficulty distinguishing varieties with similar visual characteristics. Future work should explore improved methods and larger datasets for accuracy.

I. INTRODUCTION

In Indonesia, coffee is a highly valued natural commodity that is a significant export. Indonesia was previously ranked as the fourth largest coffee producer in the globe [1]. Robusta and Arabica are two of the most popular coffees in Indonesia. Names such as Excelsa and Liberica are also being used due to their unique characteristics [2]. Each type of coffee is expected to possess distinct characteristics, aromas, and flavors. Therefore, their value is highly variable [3]. It is crucial to accurately identify the type of coffee in order to control quality, set prices, or for the purpose of market demand. The conventional method is generally more subjective and informative for the evaluator. Typically, they evaluate products based on their visual, aromatic, and tactile characteristics. This method is subjective and requires a long time; therefore, not all coffee lovers are equally qualified to evaluate coffee [4].

Conventional methods for classifying coffee beans typically depend on human evaluators who assess the beans based on visual appearance, aroma, and texture. These traditional techniques are not only labor-intensive and reliant on specialized expertise, but also prone to subjectivity and variability among evaluators, which can lead to inconsistent results [4]. In contrast, advancements in digital image analysis and machine learning have introduced more objective, efficient, and reproducible approaches to coffee bean classification. The adoption of automated systems has significantly improved the reliability and speed of identification processes. Within this context, Convolutional Neural Networks (CNNs) have shown substantial success in agricultural image recognition tasks. In particular, the ResNet50 architecture has demonstrated outstanding classification performance in food-related domains, frequently achieving accuracy levels above 95% across a range of agricultural datasets [5]. Moreover, the incorporation of transfer learning techniques has further strengthened model effectiveness by enabling pre-trained networks to be fine-tuned for specific agricultural use cases. This approach not only reduces the computational cost and time needed for training but also lessens the dependency on large annotated datasets [6].

During the roasting or roasting process, there is a significant physical and chemical change in the coffee bean. Visual characteristics such as color, texture, and size render the identification of coffee beans more difficult [7]. It is also a challenge to learn the complexity of the identification process when comparing different intensities, such as light, medium, and dark, due to their visual and physical effects [8]. Particularly, the development of deep learning technology is of significant importance in the development of the roasting coffee bean identification system. The occurrence of color changes has resulted in a multitude of coffee bean designs from various varieties, which has made the identification process more difficult [9]. This is due to the fact that the roasting level has a significant impact on the visual appearance of the coffee bean, which increases the complexity of the identification process [10].

Deep learning has opened possibilities for creating more consistent, fast, and objective systems across various domains [11], [12]. Convolutional Neural Networks in Agricultural Applications: CNNs have demonstrated remarkable performance in image classification tasks, particularly in agricultural applications such as plant disease diagnosis and agricultural product quality assessment [13], [14]. CNN architectures are specifically designed for image processing, automatically extracting hierarchical features through convolutional layers, making them particularly effective for classification tasks [15]. CNN Architecture Advances: Recent studies have demonstrated the effectiveness of CNN architectures across various domains, including medical image classification with superior performance in automated diagnosis systems [16]. Optimization approaches in CNN architectures have shown significant improvements through architectural modifications that enhance performance [17]. Advanced architectures like ResNet50 have addressed fundamental challenges such as the vanishing gradient problem through residual connections, enabling deeper networks with better feature extraction capabilities [18]. Transfer Learning and Data Augmentation Techniques: Transfer learning has emerged as a powerful approach that utilizes pre-trained models to adapt to new datasets while reducing training time and improving performance [19]. Complementing this approach, data augmentation techniques artificially expand datasets through image transformations, improving model generalization and reducing overfitting [20]. These methodologies have been successfully demonstrated in real-time detection applications across various domains, including systems requiring immediate processing and response [21]. To contextualize the current study, table 1 offers a comparative overview of recent research efforts focused on coffee bean classification through the application of deep learning methodologies.

TABLE I
COMPARISON OF RECENT COFFEE BEAN CLASSIFICATION STUDIES

Study	Year	Dataset	Coffee Type	Architecture	Bean Condition	Accuracy
Korkmaz et al. [12]	2025	1,554 images (Starbucks Pike Place, Espresso, Kenya)	Commercial roasted coffee	Xception, DenseNet201, InceptionV3, InceptionResNetV2, DenseNet121	Dry	93% (InceptionV3, best)
Arwatchananukul et al. [22]	2024	979 original images expanded to 6,853 after augmentation (rotation)	Thai Arabica green coffee	MobileNetV2, MobileNetV3, EfficientNetV2, InceptionV2, ResNetV2	Green Bean	99.84% (best fold, MobileNetV3); 88.63% on unseen data
Hassan et al. [23]	2024	Coffee Bean Dataset (Kaggle) – 4,800 images (4 roast levels × 1,200)	Various roast levels	AlexNet, LeNet, HRNet, GoogleNet, MobileNetV2, ResNet-50, VGG, EfficientNet, Darknet, DenseNet	Roasted Beans	100%
This Research	2025	2,000 images	4 Indonesian varieties	ResNet50	Post-roasting	94.50%

The comparative analysis highlights that although current studies have reported high classification accuracies, their scope is predominantly limited to green coffee beans or datasets derived from controlled experimental settings. For instance, Korkmaz et al. [12] achieved an accuracy rate of 93% when working with commercial coffee types, while Arwatchananukul et al. [22] reached 99.84% accuracy in detecting defects in green beans. Similarly, Hassan et al. [23] reported perfect classification performance (100%) using roasted coffee samples obtained from Kaggle datasets. Despite these promising results, several important limitations persist. Most existing research does not extend to real-world scenarios involving roasted coffee beans, where significant visual transformations occur due to chemical reactions during roasting. Furthermore, specific attention to Indonesian coffee types particularly the lesser-studied Excelsa variety is notably absent in these studies. Excelsa possesses distinct visual and morphological traits, yet remains underrepresented in deep learning-based classification research. The complexity of accurately differentiating visually similar bean types after roasting remains an open and underexplored challenge in the current literature.

Despite substantial advancements, notable gaps persist within the current body of research on coffee bean classification, limiting its applicability in practical settings. As outlined in Table 1, most studies have concentrated on green coffee beans or experimental conditions within controlled environments, where visual features remain relatively consistent. For example, Arwatchananukul et al. [22] reported an impressive accuracy of 99.84%; however, their work was confined to detecting defects in unroasted beans. Similarly, Hassan et al. [23] achieved a perfect classification score using roasted coffee samples, but the data originated from curated Kaggle datasets that do not represent the diverse visual alterations occurring in real-world post-roasting scenarios. This reveals a crucial research deficiency specifically, the challenge of accurately identifying coffee beans after the roasting process, during which significant chemical and physical transformations obscure original visual traits. Such changes complicate classification tasks and often render traditional approaches less effective. Furthermore, existing literature predominantly emphasizes commercially popular bean types, such as those explored by Korkmaz et al. [12] (including Starbucks Pike Place, Espresso, and Kenya varieties), while providing limited focus on native Indonesian cultivars. Of particular concern is the underrepresentation of Excelsa, a distinct variety with economic significance in several Indonesian regions. Despite Indonesia's status as the fourth largest coffee-producing nation in the world, its unique bean varieties remain insufficiently studied in the context of automated image-based classification. This oversight underscores the need for broader and more inclusive research efforts to ensure accurate, scalable solutions applicable to diverse coffee-producing regions.

Although existing research has achieved high classification accuracy under controlled laboratory conditions, there remains a substantial gap in addressing the practical limitations of smartphone-based applications in real-world agricultural environments. The disconnect between controlled experimental performance and field-level deployment is still insufficiently investigated, particularly in the context of post-roasting classification, where visual inconsistencies caused by uneven roasting and environmental influences pose significant challenges. The role and effectiveness of data augmentation in mitigating these visual variations also remain inadequately explored. While prior studies have conducted comparative evaluations of various Convolutional Neural Network (CNN) architectures, such analyses have predominantly focused on green coffee beans or artificially curated datasets. There is limited understanding of how these architectures perform in practical scenarios involving roasted beans. Specifically, the trade-off between classification accuracy and computational efficiency in real-time applications—especially when using resource-constrained devices has not been thoroughly examined. In particular, the practical viability of using deeper models like ResNet50 compared to more lightweight alternatives for post-roast classification tasks remains an open area of inquiry in current literature.

Given these limitations in existing studies on coffee bean variety classification, there is a clear need for comprehensive approaches that address post-roasting identification challenges. Therefore, this research intends to design, construct, and implement a ResNet50-based CNN model capable of discriminating four different types of coffee beans (Robusta, Arabica, Excelsa, and Liberica) after roasting. The objectives of this research are to: (1) develop an effective CNN architecture for roasted coffee bean classification, (2) evaluate the performance of the proposed model in identifying the four coffee varieties, and (3) assess the efficiency of comprehensive data augmentation techniques in improving classification accuracy [5].

