Journal of Applied Informatics and Computing (JAIC)
Vol.9, No.4, August 2025, pp. 1354~1362
e-ISSN: 2548-6861

Visual Segmentation and Classification of Coffee Beans After Roasting

Firdaus Alamanda [1]*, Rudy Susanto [2], Wiji Lestari [3]****
* Informatics Engineering Study Program, Duta Bangsa University Surakarta
** Faculty of Computer Science, Duta Bangsa University Surakarta
210103016@mhs.udb.ac.id [1], rudi_susanto@udb.ac.id [2], wiji_lestari@udb.ac.id [3]

Article Info

Article history:

Received 2025-06-05
Revised 2025-07-05
Accepted 2025-07-07

Keyword:

Coffee Bean,
Roasting Level,
Image Segmentation,
Image Classification,
Deep Learning.

ABSTRACT

This research aims to develop an image-based system for segmenting and classifying coffee beans after roasting using deep learning. A U-Net architecture was applied to isolate coffee beans from the background with high spatial accuracy, achieving a mean Intersection over Union (IoU) of 0.8833 and Dice Coefficient of 0.9375. The segmented images were then classified into six roasting levels green, light, light to medium, medium, medium to dark, and dark using a modified ResNet-50 model, which reached an overall classification accuracy of 86%. The system demonstrates strong performance for clear categories but shows overlapping predictions for visually similar classes such as "medium" and its neighboring levels, indicating that boundaries between roasting stages can be ambiguous. This study provides an objective and automated alternative for roast quality inspection, reducing reliance on subjective human assessment. However, to meet industrial standards, further improvements are needed, such as integrating additional image features or ensemble models to increase discrimination power. This two-stage system serves as a promising foundation for future developments in automated coffee quality control.

I. INTRODUCTION

The growing interest in specialty and premium-grade coffee has driven the global coffee industry to improve quality assurance in every stage of production, especially roasting. The roasting process defines not only the aroma and taste but also the final appearance of coffee beans, making it a critical step in determining overall quality. However, in many traditional practices, the assessment of roast quality remains subjective, relying heavily on the visual experience of skilled roasters. This method often leads to inconsistent outcomes and difficulty in maintaining uniform product standards [1], [2]. To overcome this limitation, digital image processing has been explored as a more objective approach for evaluating coffee bean characteristics. By analyzing visual features such as shape, color intensity, and texture, image-based techniques allow for consistent assessment of roasting levels. Several studies have implemented feature extraction methods such as HSV color space, Principal Component Analysis (PCA), and texture analysis to classify coffee beans based on visual cues related to their roasting stage [3], [4], [2].

In more recent advancements, machine learning and deep learning have emerged as powerful tools in the classification of agricultural products, including coffee. Models such as Backpropagation Neural Networks [1], Decision Trees [2], and MobileNet [5] have been applied to classify roasted coffee beans, yielding promising results in both accuracy and efficiency. Among these, Convolutional Neural Networks (CNNs) have demonstrated superior performance in learning complex visual patterns directly from image data [6]. Furthermore, deep architectures such as Residual Networks (ResNet) enable better training of deeper models by addressing the vanishing gradient problem and have been successfully applied in various classification tasks [7].

In light of these developments, this study proposes a deep learning-based system that combines the U-Net architecture for segmenting roasted coffee beans from their background and the ResNet-50 architecture for classifying them into six defined roasting levels: green, light, light to medium, medium, medium to dark, and dark. This integrated approach aims to deliver an objective, automated, and accurate method for evaluating post-roasting coffee bean quality using digital images.

II. METHOD

This study adopts an experimental approach to develop a segmentation and classification system for post-roasting coffee beans, following five main stages: Analysis, Data Collection, Preprocessing, Model Training, and Evaluation, as illustrated in Figure 1.

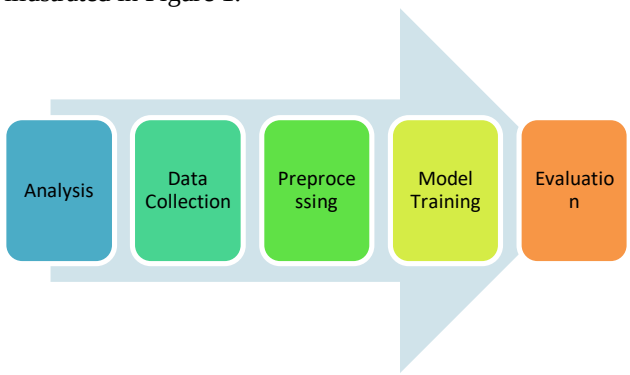

Figure 1. Phases of Research

A. Analysis

The objective of this study was to formulate a method that could objectively assess the roast level of coffee beans through image-based learning. The investigation began with a critical review of image-analysis and deep-learning literature. Prior work has demonstrated that encoder–decoder architectures such as U-Net are highly effective for pixel-level delineation in various domains, including liver CT segmentation and road-surface inspection [8], [9]. due to their ability to preserve spatial information during segmentation. For the classification stage, deep residual networks such as ResNet-50 have been shown to outperform conventional CNNs by mitigating vanishing-gradient issues and enabling deeper network structures [7], [10]. Drawing on these insights, the U-Net architecture was selected for the segmentation of coffee beans, while a modified ResNet-50 backbone with six output nodes was employed to classify the roast level.

