Smart Agricultural Technology 9 (2024) 100680

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Implementing a deep learning model for defect classification in Thai Arabica green coffee beans

Sujitra Arwatchananukul [a,c], Dan Xu [d], Phasit Charoenkwan [e], Sai Aung Moon [b], Rattapon Saengrayap [b,c,f,*]

[a] School of Applied Digital Technology, Mae Fah Luang University, Chiang Rai, Thailand
[b] School of Agro-Industry, Mae Fah Luang University, Chiang Rai, Thailand
[c] Integrated AgriTech Ecosystem Research Group (IATE), Mae Fah Luang University, Thailand
[d] School of Information Science and Engineering, Yunnan University, Kunming, PR China
[e] Modern Management and Information Technology, College of Arts, Media and Technology, Chiang Mai University, Chiang Mai, Thailand
[f] Coffee Quality Research Group (CQR), Mae Fah Luang University, Thailand

ARTICLE INFO

Keywords:
Green Bean Defects
CNN
MobileNetV3
Artificial intelligence
Non-Destructive
Web Application

ABSTRACT

Arabica coffee is a significant economic driver in Northern Thailand and has substantial opportunities for market growth. However, the Thai coffee business must ensure consistent quality standards and is currently heavily dependent on manual labor, to first identify, and then remove substandard unroasted coffee beans. This research developed a classification model based on a Convolutional Neural Network to detect 17 types of defects in green coffee beans. The image augmentation phase was enhanced by rotating images at 45, 90, 135, 180, 225, and 270° and expanding the dataset from 979 original images across 17 defect types to a robust 6,853 images. Several architectures including MobileNetV2, MobileNetV3, EfficientNetV2, InceptionV2, and ResNetV2 were assessed. Following extensive evaluations, MobileNetV3 emerged as the best-performing model and underwent further fine-tuning, achieving significant accuracy improvements through hyperparameter optimization. The model's robustness and generalizability were validated via 5-fold cross-validation, with accuracy ranging from 98.78 % to 99.84 % across all defect types. When tested with unseen data, the model achieved an accuracy of 88.63 %. A web application prototype was also developed for real-time coffee bean defect classification and its usability was tested. Seven farmers reported high satisfaction with the ease of use and effectiveness of the application in classifying coffee bean defects, with 71.4 % expressing a strong likelihood of recommending the application to others. These promising results demonstrate the practical utility of the model in enhancing quality sorting processes in the coffee industry.

1. Introduction

Arabica coffee is an important economic crop in Northern Thailand and the most popular species for cultivation and consumption. This species thrives in varied conditions of humidity, temperature, wind, rain, and altitude, allowing the harvest of many varieties [1]. Arabica coffee shows potential for expanding into domestic and international markets; however, the Thai coffee market lacks integrated knowledge to establish quality standards and identity. Most coffee drinkers are unaware of mycotoxins, natural poisons that are highly toxic to the kidneys and possibly carcinogenic, produced by certain fungi. These toxins are found on raw or green coffee beans and remain after roasting due to their thermoresistant nature [2].

On a global scale, the coffee trade has encountered notable variations in production and consumption. However, recent estimates suggest a recovery in coffee production during 2022–2023. The International Coffee Organization (ICO) reported a growth of 1.7 % in coffee production to 171.3 million bags, with a significant 4.2 % increase in consumption at 175.6 million bags in 2021–2022 [3], highlighting the increasing importance of coffee as a crucial commodity within the beverage industry. The interplay between coffee production and demand dynamics significantly influences worldwide trade patterns. To uphold the integrity and competitiveness of the coffee business, it is crucial to meticulously classify and identify each green coffee bean to

* Corresponding author.
E-mail address: rattapon.sae@mfu.ac.th (R. Saengrayap).

https://doi.org/10.1016/j.atech.2024.100680
Received 26 August 2024; Received in revised form 11 November 2024; Accepted 29 November 2024
Available online 30 November 2024

ensure the highest quality. This requires a detailed evaluation of flaws and overall standards to achieve increased production by optimizing manufacturing schedules and enhancing quality control protocols [4]. This study explored the crucial significance of precise categorization and recognition of unroasted coffee beans to guarantee the highest level of excellence, productivity, and regulation within the coffee sector.

Modern technology now plays a crucial role in enhancing and improving the efficiency of various aspects of the agricultural industry. Technologies such as Artificial Intelligence (AI) for analyzing plant growth data and the Internet of Things (IoT) for monitoring and controlling environmental cultivation conditions have significantly contributed to increased productivity by reducing losses through the efficient use of resources [5–7]. Machine learning and deep learning applications are now widely used to detect and classify defects in agricultural products [8], with convolutional neural networks (CNN) used to analyze images, detect defects, and classify foreign objects. These technologies greatly enhance accuracy and reduce the time required for sorting and quality control. High-quality coffee beans are essential for producing coffee with good taste and the classification of defective coffee beans is a crucial step in the coffee production process [9,10]. However, traditional methods relying on manual labor and visual inspection have limited accuracy and efficiency. The application of modern technology, particularly research in information technology, has played a crucial role in developing systems for classifying large quantities of coffee beans. Deep learning techniques, specifically CNN [11, 12], can be used to accurately classify different types of coffee bean defects with significant improvements in accuracy and efficiency compared to manual evaluation [13,14].

Several studies demonstrated that using CNN technology significantly improved the classification of coffee beans. Chang and Huang [13] utilized CNN to distinguish defective coffee beans with highly accurate results, while [14] applied the k-nearest neighbor (KNN) algorithm to categorize high-quality green coffee beans and detect sour, very long-berry, black, broken, and small defects. This was achieved using image processing and machine learning techniques which analyzed bean color, morphology, shape, and size. This artificial vision technique improved the process of discerning high-quality coffee beans and enhanced production by improving quality management. Machine vision systems identify defects in coffee beans by specific defect type including sour, very long berry, black, tiny, or broken. Several machine vision systems have been proposed for selecting green coffee beans and assessing their quality based on various physical criteria. The utilization of deep learning models in computer vision has proven to be advantageous in the identification and categorization of coffee fruits based on their ripeness levels, namely unripe, ripe, and overripe. This technology maps the maturation stage during the harvest process, hence facilitating efficient data collection, effective crop management, and successful integration of coffee harvesters into agricultural practices [15]. In 2022, [16] demonstrated that utilizing gray-level co-occurrence matrices (GLCM) and the KNN algorithm on a web-based platform gave a 90 % success rate in classifying defects in coffee beans. This approach offers potential benefits to coffee farmers by assessing levels of coffee bean defects through image input. Artificial neural networks (ANN) and the KNN algorithms were used to categorize coffee bean species from various geographical origins, demonstrating the efficacy of these image approaches as reliable ways for classifying coffee bean species. A study by [17], showed that the ANN method outperformed the KNN strategy in coffee bean identification when utilizing the created dataset. Hyperspectral imaging was also used to categorize the qualitative characteristics of coffee beans affected by roasting defects [18]]. These defects encompassed medium roasting, underdeveloped roast level, and oven roasted level as well as interior underdeveloped and interior scorching levels. Correlations were found between these defect levels and alterations in coffee bean moisture content, browning index, chlorogenic acid, trigonelline, and caffeine content. Image processing and data augmentation techniques were also examined in conjunction with deep learning methodologies, specifically CNN, to analyze picture data and extract relevant information [19]. Grayscale images are susceptible to the influence of ambient light due to the limited retention of one-dimensional image information. Connecting the model with an IP camera can promptly discern the quality attributes of green coffee beans, reducing the labor-intensive aspects associated with this process. A CNN model was used to classify coffee bean defects into six categories: black, sour, faded, peaberry, damaged, and normal [20]]. High classification accuracy (over 90 %) was achieved for black and sour beans due to their distinct coloration while other categories had lower accuracy (72 %). The exact factors for lower accuracy were unclear but color characteristics significantly influenced the results. In 2023, researchers [21] classified 14 dry bean cultivars using deep CNN features and an extreme learning machine (ELM) model optimized using the salp swarm algorithm (SSA). Utilizing GoogLeNet for transfer learning, the SSA-ELM model achieved a 91.43 % success rate, outperforming traditional machine learning algorithms like support vector machines (SVM) and KNN. The results highlighted the effectiveness of using deep feature extraction and optimized ELM in agricultural applications.

The categorization of imperfections in coffee beans has the utmost significance for various pivotal rationales by exerting a direct influence on their general caliber and economic worth. Various defects including insect damage, mold, and overfermentation negatively impact the sensory attributes, namely flavor, fragrance, and appearance of the resultant coffee beverage. Precise categorization guarantees that all beans fulfill established quality criteria. This process protects the reputation of coffee producers and upholds the faith placed in them by consumers in the industry. The classification of defects contributes to the optimization of production processes. Early identification and classification of problems within the supply chain allows farmers and processors to make well-informed decisions on the sorting, processing, and implementation of quality control measures. This results in enhanced operational efficiency and mitigates the squandering of resources, time, and exertion. The precise categorization of defects also promotes fair and just trading procedures. Quality grades are frequently used as a basis for trading coffee beans in the worldwide coffee market. The implementation of accurate categorization practices plays a crucial role in establishing equitable pricing mechanisms and safeguards coffee farmers from exploitation. A proper classification process improves economic viability and long-term sustainability for farmers by effectively differentiating between coffee beans of different grades. Proper categorization of defects also facilitates research and development endeavors. Correct identification of common faults and their underlying causes enables researchers to devise focused interventions, thereby enhancing farming practices, processing methods, and the overall quality of the beans. The precise categorization of coffee bean imperfections is important to uphold product excellence and enhance production methodologies, fostering equitable trading practices, and propelling advancements in research endeavors. Accurate coffee bean categorization also ensures sustainable and profitable coffee production to meet customer expectations and adhere to industry standards.