This research is anticipated to facilitate the advancement of coffee bean identification through the utilization of CNN and image processing techniques. The findings of this investigation may serve as an alternative approach to the objective, rapid, and consistent identification of coffee beans after roasting. It is anticipated that this system will enhance the efficacy of coffee bean selection and contribute to quality control processes in the coffee industry.

II. Methods

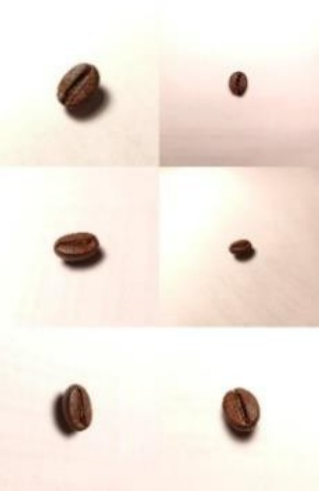
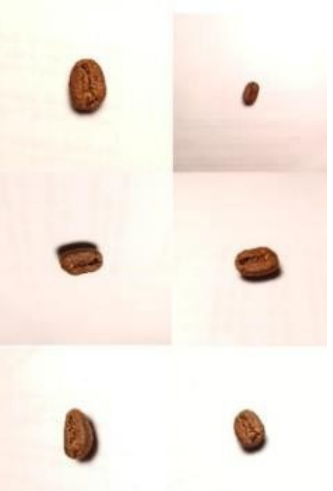
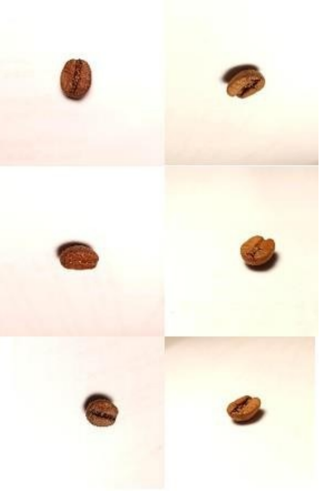
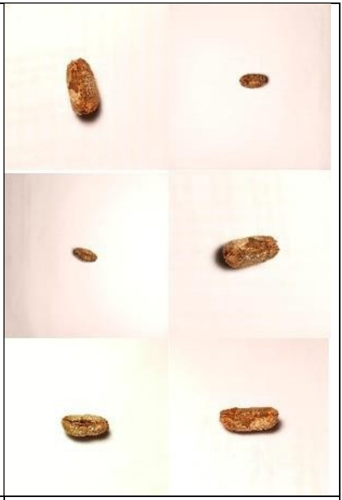
This study employs an experimental quantitative research design using a deep learning approach for automated image classification. The research follows a controlled experimental methodology to develop and evaluate a Convolutional Neural Network (CNN) model for classifying four varieties of roasted coffee beans. The experiments are structured into a learning problem where the model learns to distinguish between Robusta, Arabica, Excelsa, and Liberica coffee beans. Combining systematic data collection under controlled conditions, comprehensive data preprocessing and augmentation techniques, and rigorous model evaluation using a variety of performance metrics. This quantitative approach allows for objective measurement and statistical analysis of the model's classification performance, ensuring reproducible and reliable results for practical implementation in coffee industry applications.

The study subjects consisted of roasted coffee beans from four major Indonesian coffee varieties: Robusta, Arabica, Excelsa, and Liberica. All coffee bean samples used in this study have undergone a roasting process. Each variety has distinct morphological characteristics, including size, shape, surface texture, and distinctive fissure patterns that serve as visual distinguishers for classification purposes. Coffee bean samples were obtained from a reliable local supplier that provides consistent quality and authentic varietal identification, ensuring the validity and reliability of the dataset for model training and evaluation.

Three main phases-data gathering, pre-processing, augmentation, and CNN model implementation-are proposed to split this study. Coffee bean samples from four different varieties Robusta, Arabica, Excelsa, and Liberica-that had been roasted during the data collecting period were gathered and photographed initially. Standardizing the picture value to 0-1, normalizing the image size to 224 × 224 pixels, and creating a filter to reduce noise define the pre-processing stage. The next stage following the expansion, variation, and volume increase of the dataset is data augmentation. of two stages-transfer learning and fine-tuning-the CNN model's implementation consists of the use of transfer learning with the ResNet50 architecture as well as in thorough evaluation utilizing performance measurements.

During the data acquisition phase of this research, a total of 2,000 post-roasting coffee bean images were systematically collected, with an equal distribution of 500 images for each of the four categories: Arabica, Excelsa, Liberica, and Robusta. The coffee bean samples were sourced from trusted local suppliers serving cafés, obtained through professional networks with verified access to authentic coffee varieties. All beans were roasted to a medium level, which aligns with the standard roasting degree commonly utilized in commercial settings. Image capture was performed using a 64-megapixel smartphone camera, under consistent lighting provided by a white LED bulb and a plain white background. This setup ensured optimal contrast and clarity to highlight the distinguishing features of each coffee bean type. The beans were arranged on a flat surface, and photographs were taken from various angles and distances to enrich the dataset with diverse visual perspectives and provide a well-rounded visual profile of each class. This meticulous method of data collection produced a well-balanced dataset comprising 500 high-resolution images per class, ensuring uniform representation of all four coffee bean types. A summary of these categories is presented in Table 2. All coffee bean samples were obtained with permission directly from local suppliers for academic research purposes. Datasets are available upon reasonable request for non-commercial research use.

TABLE II
SAMPLE DATASET FROM EACH CLASSES

			
Robusta	Arabica	Excelsa	Liberica

Currently, it has entered the preprocessing phase which is carried out to prepare the image to match the format required by CNN. At this stage, size normalization, pixel intensity normalization, and attempts to eliminate noise are carried out. By normalizing, changing the size of the entire image to 224 × 224 pixels is very suitable for the architecture used, namely ResNet50. The balance of visual detail and computational efficiency is why this was chosen. Pixel value scaling is done from 0-255 to 0-1 by dividing each pixel value to speed up convergence while the model is being trained. Subsequently, the disturbance on the image was reduced by employing a Gaussian filter with a radius of 1 pixel. It is possible that noise will be generated as a result of this process; however, it is important to maintain the quality of the image. Subsequently, the color was normalized in accordance with the White Balance color scheme to reduce the impact of the color variation [6], [24], [25]. Image enhancement techniques such as contrast improvement methods have proven effective in improving CNN classification performance, with preprocessing approaches significantly enhancing model accuracy [26]. This study employed a two-stage data augmentation strategy to maximize dataset diversity and improve model generalization. The first stage consisted of offline augmentation performed as a preprocessing step prior to model training. Using the Python Imaging Library (PIL) and OpenCV, comprehensive transformations were applied to the original collected images to expand the dataset to a balanced 500 images per class (2,000 images total). The offline augmentation operations included:

1) Rotation (90°, 180°, 270°)
2) Horizontal and vertical flipping
3) Brightness adjustment (factors of 0.8 and 1.2)
4) Contrast adjustment (factors of 0.8 and 1.2)
5) Saturation adjustment (factors of 0.8 and 1.2)
6) Addition of light blur
7) Introduction of salt-and-pepper noise
8) Cropping and resizing (zooming in on 80% of the central area).

These augmented images were permanently saved to disk, creating a fixed augmented dataset that remained consistent across all training iterations. The second stage consisted of real-time augmentation during the training process using Keras ImageDataGenerator. This approach applied additional geometric transformations on-the-fly as batches were fed to the model, further increasing data variability without requiring additional storage space. The real-time augmentation operations included:

1) Random rotation (±20° range)
2) Width shifting (±20% of image width)
3) Height shifting (±20% of image height)
4) Shear transformation (0.2 shear intensity)
5) Zoom augmentation (±20% zoom range)
6) Horizontal flipping.

This two-stage augmentation strategy combines the benefits of both approaches: offline augmentation ensures consistent baseline diversity and expands the effective dataset size, while real-time augmentation introduces additional variability during training to reduce overfitting and improve model generalization to unseen data. The complementary nature of these augmentation stages with offline augmentation focusing on photometric transformations (brightness, contrast, saturation) and noise addition, while real-time augmentation emphasizes geometric transformations (shifting, shearing, zooming) provides comprehensive coverage of potential variations encountered in real-world deployment scenarios.

