

Random Forest Implementation for Indonesian Coffee Identification Based on Digital Image Feature Extraction

Widya Lelisa Army^{1*}, Abdul Fadlil², Sunardi³

^{1,2,3}Doctoral Program in Informatics, Universitas Ahmad Dahlan, Yogyakarta, Indonesia 55191

¹Department of Information System, Universitas Pertiwi, Bekasi, Indonesia

¹2537083004@webmail.uad.ac.id, widya.lelisa@pertiwi.ac.id, ²fadlil@uad.ac.id, ³sunardi@uad.ac.id

Submitted : Jul 14, 2026 | **Accepted** : Aug 17, 2026 | **Published** : Oct 5, 2026

Abstract: Accurate identification of coffee types is a critical challenge in Indonesia's coffee industry, which relies heavily on subjective human visual inspection. This study implements the Random Forest ensemble learning algorithm to classify three commercially important coffee species Arabica (*Coffea arabica*), Liberica (*Coffea liberica*), and Robusta (*Coffea canephora*) using digital image features. The dataset consists of 1,913 real coffee bean images from Roboflow Universe (Coffee Bean Type v1, CC BY 4.0), split into train (1,530), valid (287), and test (96). Feature extraction produces an 11-dimensional vector combining five Gray Level Co-occurrence Matrix (GLCM) texture features (Contrast, Energy, Homogeneity, Correlation, Dissimilarity) computed at four angles and six RGB color statistical features (mean and standard deviation per channel). The model is trained on 1,817 combined train-valid samples with optimal hyperparameters ($n_estimators=200$, $max_depth=None$, $criterion=gini$) selected via Grid Search with 5-fold stratified cross-validation. Evaluation on 96 held-out test samples yields accuracy 98.96%, macro precision 99.05%, macro recall 98.67%, macro F1-score 98.84%, and ROC-AUC 0.9996, with only one misclassification. Cross-validation confirms model stability at $98.07\% \pm 0.60\%$. Feature importance analysis identifies GLCM Energy (21.98%) and Homogeneity (14.57%) as the most discriminative features. The Robusta class achieves perfect classification (100% all metrics). Manual verification confirms that the reported metrics are consistent with the confusion matrix. These results demonstrate that Random Forest with GLCM and RGB color statistical features is effective for image-based coffee classification, achieving higher accuracy than several prior KNN and neural-network-based studies, although those studies differ in task, dataset, and class composition and are therefore not directly comparable.

Keywords: coffee bean classification; ensemble learning; feature importance; GLCM; random forest

INTRODUCTION

Indonesia is 4th ranks among global coffee producers, with export revenues of USD 929 million in 2023 (Direktorat Jenderal Perkebunan, 2023). The three commercially cultivated species *Coffea arabica* (Arabica), *Coffea liberica* (Liberica), and *Coffea canephora* (Robusta) differ substantially in flavor profile, market value, and physical bean morphology, making accurate classification critical for quality assurance, pricing, fraud prevention, and origin traceability (Rahardjo, 2015; Motta et al., 2025). Manual identification by sensory experts remains the dominant practice, despite being subjective, time-consuming, and not scalable. Computer vision and machine learning offer objective and reproducible alternatives. Random Forest (RF), introduced by Breiman (2001), is a non-parametric ensemble method that builds multiple decision trees using bootstrap aggregating (bagging) and random feature selection. Its key advantages include robustness to overfitting, effectiveness on moderate-sized datasets, no requirement for feature normalization, and built-in feature importance via Mean Decrease Impurity (MDI) making it highly interpretable compared to deep learning approaches (Liaw & Wiener, 2002). Prior studies on coffee classification have generally been limited to binary bean-type classification using small local datasets and classical texture-based features (Jatmoko & Sinaga, 2022), or to related but distinct tasks such as electronic-

*name of corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.



nose-based multi-class classification (Triyono et al., 2023) and leaf-disease identification (Fatchurrachman & Udjulawa, 2023). Deep learning approaches achieve high accuracy but require large datasets, GPU resources, and produce black-box models (Hakim et al., 2020; Metha et al., 2024). A study using RF on a real, publicly available three-class coffee dataset with full metric verification remains absent in the literature. This study addresses the gap with three contributions: (1) evaluation of RF on 1,913 real coffee bean images including Liberica a class rarely studied; (2) feature importance analysis identifying the most discriminative visual characteristics; and (3) complete manual metric verification from the confusion matrix for methodological transparency. The methodological novelty of combining Random Forest with GLCM and RGB color statistics is modest, as this combination is well established in image classification; the primary contribution of this study lies in its empirical scope (a real, three-class, publicly available dataset including Liberica) and its transparent, fully verified evaluation rather than in a new algorithm.