B. Data Collection

A This study utilized a publicly available dataset sourced from the Kaggle platform, comprising a total of 1,800 RGB images of roasted coffee beans. The dataset consisted of a total of 2,934 images, split into 2,416 images for training and 518 for validation. An additional independent test set of 300 images was used to evaluate real-world generalization. This ensured a balanced distribution of classes, eliminating the need for additional data balancing techniques such as SMOTE or oversampling. All images were manually labeled by a certified coffee cupper to ensure that the roasting level categories reflected standard cupping guidelines. Visual labeling was done under consistent lighting conditions to minimize annotation bias. The dataset was randomly split into three subsets: 70% for training, 15% for validation, and 15% for testing, ensuring that each subset preserved class balance. The dataset includes images captured under varying lighting intensities and contains multiple bean varieties with different roasting techniques ranging from drum roasting to hot air roasting, to reflect realistic post-roasting scenarios.

C. Digital Image

All images were resized to a resolution of 256×256 pixels to ensure consistency in input dimensions across the model. To enhance the robustness of the model and mitigate overfitting, on-the-fly data augmentation was implemented using an image generator during training. The augmentation parameters included:

1. *rescale=1./255:* Normalizes pixel values from the range [0, 255] to [0, 1] to aid model convergence.
2. *rotation_range=30:* Applies random rotations up to ±30 degrees.
3. *width_shift_range=0.2 and height_shift_range=0.*2: Shifts the image horizontally or vertically by up to 20% of the total dimension.
4. *shear_range=0.2*: Applies shear transformations to simulate angular distortions.
5. *zoom_range=0.2*: Performs random zooming in or out up to 20%.
6. *horizontal_flip=True*: Randomly flips the image horizontally.
7. *brightness_range=[0.8, 1.2]*: Randomly adjusts brightness within a defined range.
8. *fill_mode='nearest'*: Fills empty pixels after transformation using the nearest valid pixel values.

These techniques were applied in real time during training to simulate various imaging conditions and increase the diversity of input data. No additional normalization (e.g., histogram equalization) was used, as the dataset was already captured under uniform lighting and background conditions.

D. Model Training

The model development followed a two-stage pipeline comprising segmentation and classification. In the first stage, a U-Net architecture was implemented to segment coffee beans from the background in each image. This preprocessing step was critical for reducing background noise and enhancing the model's focus on the relevant morphological features of the beans. The U-Net was trained for 50 epochs with 79 batches per epoch, using the Dice Loss function, which is particularly effective for evaluating the overlap between predicted masks and ground truth in binary segmentation tasks. Dice Loss also guided segmentation refinement throughout training, resulting in more accurate boundary delineation [8], [9].

In the second stage, the segmented images were passed into a modified ResNet-50 network, in which the final fully connected layer was replaced with a six-node softmax output corresponding to the six roast-level classes. This deep residual architecture was selected due to its ability to maintain stable gradient flow across layers, thereby mitigating vanishing-gradient issues and enabling deeper feature learning

compared to conventional convolutional neural networks (CNNs) [7], [10]. Residual shortcuts further facilitated faster convergence, as demonstrated in prior applications involving plant disease identification and object detection.

Classification was conducted in two stages. In Stage 1, only the top layers of the network were trained while the base layers of ResNet-50 remained frozen. This stage was run for 20 epochs with 76 batches per epoch. In Stage 2, fine-tuning was performed by unfreezing selected deeper layers of the network, allowing for end-to-end optimization. This stage also involved 20 epochs with 76 batches per epoch. The Adam optimizer was used with a learning rate of 1×10^{-4}, and categorical cross-entropy served as the loss function due to its suitability for multi-class classification tasks. Model performance was continuously monitored on a validation set, and early stopping was applied to prevent overfitting by terminating training if no improvement in validation loss was observed over a predefined number of epochs.

E. Evaluation

Segmentation performance was measured using Intersection over Union (IoU) and Dice Coefficient, while classification performance was evaluated using accuracy, precision, recall, and F1-score. Visual examples of segmentation outputs and a confusion matrix were included to interpret the model's behavior. It should be noted that the model was validated on an in-distribution test set. Future evaluations are planned using out-of-distribution (OOD) data to assess generalization across various imaging conditions and sources.

.

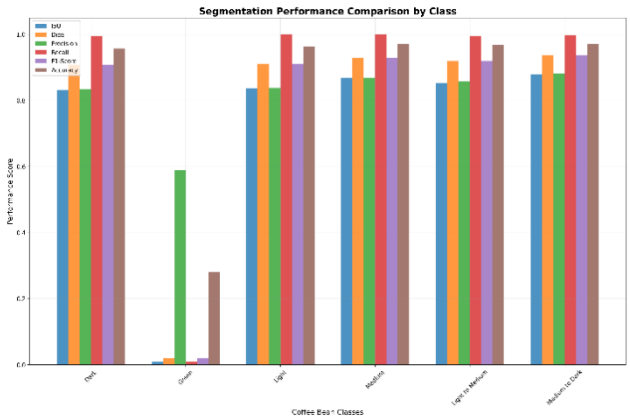

Figure 2. Segmentation class comparison

Figure 2, shows a comparative analysis of segmentation performance across all six coffee bean classes. Metrics including IoU, Dice, Precision, Recall, F1-Score, and Accuracy indicate consistent performance, with most classes achieving scores above 0.85. However, the 'Green' class shows reduced segmentation metrics, suggesting visual variability or insufficient representation in segmentation-specific preprocessing.