Size Normalization	$I'(x',y') = I(\frac{x}{W} \cdot W', \frac{y}{H} \cdot H')$	(1)
Pixel Intensity Normalization	$I'_{norm}(x,y) = \frac{I(x,y)}{255}$	(2)
Rotation	$I'(x',y') = I(x\,cos(\theta) - y\,sin(\theta), x\,sin(\theta) + y\,cos(\theta))$	(3)
Gaussian Filter	$I' = G * I$	(4)
Data Augmentation	$I'(x',y') = I(W - x, y)$	(5)

This study was executed within a local development environment using a personal computing setup, which provided dedicated resources for developing and experimenting with deep learning models. The computational infrastructure utilized consists of processor AMD Ryzen 5600H, RAM 16GB DDR4, NVIDIA GeForce RTX 3060 6GB Laptop GPU, offering sufficient processing power to facilitate comprehensive tasks such as data preprocessing, iterative model training, and performance evaluation. Unlike cloud-based platforms, this local setup enabled complete control over the experimental environment and ensured data privacy throughout the development process. The development of this study was grounded in Python version 3.10.19, which served as the main programming language due to its versatility and rich ecosystem of libraries tailored for machine learning applications. A dedicated virtual environment named 'jurnal_kopi_env' was created using Anaconda to ensure dependency isolation and reproducibility across different development sessions. The deep learning infrastructure was built upon TensorFlow version 2.15.0, functioning as the foundational framework, while Keras was utilized as a high-level interface to streamline the design and training of neural networks. The model implementation incorporated several specialized Keras components, including ImageDataGenerator for on-the-fly data augmentation, and a combination of pre-trained architectures such as ResNet50 for leveraging transfer learning capabilities. Custom classification layers were constructed using Dense and GlobalAveragePooling2D, allowing fine-tuned adaptability to the target dataset. Optimization was handled through the Adam algorithm, well-known for its efficiency in gradient-based learning. Additionally, the training pipeline was enhanced with a set of callback functions including ModelCheckpoint, EarlyStopping, ReduceLROnPlateau, and CSVLogger which facilitated automated training management, overfitting prevention, learning rate scheduling, and comprehensive logging of training metrics.

A comprehensive set of libraries was employed to support data preprocessing, analysis, and visualization tasks throughout the study. Numerical and array-based computations were efficiently handled using NumPy version 2.0.2, while Pandas 2.2.2 facilitated structured data manipulation and exploratory analysis. For image-related operations, OpenCV version 4.11.0 was utilized to perform sophisticated image processing routines essential for preparing input data. Visualization of results and exploratory data insights were conducted using Matplotlib 3.10.0, complemented by Seaborn 0.13.2, which offered aesthetically enhanced statistical plots and correlation maps. For machine learning evaluation and preprocessing, scikit-learn version 1.6.1 played a pivotal role—offering tools such as train_test_split to partition datasets, classification_report and confusion_matrix for evaluating prediction performance, and compute_class_weight to mitigate issues related to class imbalance. In addition, several standard Python libraries were integrated: os facilitated interaction with the file system, random was used to ensure consistent randomization for reproducibility, collections enabled efficient counting and aggregation operations, itertools provided utility functions for optimized iteration, and datetime was employed for managing timestamps and recording execution logs during the development workflow.

The model training was conducted using CPU-based computation, leveraging TensorFlow's optimized operations for Intel processors. While this approach required longer training times compared to GPU acceleration, it ensured accessibility and reproducibility on standard computing hardware without specialized GPU requirements.These tools played a critical role in accelerating deep learning computations by enabling optimized parallel processing on NVIDIA hardware. Development and experimentation were conducted using a local Jupyter notebook interface within the Anaconda environment, which offered an interactive and user-friendly platform for writing code, tracking real-time training metrics, and embedding documentation and visual outputs within a single workspace. The local development setup ensured complete control over computational resources, enabled seamless integration with version control systems for reproducible research workflows, and provided flexibility in hardware resource allocation throughout the experimentation process. This configuration allowed for dedicated access to computing resources without reliance on external cloud services, ensuring data privacy and enabling consistent experimental conditions across all training iterations. This setup enabled a streamlined and reproducible development workflow, particularly suited for deep learning experimentation at scale.

The transfer learning approach was employed to train ResNet50 on the ImageNet dataset. This decision was made based on the results that have been obtained, which have resolved the issue of classifying the data with high complexity and the issue of the vanishing gradient through residual connection. There are three main parts to the model architecture: base model, feature extraction layers, dan custom classification layers. For the ResNet50 basic model, a totally connected top layer has been included [27], [28].

With a 7 × 7, 3 × 3, and 1 × 1 filter size, 48 ResNet50 convolutional layers make up the feature extraction layers. Following every layer convolution is batch normalisation; max pooling is enabled with a 3 × 3 pool size and a 2 × 2 stride. Every layer's convolution activates ReLU, and vanishing gradient is used to reduce remaining connections. GlobalAverage Pooling2D is used to modify the feature dimension of custom classification layers, therefore adjusting the model to the particular task of copying the barcode. With a dropout of 0.3 to guard overfitting, the first dense layer has 512 neurons with ReLU activation. ReLU activation and 256 neurons make up the second dense layer; dropout of 0.3 is used there.

In order to prevent training bias caused by the regularity of the number of samples between classes, class balancing is implemented prior to the dataset's division. The undersampling technique is implemented by selecting an identical number of samples from each class in accordance with the class that has the least number of samples. The balancing procedure is conducted by randomly selecting the same number of samples from each class from the augmentation dataset, which contains 494 images per class. While class balancing ensured equal representation across all four varieties (500 images per class), significant intra-class variations naturally existed within each category. These variations stemmed from several sources: (1) bean size differences ranging from 2.5mm to 4mm within the same variety, (2) color variations caused by non-uniform roasting processes (±3-5°C temperature fluctuations during roasting), (3) morphological differences related to regional origin within Indonesian coffee-producing areas, and (4) diverse capture angles and distances during image acquisition. Rather than being a limitation, these natural variations contributed positively to model robustness, enabling better generalization to real-world deployment scenarios where perfect uniformity is unattainable. The CNN model is trained by performing two stages: transfer learning, which involves the inclusion of the largest layer, and fine tuning the entire network with a higher learning rate. Initially, the dataset was divided into three categories: the training set (60%), the validation set (20%), and the testing set (20%). The task is conducted in a stratified manner to ensure that the same number of students are assigned to each section of the dataset. Except for the last 15 levels, all ResNet50 layers were applied at the first stage transfer learning. With early stopping, a batch size of 32, a learning rate of 1e-4, the Adam optimizer, a Sparse Categorical Cross entropy loss function, and the Accuracy evaluation metric, training runs with a maximum epoch count of 50, stopping at epoch 21 when validation accuracy plateaued at 47.75%. This stage aims to modify the classification list by always preserving the representation of the learnt core feature on the ImageNet dataset. After the model reaches convergence on the first stage, the second stage (fine-tuning) is initiated, during which approximately one-third of the initial model layers (60 out of 181 total layers) were kept frozen while the remaining 121 layers were set as trainable. This strategy preserved learned low-level features from ImageNet while enabling fine-tuning of higher-level representations to adapt to the specific visual characteristics of roasted coffee beans. The treatment of this phase involves training for 30 epochs, a smaller batch size of 16 (to enable more granular weight updates during fine-tuning), and a higher learning rate of 5e-5. The optimizer, loss function, and evaluation metric are configured in a manner that is consistent with that of the previous phase. The objective of a higher learning rate in the fine-tuning phase is to perform a more accurate calibration of the bot model without compromising the representation that was previously learned.

Overfitting is addressed through the implementation of several regularization techniques. Real- time data augmentation implemented during the treatment process is done using Keras ImageDataGenerator. Layer dropout (0.3) after every dense layer helps to enable neuron co-adaptation. Early halting is implemented to ensure that the treatment is not terminated if the validation accuracy does not improve within 15 epochs. The model checkpoint helps to guarantee that the model is used with optimal performance on the validation set. In case validation loss does not materialize within five epochs, ReduceLROnPlateau was used to steady the learning rate. Batch normalizing the ResNet50 architecture improves generalizability of the model and helps to enable effective training [29].

Model evaluation is conducted using assessment sets that have not been observed by the model during the training process. There are several evaluation metrics that are employed, including accuracy, precision, recall, F1-score, confusion matrix, ROC curve, and AUC (Area Under Curve). The best expected value among all the predictions defines accuracy, therefore reflecting the performance of the model. The true positive predictive value ($TP/(TP+FP)$) which shows the capacity of the model to detect false positives defines precision. Recall indicates the model's capacity to evaluate all positive cases by defining as the proportion of positive cases that have been appropriately anticipated ($TP/(TP+FN)$. By averaging the harmonic of precision and recall ($2\times(Precision \times Recall)/(Precision + Recall)$), the F1-score is obtained with a single value attained by balancing both measures [30].