LITERATURE REVIEW

Random Forest

Random Forest is an ensemble learning method that constructs multiple decision trees in parallel using bootstrap samples of the training data. At each node split, a random subset of features is evaluated instead of considering all available features. This random feature selection increases diversity among trees, reduces correlation between them, and improves generalization performance (Breiman, 2001).

$$\hat{y}(x) = \text{mode}\{T_1(x), T_2(x), \dots, T_h(x)\} \quad (1)$$

Where :

$\hat{y}(x)$: predicted class label for observation x ;

$T_i(x)$: prediction generated by the i -th decision tree;

h : total number of trees in the forest.

As shown in Equation (2), the final class label is determined based on the majority vote of all trees. Feature importance is measured via MDI: the weighted average reduction in Gini impurity attributable to each feature across all trees and nodes (Breiman, 2001; Liaw & Wiener, 2002).

$$Gini(t) = 1 - \sum_{i=1}^C p(i|t)^2 \quad (2)$$

where

$Gini(t)$: impurity of node t ;

$p(i|t)$: the proportion (or probability) of samples in node t that belong to class i ;

C : total number of classes.

Gray Level Co-occurrence Matrix (GLCM)

Gray Level Co-occurrence Matrix (GLCM), developed by Haralick et al. (1973), is a statistical method used to describe the spatial relationship between neighboring pixels in a grayscale image. GLCM captures the frequency of occurrence of pixel pairs with specific gray-level values at a given distance and orientation. Each element of the GLCM is represented by Equation (3).

$$P(i, j | d, \theta) = \frac{N(i, j | d, \theta)}{\sum_{i,j} N(i, j | d, \theta)} \quad (3)$$

where:

$P(i, j | d, \theta)$: normalized probability of occurrence of a pixel pair with gray levels i and j ;

$N(i, j | d, \theta)$: number of occurrences of pixel pairs with gray levels i and j ;

i : gray-level intensity of the reference pixel;

j : gray-level intensity of the neighboring pixel;

d : displacement distance between the pixel pairs;

θ : orientation angle used for the co-occurrence calculation.

Equation (3) represents the probability of finding two pixels with gray levels i and j , separated by distance d at orientation angle θ . Based on the normalized GLCM, several texture features can be extracted. In this study, five texture features were utilized: Contrast, Energy, Homogeneity, Correlation, and Dissimilarity. Five texture features are derived from the normalized GLCM:

Contrast :

*name of corresponding author

This is an Creative Commons License This work is licensed under a
Creative Commons Attribution-NonCommercial 4.0 International
License.



$$Contrast = \sum_{i,j} (i - j)^2 p(i, j) \quad (4)$$

where:

Contrast : texture feature measuring local intensity variation;

i : gray-level intensity of the reference pixel;

j : gray-level intensity of the neighboring pixel;

$P(i, j)$: normalized probability of occurrence of the gray-level pair (i, j) .

Energy :

$$Energy = \sum_{i,j} p(i, j)^2 \quad (5)$$

where:

Energy : texture feature measuring image uniformity;

$P(i, j)$: normalized probability of occurrence of the gray-level pair (i, j) ;

i : gray-level intensity of the reference pixel;

j : gray-level intensity of the neighboring pixel.

Homogeneity :

$$Homogeneity = \sum_{i,j} \frac{p(i, j)}{1 + |i - j|} \quad (6)$$

where:

Homogeneity : texture feature measuring the similarity of gray levels;

$P(i, j)$: normalized probability of occurrence of the gray-level pair (i, j) ;

i : gray-level intensity of the reference pixel;

j : gray-level intensity of the neighboring pixel.

Correlation, and Dissimilarity. GLCM computed at four angles (0° , 45° , 90° , 135°) and averaged provides rotation-invariant texture representation (Gonzalez & Woods, 2018).

Coffee Classification Studies

Jatmoko and Sinaga (2022) applied K-Nearest Neighbors (KNN) with GLCM features on a local dataset of coffee bean images, achieving up to 95.00% accuracy for binary Arabica-Robusta bean classification. Triyono et al. (2023) used an Artificial Neural Network (ANN) backpropagation model on electronic nose data across five coffee classes, yielding 88.60% accuracy. Hakim et al. (2020) applied Convolutional Neural Network (CNN) MobileNetV2 for roast degree classification achieving 91.50%, while Metha et al. (2024) reported 92.10% using a synchronous CNN approach. Motta et al. (2025) conducted a comprehensive review of Machine Learning (ML) techniques for coffee classification, confirming the dominance of CNN-based methods but noting interpretability limitations. Mehta et al. (2025) proposed EffResViT-SE FusionNet, a hybrid CNN+ViT model, achieving 99.00% on leaf disease classification with 58,555 images. None of these studies evaluated RF on a publicly available multi-class dataset including Liberica with full metric transparency.

METHODS

Dataset

The dataset used is Coffee Bean Type v1 from Roboflow Universe (Iskandar, 2024), licensed CC BY 4.0, containing 1,913 real coffee bean images at 640×640 pixels across three classes. Roboflow's default split is adopted for reproducibility. Table 1 shows the distribution.