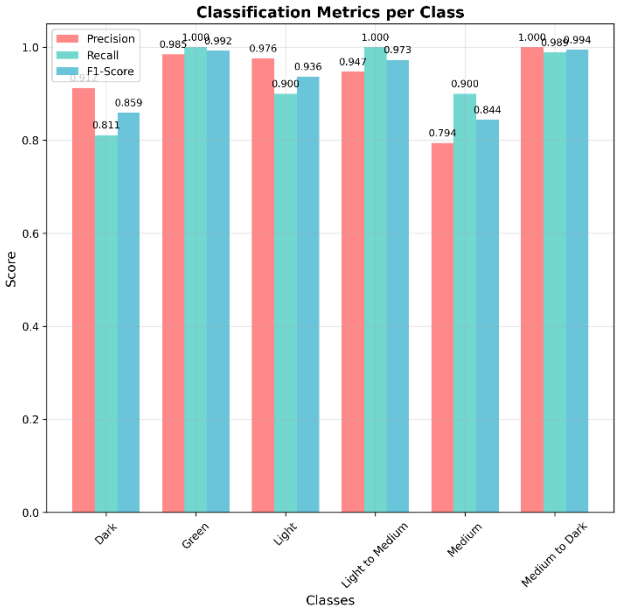

Figure 3. Classification Metrics per Class

Figure 3, provides a detailed view of the classification performance across all six classes. The left chart shows class-wise Precision, Recall, and F1-Score, highlighting that 'Green' and 'Light' categories achieve nearly perfect scores. The right chart illustrates the number of test samples per class, confirming balanced representation. While most classes perform well, the 'Medium' category shows slightly reduced performance, possibly due to its visual similarity with adjacent roast levels.

F. Model Architecture

1). Digital Image: A digital image represents a visual object using numerical values arranged in two-dimensional space, where each value corresponds to a pixel that carries intensity or color information. These pixel-based representations are essential in various computer vision applications, especially in systems supported by artificial intelligence for pattern classification and clustering [11]. In this research, digital image processing serves as the foundation for extracting relevant visual cues from coffee beans after roasting.

Common image processing techniques, including color space transformation (such as HSV), edge detection, and thresholding, are widely applied to identify visual differences in roasted coffee beans [12]. These methods allow systems to highlight texture, shape, and color contrast, which are crucial indicators of roasting levels. For example, Prastyaningsih and Kusrini [13] employed HSV-based feature extraction and a similarity measure using Minkowski distance to classify coffee bean roast levels, demonstrating the effectiveness of visual features in differentiating roasting maturity. Although the study did not incorporate deep learning, the role of color-based descriptors remained central in modeling post-roasting appearances.

2). Deep Learning: A branch of artificial intelligence, has become increasingly relevant in automated visual analysis tasks. It employs multi-layered neural networks capable of learning complex patterns directly from data without relying on manually engineered features. These networks simulate aspects of human cognitive processes by recognizing patterns in both structured and unstructured datasets [11].

The effectiveness of deep learning has been demonstrated across various fields, including image classification, object detection, and medical diagnostics. Its layered architecture allows the model to gradually abstract low-level features (e.g., edges, colors) into high-level representations, which are critical for accurate decision-making [14]. In the context of coffee bean analysis, deep learning enables robust feature extraction from post-roasting images, eliminating the need for hand-crafted rules or filters. Prior work has shown that deep architectures significantly outperform traditional machine learning approaches in terms of flexibility, scalability, and accuracy [15].

3). Convolutional Neural Network (CNN): Convolutional Neural Networks (CNNs) are a class of deep learning models specifically designed for visual data processing. By using convolutional layers, CNNs are able to detect spatial hierarchies of features such as edges, patterns, and textures within an image. These features are progressively abstracted through multiple layers to enable effective classification or recognition tasks [15].

A typical CNN architecture includes convolutional layers for feature extraction, activation functions (such as ReLU) for introducing non-linearity, pooling layers for dimensionality reduction, and fully connected layers that serve as the decision-making stage. The integration of these components allows CNNs to learn relevant patterns directly from image data, which eliminates the need for manual feature engineering. In this study, CNN serves as the foundational concept for both segmentation and classification networks, making it suitable for identifying roasting levels in post-roasting coffee beans [8].

4). U-Net: U-Net is a specialized type of Convolutional Neural Network (CNN) commonly applied in segmentation tasks, particularly in the medical imaging domain. Its architecture is composed of two main sections: the encoder (also known as the contracting path), which compresses spatial information to extract deeper semantic features; and the decoder (or expansive path), which restores the original image dimensions to produce a segmentation output [8]. In this research, the implemented U-Net structure is illustrated in Figure 2. The encoder is positioned on the left and the decoder on the right, with gray blocks indicating the input/output layers, green for pooling operations, blue for upsampling through transposed convolutions, and orange for standard convolution layers.

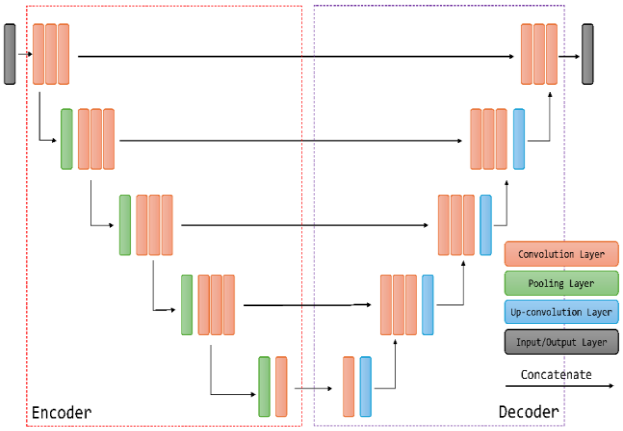

Figure 4. U-Net architecture

U-Net is a deep learning architecture initially developed for biomedical image segmentation but has since proven effective in tasks requiring precise localization of objects. It features a symmetric encoder-decoder structure, where the encoder compresses the spatial information into abstract representations, and the decoder reconstructs these features back to the original resolution. Crucially, U-Net incorporates skip connections that link encoder layers to their corresponding decoder layers, preserving fine-grained spatial details [8], [9].