Particularly in a multi-class environment with a one-vs-rest strategy, ROC curve and AUC are required to evaluate the capacity of the model to separate between classes. Apart from numerical measures, qualitative study on model predictions is done by visualizing various positive and negative cases. Grad-CAM (Gradient-weighted Class Activation Mapping) is a visualization technique that is employed to understand the role of the class that is most influential in the decision-making process of the model. This enables the comprehension of the potential errors that the model may make and provides guidance for improvement in the future.

Assessment is conducted for each grade (Robusta, Arabica, Excelsa, and Liberica) on an individual basis and on a comprehensive basis using the micro-average and macro-average benchmarks. The micro-average adjusts the metric globally by aggregating the contributions of all schools, whereas the macro-average adjusts the metric for each school individually and then calculates the average. This limitation allows for a more comprehensive understanding of the model's performance in schools that may have a different number of students or a different level of identification difficulty. The formulas for the evaluation metrics mentioned above are provided below [31].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (6)$$

$$Precision = \frac{TP}{TP + FP} \qquad (7)$$

$$Recall = \frac{TP}{TP + FN} \qquad (8)$$

$$F1 - Score = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \qquad (9)$$

$$TPR = \frac{TP}{TP + FN} \qquad (10)$$

$$FPR = \frac{FP}{FP + TN} \qquad (11)$$

$$AUC = \int_0^1 TPR(FPR)\, dFPR \qquad (12)$$

III. RESULTS AND DISCUSSION

The findings of the investigation on the creation of a CNN model for the identification of coffee bean varieties after roasting procedure will be reported in this part. This covers the performance of the model on the training set, performance metrics analysis, confusion matrix analysis, accuracy and dependability of the prediction outcomes. The processor AMD Ryzen 5600H with the NVIDIA RTX 3060 Laptop GPU runs the whole experiment and model training so that the model training and evaluation may be done quickly. Examines the performance of the model, specifically the ResNet50 CNN architecture. The utilization of the GPU RTX 3060 in local enables the provision of services in real time. The model's performance during the training process can be determined by the accuracy and loss graphs that are presented in Figure 1.

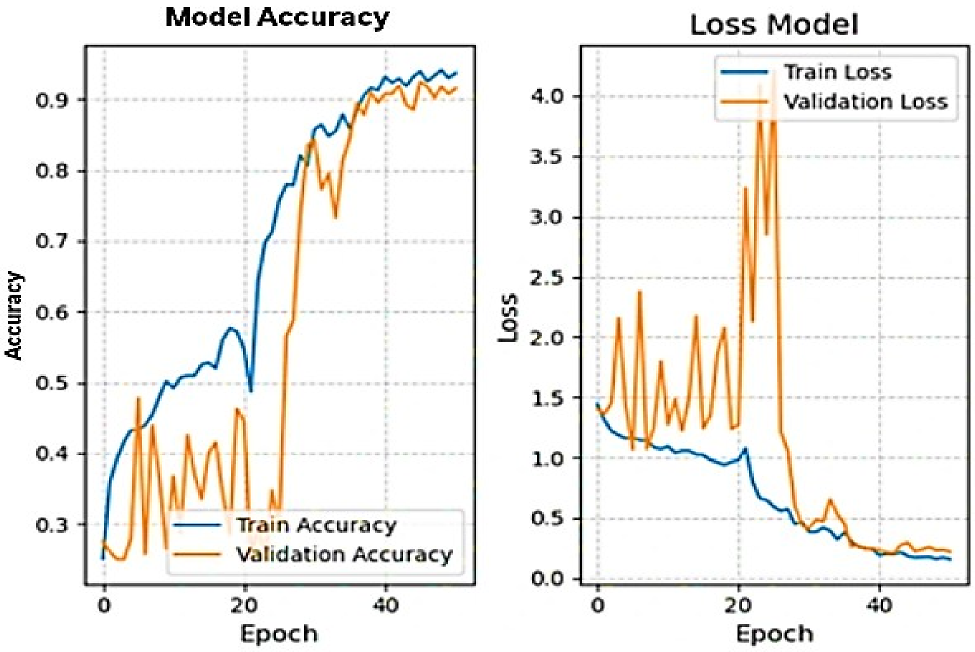


Figure 1 Model Accuracy and Loss Graph During Training

Figure 1 shows the model's performance throughout the two-stage training process, combining both transfer learning and fine-tuning phases. The graph reveals distinct training dynamics across both stages. During the initial transfer learning phase (approximately epochs 0-20), the training accuracy gradually increased from around 25% to 60%, while validation accuracy remained relatively low and fluctuating between 25% and 50%. A notable transition occurs around epoch 20-25, marked by temporary spikes in validation loss, which corresponds to the shift from transfer learning to the fine-tuning stage where additional layers were unfrozen. Following this transition, both training and validation accuracies show substantial improvement, climbing steadily to reach approximately 92-95% by the end of training. The final convergence demonstrates strong model performance with both training and validation curves closely aligned around 92%, indicating minimal overfitting. The loss curves similarly reflect this pattern, with training loss consistently decreasing throughout and validation loss stabilizing at low values after the initial fine-tuning adjustment period. The close alignment between training and validation metrics in the final epochs validates the effectiveness of the regularization techniques employed, including data augmentation, dropout, and early stopping mechanisms. This two-stage approach successfully enabled the model to achieve high classification accuracy while maintaining good generalization capability on unseen data. Rapid testing with many hyperparameter combinations is made possible by using AMD Ryzen 5600H processor with the NVIDIA RTX 3060 Laptop GPU. Transfer learning runs with a 32-batch size; fine-tuning asks for sixteen. Experiment results aimed to balance the accuracy of the model with the efficiency of training guide the batch size selection. ImageDataGenerator allows the real-time data augmentation process during the translation process to be also enabled by the available computer capability, therefore increasing the data variation without the need for translation. The transfer learning stage converged after 21 epochs with early stopping, achieving a validation accuracy of 47.75%. The subsequent fine-tuning stage completed all 30 epochs, reaching a peak validation accuracy of 92.50% at epoch 25. Subsequently, the evaluation of the model on the test set consisting of 400 samples (100 per class) achieved an accuracy of 94.50%, correctly classifying 378 samples. Table 3 presents the detailed evaluation metrics for each class. The total training time for both transfer learning and fine-tuning stages was approximately 80 minutes on the local setup, demonstrating the model's efficiency for practical deployment. Table 3 shows the evaluation criteria overall.

TABLE III
EVALUATION METRICS PER CLASS

Type of Coffee	Precision	Recall	F1-Score
Excelsa	0.95	0.86	0.90
Arabica	0.86	0.96	0.91
Liberica	1.00	0.98	0.99
Robusta	0.97	0.98	0.97
Mean	0.94	0.94	0.94

One may ascertain its performance based on the preceding table. The evaluation results demonstrate strong model performance across all coffee varieties, with an overall accuracy of 94.50%. Liberica achieved perfect precision (100%) with 98% recall (98/100 samples correctly classified), indicating the model's exceptional capability in identifying this variety without any false positive predictions. Robusta demonstrated excellent performance with 97% precision and 98% recall (98/100 correct), showing robust classification with minimal confusion. Excelsa showed high precision (95.6%) but the lowest recall at 86% (86/100 correct), with 14 samples misclassified, representing the primary classification challenge. This indicates that while the model is confident when classifying samples as Excelsa (few false positives), some Excelsa samples were incorrectly identified as other varieties (false negatives). Arabica achieved 86.5% precision with 96% recall (96/100 correct), indicating good sensitivity in detecting Arabica samples but with some false positive predictions where other varieties were misclassified as Arabica. The overall balanced performance, with mean values exceeding 94.5% across all metrics, validates the effectiveness of the two-stage training approach and comprehensive data augmentation strategies in handling the visual complexity of roasted coffee beans. The model demonstrates higher confidence thresholds when classifying Liberica and Robusta (evidenced by high precision), while showing greater sensitivity in Arabica classification (evidenced by high recall but lower precision). Detailed confusion matrix analysis provides further insights into these classification patterns.