Thus, this study developed a database for 17 types of Arabica green coffee bean defects: Broken, Cut, Dry Cherry, Fade, Floater, Full Black, Full Sour, Fungus Damage, Husk, Immature, Parchment, Partial Black, Partial Sour, Severe Insect Damage, Shell, Slight Insect Damage, and Withered. Deep learning techniques were used to classify these defects and the model was integrated into a web application. The paper is organized as follows: Section 2, Dataset Creation, details how the coffee bean database was constructed. Section 3, Methodology, discusses the research steps. Section 4 describes the experiments conducted and discusses the results while Section 5 presents the study conclusions

2. Arabica green coffee bean defects dataset

The unique features of the 17 types of Thai Arabica green coffee bean defects are illustrated in Fig. 1, which represents one of the 17 types in

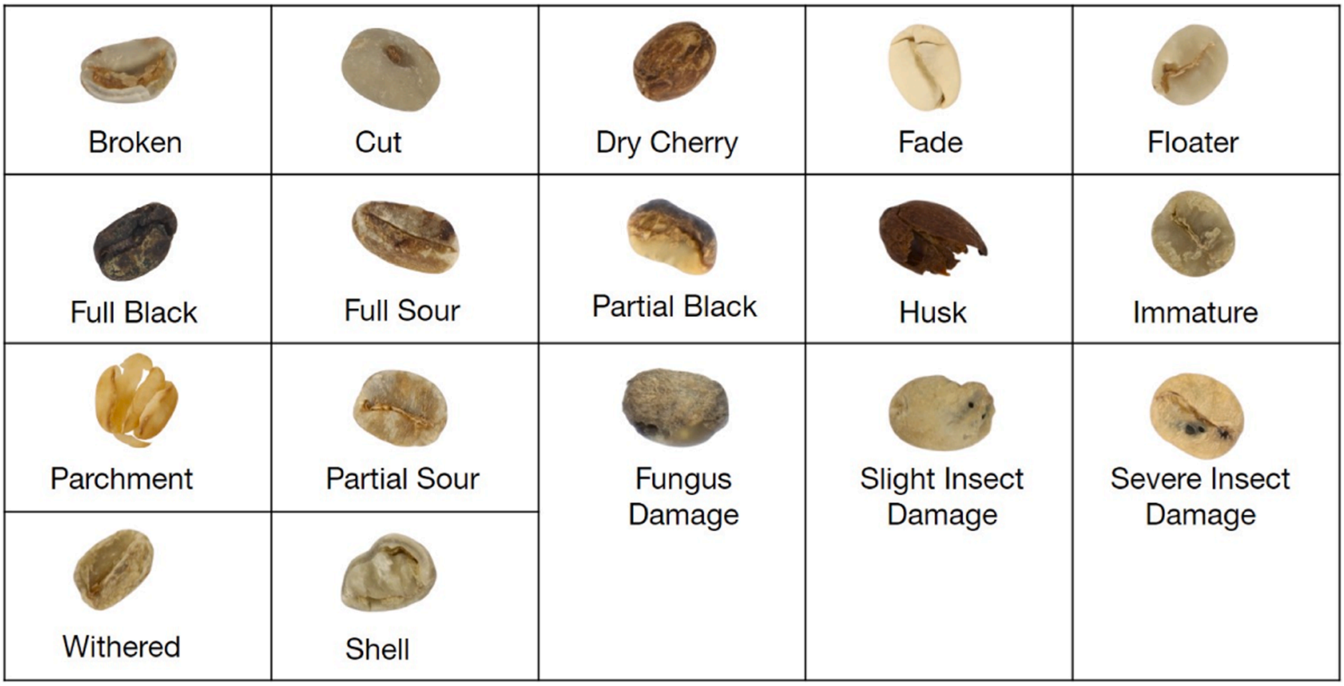

Fig. 1. The 17 Types of Arabica green coffee bean defects.

the dataset, including broken, cut, dry cherry, fade, floater, full black, full sour, fungus damage, husk, immature, parchment, partial black, partial sour, severe insect damage, shell, slight insect damage, and withered.

2.1. Image characteristics of Arabica green coffee bean defects

The initial stage in preserving optimal green coffee beans involves eliminating beans with defects. Defective green coffee beans produced substandard cup quality, deemed unacceptable by Q-graders. These defects significantly impact the low-cost market and can adversely affect human health [22]. Physical defects in green coffee beans can arise from cultivation, harvesting, processing, storage, and insect exposure. Observed defects include full black, full sour, fungus damage, severe insect infestation, dried cherry, partial black, immature, partial sour, slight insect damage, broken, cut, wet mill, withered, shell, floater, parchment, and husk [23]. Physical defects in green coffee beans arise from cultivation, harvesting, processing, storage, and insect exposure. Observed defects include full black, full sour, fungus damage, severe insect infestation, dried cherry, partial black, immature, partial sour, slight insect damage, broken, cut, wet mill, withered, shell, floater, parchment, and husk [23]. Based on the grading procedures outlined by the Specialty Coffee Association of America (SCAA), green coffee beans with category 1 faults are considered unacceptable. These defects include full black, full sour, dried cherry, fungus damage, foreign matter, and severe insect damage. Category 2 defects include partial black, partial sour, parchment, floater, immature/unripe, withered, shell, broken/chipped/cut, hull/husk, and slight insect damage. Therefore, after dehulling and obtaining the green coffee beans, all defective beans must be meticulously separated to uphold the highest coffee quality standards.

3. Methodology

The research methodology was divided into five key steps, as depicted in Fig. 2. Initially, images of the 17 types of Thai Arabica green coffee bean defects were collected and accurately labeled. Second, image augmentaion involved resizing and rotating the images at 45-degree intervals, expanding the dataset from 979 to 6853 images. The data were then partitioned into training (70 %), validation (20 %), and testing (10 %) sets. Third, model building utilized CNN architectures including MobileNetV2, MobileNetV3, EfficientNetV2, InceptionV2, and ResNetV2. Fourth, the model was validated by 5-fold cross-validation and hyperparameter tuning to ensure robustness. Finally, the best model was deployed via a Flask API, enabling users to upload images and receive classification results.

3.1. Image data collection and labeling

The first step in preparing a dataset for machine learning involved classifying the green coffee beans into 17 types: broken, cut, dry cherry, fade, floater, full black, full sour, fungus damage, husk, immature, parchment, partial black, partial sour, severe insect damage, shell, slight insect damage, and withered. Each type was photographed using a controlled setup with a white background and uniform lighting to ensure consistency across images. After photographing all the types of green coffee beans, each image was cropped to 500-pixel by 500-pixel size to standardize the dataset. This process was efficiently executed using Photoshop to handle the large volume of images. The original dataset, comprising images of green coffee bean defects, is available on Kaggle at https://www.kaggle.com/datasets/sujitraarw/coffee-green-bean-with-17defects-original. The dataset includes images meticulously labeled and prepared as described, facilitating robust machine learning model training. The original number of images for each green coffee bean defect type is shown in Table 1.

3.2. Image augmentation

Image Augmentation is a crucial step in preparing data for machine learning models, especially to address challenges related to data quality and imbalance. One effective technique is image rotation, a form of data augmentation that increases the diversity of the training set without having to collect new data. The images are rotated at specified angles, creating new perspectives and enhancing the model's ability to generalize from different orientations. This technique is particularly beneficial for datasets where the subject's orientation may vary, such as in

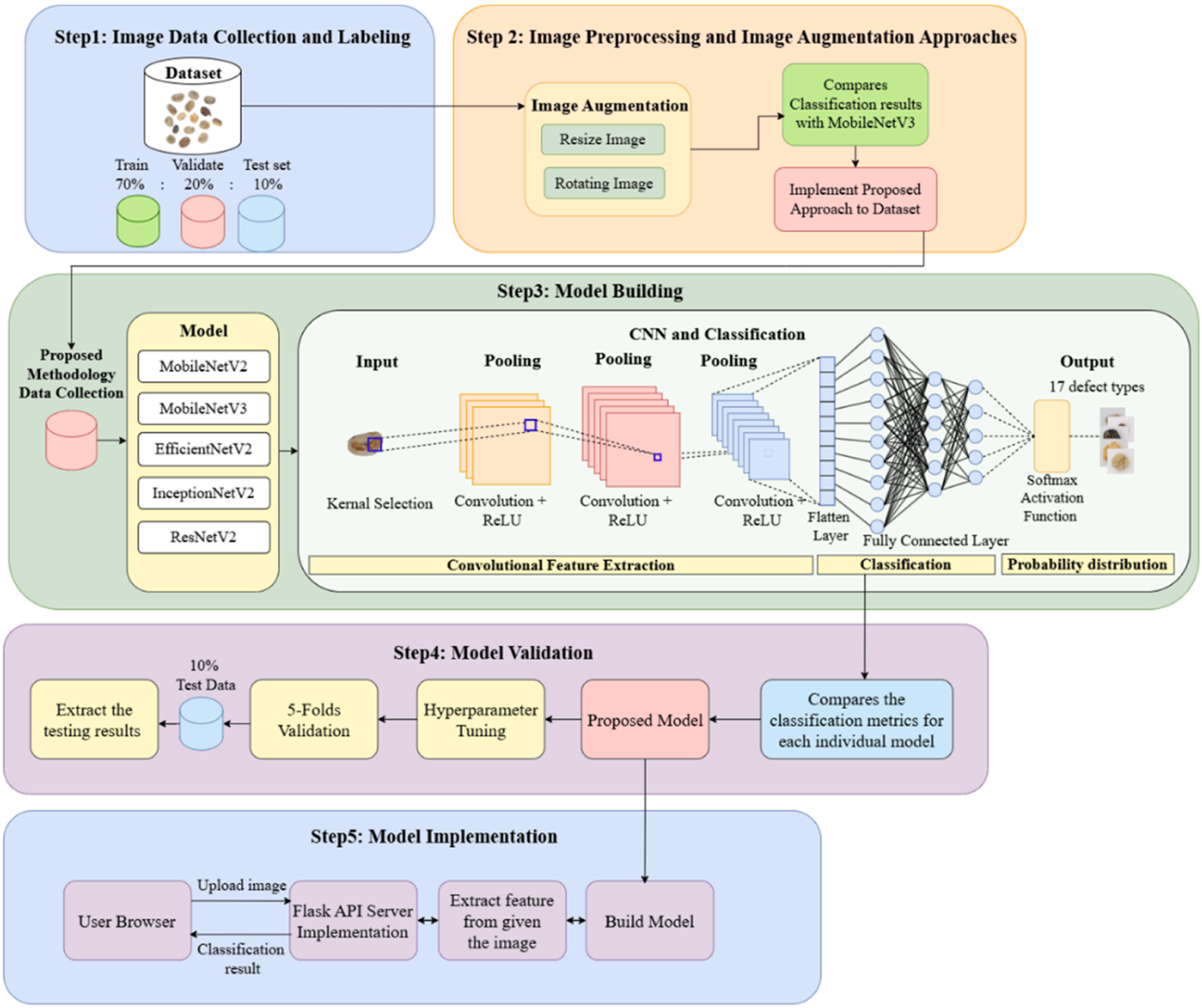

Fig. 2. Methodology workflow for green Coffee bean defect classification.

Table 1
The original dataset of Green Coffee bean defects.

No.	Defect Type	Number of Image
1	Broken	62
2	Cut	66
3	Dry Cherry	54
4	Fade	35
5	Floater	48
6	Full Black	41
7	Full Sour	75
8	Fungus Damage	75
9	Husk	53
10	Immature	78
11	Parchment	54
12	Partial Black	65
13	Partial Sour	50
14	Severe Insect Damage	57
15	Shell	57
16	Slight Insect Damage	55
17	Withered	54

agricultural product classification [24]. Rotation addresses issues of viewpoint variability and helps the model learn invariant features, leading to improved robustness and performance. By rotating images, the model is exposed to various orientations of the same object, which enhances its ability to correctly identify and classify objects regardless of their angle. This is particularly important in real-world applications where objects may not always be presented in a standardized orientation.