In this research, U-Net is employed to isolate coffee beans from their backgrounds before classification. Its ability to maintain spatial consistency is essential for small object segmentation, such as individual coffee beans with varying surface textures. By integrating a pretrained ResNet50 encoder, the segmentation process benefits from faster convergence and improved accuracy, especially when handling visually complex samples like roasted beans [9].

5). ResNet: Residual Networks (ResNet) are a family of deep CNN architectures designed to address the degradation problem in very deep networks. The core innovation in ResNet is the use of shortcut connections, which allow gradients to flow through the network more efficiently by skipping one or more layers. This residual learning mechanism helps preserve important information across layers and facilitates training of much deeper networks [7], [10].

In this study, ResNet-50 is adopted for classifying segmented coffee bean images into six distinct roasting levels. By leveraging deep residual blocks, the model is capable of learning complex visual features associated with different degrees of roasting. The shortcut connections not only improve model convergence but also support better generalization by maintaining stability during backpropagation. This makes ResNet-50 a suitable backbone for post-roasting quality assessment using image-based input [10].

III. RESULT AND DISCUSSION

A. Segmentation Result

Figure 5 presents a heatmap illustrating the segmentation performance of the U-Net model across all coffee bean classes, evaluated using six key metrics: Intersection over Union (IoU), Dice Coefficient, Precision, Recall, F1-Score, and Accuracy. The matrix reveals strong segmentation performance for most classes, with average scores exceeding 0.85, except for the 'Green' class which significantly underperforms across all metrics.

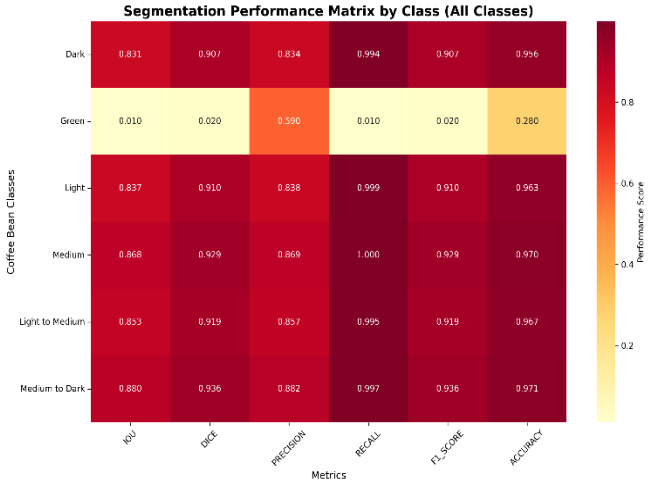

Figure 5. Segmentation Performance Matrix

The highest segmentation quality is observed in the 'Medium to Dark' and 'Light to Medium' classes, where IoU values exceed 0.83 and Dice coefficients approach 0.94. These results reflect the model's ability to consistently isolate coffee beans with clearly distinguishable surface features and color intensity. Notably, the Recall values for all classes except 'Green' approach 0.99 or even reach 1.0, indicating that the model effectively captures the complete object regions without significant under-segmentation.

However, the 'Green' class exhibits critically low segmentation scores, with an IoU of 0.01 and Dice of 0.02, accompanied by low precision and recall. This underperformance can likely be attributed to two factors: (1) visual similarity between green beans and the image background due to minimal roasting coloration, and (2) potential inconsistencies or noise in the ground truth masks during training for this specific class. Such disparities highlight the challenges faced in segmenting objects with subtle visual boundaries, suggesting a need for tailored preprocessing or attention-based segmentation refinement for underrepresented or visually ambiguous classes.

In summary, the segmentation model achieves high accuracy and robustness for most roasting categories, but shows clear limitations in handling visually low-contrast classes such as 'Green'. These findings support the use of the current U-Net configuration for general-purpose segmentation while also motivating future research into hybrid segmentation approaches or additional feature enhancement techniques to improve segmentation reliability for visually subtle bean categories.

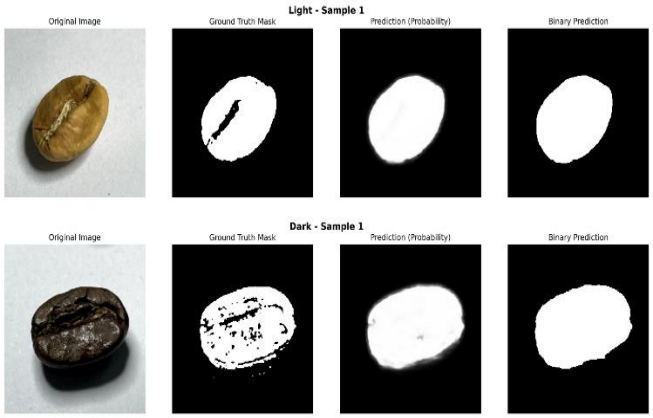

Figure 6. Dark beans segmentation with ground truth

Figure 6 presents visual samples comparing the original image, ground truth mask, predicted probability map, and binarized segmentation output for two coffee bean classes: *Light* and *Dark*. These visualizations serve to demonstrate the performance of the U-Net segmentation model across distinct roasting levels, while also addressing the need for direct comparison between model output and reference annotations (ground truth).

In the *Light Sample* image, the predicted mask exhibits slight boundary irregularities when compared to the ground truth. While the core region of the coffee bean is successfully identified, minor noise appears at the contour level, particularly in the lower-left segment. The probability map also shows reduced confidence near the edge areas, which might be attributed to the relatively low contrast between the bean and the background. However, the binarized result still aligns reasonably well with the annotated mask, indicating acceptable segmentation despite some uncertainty in edge delineation.