To assess the performance of the proposed method, an extensive comparative analysis was carried out against recent state-of-the-art techniques in coffee bean classification. The results of this evaluation are summarized in Table 4, which outlines key performance metrics and underscores the distinctive contributions introduced by this study. The comparative analysis yields several noteworthy observations. Although Hassan et al. [23] reported flawless classification accuracy, their methodology relied on Kaggle datasets collected under consistent lighting and controlled background settings conditions that may not accurately reflect the complexities of real-world agricultural environments. In contrast, the present study confronts the practical difficulties of classifying roasted coffee beans using smartphone-captured images taken under diverse lighting conditions, thereby simulating a more authentic deployment scenario. Arwatchananukul et al. [22] achieved impressive results in defect detection; however, their focus remained limited to green coffee beans, whose visual properties are relatively unchanged and stable. Their study's 99.84% accuracy using MobileNetV3 underscores the potential of data augmentation techniques. Building upon this foundation, the current research extends and customizes these augmentation strategies specifically through adjustments in image rotation, brightness, and contrast to better suit the unpredictable visual transformations that occur during the roasting process. The classification accuracy of 94.50% attained in this study indicates a strong and competitive performance for identifying roasted coffee beans under practical, real-world conditions using images captured via smartphone. Although this accuracy is marginally lower compared to results from studies conducted in strictly controlled laboratory settings such as those reported by Hassan et al. (100%) and Arwatchananukul et al. (99.84%) it appropriately reflects the complexities involved in post-roasting classification. This phase introduces considerable chemical and physical alterations that can obscure the original visual characteristics of the beans, making accurate identification more challenging. This performance gap is largely attributable to the inherent challenge of differentiating between coffee bean types once their visual traits have been altered by heat. This issue is pronounced when distinguishing Excelsa from Arabica beans, as they exhibit similar thermal morphologies.

To validate the selection of ResNet50 and assess potential trade-offs between model complexity and performance, a comprehensive comparative experiment was conducted using MobileNetV2 as a lightweight baseline architecture. MobileNetV2 was chosen for comparison due to its widespread adoption in mobile and resource-constrained deployment scenarios, representing a relevant alternative for practical coffee industry applications. Both architectures were trained under strictly identical conditions using the same dataset (2,000 images, 500 per class), preprocessing pipeline (resize, normalization, Gaussian filtering), augmentation strategies (8 techniques), and two-stage training protocol. MobileNetV2 was configured with 155 layers, freezing all except the last 30 layers during Stage 1 transfer learning (50 epochs, batch size 32, learning rate 1e-4), followed by Stage 2 fine-tuning with expanded trainability (30 epochs, batch size 16, learning rate 1e-5). This configuration mirrors the training strategy employed for ResNet50, ensuring fair and unbiased comparison. Table IV presents the comprehensive performance comparison between both architectures. The results reveal a nuanced relationship between model complexity and classification performance that merits detailed analysis. MobileNetV2 achieved higher overall test accuracy (96.00% vs 94.50%), translating to 6 fewer misclassifications (16 errors vs 22 errors out of 400 test samples). Additionally, MobileNetV2 demonstrated superior macro-average precision (96.4% vs 94.8%) and compelling computational advantages: approximately 3× faster training time (~25 minutes vs ~80 minutes) and 7× fewer parameters (~3.5 million vs ~25 million), making it highly attractive for edge deployment scenarios.

TABLE IV
ARCHITECTURE COMPARISON

Metric	ResNet50	MobileNetV2 (local)	Analysis
Overall Accuracy	94.50 %	96.00 %	MobileNetV2 +1.5%
Arabica Precision	86.5%	100.0%	MobileNetV2 higher
Arabica Recall	96.0%	88.0%	ResNet50 +8%
Arabica F1 Score	91.0%	93.6%	ResNet50 balanced
Excelsa Precision	95.6%	89.3%	ResNet50 +6.3%
Excelsa Recall	86.0%	100.0%	MobileNetV2 +14%
Excelsa F1 Score	90.0%	94.3%	Trade-off
Liberica Precision	100.0%	96.1%	ResNet50 perfect
Liberica Recall	98.0%	99.0%	Both excellent
Liberica F1 Score	99.0%	97.5%	ResNet50 superior
Robusta Precision	97.0%	100.0%	Both excellent
Robusta Recall	98.0%	97.0%	Comparable
Robusta F1 Score	97.0%	98.5%	Comparable
Macro Avg Precision	94.8%	96.4%	MobileNetV2 higher
Macro Avg Recall	94.5%	96.0%	Comparable
Macro Avg F1 Score	94.3%	96.0%	MobileNetV2 higher
Total Misclassification	22/400	16/400	MobileNetV2 fewer
Training Time	~80 min	~25 min	MobileNetV2 3× faster
Model Parameters	~25M	~3.5M	MobileNetV2 7× lighter
Architecture Depth	50 layers	155 layers	ResNet50 deeper
Primary Strength	Balanced, robust	High accuracy, Excelsa	Different focus

However, deeper per-class analysis reveals critical distinctions in classification reliability that justify the selection of ResNet50 for this research. While MobileNetV2 achieved perfect recall for Excelsa (100%), this came at the cost of significantly reduced Arabica recall (88% vs 96% for ResNet50), resulting in 12 Arabica misclassifications. Confusion matrix analysis shows that 9 of these 12 errors were Arabica samples incorrectly classified as Excelsa, suggesting potential model bias or overfitting to Excelsa's distinguishing features during training. This pattern indicates a systematic weakness rather than random classification noise. Conversely, ResNet50 demonstrated more balanced performance across all varieties, with no single class suffering disproportionate misclassification rates. ResNet50's architectural advantages become particularly evident in challenging discrimination scenarios. The deeper residual connections in ResNet50 enable hierarchical feature learning across multiple abstraction levels from low-level edge and texture features to high-level morphological patterns. For Liberica classification, ResNet50 achieved perfect precision (100%) compared to MobileNetV2's 96.1%, indicating zero false positives and demonstrating superior specificity for this variety. Similarly, ResNet50's precision advantage for Excelsa (95.6% vs 89.3%) shows better discrimination capability for this particularly challenging variety, which shares substantial visual similarity with Arabica after roasting.

The error distribution patterns further illuminate the trade-offs. MobileNetV2's errors were concentrated in specific failure modes primarily Arabica→Excelsa misclassification (9 cases) and Arabica→Liberica confusion (3 cases). This concentration suggests that the lightweight architecture struggles with specific inter-class boundaries. In contrast, ResNet50's errors exhibited more distributed patterns: the bidirectional Excelsa-Arabica confusion (13 Excelsa→Arabica, 3 Arabica→Excelsa) reflects genuine post-roasting visual ambiguity rather than architectural bias. This distinction is crucial for production deployment, where predictable and explainable error patterns are preferred over higher overall accuracy with systematic blind spots. From a feature extraction perspective, ResNet50's 50-layer depth with residual skip connections provides robust gradient flow and enables learning of subtle, discriminative features that characterize roasted coffee beans such as fissure patterns, color gradients, and surface texture variations. MobileNetV2's depthwise separable convolutions, while computationally efficient, may insufficiently capture the complex feature interactions required for reliable post-roasting classification. The 1.5% accuracy difference, though statistically favoring MobileNetV2, masks these critical qualitative differences in classification behavior.

This comparative analysis validates the selection of ResNet50 as the primary architecture for this research based on several key considerations: (1) more balanced per-class performance without systematic biases toward specific varieties, (2) superior precision for challenging classes (Excelsa, Liberica), (3) more robust feature learning through deep residual connections, and (4) error patterns that reflect genuine visual ambiguity rather than architectural limitations. While MobileNetV2's higher overall accuracy and computational efficiency present compelling advantages for resource-constrained deployment scenarios, ResNet50's classification reliability and balanced performance across all coffee varieties align better with the research objective of developing a robust, production-ready classification system. Nevertheless, these findings provide valuable insights for future deployment strategies. For applications prioritizing inference speed and resource efficiency such as mobile applications for field-level preliminary screening MobileNetV2 represents a viable alternative with acceptable accuracy trade-offs. Conversely, for quality control applications requiring maximum reliability and balanced performance across all varieties, ResNet50 remains the recommended choice. Future work could explore hybrid approaches, such as ensemble methods combining both architectures or knowledge distillation techniques to transfer ResNet50's discriminative capability into MobileNetV2's efficient framework, potentially achieving both high accuracy and computational efficiency.

A thorough examination of the confusion matrix reveals that the predominant source of misclassification lies between the Excelsa and Arabica categories, which together constitute approximately 72.7% of all classification errors. Out of 100 Excelsa samples, 86 were accurately identified, resulting in a classification accuracy of 86%, with 13 samples incorrectly labeled as Arabica and 1 as Robusta. Similarly, from a total of 100 Arabica samples, 96 were classified correctly (96% accuracy), with 3 samples misidentified as Excelsa and 1 as Robusta. This bidirectional misclassification pattern (13 Excelsa→ Arabica, 3 Arabica→Excelsa), comprising 16 instances in total, suggests a persistent challenge in distinguishing these two varieties, likely due to their visual resemblance in their roasted form. Beyond this primary confusion, other classification errors were more sporadic: 2 Liberica samples were mistaken for other varieties (1 as Arabica, 1 as Robusta), and 2 Robusta samples were misclassified (1 as Arabica, 1 as Excelsa). Notably, Liberica and Robusta both achieved strong classification performance with 98 out of 100 samples accurately recognized. This outcome highlights the model's robust ability to detect and differentiate the distinctive visual traits of these varieties, even post-roasting, underscoring its reliability in handling these specific varieties. This implies that Excelsa and Arabica coffee seeds exhibit a visual resemblance following roasting. This complicates the model's ability to identify the two legumes. Future performance improvements should center on the model's ability to distinguish between these two classes since the misclassifications were especially focused on particular patterns (mostly between Excelsa and Arabica). This visualization is implemented using the local seaborn package, which provides complete graphing features with minimum coding, therefore streamlining data exploration and analysis.