Image rotation was applied to augment the Arabica green coffee bean defects dataset. Fig. 3 shows the original images were rotated at (a) 45°, (b) 90°, (c) 135°, (d) 180°, (e) 225°, and (f) 270° This augmentation process expanded the dataset, as shown in Table 2, and enhanced the model's ability to recognize defects from different perspectives. In summary, the framework for identifying the most suitable augmentation techniques for the coffee bean dataset encompasses two distinct sets: the original dataset and the augmentation dataset with rotation technique. The optimal augmentation technique tailored to the dataset's specific needs will be determined following the comprehensive evaluation of these sets using MobileNetV2 and subsequent analysis via a classification report.

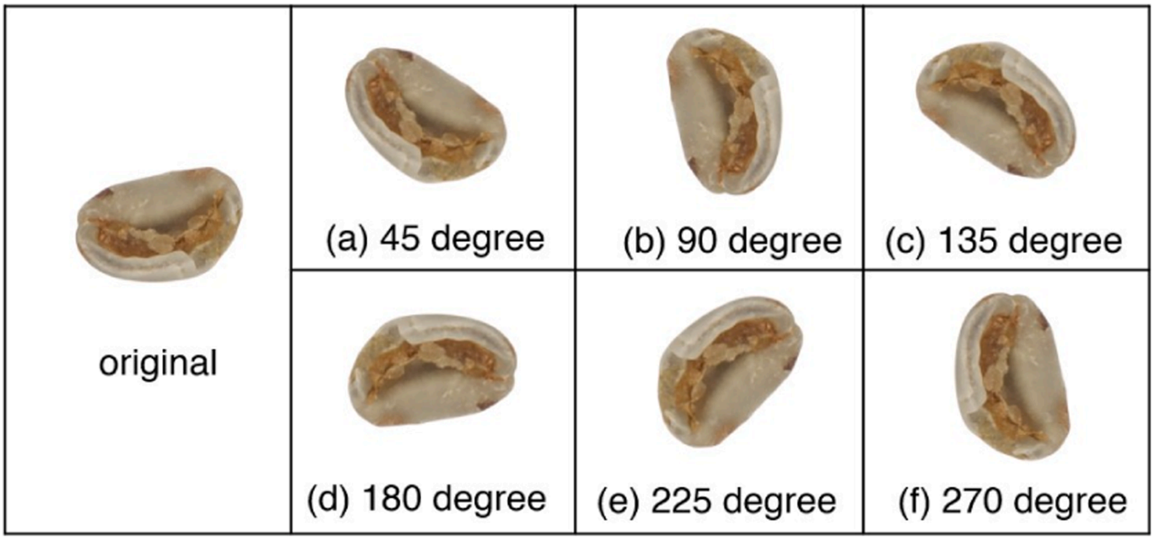

Fig. 3. Augmentation of Arabica green coffee bean defect images through rotations at various degrees: (a) 45°, (b) 90°, (c) 135°, (d) 180°, (e) 225°, and (f) 270°

Table 2
Number of images of Green Coffee bean defects.

Defect Type	Original	After Rotation	Total
Broken	62	372	434
Cut	66	396	462
Dry Cherry	54	324	378
Fade	35	210	245
Floater	48	288	336
Full Black	41	246	287
Full Sour	75	450	525
Fungus Damage	75	450	525
Husk	53	318	371
Immature	78	468	546
Parchment	54	324	378
Partial Black	65	390	455
Partial Sour	50	300	350
Severe Insect Damage	57	342	399
Shell	57	342	399
Slight Insect Damage	55	330	385
Withered	54	324	378
Total:	979	5874	6853

3.3. Model building

Following the collection and augmentation of the dataset through rotation techniques, we organized the data into three primary segments: 70 % was allocated to the training set, with the remaining 30 % equally divided between validation (20 %) and testing sets (10 %). Convolutional neural networks (CNN) can effectively capture image spatial hierarchies and process data in multiple layers, replicating the function of the human visual cortex. CNN can recognize patterns and characteristics at multiple levels of abstraction to attain a deeper understanding of the processed images [25,26]. This study employed five advanced CNN architectures to take advantage of their specific strengths in feature extraction and computational efficiency including:

- MobileNetV2 and MobileNetV3: These models are renowned for their optimal balance of accuracy and efficiency, making them suitable for deployment on mobile devices where computational resources are limited [27,28].
- EfficientNetV2: This architecture is designed to scale the model size efficiently, leading to enhanced performance and reduced resource usage [29].
- InceptionV2: By incorporating inception modules, this network captures information at varying scales, which enhances its ability to generalize and adapt to new data [30].
- ResNetV2: Known for its implementation of residual learning, ResNetV2 facilitates the training of significantly deeper networks than previously possible by simplifying the learning process [31].

For the CNN architectures in this study, all models—MobileNetV2, MobileNetV3, EfficientNetV2, InceptionV2, and ResNetV2—using their default pre-trained weights from the ImageNet dataset, without any additional changes to the internal architecture, allowing for a direct assessment of their baseline performance on our dataset. These default pre-trained models provided a robust foundation by leveraging general visual features learned from a large, diverse dataset, reducing the need for extensive task-specific training data and computational resources. The architecture and transfer learning process for these models are illustrated in Fig. 4. After testing each model's performance, the model with the highest accuracy was selected for further hyperparameter tuning.

3.4. Hyperparameter tuning and model validation

During this stage, the study concentrated on enhancing the optimal model by fine-tuning the hyperparameters and performing cross-validation to validate robustness. Hyperparameter tuning optimizes the settings of a machine learning model that are not learned directly from the data but significantly affect its learning dynamics. This study optimized model performance by experimenting with different numbers of epochs and setting the learning rate at 0.01.

A 5-fold cross-check validation was used to verify model reliability and generalizability. This method tested the model on different subsets of data by rotating the validation set across five distinct folds. Performance metrics like validation accuracy, loss, precision, recall, and F1score were averaged to achieve a reliable assessment of model effectiveness. This technique ensured that the model delivered consistent performance across various data scenarios without being overly fitted to the training dataset.

3.5. Web application model

The most effective model was implemented into a web application designed for practical use that enabled users to upload images of green coffee beans and receive real-time classification results, thereby

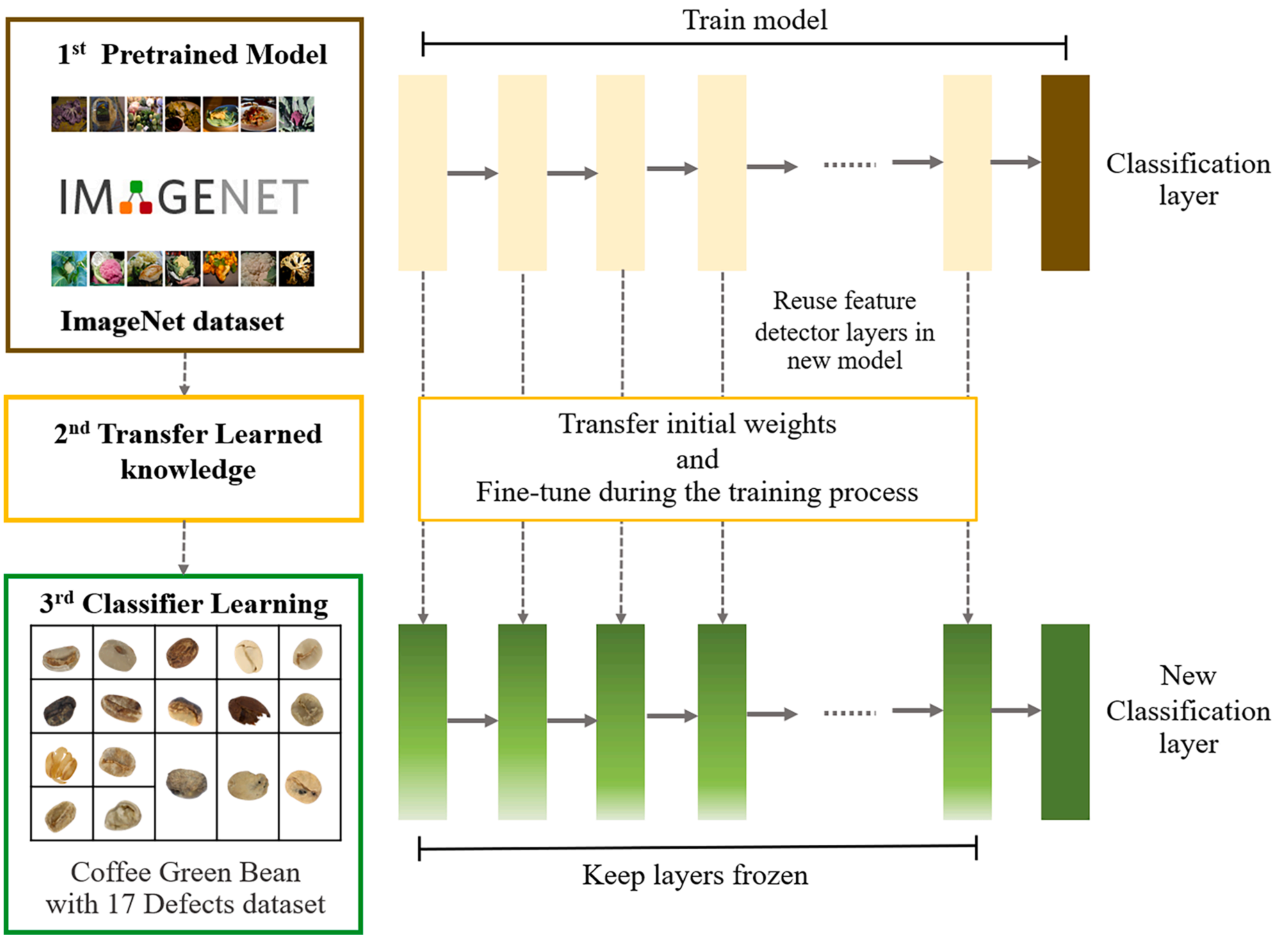

Fig. 4. Transfer learning architecture for Coffee Bean defect classification.