In contrast, the *Dark Sample* image reveals a significantly more accurate segmentation. The predicted mask closely aligns with the ground truth, both in shape and coverage. The probability map demonstrates high confidence across the entire object, suggesting that the model is particularly effective when dealing with darker roasted beans, likely due to higher color contrast and more defined surface features. The binarized prediction preserves the contours well, minimizing noise and confirming strong model generalization in this class. These visualizations confirm that the segmentation stage of the system is not only quantitatively validated (as shown in Figure 5) but also qualitatively interpretable. They highlight both the strengths of the U-Net model in handling high-contrast classes (e.g., *Dark*) and the challenges it faces with low-contrast classes (e.g., *Light*). This insight supports further investigation into post-processing or

attention mechanisms to improve boundary detection in visually ambiguous samples.

B. Classification Result

The model addresses visual ambiguity between adjacent roasting levels using multiple implicit strategies. First, the ResNet-50 architecture extracts deep, hierarchical features that capture subtle distinctions in texture and surface shading. Second, probability-based outputs allow the model to express uncertainty, particularly when predictions fall between visually similar categories, such as medium and light-to-medium. Additionally, extensive data augmentation during training improves robustness to visual variability, encouraging the model to focus on essential discriminative patterns. Finally, a balanced class distribution ensures that each roast level is equally represented during learning, reducing the risk of bias toward more distinct categories.

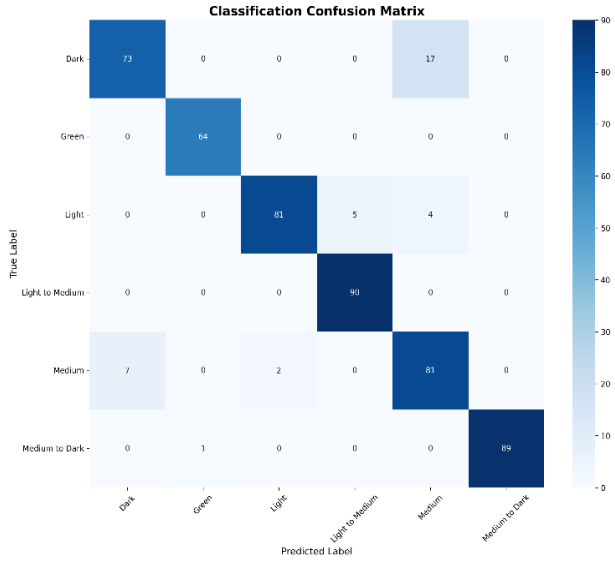

Figure 7. Classification Confusion Matrix

Figure 7 illustrates the classification confusion matrix for the ResNet-50 model applied to six roasting level categories: Green, Light, Light to Medium, Medium, Medium to Dark, and Dark. Each cell indicates the number of test samples predicted by the model for a specific class, given the true label on the vertical axis.

The model achieves near-perfect classification performance for several categories, including Green (64/64) and Light to Medium (90/90), reflecting strong discriminative capability when roast level features are visually distinct. The Medium to Dark and Light classes also exhibit high classification accuracy, with only one and nine misclassifications respectively, most of which occurred within adjacent roast levels.

However, the *Medium* and *Dark* classes show greater confusion. For *Dark*, 17 samples were incorrectly classified as *Medium*, indicating substantial visual overlap between the darker roast categories. Similarly, 7 *Medium* samples were misclassified as *Dark*, with a few others scattered across nearby categories. This misclassification trend highlights the visual ambiguity between *Medium* and *Dark* beans, which often share similar surface textures and tones. The main cause of misclassification is the subtle similarity in surface color and texture between adjacent roast levels. Some beans also showed visual defects such as uneven surface burning, chipping, or color inconsistency, which further reduced model confidence.

Overall, the confusion matrix reveals that most classification errors occur between adjacent roasting levels, a phenomenon that aligns with the subjective difficulty humans also face in differentiating between subtly varying roast profiles. This suggests potential avenues for improvement, such as integrating additional texture features or using ensemble classifiers to improve the model's ability to resolve fine-grained distinctions between borderline classes.

Figure 7 illustrates two sample predictions generated by the ResNet-50 classification model, demonstrating both the raw input images and the corresponding prediction probability distributions across the six roasting categories.

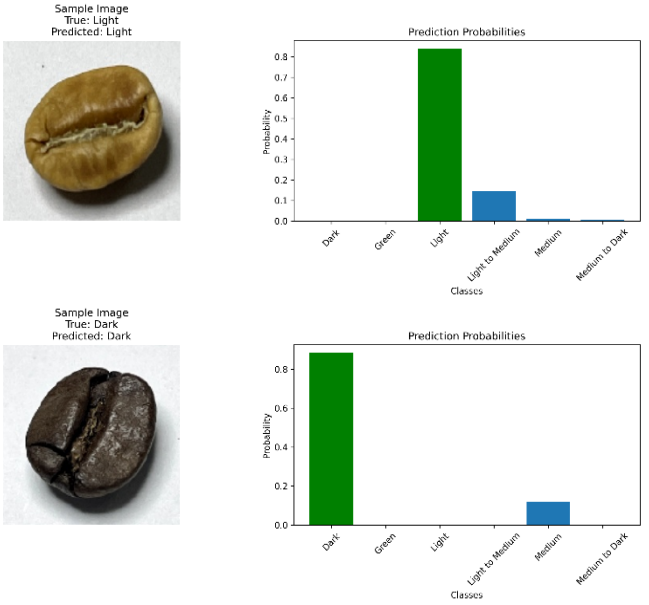

Figure 8. Example predictions for Light and Dark roasted coffee beans

In the first example (top), the true label is Light, and the model correctly predicts the same class with high confidence. The bar chart on the right shows that the predicted probability for the "Light" category exceeds 80%, with minimal probabilities assigned to other categories. This sharp peak in confidence indicates strong model certainty and reliable feature extraction for this class. The morphological features of light-roasted beans, such as a relatively light surface color and pronounced crease line, appear to be consistently learned by the model.