Table 1. Dataset Distribution Coffee Bean Type v1 (Roboflow, CC BY 4.0)

No	Coffee Type	Species	Train	Valid	Test	%
1	Arabica	<i>Coffea arabica</i>	510	98	25	~33%
2	Liberica	<i>Coffea liberica</i>	502	103	34	~34%
3	Robusta	<i>Coffea canephora</i>	518	86	37	~33%
	Total	3 Species	1,530	287	96	100%

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

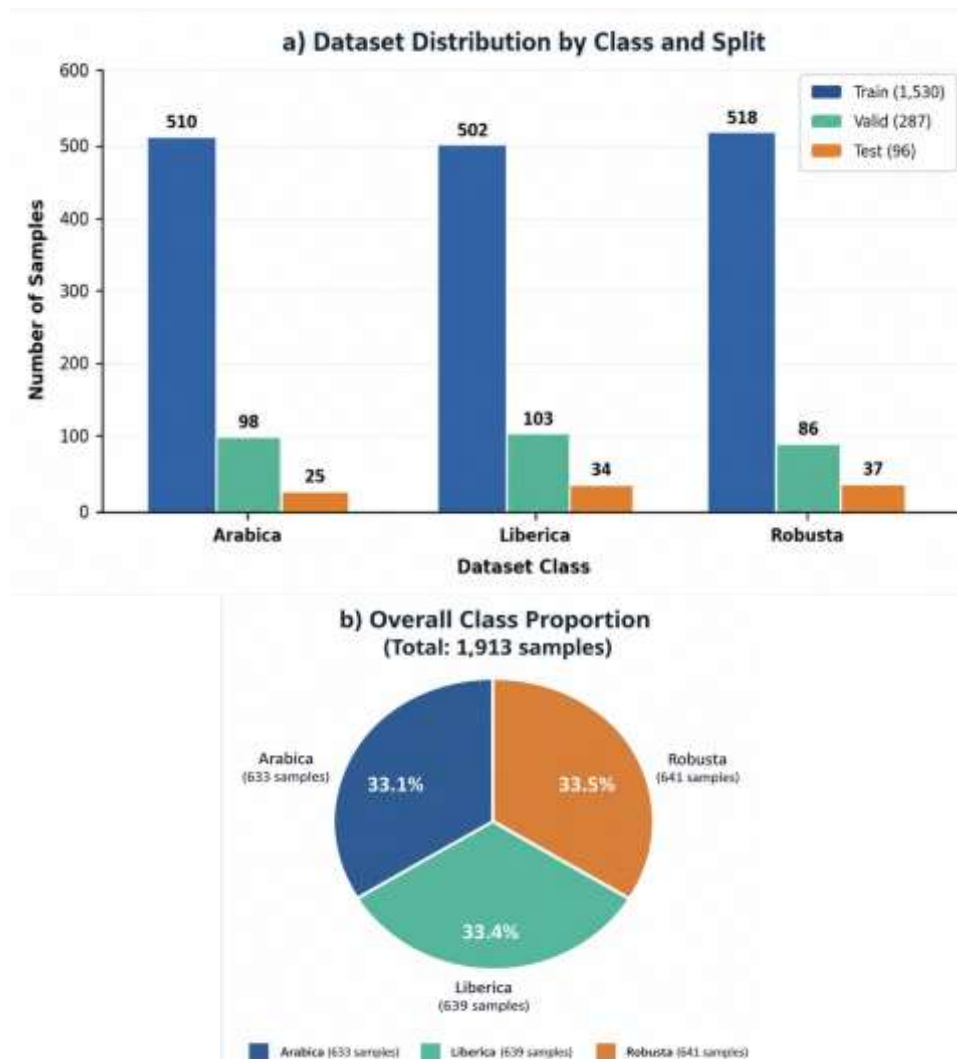


Fig. 1. Dataset distribution (a) per class and split and (b) overall class proportion

Preprocessing

Each image undergoes: (1) resize to 256×256 pixels using bilinear interpolation; (2) conversion to Grayscale (ITU-R BT.601: $Y=0.299R+0.587G+0.114B$) for GLCM extraction; (3) normalization to [0,1] for color features. No augmentation is applied, as RF does not benefit from training-time augmentation.

Feature Extraction

An 11-dimensional feature vector is extracted per image combining GLCM texture features (T) and RGB color statistical features (C) the mean and standard deviation of each channel, not a full color histogram as detailed in Table 2. GLCM is computed at four angles ($\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$), $d=1$, symmetric and normalized, then averaged for rotation invariance.