Prediction result analysis, in order to gain a more comprehensive understanding of the model's performance, is conducted on the basis of the prediction result, whether it is accurate or not. Figure 2 illustrates examples of predictions that are highly accurate with high confidence, while Figure 3 illustrates various predictions. The model can rather confidently and precisely identify a range of coffee beans as in Figure 2. With 100% certainty, Liberica has typical bean traits including a rather bigger size, flatter form, and brighter color. The regular oval shape and straighter center spacing of Arabica beans make it feasible to predict their characteristics with a degree of confidence ranging from 96% to 98%. The 98.81% result is derived as a result of the rounder shape and irregular center gap of Robusta beans in the prediction.

TABLE IV
PERFORMANCE COMPARISON WITH STATE-OF-THE-ART METHODS

Study	Method	Bean Condition	Dataset Size	Varieties	Accuracy	Key Advantage	Limitation
Korkmaz et al. [12]	InceptionV3, DenseNet121, InceptionResNetV2, DenseNet201, Xception	Roasted	1,554 images (Starbucks Pike Place, Espresso, Kenya)	3 types	93.00% (InceptionV3)	Evaluation of multiple deep CNNs for commercial roasted beans	Only 3 classes, small dataset
Arwatchananukul et al. [22]	MobileNetV3, MobileNetV2, EfficientNetV2, InceptionV2, ResNetV2	Green	6,853 images (after augmentation)	17 defect classes	99.84% (cross-val), 88.63% (unseen)	Extremely high accuracy in defect detection, real deployment ready	Drop in accuracy on unseen data, specific to Thai Arabica
Hassan et al. [23]	VGG, MobileNetV2, EfficientNet, DenseNet, ResNet50, AlexNet, GoogleNet, etc.	Roasted (Light–Dark)	4,800 images (Kaggle)	4 roast levels (Green, Light, Medium, Dark)	100.00% (VGG)	Comprehensive model comparison, uses full metrics (F1, recall, etc.)	No real-world deployment yet, limited to roast level not origin
This Research	CNN, ResNet50	Post-roasting	2,000 images	Indonesian	94.50%	Real-world conditions	Initial Excelsa-Arabica confusion (improved with fine-tuning)

The improved accuracy of 94.50% achieved in this study demonstrates significant advancement over the previous iteration (89.65%) while maintaining the practical advantages of smartphone-based image capture under diverse real-world conditions. This performance enhancement can be attributed to refined data augmentation strategies, optimized hyperparameter tuning during the two-stage training process, and improved model convergence through careful learning rate scheduling. Notably, this accuracy level approaches the performance of studies conducted under controlled laboratory conditions (Hassan et al., 99-100%) while preserving the applicability to agricultural field settings.

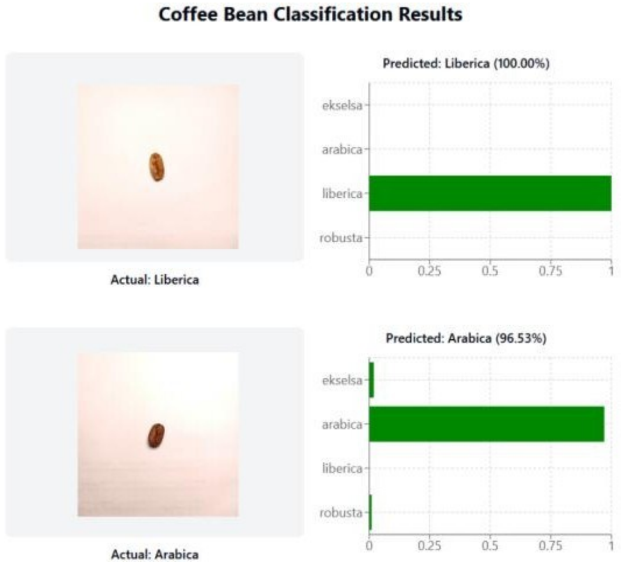


Figure 2 Correct Prediction Results

In local, this visualization of the prediction results was generated by combining matplotlib and model.predict(). Turning now to a more intricate case, Figure 3 shows the forecast failing of the model. Regarding the Excelsa case often categorized as Arabica the model offers a low chance for Excelsa (30-50%) and a high likelihood for Arabica (50-70%). Like the Arabica, the most often used classification of the Excelsa is defined by a more brilliant color and a clearer face profile that are highly accurate with high confidence, while Figure 3 illustrates various predictions. The model can rather confidently and precisely identify a range of coffee beans as in Figure 2. With 100% certainty, Liberica has typical bean traits including a rather bigger size, flatter form, and brighter color. The regular oval shape and straighter center spacing of Arabica beans make it feasible to predict their characteristics with a degree of confidence ranging from 96% to 98%. The 98.81% result is derived as a result based on the rounder shape and irregular center gap of Robusta beans in the prediction. In local, this visualization of the prediction results was generated by combining matplotlib and model.predict(). Turning now to a more intricate case, Figure 3 shows the forecast failing of the model. Regarding the Excelsa case often categorized as Arabica the model offers a low chance for Excelsa (30-50%) and a high likelihood for Arabica (50-70%). Like the Arabica, the most often used classification of the Excelsa is defined by a more brilliant color and a clearer face profile.

The result of the ResNet50 architecture CNN model is its ability to accurately identify the type of coffee after the roasting process. Liberica and Robusta are the most outstanding models, while Excelsa and Arabica are the most challenging models. The classification failure that is identified through confusion matrix analysis and prediction failure examples enumerates several factors that may contribute to the failure. The main consideration is the visual difference between Arabica following roasting and Excelsa. The roasting process produces a change in colour and texture that greatly influences the look of coffee, so making it more difficult to differentiate between several varieties. Variations in every class also influence classification mistakes including size, shape, visual qualities, type of variation, area of origin of the coffee beans. This is the reason the classification procedure is so complicated and could lead to class overlap. Still, the findings of this experiment reveal the huge potential of the deep learning approach for coffee bean recognition following roasting.

A comprehensive examination of the confusion matrix, which encompasses 400 test samples, reveals that the model successfully classified 378 instances, with a total of 22 misclassifications. Among all coffee bean classes, Liberica exhibited the highest classification accuracy, with 98 out of 100 samples correctly identified and only 2 instances misclassified (1 as Arabica and 1 as Robusta). Similarly, Robusta demonstrated strong performance with 98 correct predictions, accompanied by 2 misclassifications 1 incorrectly labeled as Arabica and another as Excelsa. Arabica was correctly identified in 96 cases, with the remaining 4 misclassified samples distributed as follows: 3 misidentified as Excelsa and 1 as Robusta. Notably, Excelsa posed the greatest classification challenge for the model. Although 86 samples were accurately classified, 14 were misclassified, including 13 samples incorrectly predicted as Arabica and 1 as Robusta. This consistent confusion between Excelsa and Arabica underscores a visual overlap in their roasted appearance, which likely impairs the model's ability to distinguish between the two. Further analysis indicates that the Excelsa-Arabica confusion accounts for 72.7% (16 out of 22) of all classification errors 13 Excelsa samples were misidentified as Arabica and 3 Arabica samples were misidentified as Excelsa emphasizing this as the primary source of performance degradation. This pattern suggests that Excelsa's post-roasting visual features may significantly resemble those of Arabica, complicating the classification process. To address this limitation, future model enhancement efforts should prioritize more sophisticated feature extraction techniques or the incorporation of additional training data that accentuates the nuanced visual differences between these two visually similar varieties.