In the second example (bottom), the true class is Dark, and the model again produces a correct classification with over 85% confidence. The histogram shows a highly skewed

prediction towards the "Dark" class, with only a minor response in the "Medium" class, which is visually adjacent in the roasting spectrum. The subtle visual similarity between dark and medium-dark beans could explain the low but non-zero response in the neighboring class. Nonetheless, the model successfully distinguishes the defining characteristics of the dark roast, such as deep brown coloration, oily surface, and broader crease pattern.

These visualizations demonstrate that the model is capable not only of accurate classification but also of assigning prediction confidence that aligns well with visual distinctiveness among classes. However, they also reveal the potential challenge of interclass ambiguity particularly between visually adjacent roast levels like "Light to Medium" and "Medium" or "Medium" and "Medium to Dark." This underscores the importance of incorporating confidence-based evaluation and potentially hybrid decision strategies (e.g., ensemble classifiers or human-in-the-loop verification) for borderline predictions in critical quality control applications.

C. Segmentation Model Evaluation

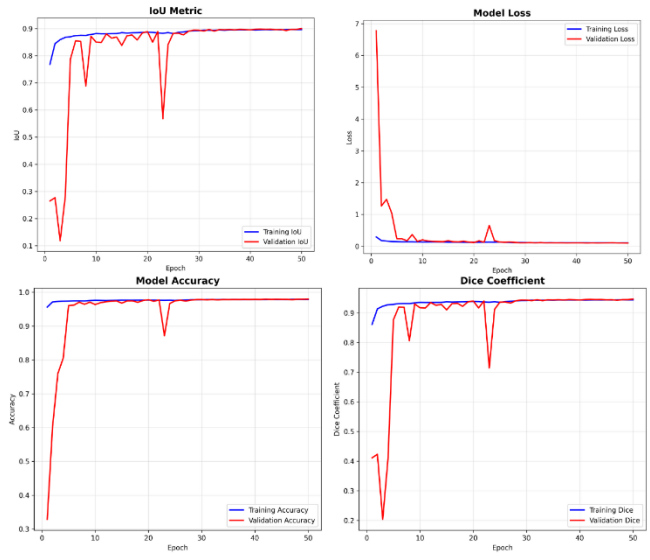

Figure 9. U-Net Segmentation Model Evaluation

Figure 9 illustrates the training curves of the U-Net segmentation model across four performance metrics: Intersection over Union (IoU), model loss, classification accuracy, and Dice Coefficient, recorded over 50 epochs. These plots provide insights into how the model's learning progressed on both the training and validation datasets. Additional tests were conducted with sample images taken under lower and higher ambient lighting to examine model robustness. The IoU and classification accuracy dropped slightly (by about 3–5%), showing that lighting consistency remains important for optimal performance.

The IoU curve shows that the model quickly achieved a high overlap between the predicted and ground truth masks within the first 10 epochs, with training IoU stabilizing above 0.9. Although the validation IoU experienced slight fluctuations most notably a temporary drop around epoch 20 it recovered and returned to stable values, indicating that the model retained strong generalization capacity throughout the training process. The loss curves for both training and validation exhibit a steep decline during the early epochs. Training loss approached near-zero values, while validation loss stabilized around a relatively low point after a brief period of volatility. These patterns suggest that the model effectively minimized prediction errors and maintained consistent performance when exposed to unseen data.

Accuracy trends similarly show a sharp rise during the initial epochs. Both training and validation accuracy surpassed 90% early on and remained relatively constant, further confirming that the model did not suffer from overfitting and continued to generalize well. The Dice Coefficient curves, which measure the similarity between predicted and actual object masks, aligned closely with the IoU trends. The training Dice remained high throughout, and while the validation Dice exhibited brief instability, it ultimately returned to levels indicative of strong segmentation performance.

Collectively, these curves demonstrate that the U-Net model converged efficiently, learned effectively from the training data, and was able to maintain stable and accurate segmentation performance on the validation set. These results support the model's reliability for the segmentation of roasted coffee beans in an image-based evaluation framework.

D. Classification Model Evaluation

The evaluation of the ResNet classification model training process is illustrated in Figure 7, which presents the progression of training and validation accuracy over 50 epochs. At the beginning of the training phase, the training accuracy increased sharply from approximately 0.45 to above 0.6 within the first few epochs. Although the validation accuracy exhibited considerable fluctuations during the early stages, it gradually stabilized and continued to improve, reaching approximately 0.8 by the end of the training process.

This indicates that the model was progressively able to adjust its network weights in response to the training data. The validation accuracy exhibited a different trend initially. The curve showed sharp fluctuations during the first 20 epochs, likely due to the complexity of the validation data distribution or class imbalance in the number of samples. However, beyond this point, the validation accuracy began to demonstrate a more stable upward trend, eventually aligning with the training accuracy in the range of 0.75 to 0.80. This suggests that the model was not only learning from the training data but was also able to generalize effectively to unseen data. The convergence tendency of both curves toward the end of the training is a positive indicator that the model did not suffer from significant overfitting. This is particularly crucial in coffee bean classification systems, considering the visual similarity among different roasting levels, which poses a challenge during the model learning process.