Table 2. Feature Specification: 11-Dimensional Feature Vector

No	Category	Feature Name	Description
1	Texture	Energy	Texture uniformity; top contributor MDI=21.98%
2	Texture	Homogeneity	GLCM diagonal proximity; MDI=14.57%
3	Texture	Dissimilarity	Inter-pixel difference; MDI=5.20%
4	Texture	Correlation	Linear inter-pixel dependency; MDI=5.14%
5	Texture	Contrast	Local intensity variation; MDI=3.11%
6	Color	Mean Red ($\bar{R}$)	Mean intensity of Red channel; MDI=13.57%

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

No	Category	Feature Name	Description
7	Color	Mean Blue ($\bar{B}$)	Mean intensity of Blue channel; MDI=13.67%
8	Color	Mean Green ($\bar{G}$)	Mean intensity of Green channel; MDI=4.13%
9	Color	Std-G (σG)	Std deviation of Green channel; MDI=12.07%
10	Color	Std-B (σB)	Std deviation of Blue channel; MDI=3.71%
11	Color	Std-R (σR)	Std deviation of Red channel; MDI=2.85%

Random Forest Configuration

The model is trained on the combined train+valid set (1,817 samples). Hyperparameter optimization is performed via GridSearchCV with 5-fold stratified cross-validation. Table 3 shows the search space and optimal values. Implementation uses RandomForestClassifier from scikit-learn 1.8.0.

Table 3. Hyperparameter Search Space and Optimal Configuration

Parameter	Search Range	Optimal	Description
<i>n_estimators</i>	50, 100, 150, 200, 300	200	Number of trees in forest
<i>max_depth</i>	10, 15, 20, None	None	No depth limit — sufficient data
<i>min_samples_split</i>	2, 5	2	Min samples to split a node
<i>min_samples_leaf</i>	1, 2	1	Min samples at leaf node
<i>max_features</i>	sqrt, log2	sqrt	$\sqrt{11} \approx 3$ features per split
<i>criterion</i>	gini, entropy	gini	Split quality function

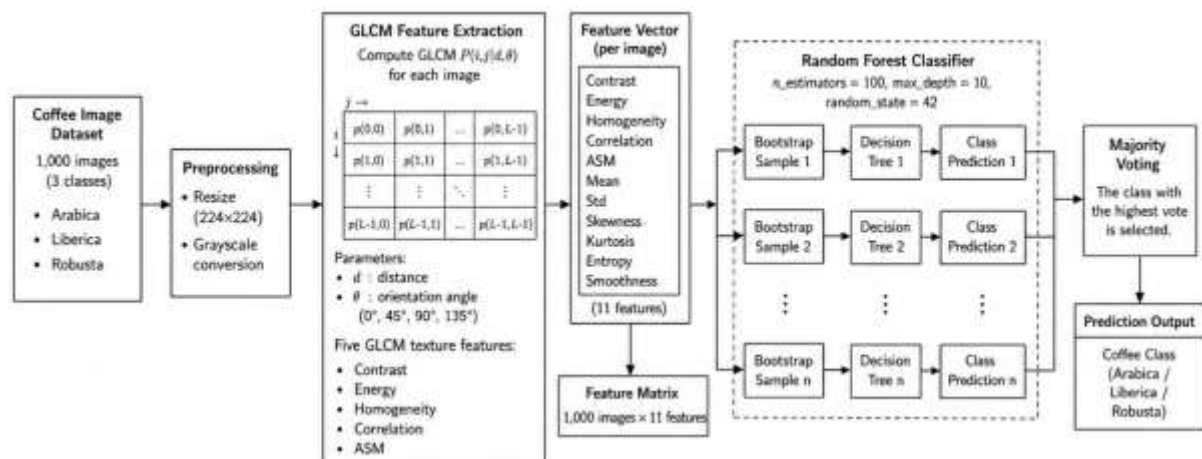


Fig. 2. Framework of the proposed GLCM-Random Forest Classification method

Testing Strategy

Three testing strategies are applied: (1) Hold-out test set 96 Roboflow-default test samples, completely isolated during training; (2) 5-fold stratified cross-validation (StratifiedKFold, *n_splits*=5, *shuffle*=True, *random_state*=42) on 1,817 train+valid samples, preserving class proportions per fold; (3) ROC-AUC One-vs-Rest each class treated as positive against the others, macro-averaged.

Evaluation Metrics

Four metrics are computed per class $k \in \{\text{Arabica}, \text{Liberica}, \text{Robusta}\}$ from confusion matrix values TP^k , FP^k , FN^k :

Precision

Precision for each class is calculated using Equation (7).

$$\text{Precision}^k = \frac{TP^k}{TP^k + FP^k} \quad (7)$$

*name of corresponding author

This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.



Where:

Precision^k : Precision value for class k .

TP^k (True Positive) : The number of samples from class k that are correctly classified as class k .

FP^k (False Positive) : The number of samples from other classes that are incorrectly classified as class k .

Recall

Recall is computed using Equation (8).

$$\text{Recall}^k = \frac{TP^k}{TP^k + FN^k} \quad (8)$$

Where:

Recall^k : Recall value for class k .