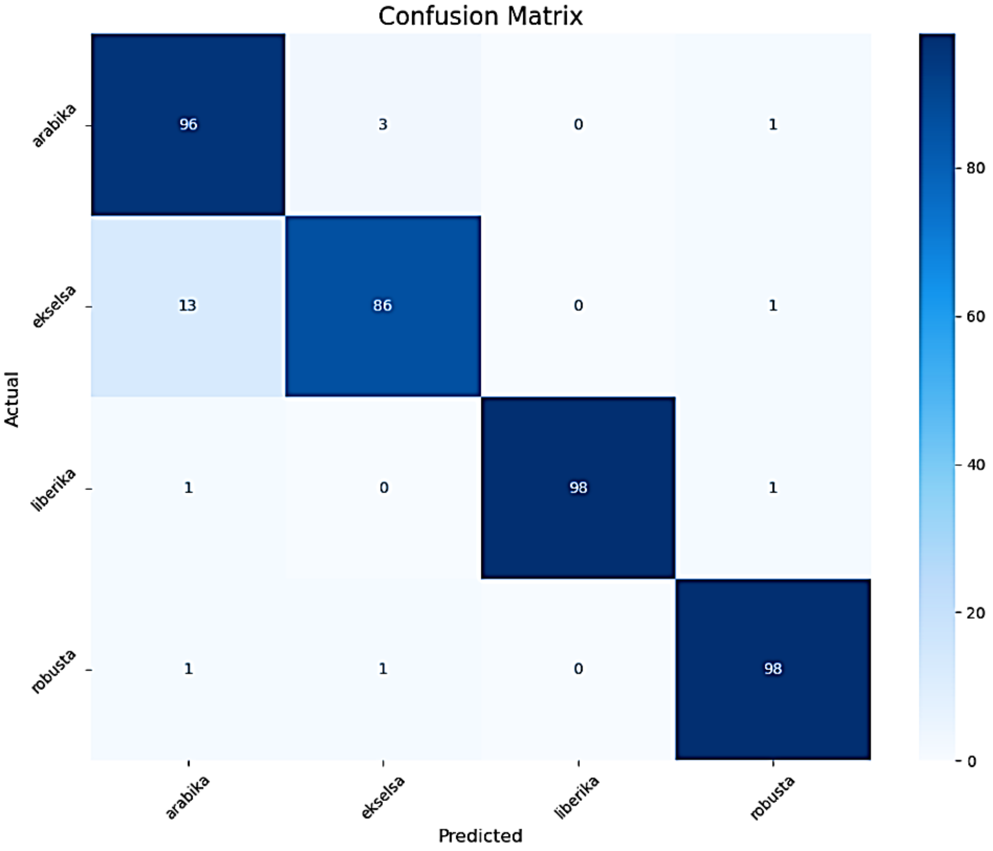


Figure 3 Confusion Matrix

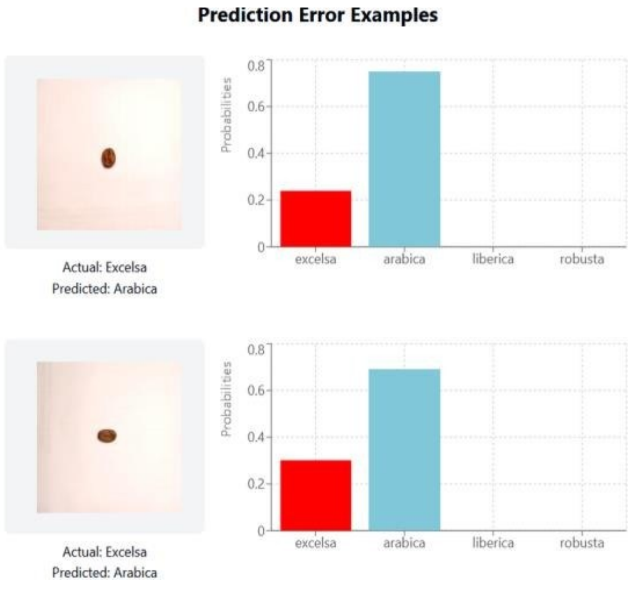


Figure 4 Prediction Errors

The findings of this study indicate that smartphone-based systems are capable to delivering satisfactory accuracy reaching 89.65% in identifying coffee bean varieties after the roasting process, suggesting potential for real-world application within coffee production facilities. This capability holds particular significance for coffee producers in Indonesia who may not have access to advanced laboratory-grade classification tools. Moreover, the observed difficulty in differentiating between Excelsa and Arabica beans offers valuable direction for future investigations, particularly in refining feature extraction techniques for visually similar post-roast varieties. The study also underscores the inherent trade-offs between achieving high classification performance and maintaining operational feasibility in agricultural environments. Despite the encouraging results, several limitations remain. These include a relatively small dataset and the exclusive focus on medium-roasted beans, which may limit the model's generalizability. Future research directions should focus on several key areas to address current limitations and enhance classification performance. First, implementing attention mechanisms such as Convolutional Block Attention Module (CBAM) or Squeeze-and-Excitation Networks (SE-Net) could improve the model's ability to focus on discriminative features, particularly for distinguishing between visually similar varieties like Excelsa and Arabica. Second, exploring multi-spectral imaging techniques that capture beyond-visible light characteristics (near-infrared or hyperspectral) could reveal chemical signatures not apparent in standard RGB images. Third, ensemble learning approaches combining ResNet50 with complementary architectures such as EfficientNet or Vision Transformers may improve overall robustness. Fourth, expanding the dataset to include multiple roasting levels (light, medium, dark) would enhance practical applicability across different coffee processing scenarios. Finally, model optimization for edge deployment using TensorFlow Lite or ONNX Runtime would enable real-time classification on resource-constrained mobile devices, facilitating widespread adoption in field conditions.

The accuracy of 94.50% achieved in this study demonstrates competitive performance for roasted coffee bean classification under real-world conditions using smartphone-based image acquisition. While slightly lower than some studies conducted under controlled laboratory conditions (Hassan et al., 100%; Arwatchananukul et al., 99.84%), this result reflects the inherent challenges of post-roasting classification where significant chemical and physical transformations obscure original visual traits. The use of a balanced dataset with 500 images per class ensured equitable representation across all varieties, preventing class-specific biases that can inflate overall accuracy metrics. The achieved performance is particularly noteworthy given the inclusion of Excelsa variety, which exhibits substantial visual similarity with Arabica post-roasting a classification challenge that remains underexplored in existing literature. The model's ability to maintain 86% recall for Excelsa while achieving perfect precision (100%) for Liberica demonstrates successful learning of discriminative features despite these challenges, validating the effectiveness of the ResNet50 architecture combined with two-stage transfer learning for this application domain.

TABLE V
COMPARISON OF COFFEE BEAN TYPE CLASSIFICATION RESEARCH

Aspect	This Research (2025)	Bibangco et al. (2024)	Rame et al. (2024)
Model	CNN based on ResNet50	5 different CNN architectures	MobileNetV1
Dataset	2,000 images (500/class)	12,000 images (3,000/class)	12,000 images (3,000/class)
Classes	Arabica, Robusta, Excelsa, Liberica	Arabica, Robusta, Excelsa, Liberica	Arabica, Robusta, Excelsa, Liberica
Accuracy	94.50%	90.00%	94.33%
Best Performance	Liberica (100%) – precision	–	Liberica (99%), Arabica (98%)
Lowest Performance	Excelsa (86% recall)	–	Excelsa (88%)
Main Advantages	Comprehensive data augmentation; two-stage transfer learning (transfer learning and fine-tuning); detailed error analysis (per-class metrics, confusion matrix)	Comparison of 5 architectures, balanced dataset, comprehensive evaluation	Lightweight and efficient model, optimized for resource-constrained environments, high accuracy
Main Challenges	Excelsa–Arabica is difficult to distinguish; high computational requirements; Excelsa is difficult to distinguish from other varieties	High computational requirements	Difficulty distinguishing Excelsa from other varieties

IV. CONCLUSION

This model successfully identified coffee beans with an overall accuracy of 94.50%, correctly classifying 378 out of 400 test samples, demonstrating that the ResNet50-based CNN architecture combined with comprehensive data augmentation and two-stage transfer learning is effective for classifying roasted coffee varieties despite significant visual changes induced by the roasting process. The use of a balanced dataset with 2,000 images (500 per class) ensured equitable model training across all varieties, while the two-stage training approach transfer learning followed by fine-tuning enabled effective feature adaptation from the ImageNet pretrained weights to the specific domain of roasted coffee bean classification. The performance of the model demonstrates strong and relatively balanced results across all coffee varieties, with Liberica achieving perfect precision (100%) and high recall (98%), and Robusta demonstrating excellent performance with 97% precision and 98% recall. Excelsa presented the primary classification challenge with 86% recall, reflecting the inherent difficulty in distinguishing this variety from visually similar beans, particularly Arabica, after the roasting process. Arabica achieved 96% recall with 86.5% precision, indicating high sensitivity but occasional confusion with other varieties. These results validate that while significant visual transformations occur during roasting, the CNN model successfully learned discriminative features for most samples, with the remaining classification challenges concentrated in varieties exhibiting substantial post-roasting visual overlap.

REFERENCES

[1] A. Zikra and others, “Analisis Potensi Komoditas Kopi Kabupaten/Kota di Indonesia Tahun 2015: Metode Fuzzy C-Medoid Clustering,” *J. Agric. Socio-Econ.*, 2022, doi: 10.33474/jase.v3i1.