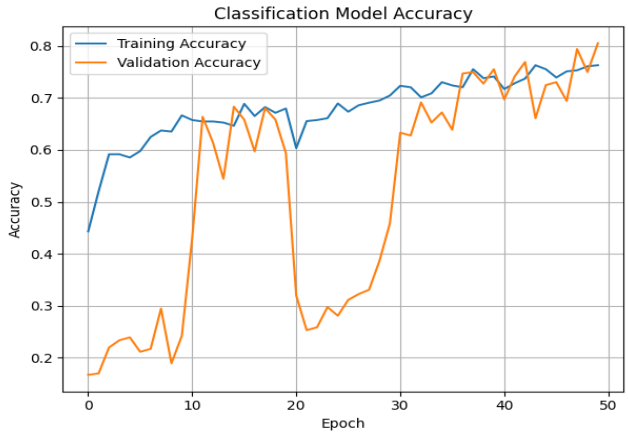

Figure 10. ResNet accuracy and validation

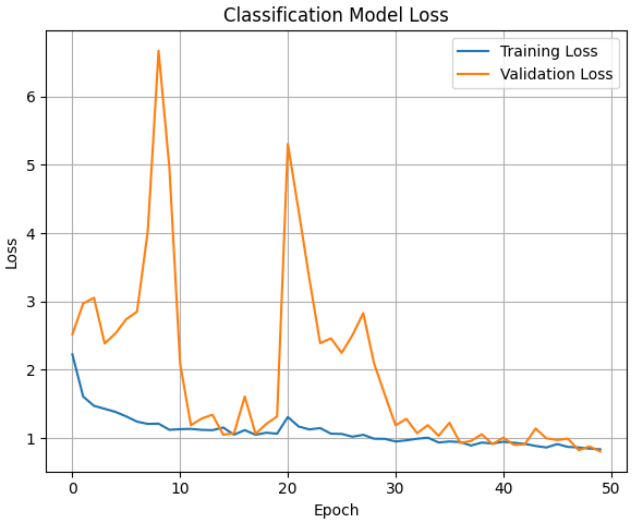

Figure 11. Loss and ResNet validation

Figure 11 presents the progression of loss values throughout the training and validation stages. The training loss demonstrates a steady downward trend, dropping from an initial value near 1.8 to approximately 0.9 by the conclusion of the training process. This pattern reflects the model's continuous improvement in minimizing prediction errors on the training set, which corresponds with the observed increase in classification accuracy reported earlier.

However, the validation loss curve exhibits a more dynamic pattern. Several sharp spikes occurred, particularly around epochs 8 and 20, where the validation loss increased significantly to values exceeding 6. This is most likely due to the presence of outliers or noise within the validation data, or imbalanced class distributions within certain batches. After this phase, the validation loss gradually declined and stabilized around 1.0 toward the end of training. This stability reinforces the notion that the model successfully transitioned out of the exploratory phase and began making more consistent and accurate predictions. Although a gap remains between the training and validation loss, the difference is not substantial, suggesting that the model possesses a reasonable level of generalization capability for the task of automated coffee roasting level classification.

In addition to the confusion matrix, the model's performance was quantitatively evaluated using precision, recall, and F1-score metrics, as presented in Table 1. Precision indicates how accurate the model is when predicting a particular class. For instance, a precision score of 0.95 for the Light class implies that 95% of the Light class predictions were correct. Recall reflects the model's sensitivity in identifying instances of that class; a recall of 0.70 means that only 70% of actual Light samples were correctly recognized. The F1-score, which is the harmonic mean of precision and recall, provides a balanced measure between the two. The Medium class yielded the lowest F1-score (0.69), reinforcing the findings from the confusion matrix that this class experienced the highest rate of misclassification. In contrast, the Light to Medium and Medium to Dark classes achieved the highest F1-scores (0.96), indicating the model's strong consistency in recognizing patterns associated with these two roasting levels.

TABLE I.
SUMMARY OF RESNET MODEL EVALUATION BY CLASS

Kelas	Precision	Recall	F1-score	Support
Dark	0.71	0.93	0.81	300
Green	0.87	0.97	0.92	300
Light	0.95	0.70	0.80	300
Light to Medium	0.97	0.95	0.96	300
Medium	0.74	0.65	0.69	300
Medium to Dark	0.96	0.96	0.96	300
accuracy	0.86			1800
Macro avg	0.87	0.86	0.86	1800
weighted avg	0.87	0.86	0.86	1800

The overall accuracy of the model reached 86%, which is considered high for image-based multi-class classification involving six distinct labels. Additionally, both the macro average and weighted average F1-scores were approximately 0.86, indicating balanced performance despite the presence of classes with relatively lower scores. This comprehensive evaluation confirms that the ResNet-based classification system developed in this study demonstrates strong stability and effectiveness in recognizing variations in coffee bean roasting levels.

IV. CONCLUSION

This This study successfully developed an image-based segmentation and classification system for post-roasting coffee beans using a two-stage deep learning pipeline. The U-Net model demonstrated high performance in isolating coffee beans from the background, achieving an Intersection over Union (IoU) of 0.8833 and Dice Coefficient of 0.9375, indicating effective spatial localization. The classification stage, powered by a modified ResNet-50 architecture, achieved an overall accuracy of 86%, with particularly high performance observed in the "light to medium" and "medium to dark" categories. Through detailed analysis of confusion matrices, prediction probability visualizations, and class-wise performance metrics, it was observed that overlapping visual features between adjacent roasting levels particularly between "light to medium" and "medium," or "medium" and "medium to dark" led to a number of misclassifications. These ambiguities suggest that the visual boundaries between roasting classes are inherently fuzzy, making strict class separation challenging even for deep learning models. This visual overlap could potentially be addressed in future work by incorporating additional image features, such as handcrafted color histograms or texture descriptors, and by exploring ensemble architectures that combine multiple models or attention mechanisms to improve discriminative power.