TP^k (True Positive) : The number of samples from class k that are correctly classified as class k .

FN^k (False Negative) : The number of samples from class k that are incorrectly classified as another class.

F1-Score

The F1-score is obtained using Equation (9).

$$F1^k = \frac{2 \times \text{Precision}^k \times \text{Recall}^k}{\text{Precision}^k + \text{Recall}^k} \quad (9)$$

Where:

F1^k : F1-score for class k .

Precision^k : Precision value for class k .

Recall^k : Recall value for class k .

Accuracy

Overall accuracy is calculated using Equation (10).

$$\text{Accuracy} = \frac{\sum TP^k}{N} \quad (10)$$

Where:

Accuracy : Overall classification accuracy.

TP^k (True Positive) : Number of correctly classified samples in class k .

$\sum TP^k$: Total number of correctly classified samples across all classes.

N : Total number of samples in the dataset.

Macro Average

Macro-averaged metrics are computed using Equation (11).

$$\text{Metric}_{\text{macro}} = \frac{\sum \text{Metric}^k}{K} \quad (11)$$

Where:

Metric_{macro} : Macro-averaged evaluation metric.

Metric^k : Evaluation metric value (Precision, Recall, or F1-score) for class k .

$\sum \text{Metric}^k$: Sum of metric values across all classes.

K : Total number of classes.

RESULT

Feature Importance and Cross-Validation

Fig. 3 presents the MDI feature importance ranking and 5-fold CV results. GLCM Energy (21.98%) and Homogeneity (14.57%) are the top contributors counterintuitively, not Contrast. This finding indicates that texture uniformity of the bean surface is more discriminative than local intensity variation under real imaging conditions. The top five features (Energy, Homogeneity, Mean-Blue, Mean-Red, Std-Green) account for 75.86% of total importance, confirming the complementarity of texture and color information. CV results show mean 98.07%±0.60% across five folds. Because this 5-fold split was also used by GridSearchCV to select hyperparameters, this figure reflects the same data used for model selection rather than a fully independent generalization estimate; it is reported here as an indicative stability measure, and a nested or repeated cross-validation on a separately held-out configuration is recommended to obtain an unbiased estimate. Because MDI can be biased toward continuous features and features that are correlated with one another, the

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Energy/Homogeneity ranking should be treated as indicative rather than definitive; permutation importance or SHAP values would provide a more robust cross-check and are recommended for future work.

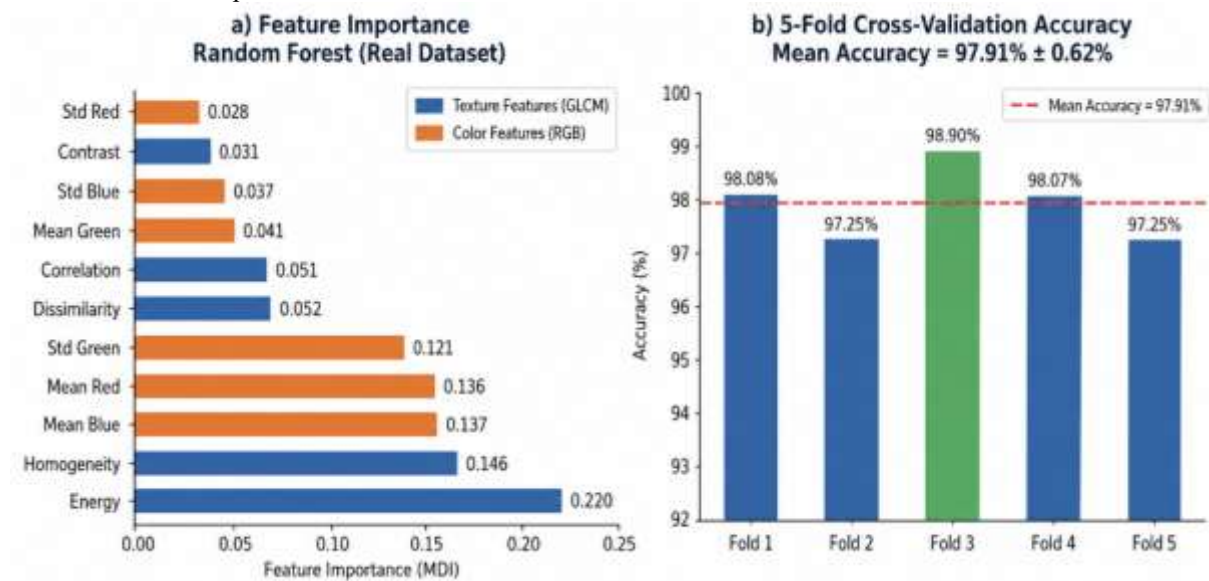


Fig. 3. (a) MDI feature importance and (b) 5-fold stratified cross-validation accuracy

Evaluation on Test Set

Table 4 presents the per-class evaluation results. The model achieves 98.96% overall accuracy with only one misclassification (one Arabica sample predicted as Liberica) out of 96 test samples; because the test set is small, this single error already accounts for approximately 1.04 percentage points of the overall accuracy. The rightmost "Accuracy (1-vs-Rest)" column reports each class treated as the positive class against the other two combined, and is separate from the single overall accuracy value; Robusta achieves a perfect one-vs-rest score (100% all metrics). ROC-AUC of 0.9996 indicates near-perfect class separability.