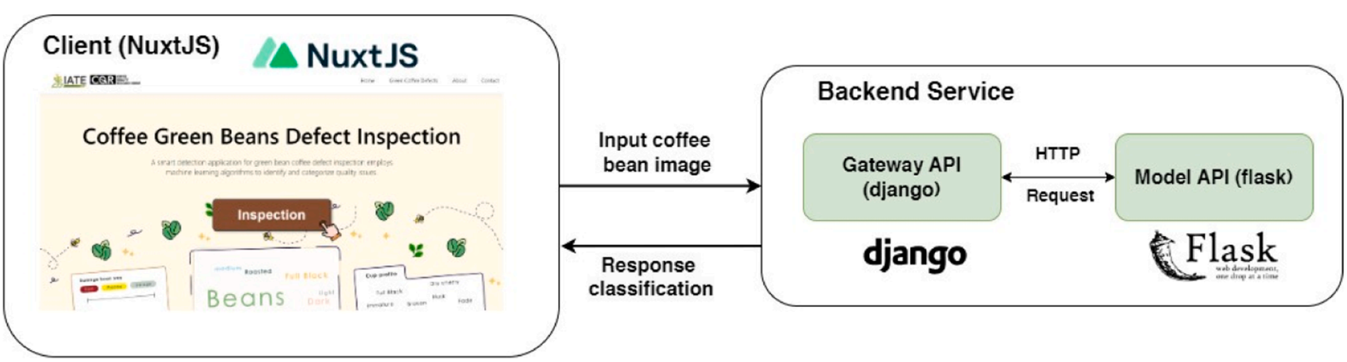

Fig. 5. Web application architecture for Green Coffee bean defects classification.

boosting the efficiency of sorting and quality control. The technological framework selected for deployment comprised Nuxt.js, Django, and Flask, as illustrated in Fig. 5. Each component fulfilled a critical and collaborative role:

- Nuxt.js: for User Interface: Utilized for its server-side rendering capabilities, Nuxt.js improved the user experience by delivering faster page loads and better enhanced search engine optimization, making the application user-friendly and easily discoverable.
- Django: as API Content Management System: Chosen for its robustness and versatility, Django managed the application's content and data requests efficiently and supported a secure and structured API system enhanced by a rich library ecosystem.
- Flask: for Model Integration: Flask's lightweight and flexible nature facilitated the embedding of the machine learning model and ensured smooth communication with the Django-based API. Flask simplified real-time interactions between the user interface and the model.

The combination of these technologies ensured optimal performance and user experience. Nuxt.js enhanced frontend responsiveness, Django ensured reliable data management, and Flask provided efficient backend model integration. Together, they created a responsive, reliable, and intuitive web application that effectively demonstrated the model's capabilities to users.

4. Experimental results and discussion

This study systematically investigated four key components essential for developing a robust coffee bean defect classification system. First, the research focused on image augmentation techniques, comparing the effectiveness of methods applied to original and augmented datasets, with a particular emphasis on rotation augmentation. The study evaluated the classification performance of five distinct model architectures—MobileNetV2, MobileNetV3, ResNetV2, InceptionV2, and EfficientNetV2—to identify the optimal framework for detecting green coffee bean defects. Second, the selected model architecture was fine-tuned to maximize classification accuracy. This process involved rigorous hyperparameter tuning and optimization strategies to enhance model performance. Third, a 5-fold cross-validation strategy was employed to ensure the reliability and generalizability of the approach. This method provided a robust assessment of model performance across diverse subsets of the dataset. Finally, the optimized model was integrated into a web application, enabling practical use for real-time classification of coffee bean defects.

4.1. Comparison of pre-trained and fine-tuned model performance with augmentation

The initial experiment in this study was to determine the most effective image augmentation techniques for classifying defects in coffee beans. By utilizing the original dataset (pre-trained) and the augmented dataset (after transfer learning), the experiment assessed the performance of different image augmentation approaches by comparing model accuracies. For this study, all models were initialized with default pre-trained weights from the ImageNet dataset, without any additional changes to the internal architecture. This study employed five advanced CNN architectures evaluated including MobileNetV2, MobileNetV3, EfficientNetV2, InceptionV2, and ResNetV2.

Results demonstrated that the augmented dataset yielded significantly higher accuracy across all models evaluated. Specific improvements in accuracy for each model are shown in Table 3:

- MobileNetV2: Accuracy increased from 66.23 % (pre-trained) to 84.87 % (augmented).
- EfficientNetV2: Accuracy increased from 63.28 % (pre-trained) to 78.61 % (augmented).
- InceptionV2: Accuracy increased from 66.10 % (pre-trained) to 73.60 % (augmented).
- ResNetV2: Accuracy increased from 57.63 % (pre-trained) to 76.42 % (augmented).
- MobileNetV3: Accuracy increased from 73.45 % (pre-trained) to 90.19 % (augmented).

The accuracy of MobileNetV3 on the augmented dataset (90.19 %) was significantly higher compared to the original dataset (73.45 %). This trend was consistent across all models tested, suggesting that the augmented dataset had substantial potential for enhancing the model performance. The MobileNetV3 architecture achieved the highest accuracy among the models tested, establishing a robust baseline for subsequent experiments.

Results indicated that augmentation, particularly rotation, significantly enhanced classification accuracy across all the models. The superior performance of MobileNetV3 underscored its suitability for ongoing research and application in coffee bean defect classification.

Table 3
Comparison of accuracy before and after transfer learning.

Model	Pre-trained Accuracy (%)	After Transfer Learning Accuracy (%)
MobileNetV2	66.23	84.87
EfficientNetV2	63.28	78.61
InceptionV2	66.10	73.60
ResNetV2	57.63	76.42
MobileNetV3	73.45	90.19

4.2. Result of each model architecture selection

The results of evaluating various model architectures to identify the optimal framework for learning coffee bean defect patterns are presented. The study utilized the selected augmented dataset, previously demonstrated to enhance accuracy, and assessed the performance of MobileNetV2, EfficientNetV2, InceptionV2, ResNetV2, and MobileNetV3 architectures.

4.2.1. MobileNetV2

The training and validation accuracy (Table 4) for MobileNetV2 showed that the training improved steadily from 75.0 % at 10 epochs to 95.0 % at 100 epochs indicating effective learning from the training data. By contrast, validation accuracy increased rapidly at first but then fluctuated at around 82.0 %, suggesting good generalization but also potential overfitting, with no improvement alongside the training accuracy. The training and validation loss data further supported this observation. The training loss decreased steadily from 60.0 % at 10 epochs to 15.0 % at 100 epochs, confirming the model's effective learning. The validation loss initially decreased from 60.0 % to 50.0 % by the 10th epoch and then Remained at 50 % from 20 to 100 epochs.

This pattern is typical when a model starts to overfit the training data, as indicated by the validation loss not decreasing further and starting to rise slightly. These results implied that while MobileNetV2 learned well from the training data, there was a need to address overfitting. Potential solutions include implementing techniques such as dropout, regularization, or further data augmentation.

The confusion matrix in Fig. 6 provides a detailed evaluation of MobileNetV2's performance in classifying coffee bean defects, with an overall accuracy of 84.87 %. The model exceled in several categories, achieving high classification accuracy. Categories with the five highest classification accuracies were Dry Cherry (100 % accuracy, 63 out of 63), Parchment (98.61 % accuracy, 71 out of 72), Husk (98.48 % accuracy, 65 out of 66), Full Black (98.46 % accuracy, 83 out of 84), and Shell (87.69 % accuracy, 57 out of 65). Conversely, the categories with the lowest classification accuracies were Partial Black (77.27 % accuracy, 68 out of 88), Fade (74.55 % accuracy, 41 out of 55), and Partial Sour (64.91 % accuracy, 37 out of 57). These results indicated that while MobileNetV2 performed exceptionally well in certain defect categories, there was room for improvement in others, particularly those with lower accuracy. Strategies such as additional data augmentation, dropout, and regularization could help mitigate overfitting and enhance classification accuracy across all defect categories.

4.2.2. EfficientNetV2

The training and validation accuracy and loss data for

Table 4
Training and validation accuracy and loss for MobileNetV2 at less than or equal to 100 epochs.

Epoch	Training Accuracy (%)	Validation Accuracy (%)	Training Loss (%)	Validation Loss (%)
10	75.0	72.0	60.0	60.0
20	83.0	77.0	45.0	50.0
30	87.0	79.0	35.0	50.0
40	89.0	80.0	30.0	50.0
50	91.0	81.0	25.0	50.0
60	92.0	82.0	20.0	50.0
70	93.0	82.0	18.0	50.0
80	94.0	83.0	17.0	50.0
90	95.0	83.0	16.0	50.0
100	95.0	82.0	15.0	50.0

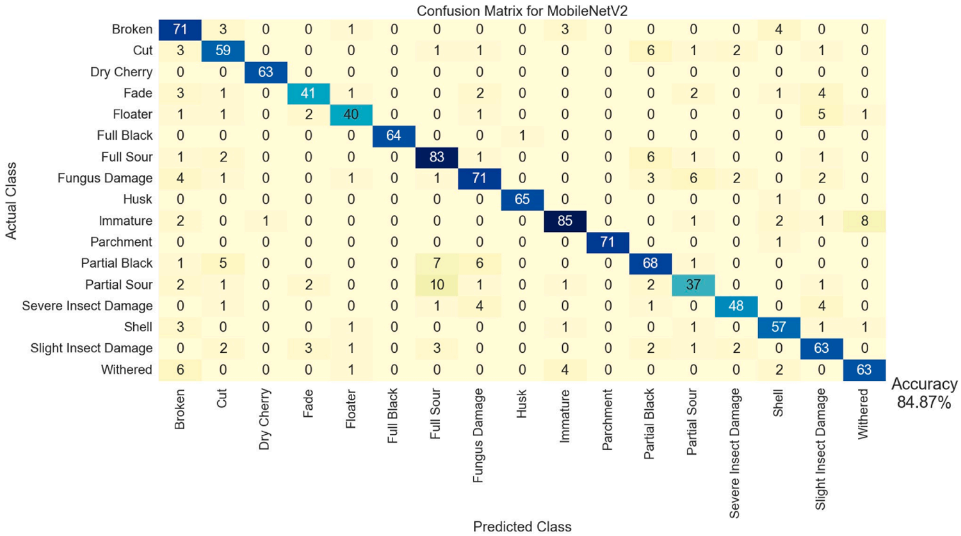

Fig. 6. Confusion matrix of MobileNetV2 models on the validation set.

Table 5
Training and validation accuracy and loss for efficientNetV2 less than equal 100 Epochs.