Although the current system shows strong potential, the achieved accuracy may still be considered moderate for industrial-grade quality control systems, which often demand near-perfect consistency. As such, the model is best positioned as a decision-support tool that can aid human inspectors in roast level classification, particularly in semi-automated workflows. To ensure reliability in real-world applications, early implementations should include human validation, especially when model confidence is low or when ambiguity between classes is detected.

In conclusion, this research presents a viable foundation for automating visual quality assessment in coffee roasting. Future improvements can target domain generalization, integration with multimodal data (e.g., aroma or temperature), and deployment via mobile or web platforms for broader usability across the coffee supply chain.

BIBLIOGRAPHY

[1] A. Taslim, S. Sudin, M. Dzikrullah Suratin Identifikasi Mutu Roasting Biji Kopi Menggunakan Fitur Warna Dan Backpropagation, and M. Dzikrullah Suratin, "Identifikasi Mutu Roasting Biji Kopi Menggunakan Fitur Warna Dan Backpropagation Identification Of Roasting Quality Coffee Beans Using Color And Backpropagation Features," 2023.

[2] Purnomo Haykal Mohammed, Raharjo Jangkung, and Magdelena Rita, "Deteksi Kualitas Biji Kopi Melalui Pengolahan Citra Digital Dengan Metode Adaptive Region Growing Dan Klasifikasi Decision Tree (Coffee Bean Quality Detection Through Digital Image Processing With Adaptive Region Growing Method And Decision Tree Classification)," 2021.

[3] D. Aditya Nugraha and A. Sartika Wiguna, "Seleksi Fitur Warna Citra Digital Biji Kopi Menggunakan Metode Principal Component Analysis Digital Image Selection of Coffee Seed Using Component Analysis Method," 2020.

[4] J. Khatib Sulaiman, N. Amelia, M. Garonga, J. Rusman, and I. Artikel Abstrak, "Penerapan Metode K-Nearest Neighbor (Knn) Untuk Klasifikasi Kematangan Buah Kopi," *Indonesian Journal of Computer Science Attribution*, vol. 12, no. 2, 2023.

[5] Firmansyah Tegar, Kurniawan Rudi, and Hidayat Toyib Asep, "Klasfikasi Tingkat Kematangan Roasting Biji Kopi Berbasis Deep Learning dengan Arsitektur MobileNet," 2025, doi: 10.47065/josh.v6i2.6811.

[6] R. Janandi and T. W. Cenggoro, *An Implementation of Convolutional Neural Network for Coffee Beans Quality Classification in a Mobile Information System*. 2020. doi: 10.1109/ICIMTech50083.2020.9211257.

[7] A. Ridhovan *et al*., "Penerapan Metode Residual Network (RESNET) Dalam Klasifikasi Penyakit Pada Daun Gandum," 2022.

[8] M. A. Djohar *et al*., "Liver Segmentation Using Convolutional Neural Network Method with U-Net Architecture," *Journal of Informatics And Telecommunication Engineering*, vol. 6, no. 1, pp. 221–234, Jul. 2022, doi: 10.31289/jite.v6i1.6751.

[9] A. Di Benedetto, M. Fiani, and L. M. Gujski, "U-Net-Based CNN Architecture for Road Crack Segmentation," *Infrastructures (Basel)*, vol. 8, no. 5, May 2023, doi: 10.3390/infrastructures8050090.

[10] A. I. Mohammed and A. AK. Tahir, "A New Optimizer for Image Classification using Wide ResNet (WRN)," *Academic Journal of Nawroz University*, vol. 9, no. 4, p. 1, Sep. 2020, doi: 10.25007/ajnu.v9n4a858.

[11] A. Kaur, Y. Singh, N. Neeru, L. Kaur, and A. Singh, "A Survey on Deep Learning Approaches to Medical Images and a Systematic Look up into Real-Time Object Detection," Jun. 01, 2022, *Springer Science and Business Media B.V.* doi: 10.1007/s11831-021-09649-9.

[12] J. Jumadi and D. Sartika, "Pengolahan Citra Digital Untuk Identifikasi Objek Menggunakan Metode Hierarchical Agglomerative Clustering," 2021.

[13] Y. Prastyaningsih, W. Kusrini, P. Negeri Tanah Laut, J. A. Yani KM, D. Panggung KecPelaihari KabTanah Laut, and K. Selatan, "Sistem Temu Kembali Citra Pada Level Roasting Biji Kopi Menggunakan Ekstraksi Fitur Warna," vol. 6, no. 2, 2021.

[14] Y. Hafifah, K. Muchtar, A. Ahmadiar, and S. Esabella, "Perbandingan Kinerja Deep Learning Dalam Pendeteksian Kerusakan Biji Kopi," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 6, p. 1928, Dec. 2022, doi: 10.30865/jurikom.v9i6.5151.

[15] B. T. W. Putra, R. Amirudin, and B. Marhaenanto, "The Evaluation of Deep Learning Using Convolutional Neural Network (CNN) Approach for Identifying Arabica and Robusta Coffee Plants," *Journal of Biosystems Engineering*, vol. 47, no. 2, pp. 118–129, 2022, doi: 10.1007/s42853-022-00136-y.