Table 4. Per-Class Evaluation Results (96 Test Samples)

Coffee Class	Precision (%)	Recall (%)	F1-Score (%)	Support	Accuracy
Arabica	100.00	96.00	97.96	25	98.96%
Liberica	97.14	100.00	98.55	34	98.96%
Robusta	100.00	100.00	100.00	37	100%
Macro Average	99.05	98.67	98.84	96	
Overall Accuracy	98.96%				
AUC	0.9996				

Confusion Matrix

Table 5 and Fig. 4 present the confusion matrix. The single error (Arabica to Liberica) may plausibly relate to overlapping GLCM Energy values between the two classes, since Energy is the top MDI feature overall; however, this explanation is a hypothesis rather than a demonstrated cause, as the Energy value of the misclassified sample was not examined against the class-wise Energy distribution, and no local explanation method (e.g., SHAP) was applied to this specific instance. Confirming this would require sample-level feature inspection or local explainability analysis, which is left for future work.

Table 5. Confusion Matrix of Random Forest (96 Test Samples)

Actual \ Predicted	Arabica	Liberica	Robusta
Arabica	24	1	0
Liberica	0	34	0
Robusta	0	0	37

*name of corresponding author

This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.



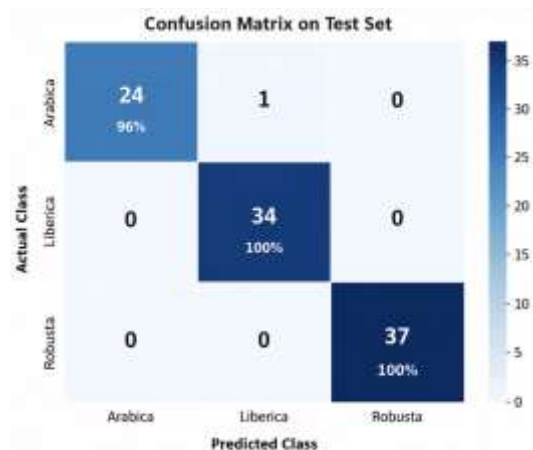


Fig. 4. Confusion matrix heatmap (accuracy: 98.96%, 96 test samples)

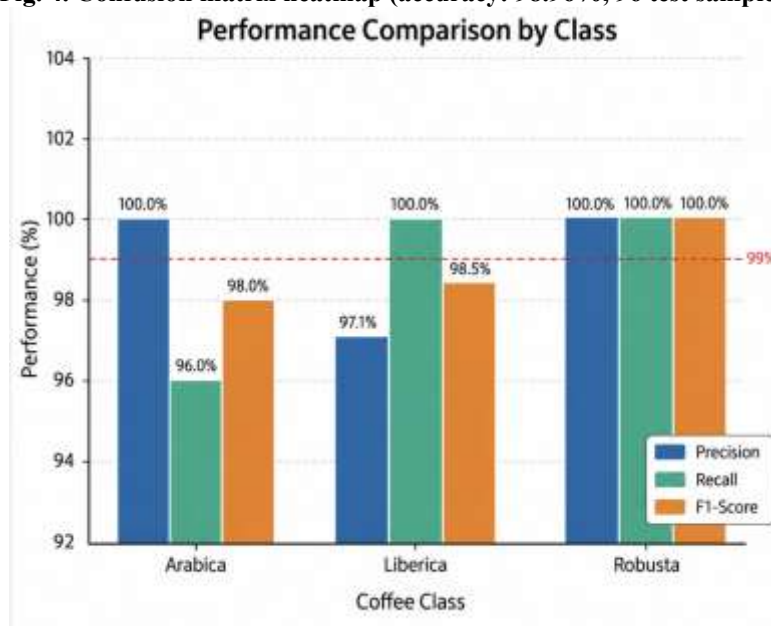


Fig. 5. Per-class performance metrics (Precision, Recall, F1-Score)

DISCUSSION

Manual Metric Verification

Precision, Recall, F1-score, and overall Accuracy are verified manually from the confusion matrix using Equations (7)(10). Note that this verification checks the arithmetic consistency of the reported metrics with the confusion matrix; it does not independently verify the feature extraction pipeline, absence of data leakage, or the GridSearchCV procedure. The per-class "Accuracy" figures reported below (Arabica, Liberica, Robusta) are one-vs-rest accuracies each class treated as the positive class against the other two combined and are distinct from the single overall Accuracy value reported in Table 4; they should not be interpreted as three separate overall-accuracy values.