Epochs	Training Accuracy (%)	Validation Accuracy (%)	Training Loss (%)	Validation Loss (%)
10	75.0	72.0	60.0	80.0
20	83.0	77.0	45.0	60.0
30	87.0	79.0	35.0	55.0
40	89.0	80.0	30.0	55.0
50	91.0	81.0	25.0	55.0
60	92.0	82.0	22.0	55.0
70	93.0	82.0	20.0	53.0
80	94.0	83.0	18.0	53.0
90	95.0	83.0	17.0	54.0
100	95.0	82.0	16.0	55.0

EfficientNetV2, presented in Table 5, revealed key insights into the model's learning process over 100 epochs. The training accuracy improved steadily from 75.0 % at 10 epochs to 95.0 % at 100 epochs indicating effective learning from the training data. By contrast, the validation accuracy increased to 83.0 % by the 80th epoch and then decreased to 82.0 % by the 100th epoch, suggesting a stable but slightly lower performance compared to the training accuracy. The training and validation loss data further supported this observation. The training loss decreased steadily from 60.0 % at 10 epochs to 16.0 % at 100 epochs, confirming the model's effective learning. However, the validation loss remained at around 55.0 % throughout the training process, indicating a slight overfitting as the validation loss was higher than the training loss. These results suggested that while EfficientNetV2 learned well from the training data, there was a need to address overfitting. Potential solutions include implementing techniques such as dropout, regularization, or further data augmentation.

These results indicated that while EfficientNetV2 effectively learned from the training data, the validation performance suggested a degree of overfitting. The plateau in validation accuracy and the relatively higher validation loss compared to the training loss underscored this issue. Implementing strategies such as dropout, regularization, or further data augmentation may help mitigate overfitting and improve the model's generalization to new, unseen data.

The confusion matrix for EfficientNetV2, shown in Fig. 7, highlighted the model's performance across various categories, showcasing correct

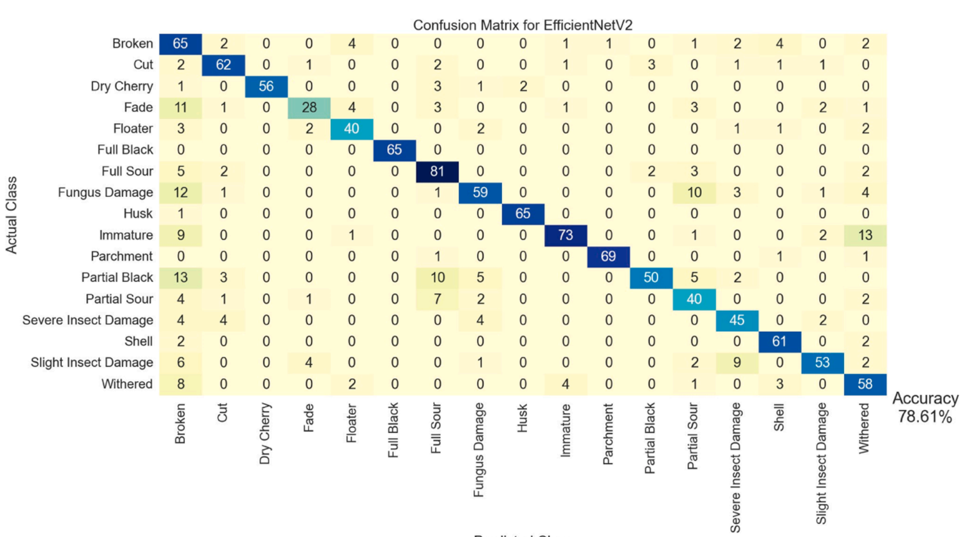

Fig. 7. Confusion matrix of efficientNetV2 models on the validation set.

classifications and common misclassifications. Achieving an overall accuracy of 78.61 %, the model demonstrated strong performance with some room for improvement. The model exceled in certain categories such as "Full Black" with 100 % accuracy (65 out of 65), "Parchment" with 98.57 % accuracy (69 out of 70), and "Dry Cherry" with 90.32 % accuracy (56 out of 62). Other categories like "Broken" with 78.31 % accuracy (65 out of 83), "Cut" with 84.93 % accuracy (62 out of 73), and "Floater" with 83.33 % accuracy (40 out of 48) also showed high performance. However, notable misclassifications occurred particularly "Partial Black" with 56.82 % accuracy (50 out of 88), "Partial Sour" with 58.82 % accuracy (40 out of 68), and "Severe Insect Damage" with 76.27 % accuracy (45 out of 59), which are frequently confused with other similar defects. These misclassifications indicated that the model struggled to distinguish between similar defect categories, suggesting areas for further improvement. Overall, the EfficientNetV2 model demonstrated strong performance but also highlighted specific areas for improvement in classification accuracy. Potential strategies to enhance model performance include additional data augmentation, regularization, and dropout techniques to mitigate overfitting and improve generalization to new, unseen data. Results suggested that this model was not the most suitable for green coffee bean defect classification and further refinement was needed.

4.2.3. InceptionV2

The training and validation accuracy and loss data for InceptionV2, presented in Table 6, revealed important insights into the model's learning process over 100 epochs. Training accuracy improved steadily from 75.0 % at 10 epochs to 97.0 % at 100 epochs indicating effective learning from the training data. However, validation accuracy plateaued at 74.0 %, suggesting a significant gap between the training and validation performance, indicative of potential overfitting. The training and validation loss data further supported this observation. The training loss consistently decreased from 80.0 % at 10 epochs to 16.0 % at 100 epochs, confirming the model's effective learning. However, the validation loss remained at around 76.0 % throughout the training process, showing signs of overfitting as it did not decrease significantly compared to the training loss.

These results indicated that while InceptionV2 effectively learned from the training data, as evidenced by the substantial improvement in training accuracy and reduction in training loss, there was a significant gap in its performance on the validation set. The consistent validation accuracy and loss values suggested overfitting, where the model performed well on the training data but failed to generalize to new, unseen data. Strategies such as regularization, dropout, and further data augmentation should be considered to address this overfitting issue and improve the model's generalization capabilities

The confusion matrix for InceptionV2, shown in Fig. 8, provided detailed insights into the model's classification performance across various categories. The model demonstrated an overall accuracy of 73.60 %, indicating a generally strong performance with some areas needing improvement. While the model exceled in categories such as "Parchment," "Dry Cherry," "Full Black," "Husk," and "Fungus Damage," there were notable misclassifications that highlighted areas for further enhancement. The three categories with the lowest accuracy were:

- Partial Sour: 47.37 % (27 out of 57)
- Withered: 40.79 % (31 out of 76)
- Slight Insect Damage: 61.04 % (47 out of 77)

These misclassifications suggested that the model struggled to distinguish between similar defect categories. For instance, "Partial Black" was often confused with "Partial Sour" and "Full Sour," and "Withered" was frequently confused with "Immature" and "Cut". Overall, while the InceptionV2 model exhibited good classification performance, there was a need for further refinement to improve its ability to distinguish between similar defect categories. Potential strategies to enhance classification accuracy across all defect categories include additional data augmentation, regularization, and dropout techniques.

Table 6
Training and Validation Accuracy and Loss for InceptionV2 at less than or equal to 100 Epochs.

Epochs	Training Accuracy (%)	Validation Accuracy (%)	Training Loss (%)	Validation Loss (%)
10	75.0	70.0	80.0	80.0
20	80.0	72.0	60.0	80.0
30	85.0	73.0	50.0	80.0
40	88.0	74.0	40.0	78.0
50	90.0	74.0	30.0	77.0
60	92.0	74.0	25.0	76.0
70	93.0	74.0	22.0	76.0
80	94.0	74.0	20.0	76.0
90	95.0	74.0	18.0	76.0
100	97.0	74.0	16.0	76.0

4.2.4. ResNetV2

The training and validation accuracy and loss data for ResNetV2, presented in Table 7, presented in Table 8, provided insights into the model's learning process over 100 epochs. The training accuracy improved steadily from 70.0 % at 10 epochs to 97.0 % at 100 epochs indicating effective learning from the training data. However, the validation accuracy plateaued at around 79.0 %, indicating a gap between training and validation performance and suggesting potential overfitting. The training and validation loss data further supported this observation. The training loss consistently decreased from 80.0 % at 10 epochs to 16.0 % at 100 epochs, confirming the model's effective learning. However, the validation loss remained at around 80.0 % throughout the training process, showing signs of overfitting as it did not decrease significantly compared to the training loss.

These results suggested that while ResNetV2 learned effectively from the training data, achieving a high training accuracy of 97.0 % by the 100th epoch, the validation performance indicated potential overfitting. The validation accuracy plateaued at 79.0 %, and the validation loss remained consistently high at 80.0 %, highlighting the gap between training and validation performance.

The confusion matrix for ResNetV2, shown in Fig. 9, provided detailed insights into the model's classification performance across various categories. Achieving an overall accuracy of 76.42 %, the model showed strong performance in several categories such as "Immature" with 84 correct predictions out of 100 and "Broken" with 69 correct predictions out of 90. However, there were notable misclassifications such as "Withered" being confused with "Immature" and "Cut" and "Partial Black" being confused with "Partial Sour" and "Severe Insect Damage." These misclassifications indicated areas where the model could improve in distinguishing between similar categories. For instance, "Full Sour" was often confused with "Fungus Damage" and "Broken." These misclassifications suggested that while the ResNetV2 model demonstrated good classification performance, it was not the best model for green coffee bean defect classification.

4.2.5. MobileNetV3

The confusion matrix for MobileNetV3, shown in Fig. 10, provided detailed insights into the model's classification performance across various categories. The MobileNetV3 model demonstrated a high overall accuracy of 90.19 %, indicating strong performance in green coffee bean defect classification. For instance, "Immature" was correctly classified 90 times out of 100, "Parchment" had 72 correct classifications out of 73, and "Broken" was correctly identified 73 times out of 79. However, there were still notable misclassifications. Categories such as "Partial Sour" and "Withered" showed areas for improvement with lower

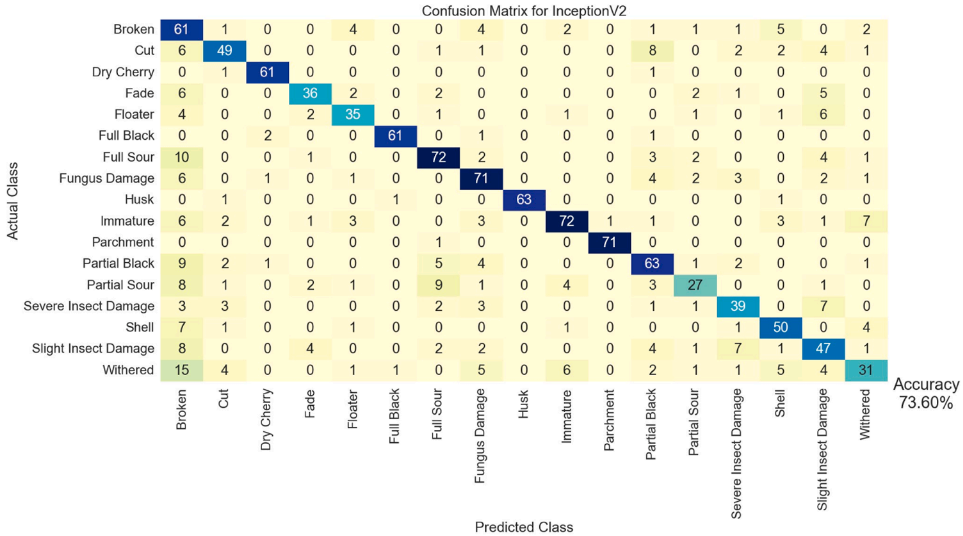

Fig. 8. Confusion Matrix of InceptionV2 models on the validation set.