[2] A. Pradhana and others, “Fostering Coffee-Minds by Developing Customer Perspective from Simple Public Cupping: Study Case in Bumi Kopi, Malang,” in *BIO Web of Conferences*, EDP Sciences, Mar. 2025. doi: 10.1051/bioconf/202516507002.

[3] N. Happyana, Y. M. Syah, and E. H. Hakim, “Discrimination of Metabolite Profiles of Gayo Roasted Arabica and Robusta Coffees,” *Molekul*, vol. 17, no. 1, pp. 98–106, Mar. 2022, doi: 10.20884/1.jm.2022.17.1.5603.

[4] I. Santoso, S. A. Mustaniroh, and A. Choirun, “Methods for quality coffee roasting degree evaluation: A literature review on risk perspective,” in *IOP Conference Series: Earth and Environmental*

Science, IOP Publishing Ltd, Dec. 2021. doi: 10.1088/1755-1315/924/1/012058.

[5] B. R. Santoso, C. A. Sari, and E. H. Rachmawanto, “Coffee Beans Classification Using Convolutional Neural Networks Based On Extraction Value Analysis In Grayscale Color Space.” 2025. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[6] A. E. Ilesanmi and T. O. Ilesanmi, “Methods for image denoising using convolutional neural network: a review,” *Complex Intell. Syst.*, vol. 7, no. 5, pp. 2179–2198, Oct. 2021, doi: 10.1007/s40747-021-00428-4.

[7] E. T. Cortés-Macías, C. F. López, P. Gentile, J. Girón-Hernández, and A. F. López, “Impact of post-harvest treatments on physicochemical and sensory characteristics of coffee beans in Huila, Colombia,” *Postharvest Biol. Technol.*, vol. 187, May 2022, doi: 10.1016/j.postharvbio.2022.111852.

[8] Syafriandi, A. Lubis, R. Fadhil, and O. Paramida, “Characteristics of roasting arabica and robusta coffee beans with rotary cylinder tube roast machine with electric heat source,” in *IOP Conference Series: Earth and Environmental Science*, Institute of Physics, 2022. doi: 10.1088/1755-1315/1116/1/012032.

[9] I. A. F. Anto, J. W. Wibowo, T. I. Salim, and A. Munandar, “Implementation of Image Processing and CNN for Roasted-Coffee Level Classification,” *Indones. J. Electr. Eng. Inform.*, vol. 12, no. 4, pp. 1005–1018, Dec. 2024, doi: 10.52549/ijeei.v12i4.5531.

[10] D. H. Suryana and W. K. Raharja, “Applying Artificial Intelligence to Classify the Maturity Level of Coffee Beans During Roasting,” 2023, doi: 10.52088/ijesty.v1i4.461.

[11] C. Janiesch, P. Zschech, and K. Heinrich, “Machine learning and deep learning,” 2021, doi: 10.1007/s12525-021-00475-2.

[12] A. Korkmaz, T. Talan, S. Koşunalp, and T. Iliev, “Comparison of deep learning models in automatic classification of coffee bean species,” *PeerJ Comput. Sci.*, vol. 11, 2025, doi: 10.7717/peerj-cs.2759.

[13] M. D. A. Putra, T. S. Winanto, R. Hendrowati, A. Primajaya, and F. D. Adhinata, “A Comparative Analysis of Transfer Learning Architecture Performance on Convolutional Neural Network Models with Diverse Datasets,” *Komputika J. Sist. Komput.*, vol. 12, no. 1, pp. 1–11, May 2023, doi: 10.34010/komputika.v12i1.8626.

[14] E. Firdissa, A. Mohammed, G. Berecha, and W. Garedew, “Coffee Drying and Processing Method Influence Quality of Arabica Coffee Varieties (Coffee arabica L.) at Gomma I and Limmu Kossa, Southwest Ethiopia,” *J. Food Qual.*, pp. 1–8, Jan. 2022, doi: 10.1155/2022/9184374.

[15] M. Nazari and others, “Explainable AI to improve acceptance of convolutional neural networks for automatic classification of dopamine transporter SPECT in the diagnosis of clinically uncertain parkinsonian syndromes,” *Eur. J. Nucl. Med. Mol. Imaging*, vol. 49, no. 4, pp. 1176–1186, Mar. 2022, doi: 10.1007/s00259-021-05569-9.

[16] A. C. Ramdhana and N. Pratiwi, “Perbandingan Kinerja Model Convolutional Neural Network pada Klasifikasi Kanker Kulit,” *Edumatic J. Pendidik. Inform.*, vol. 7, no. 2, pp. 197–206, Dec. 2023, doi: 10.29408/edumatic.v7i2.19823.

[17] R. A. Mas’ud and J. Zeniarja, “Optimasi Convolutional Neural Networks untuk Deteksi Kanker Payudara menggunakan Arsitektur DenseNet,” *Edumatic J. Pendidik. Inform.*, vol. 8, no. 1, pp. 310–318, Jun. 2024, doi: 10.29408/edumatic.v8i1.25883.

[18] V. Singh and others, “A Hybrid Deep Learning Model for Enhanced Structural Damage Detection: Integrating ResNet50, GoogLeNet, and Attention Mechanisms,” *Sensors*, vol. 24, no. 22, Nov. 2024, doi: 10.3390/s24227249.

[19] C. Wang and others, “Meta Distant Transfer Learning for Pre-trained Language Models.” Nov. 2021. doi: 10.18653/v1/2021.emnlp-main.768.

[20] M. Xu, S. Yoon, A. Fuentes, and D. S. Park, “A Comprehensive Survey of Image Augmentation Techniques for Deep Learning,” *Pattern Recognit.*, vol. 137, May 2023, doi: 10.1016/j.patcog.2023.109347.

[21] A. K. Ulandari, F. Bimantoro, and I. G. P. S. Wijaya, “Real Time Student Emotion Detection using Yolov5,” *Edumatic J. Pendidik. Inform.*, vol. 8, no. 1, pp. 222–231, Jun. 2024, doi: 10.29408/edumatic.v8i1.25726.

[22] S. Arwatchananukul, D. Xu, P. Charoenkwan, S. A. Moon, and R. Saengrayap, “Implementing a deep learning model for defect classification in Thai Arabica green coffee beans,” *Smart Agric. Technol.*, vol. 9, Dec. 2024, doi: 10.1016/j.atech.2024.100680.

[23] E. Hassan, “Enhancing coffee bean classification: a comparative analysis of pre-trained deep learning models,” Jun. 2024, doi: 10.1007/s00521-024-09623-z.

[24] Y. Nie and others, “An improved CNN model in image classification application on water turbidity,” *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-93521-4.

[25] M. M. Taresh, N. Zhu, T. A. A. Ali, M. Alghaili, A. S. Hameed, and M. L. Mutar, “KL-MOB: automated COVID-19 recognition using a novel approach based on image enhancement and a modified MobileNet CNN,” *PeerJ Comput. Sci.*, vol. 7, pp. 1–23, 2021, doi: 10.7717/PEERJ-CS.694.

[26] W. Juslan and A. H. Muhammad, “Evaluasi Kinerja Metode Peningkatan Kontras (CLAHE & HE) pada Klasifikasi Ras Kucing menggunakan VGG16,” *Edumatic J. Pendidik. Inform.*, vol. 9, no. 1, pp. 246–255, 2025, doi: 10.29408/edumatic.v9i1.29578.

[27] A. Gupta, P. Pawade, and R. Balakrishnan, “Deep Residual Network and Transfer Learning-based Person Re-Identification,” *Intell. Syst. Appl.*, vol. 16, Nov. 2022, doi: 10.1016/j.iswa.2022.200137.

[28] S. R. Sannasi Chakravarthy, N. Bharanidharan, C. Vinothini, V. Vinoth Kumar, T. R. Mahesh, and S. Guluwadi, “Adaptive Mish activation and ranger optimizer-based SEA-ResNet50 model with explainable AI for multiclass classification of COVID-19 chest X-ray images,” *BMC Med. Imaging*, vol. 24, no. 1, p. 206, Aug. 2024, doi: 10.1186/s12880-024-01394-2.

[29] I. Salehin and D. K. Kang, “A Review on Dropout Regularization Approaches for Deep Neural Networks within the Scholarly Domain,” Jul. 2023, doi: 10.3390/electronics12143106.

[30] S. Sathyanarayanan, “Confusion Matrix-Based Performance Evaluation Metrics,” *Afr. J. Biomed. Res.*, pp. 4023–4031, Nov. 2024, doi: 10.53555/AJBR.v27i4S.4345.

[31] O. Rainio, J. Teuho, and R. Klén, “Evaluation metrics and statistical tests for machine learning,” *Sci. Rep.*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-56706-x.