Actual \ Predicted	Arabica	Liberica	Robusta
Arabica	24	1	0
Liberica	0	34	0
Robusta	0	0	37
Total Sample			96

1. Accuracy Arabica

*name of corresponding author

This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.



Using Equation (10), the accuracy for the Arabica class was calculated as follows:
TP = 24 , FN = 1, FP = 0, TN = 71

$$Accuracy_{Arabica} = \frac{24 + 71}{96} = \frac{95}{96} = 98.96\%$$

2. Accuracy Liberica

Using Equation (10), the accuracy for the Liberica class was calculated as follows:
TP = 34, FN = 0, FP = 1, TN = 61

$$Accuracy_{Liberica} = \frac{34 + 61}{96} = \frac{95}{96} = 98.96\%$$

3. Accuracy Robusta

Using Equation (10), the accuracy for the Robusta class was calculated as follows:
TP = 37, FN = 0, FP = 0, TN = 59

$$Accuracy_{Robusta} = \frac{37 + 59}{96} = \frac{96}{96} = 100\%$$

Comparison with Prior Work

Table 6 lists the accuracy reported by related studies alongside this study's result for reference. RF in this study reaches 98.96%, numerically higher than the accuracy figures reported for KNN (95.00%), ANN (88.60%), and CNN (91.50%–92.10%). However, these studies differ from this study in classification object (coffee-leaf disease, roast degree, or electronic-nose signals rather than bean-type images), dataset, and number of classes, so these figures should be read as contextual reference points rather than a controlled benchmark, and RF cannot be claimed to "outperform" these methods in a strict methodological sense. Unlike CNN approaches, RF requires no GPU, trains in under 1 second, and provides interpretable feature importance via MDI. EffResViT-SE (2025) achieves 99.00% but on a different domain (leaf disease, 58,555 images) and is not directly comparable, as also acknowledged for the other studies above. Notably, this study includes Liberica a class absent in most prior work and uses a publicly available dataset ensuring reproducibility.

Table 6. Comparison with State-of-the-Art in Coffee Image Classification

Study	Method	Dataset	Accuracy	XAI	Classes	Note
Jatmoko & Sinaga (2022)	KNN + GLCM	Local	95.00%	No	2	KNN baseline (bean images)
Triyono et al. (2023)	ANN Backprop	Local	88.60%	No	5	E-nose data
Hakim et al. (2020)	CNN MobileNetV2	Global	91.50%	Partial	3	Deep learning
Metha et al. (2024)	CNN	Global	92.10%	No	3	Roasted beans
EffResViT-SE (2025)	CNN+ViT	Global	99.00%	No	5	Leaf disease
This Study (2026)	Random Forest	Real 1,913	98.96%	Partial (MDI)	3	Best model

Threat to Validity

The test set contains only 96 samples a limitation inherited from Roboflow's default split; with a set this small, a single additional misclassification would shift overall accuracy by roughly one percentage point, so the reported accuracy should be read together with the 5-fold CV results (98.07%±0.60% over 1,817 samples) rather than in isolation, keeping in mind that the same CV split was also used for hyperparameter selection (see Results). The dataset is sourced from Roboflow Universe under CC BY 4.0, ensuring reproducibility but potentially limiting diversity compared to multi-source field collection. The Roboflow dataset documentation does not specify that the images were captured in Indonesia or sourced from Indonesian-grown plantings; consequently, while the three species classified (Arabica, Liberica, Robusta) are the same species commercially grown in Indonesia, the study does not provide direct evidence that the images represent Indonesian coffee production conditions specifically, and the title's scope should be read as classification of these species generally rather than a claim of Indonesia-specific validation. Background, lighting, camera angle, and image-source consistency across classes were not

*name of corresponding author

This is anCreative Commons License This work is licensed under a
Creative Commons Attribution-NonCommercial 4.0 International
License.



separately analyzed; it is possible that RF partly learns incidental imaging or background characteristics rather than purely bean morphology, and no external validation using images from an independent source was performed. Future work should validate on larger, multi-source datasets including images verified to originate from Indonesian coffee-growing regions (e.g., Wamena Papua, Java Preanger) and should assess feature and prediction stability across varying imaging conditions.