Table 7
Training and validation accuracy and loss for ResNetV2 at less than or equal to 100 Epochs.

Epochs	Training Accuracy (%)	Validation Accuracy (%)	Training Loss (%)	Validation Loss (%)
10	70.0	70.0	80.0	80.0
20	80.0	75.0	60.0	80.0
30	85.0	77.0	50.0	80.0
40	88.0	78.0	40.0	80.0
50	90.0	79.0	30.0	80.0
60	92.0	79.0	25.0	80.0
70	93.0	79.0	22.0	80.0
80	94.0	79.0	20.0	80.0
90	95.0	79.0	18.0	80.0
100	97.0	79.0	16.0	80.0

Table 8
Training and validation accuracy and loss for MobileNetV3 at less than or equal to 100 Epochs.

Epochs	Training Accuracy (%)	Validation Accuracy (%)	Training Loss (%)	Validation Loss (%)
10	80.0	75.0	60.0	50.0
20	85.0	80.0	40.0	45.0
30	90.0	85.0	30.0	40.0
40	92.0	87.0	25.0	40.0
50	94.0	88.0	20.0	38.0
60	95.0	89.0	18.0	38.0
70	96.0	89.0	16.0	38.0
80	97.0	89.0	14.0	38.0
90	98.0	89.0	12.0	38.0
100	100.0	89.0	10.0	38.0

accuracies. ”Withered” was correctly classified 68 times out of 76 but was often confused with ”Immature” and ”Cut.” Similarly, ”Partial Black” was confused with ”Partial Sour” and ”Severe Insect Damage.”

Table 8 illustrates the learning process of the MobileNetV3 model over 100 epochs. The training accuracy improved steadily from 80.0 % at 10 epochs to 100.0 % at 100 epochs indicating effective learning from the training data. However, the validation accuracy plateaued at around 89.0 %, suggesting a stable but slightly lower performance compared to the training accuracy. The training loss consistently decreased, approaching zero, indicating effective learning, while the validation loss stabilized at around 38.0 %, suggesting slight overfitting.

Overall, from Table 9 presented the validation accuracy for each model architecture, highlighting that the MobileNetV3 model was the best-performing model for green coffee bean defect classification among those tested. Although EfficientNetV2, InceptionV2, and ResNetV2 are robust architectures for general image classification, they may not be ideally suited for this dataset. Their higher model complexity can make them more prone to overfitting, especially given the dataset’s limited diversity, even after augmentation. In contrast, MobileNetV2 and MobileNetV3, achieving validation accuracies of 84.87 % and 90.19 % respectively, are designed for computational efficiency and optimized for lightweight inference, demonstrating stronger performance on this dataset. MobileNetV3, in particular, outperforms the other models due to its focus on efficiency and specific activation mechanisms, making it better suited to handle this type of data. For smaller, more specialized datasets, lightweight models like MobileNetV3 tend to generalize better than more complex architectures. Despite its high accuracy, there was still room for improvement in certain categories to reduce misclassifications and further enhance overall performance.

4.3. Model tuning and evaluation for optimal performance

After selecting MobileNetV3 as the best model for classifying the 17 types of coffee bean defects, a tuning process was undertaken to optimize its performance. This involved experimenting with different numbers of epochs and setting the learning rate at 0.01. The top five results, based on validation accuracy and loss, are summarized in Table 10. The best performance was achieved with 3 epochs, reaching a validation accuracy of 97.84 % and a validation loss of 13.60 %. Other top configurations included 4 epochs with 97.73 % accuracy and 13.32 % loss, and 9 epochs with 97.73 % accuracy and 14.20 % loss. These results indicated that MobileNetV3 performed well with relatively few epochs, achieving high accuracy and low loss in classifying coffee bean defects

To ensure the robustness and generalizability of the proposed model, a 5-fold cross-validation was conducted. This experiment validated the effectiveness of the trained model in making accurate predictions across different subsets of the dataset. The cross-validation results, summarized in Table 11, show high validation accuracy and low validation loss across all folds. The validation accuracies ranged from 98.78 % to 99.84 % and the validation losses ranged from 0.96 % to 6.61 %. These results confirmed the model’s robustness and reliability in classifying coffee bean defects.

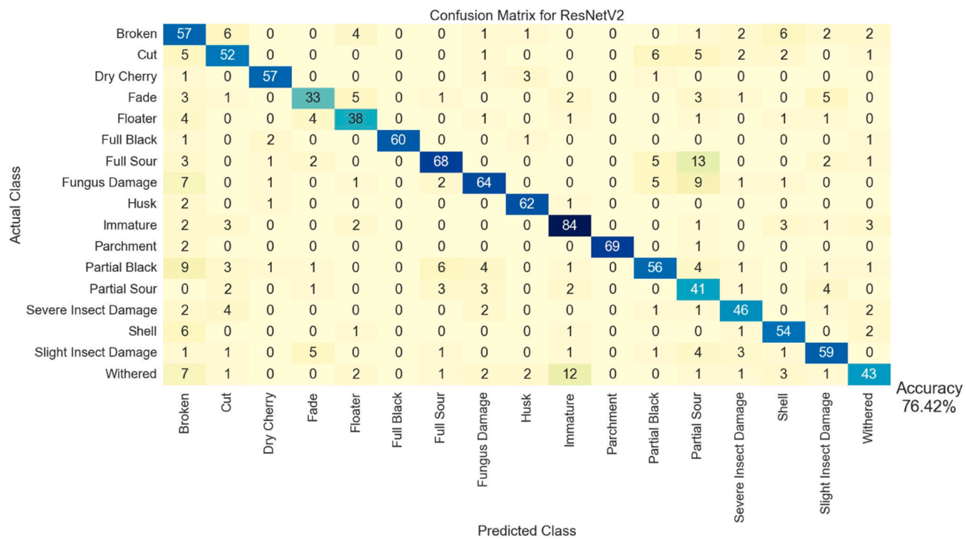

Fig. 9. Confusion matrix of ResNetV2 models on the validation set.

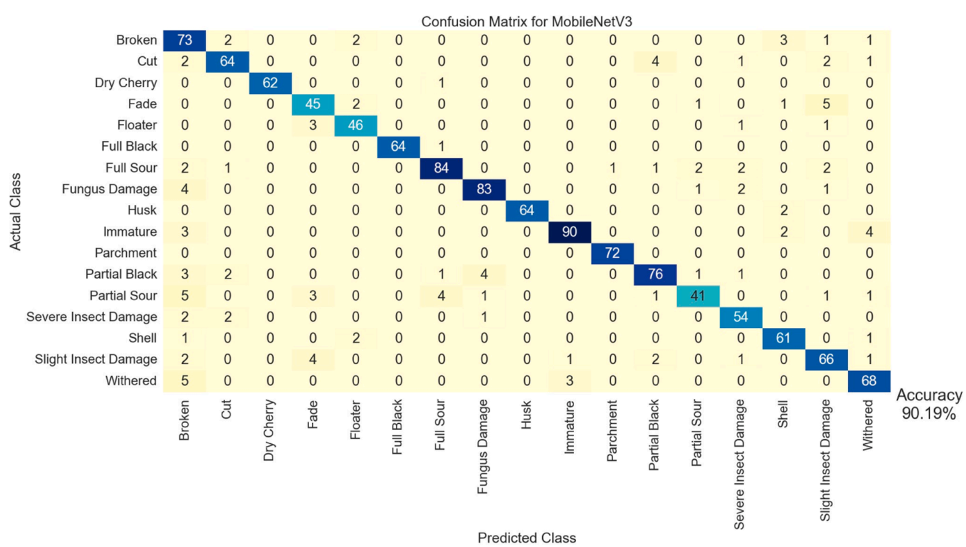

Fig. 10. Confusion matrix of mobileNetV3 models on the validation set.

Table 9
Comparison of validation accuracy for each model architecture.

Model Architecture	Validation Accuracy (%)
MobileNetV2	84.87
EfficientNetV2	78.61
InceptionV2	73.60
ResNetV2	76.42
MobileNetV3	**90.19**

Table 10
Top 5 tuned results for MobileNetV3 in classifying coffee bean defects.

Epochs	Validation Accuracy (%)	Validation Loss (%)
3	97.84	13.60
4	97.73	13.32
9	97.73	14.20
8	97.73	13.72
16	97.46	14.99

Table 11
5-Fold cross-validation results for MobileNetV3 in classifying coffee bean defects.

Fold	Validation Accuracy (%)	Validation Loss (%)
1	98.78	6.61
2	99.43	2.35
3	99.84	0.96
4	99.76	1.51
5	99.59	1.74

The confusion matrix from the test set after hyper tuning, shown in Fig. 11, provided a detailed evaluation of the MobileNetV3 model classification performance across the 17 defect categories. The model achieved an overall accuracy of 88.63 %, indicating strong performance. Categories such as ”Dry Cherry” (39 out of 39) and ”Full Black” (25 out of 25) achieved 100 % accuracy. However, there were notable misclassifications such as ”Withered” being confused with ”Slight Insect Damage” and ”Severe Insect Damage.”

The classification metrics for the test results of the tuned and cross-validated MobileNetV3 model are presented in Table 12. The table

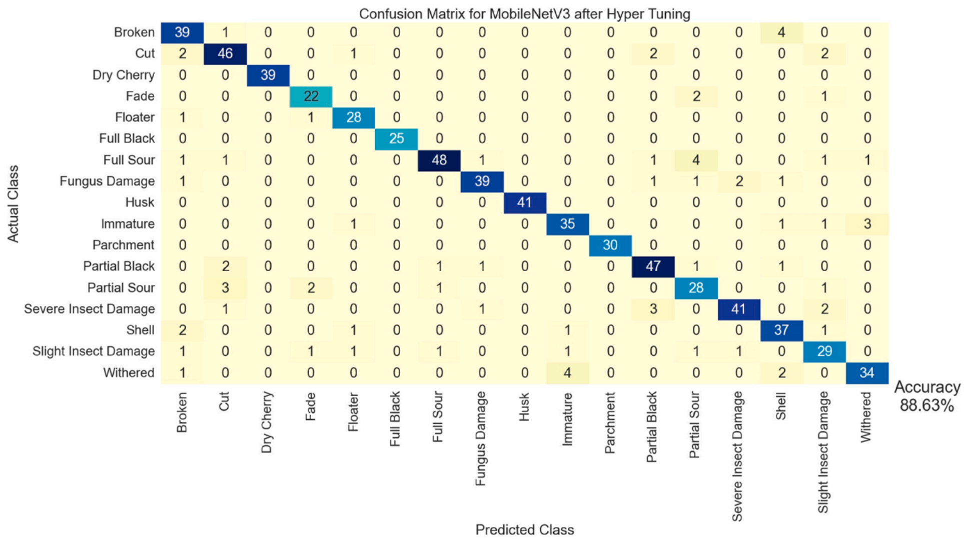

Fig. 11. Confusion Matrix of MobileNetV3 models after Hyper Tuning on the test set.