CONCLUSION

This study demonstrates that Random Forest with 11-dimensional GLCM and RGB color statistical features achieves 98.96% accuracy, macro F1-score 98.84%, and ROC-AUC 0.9996 for three-class coffee identification on real image data, on a 96-sample hold-out test set. Key findings are: (1) GLCM Energy (21.98%) and Homogeneity (14.57%) are the most discriminative features by MDI, suggesting that texture uniformity is more important than local contrast for distinguishing Arabica, Liberica, and Robusta, though MDI's known bias toward continuous and correlated features means this ranking should be corroborated with permutation importance or SHAP in future work; (2) Robusta achieves perfect one-vs-rest classification (100%), while the single error (Arabica→Liberica) is plausibly related to overlapping Energy values between the two classes, a hypothesis that has not yet been confirmed at the sample level; (3) 5-fold CV ($98.07\% \pm 0.60\%$) provides an indicative stability estimate, though it is not fully independent of the hyperparameter-selection process and should be followed up with nested cross-validation; and (4) manual verification confirms that the reported classification metrics are internally consistent with the confusion matrix, though this check does not independently validate the feature-extraction pipeline or rule out data leakage. RF achieves higher accuracy than the KNN, ANN, and CNN figures reported in several prior studies, though those studies differ in classification task, dataset, and class composition, so this comparison is contextual rather than a controlled benchmark.

ACKNOWLEDGMENT

The authors thank the Doctoral Program in Informatics, Universitas Ahmad Dahlan (UAD), Yogyakarta, for academic support. Dataset: Coffee Bean Type v1 by Jaya Iskandar, Roboflow Universe (<https://universe.roboflow.com/jaya-iskandar/coffee-bean-type>), licensed CC BY 4.0.

REFERENCES

- Ardah, H., Alrahal, M., Abd-Elhafiez, W. M., & Trabay, D. (2025). Robust coffee plant disease classification using deep learning and advanced feature engineering techniques. *PeerJ Computer Science*, e3386. <https://doi.org/10.7717/peerj-cs.3386>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Direktorat Jenderal Perkebunan, Kementerian Pertanian Republik Indonesia. (2023). *Statistik perkebunan jilid I 2022–2024*. Direktorat Jenderal Perkebunan, Kementerian Pertanian Republik Indonesia.
- Fatchurrachman, A., & Udjula, D. (2023). Identifikasi penyakit pada tanaman kopi berdasarkan citra daun menggunakan metode convolution neural network. *Jurnal Algoritme*, 3(2), 151–159. <https://doi.org/10.35957/algoritme.v3i2.3384>
- Gonzalez, R. C., & Woods, R. E. (2018). *Digital image processing* (4th ed.). Pearson Education.
- Hakim, L., Janandi, C., & Cenggoro, T. W. (2020). Automated coffee roast degree classification using MobileNetV2. In *Proceedings of the International Conference on Informatics*.
- Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3(6), 610–621. <https://doi.org/10.1109/TSMC.1973.4309314>
- Iskandar, J. (2024). *Coffee bean type dataset v1* [Data set]. Roboflow Universe. <https://universe.roboflow.com/jaya-iskandar/coffee-bean-type>
- Jatmoko, C., & Sinaga, D. (2022). Metode K-nearest neighbor dan ekstraksi fitur GLCM untuk mengklasifikasikan biji kopi Robusta dan Arabika lokal. *Seminar Nasional Teknologi dan Multidisiplin Ilmu (SEMNASTEKMU)*, 2(2), 353–366. <https://doi.org/10.51903/semnastekmu.v2i1.189>
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2(3), 18–22.
- Metha, H. S., Kusri, K., & Ariatmanto, D. (2024). Classification of types roasted coffee beans using convolutional neural network method. *Sinkron: Jurnal dan Penelitian Teknik Informatika*, 8(2), 846–851. <https://doi.org/10.33395/sinkron.v8i2.13517>
- Mehta, B., Bharany, S., Elkamchouchi, D. H., Khalaf, O. I., Garsamo, E. T., Hariharan, U., & Arora, N. (2025). EffResViT-SE FusionNet: A hybrid deep learning framework for accurate classification of coffee leaf diseases. *Food Science & Nutrition*. <https://doi.org/10.1002/fsn3.71311>
- Motta, I. V. C., Vuillerme, N., Pham, H. H., & de Figueiredo, F. A. P. (2025). Machine learning techniques for coffee classification: A comprehensive review of scientific research. *Artificial Intelligence Review*, 58. <https://doi.org/10.1007/s10462-024-11004-w>

*name of corresponding author

This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.



- Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. MIT Press.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Rahardjo, B. (2015). *Panduan budidaya dan pengolahan kopi Arabika dan Robusta*. Penebar Swadaya.
- Soh, L. K., & Tsatsoulis, C. (1999). Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. *IEEE Transactions on Geoscience and Remote Sensing*, 37(2), 780–795. <https://doi.org/10.1109/36.752194>
- Triyono, S., Laksono, R. A., & Tusi, A. (2023). Identifikasi jenis kopi menggunakan sensor e-nose dengan metode jaringan syaraf tiruan backpropagation. *Jurnal Ilmiah Rekayasa Pertanian dan Biosistem*, 11(1), 1–12.

*name of corresponding author



This is anCreative Commons License This work is licensed under a
Creative Commons Attribution-NonCommercial 4.0 International
License.