Table 12
Classification metrics for MobileNetV3 in classifying coffee bean defects.

Defect Type	Precision	Recall	F1-Score	Support
Broken	81.25	88.64	84.78	44
Cut	85.19	86.79	85.98	53
Dry Cherry	100.00	100.00	100.00	39
Fade	84.62	88.00	86.27	25
Floater	87.50	93.33	90.32	30
Full Black	100.00	100.00	100.00	25
Full Sour	94.12	82.76	88.07	58
Fungus Damage	92.86	86.67	89.66	45
Husk	100.00	100.00	100.00	41
Immature	85.37	85.37	85.37	41
Parchment	100.00	100.00	100.00	30
Partial Black	87.04	88.68	87.85	53
Partial Sour	75.68	80.00	77.78	35
Severe Insect Damage	93.18	85.42	89.13	48
Shell	80.43	88.10	84.09	42
Slight Insect Damage	76.32	80.56	78.38	36
Withered	89.47	82.93	86.08	41
Macro Avg	89.00	89.25	89.04	686

provides performance metrics for each of the 17 classes of coffee bean defects, including precision, recall, F1-score, and support (number of true instances for each label). Overall, the model shows high precision (89.00 %), recall (89.25 %), and F1-scores (89.04 %) across most classes, indicating strong performance on unseen data. Notable high scores include "Dry Cherry" with an F1-score of 100 %, "Husk" with an F1-score of 100 %, and "Parchment" with an F1-score of 100 %. Some classes, such as "Partial Sour" and "Slight Insect Damage," have slightly lower recall scores, indicating areas for potential improvement.

Overall, the MobileNetV3 model demonstrated strong performance but also highlighted specific areas for improvement in classification accuracy. While it exceled in several categories, there was room to enhance its ability to distinguish between similar defect types. Potential strategies to improve performance include additional data augmentation, regularization, and dropout techniques to mitigate overfitting and improve generalization to new, unseen data.

4.4. Model optimization techniques

This part aims to investigate the hypothesis that reducing the model size would still allow the model to be suitable for use on small devices. This research experiment showed that the optimized original MobileNetV3 model can get high accuracy with few resources when tested on training data. The quantized MobileNetV3 model shows results in Table 13. It shows training and validation accuracy that is similar to the original MobileNetV3, but the model size and inference time are much smaller. These findings underscore the potential of the optimized MobileNetV3 model for efficient and scalable applications in the coffee industry. Moreover, the MobileNetV3 pruned and MobileNetV3 quantized versions achieved similar validation accuracy to the original model, they exhibited substantial reductions in resource requirements. Specifically, the quantized model maintained high accuracy with only 5.63 MB in size and 242.76 ms in inference time, demonstrating that it can be appropriate for deployment on devices with limited resources.

Table 13
Evaluation of optimization techniques for MobileNetV3.

Model	Model Size (MB)	Parameters (M)	Inference Time (ms)	Training Accuracy (%)	Validation Accuracy (%)
Original	86.94	5.54	307.87	95.97	94.87
Pruned	21.38	5.54	485.52	99.25	95.85
Quantized	5.63	5.54	242.76	99.25	95.85

4.5. Web application integration

4.5.1. Prototype with real-time classification

The optimized model was integrated into a web application, enabling practical use for realtime classification of coffee bean defects. This web application allowed users to upload images of green coffee beans, which were then analyzed by the model to identify various types of defects. Users can test the application at http://iate-coffee-inspection.com/.

The application provides a user-friendly interface where users can easily select and upload images. Once the image is uploaded, the application processes it using the trained MobileNetV3 model and provides an instant classification result. This result helps users to identify specific defects in the Arabica green coffee beans, aiding in quality control and sorting processes. The model integration into a web application ensures accessibility and practical usability for endusers, making the technology readily available for commercial and research purposes.

The application features a simple and intuitive design for easy navigation and image upload, offering real-time processing and providing quick analysis and classification of coffee bean defects. The detailed results enhance the understanding and management of coffee bean quality. After classifying the image, the application displays the top

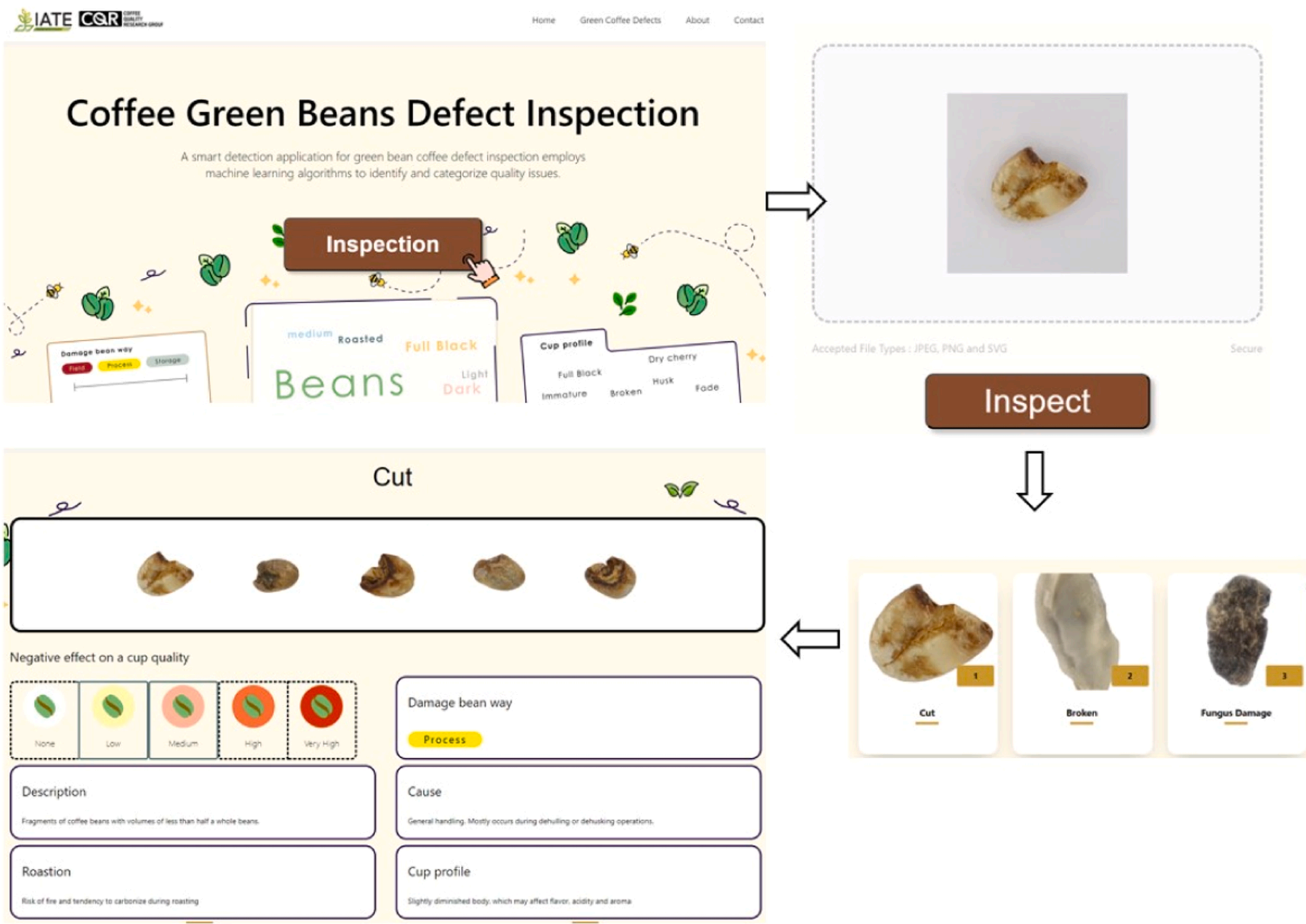

Fig. 12. Web application's interface and workflow for coffee bean defect classification.

three matching defect types, giving users a comprehensive understanding of potential defects. Finally, the user can select the details of one of the three results shown, as illustrated in Fig. 12. This integration showcases the practical application of the optimized MobileNetV3 model, bridging the gap between advanced machine learning techniques and real-world utility in the coffee industry.

4.5.2. Prototype testing with farmers

The final step involved testing the developed prototype with at least seven farmers, with results summarized below. The survey evaluated the ease of use, effectiveness in classifying coffee bean defects, and the likelihood of recommending the website to others. All responses were positive, with no ratings below 4 on a 5-point scale. The key findings were as follows:

- Ease of Use: The farmers rated their satisfaction with the web application's ease of use. Four farmers rated ease of use as 5 (57.1 %) and 3 farmers rated ease of use as 4 (42.9 %).
- Effectiveness in Classifying Defects: The farmers evaluated how well the web application helped to classify coffee bean defects, with most expressing high satisfaction. Five farmers rated effectiveness as 5 (71.4 %) and 2 farmers rated effectiveness as 4 (28.6 %).
- Likelihood to Recommend: The farmers were asked how likely they were to recommend the website to their friends. The responses indicated a strong likelihood of recommendation. Five farmers rated their likelihood to recommend as 5 (71.4 %) and 2 farmers rated their likelihood to recommend as 4 (28.6 %).

These results demonstrated that the web application was well-received by the farmers, who found it easy to use and effective in classifying coffee bean defects. They indicated a high likelihood of recommending the application to others and expressed strong overall satisfaction with the system.

4.6. Discussion

Initially, the models were evaluated on the original dataset (pre-trained) to establish baseline accuracy, with transfer learning and augmentation consistently enhancing performance across all architectures. MobileNetV3 achieved the highest validation accuracy at 90.19 %, demonstrating its strong suitability for coffee bean defect classification in ongoing research and practical applications. After hyperparameter tuning, MobileNetV3 model demonstrated exceptional performance in classifying coffee bean defects, achieving high precision, recall, and F1-scores of 89.00 %, 89.25 %, and 89.04 %, respectively across 17 defect types, as shown in Table 12. This indicated a balanced and robust performance. The model exceled in identifying defects such as 'Dry Cherry', 'Full Black', and 'Parchment', achieving perfect scores in these categories. Other classes like 'Withered' and 'Partial Black' also showed excellent performance with high F1-scores.

However, there was room for improvement in classes like 'Partial Sour', 'Slight Insect Damage', and 'Shell', which had relatively lower F1-scores. This highlighted areas where the model could benefit from further refinement. Some classes exhibited a trade-off between precision and recall. For instance, 'Full Sour' had high precision but lower recall, indicating that while the model was good at identifying true positives, it missed some instances of this defect.

The 5-fold cross-validation results further reinforced the model's robustness and its ability to generalize well to unseen data. As summarized in Table 11, the validation accuracies ranged from 98.78 % to 99.84 % and the validation losses ranged from 0.96 % to 6.61 % indicating consistent performance across different subsets of the dataset. MobileNetV3 was shown to be a promising tool for practical applications in coffee bean defect classification.

When tested with unseen data, the model achieved an overall accuracy of 88.63 %, which further confirmed its effectiveness in real-world scenarios. The model integration into a web application ensures accessibility and practical usability for end-users, making the technology readily available for commercial and research purposes. This integration showcases the practical applications of the optimized MobileNetV3 model, bridging the gap between advanced machine learning techniques and real-world utility in the coffee industry.

Furthermore, the optimization of MobileNetV3 through quantization and pruning proof of concept enhanced model efficiency, enabling it to perform well on resource-constrained devices without compromising accuracy.

4.7. Limitations and future work

This study performed well, but the dataset's variety remained limited after augmentation. The dataset of the initial 979 images, which were rotated to 6853 through augmentation, may not be diverse enough to permit generalization. Challenges that the current model has not completely encountered, such as variations in lighting, background, and bean positioning, may be introduced by real-world conditions. And also plan to investigate ensemble learning methods by combining MobileNetV3 with other architectures, like EfficientNet and ResNet, to leverage the strengths of different models for improved classification performance.

Future research will resolve this issue by extending the dataset to encompass images from a variety of environmental conditions and sources to enhance the model's generalizability and robustness. Additionally, exploring advanced augmentation techniques—such as color variations, noise, and brightness adjustments—could simulate a wider range of real-world scenarios

5. Conclusions

This research successfully developed a deep learning model to classify 17 types of Arabica green coffee bean defects using convolutional neural networks (CNN). By enhancing the dataset to 6853 images through various image rotation techniques, we ensured a robust analysis. Among the evaluated architectures, MobileNetV3 achieved the highest accuracy and was chosen for further parameter tuning. The optimized model was then integrated into a web application to facilitate practical use in coffee bean quality sorting. These experiments lay a strong foundation for developing an accurate and reliable coffee bean defect classification system. Future studies should focus on integrating additional data augmentation techniques and leveraging transfer learning from larger and more diverse datasets to further enhance model performance.

Ethics statement

Not applicable: This manuscript does not include human or animal research.

CRediT authorship contribution statement

Sujitra Arwatchananukul: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Dan Xu:** Writing – review & editing, Validation. **Phasit Charoenkwan:** Writing – review & editing, Supervision, Software, Methodology, Investigation, Formal analysis. **Sai Aung Moon:** Writing – original draft, Methodology, Investigation, Formal analysis, Data curation. **Rattapon Saengrayap:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors are grateful for the grant from Mae Fah Luang University under the National Science, Research, and Innovation Fund (NSRF) 2023: grant no 662A03049. Thanks are also due to Mae Fah Luang University, for financial support and the provision of experimental sites for this research study.

Data availability

Data will be made available on request.

References

[1] V. Sberveglieri, E. Comini, D. Zappa, A. Pulvirenti, E.Nuńez Carmona, Electronic nose for the early detection ̃of different types of indigenous mold contamination in green coffee (2013) 461–465, https://doi.org/10.1109/ICSensT. 2013.6727696.
[2] V. Sberveglieri, P. Fava, A. Pulvirenti, I. Concina, M. Falasconi, New methods for the early detection of fungal contamination on green coffee beans by an electronic nose, 2012, pp. 414–417. 10.1109/ICSensT.2012.6461711.
[3] I.C. Organization, Coffee report and outlook, URL: https://icocoffee.org/documents/cy2023-24/Coffee_Report_and_Outlook_December_2023_ICO.pdf.
[4] M. García, J.E. Candelo-Becerra, F.E. Hoyos, Quality and defect inspection of green coffee beans using a computer vision system, Appl. Sci. 9 (2019), https://doi.org/10.3390/app9194195.
[5] A. Cravero, S. Pardo, S. Sepulveda, L. Mu' noz, ̃Challenges to use machine learning in agricultural big data: a systematic literature review, Agronomy 12 (2022), https://doi.org/10.3390/agronomy12030748.
[6] E. Elbas¸ı, N. Mostafa, Z. AlArnaout, A. Zreikat, E. Cina, A. Shdefat, A. Topcu, S. Mathew, W. Abdelbaki, C. Zaki, Artificial intelligence technology in the agricultural sector a systematic literature review, IEEE Access 11 (2023) 171–202.
[7] S.O. Araujo, R.S. Peres, J.C. Ramalho, F. Lidon, J. Barata, Machine learning applications in agriculture: current' trends, challenges, and future perspectives, Agronomy 13 (2023), https://doi.org/10.3390/agronomy13122976.
[8] L. Santos, F. Neves Dos Santos, P.Moura Oliveira, P. Shinde, Deep learning applications in agriculture: a short review, 2020, pp. 139–151. 10.1007/978-3-030-35990-4_12.
[9] B.W. Ligar, Computer vision for identifying and classifying green coffee beans: a review, Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer dan Teknologi Informasi (2022).
[10] L.Y. Ke, E. Chen, C.H. Hsia, Green coffee bean defect detection using shift-invariant features and non-local block, in: 2023 IEEE 6th International conference on knowledge innovation and invention (ICKII), 2023, pp. 430–431, https://doi.org/10.1109/ICKII58656.2023.10332580.
[11] G. Vilcamiza, N. Trelles, L. Vinces, J. Oliden, A coffee bean classifier system by roast quality using convolutional neural networks and computer vision implemented in an nvidia jetson nano, 2022, pp. 1–6. doi:10.1109/CONIITI57704.2022.9953636.
[12] H.S. Metha, K. Kusrini, D. Ariatmanto, Classification of types roasted coffee beans using convolutional neural network method, Sinkron (2024). https://api.semanticscholar.org/CorpusID:268836478.
[13] S.J. Chang, C.Y. Huang, Deep learning model for the inspection of coffee bean defects, Appl. Sci. 11 (2021), https://doi.org/10.3390/app11178226.
[14] Y.C. Chou, C.J. Kuo, T.T. Chen, G.J. Horng, M.Y. Pai, M.E. Wu, Y.C. Lin, M. H. Hung, W.T. Su, Y.C. Chen, D.C. Wang, C.C. Chen, Deep-learning-based defective bean inspection with gan-structured automated labeled data augmentation in coffee industry, Applied Sciences 9 (2019), https://doi.org/10.3390/app9194166.
[15] H.C. Bazame, J.P. Molin, D. Althoff, M. Martello, Detection, classification, and mapping of coffee fruits during harvest with computer vision, Comput. Electr. Agricult. 183 (2021) 106066, https://doi.org/10.1016/j. compag.2021.106066.
[16] M. Jumarlis, M. Mirfan, A. Manga, Classification of coffee bean defects using gray-level co-occurrence matrix and k-nearest neighbor, ILKOM Jurnal Ilmiah 14 (2022) 1–9, https://doi.org/10.33096/ilkom.v14i1.910.1-9.
[17] E.R. Arboleda, A.C. Fajardo, R.P. Medina, Classification of coffee bean species using image processing, artificial neural network and k nearest neighbors, in: 2018

IEEE International conference on innovative research and development (ICIRD), 2018, pp. 1–5, https://doi.org/10.1109/ICIRD.2018.8376326.
[18] J.S. Cho, H.J. Bae, B.K. Cho, K.D. Moon, Qualitative properties of roasting defect beans and development of its classification methods by hyperspectral imaging technology, Food Chem. 220 (2017) 505–509, https://doi.org/10.1016/j.foodchem.2016.09.189.
[19] N.F. Huang, D. Chou, C. Lee, F. Wu, A. Chuang, Y. Chen, Y. Tsai, Smart agriculture: real-time classification of green coffee beans by using a convolutional neural network, IET Smart Cities 2 (2020), https://doi.org/10.1049/iet-smc.2020.0068.
[20] C. Pinto, J. Furukawa, H. Fukai, S. Tamura, Classification of green coffee bean images basec on defect types using convolutional neural network (cnn), in: 2017 International conference on advanced informatics, concepts, theory, and applications (ICAICTA), 2017, pp. 1–5, https://doi.org/10.1109/ICAICTA.2017.8090980.
[21] M. Dogan, Y.S. Taspinar, I. Cinar, R. Kursun, I.A. Ozkan, M. Koklu, Dry bean cultivars classification using deep cnn features and salp swarm algorithm based extreme learning machine, Comput. Electr. Agricult. 204 (2023) 107575, https://doi.org/10.1016/j.compag.2022.107575. URL:.
[22] S. C. A. of America (SCAA), Spcialty coffee association of american arabica green coffee defect handbook, URL: https://cdn.coffeestrategies.com/wp-content/uploads/2020/08/Green-Coffee-Defect-Handbook.pdf.
[23] J.N. Wintgens, Coffee: growing, processing, sustainable production: a guidebook for growers, processors, traders, and researchers, Internat. Coffee Organiz. (2004), https://doi.org/10.1002/9783527619627.
[24] C. Shorten, T. Khoshgoftaar, A survey on image data augmentation for deep learning, J. Big Data 6 (2019), https://doi.org/10.1186/s40537-019-0197-0.
[25] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proceed. IEEE 86 (1998) 2278–2324.
[26] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016.
[27] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L. Chen, Mobilenetv2: inverted residuals and linear bottlenecks, in: Conference on computer vision and pattern recognition (CVPR), 2018.
[28] A. Howard, M. Sandler, G. Chu, L. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q.V. Le, H. Adam, Searching for mobilenetv3, in: International conference on computer vision (ICCV), 2019.
[29] M. Tan, Q.V. Le, Efficientnetv2: smaller models and faster training, arXiv preprint arXiv:2104.00298 (2021).
[30] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Conference on computer vision and pattern recognition (CVPR), 2016.
[31] K. He, X. Zhang, S. Ren, J. Sun, Identity mappings in deep residual networks, in: European conference on computer vision (ECCV), 2016.