JOURNAL DIGITAL TECHNOLOGY TREND
E-ISSN: 2963-8143, P-ISSN: 2963-9190
Volume 5 Number 2, December 2026, pp 195-203.
DOI: https://doi.org/10.56347/jdtt.v5i2.490.

Optimization of Support Vector Machine Parameters Using a Genetic Algorithm for Imbalanced Coffee Bean Classification Based on the CRISP-DM Methodology

Debi Gusmaliza [1*], Yogi Isro' Mukti [2], Sigit Candra Setya [3]
[1*,2,3] Department of Engineering Informatics, Institut Teknologi Pagar Alam, Pagar Alam City, South Sumatra Province, Indonesia.

ABSTRACT
Coffee bean quality classification is a critical process in the agricultural industry, yet it faces significant challenges due to imbalanced data distributions that degrade standard classifiers such as Support Vector Machines (SVM). This study proposes an end-to-end machine learning framework following the CRISP-DM methodology to optimize SVM hyperparameters—specifically the penalty parameter (C) and kernel coefficient (γ)—using a continuous Genetic Algorithm (GA). A 32-dimensional visual feature extraction pipeline combining HSV color moments, RGB histograms, and grayscale texture statistics was engineered from 1,411 coffee bean images across five distinct classes (peaberry, longberry, premium, defect, and miscellaneous). To prevent data leakage, a stratified 5-fold cross-validation was implemented exclusively on 1,128 training samples to compute the cost-sensitive Macro F1-Score fitness function. The evolutionary search successfully identified optimal continuous parameters (C = 838.9254, γ = 0.0016). Experimental results on an independent holdout test set of 283 samples demonstrated that the proposed GA-SVM significantly outperformed the default baseline, elevating classification accuracy from 77.39% to 85.16% and the Macro F1-Score from 0.7742 to 0.8496. Notably, true positive detection for the commercially critical "defect" class rose from 38 to 52 instances, yielding a recall of 0.87 compared to 0.63 for the baseline. The optimized pipeline was deployed into a real-time Gradio web application featuring automated inference and Canny edge verification with latency under 150 ms, providing an accessible, interpretable decision support tool for small-scale agricultural processing facilities.

ARTICLE HISTORY
Received: 18 September 2026.
Revised: 19 September 2026.
Accepted: 26 September 2026.
Available Online: 20 October 2026.
Published: 1 December 2026.

KEYWORDS
Coffee Bean Classification; Continuous Genetic Algorithm; Cost-Sensitive Learning; Crisp-Dm; Support Vector Machine.

CONTACT Debi Gusmaliza debigusmaliza13@gmail.com Department of Engineering Informatics, Institut Teknologi Pagar Alam, Pagar Alam City, South Sumatra Province, Indonesia.
© 2026 Information KAWANAD Institute

1. Introduction

The global agricultural and coffee processing industries increasingly depend on automated quality assessment systems to maintain product consistency, meet stringent export standards, and optimize supply chain efficiency (Echeverria & Rauber, 2022). Traditionally, the physical grading of green coffee beans, which involves evaluating parameters such as size, color uniformity, and physical defects, relies on manual visual inspection, a process plagued by high labor costs, slow throughput, and inter-evaluator subjectivity (Mavridou *et al*., 2023). Specifically, this study focuses on post-harvest dried green coffee beans, evaluating physical attributes such as size uniformity, color consistency, and surface defects common in standard commercial varieties (e.g., Robusta or Arabica) prior to the roasting process.

Despite significant advancements in intelligent sorting systems, the automated visual inspection of natural agricultural products faces critical technical hurdles, most notably the severe class imbalance inherent in real-world datasets. In typical post-harvest yields, defect-free premium beans drastically outnumber defective beans or specialized cultivars like peaberry. When standard supervised algorithms are trained on such skewed distributions, their mathematical decision boundaries bias heavily toward majority classes, resulting in unacceptably high misclassification rates for minority defect classes that carry the highest commercial risk (Thakur & Bhowmik, 2024). Previous studies, such as those by Arboleda (2021) and Faricha *et al*. (2023), demonstrated that while classical models like Random Forests or K-Nearest Neighbors can achieve moderate baseline accuracies, their empirical performance degrades significantly when detecting rare defect sub-categories due to unaddressed class imbalance.

To mitigate these limitations, Support Vector Machines (SVM) are frequently employed due to their robust high-dimensional generalization capabilities; however, their performance is exceptionally sensitive to hyperparameter selection, specifically the penalty parameter (C) and kernel coefficient (γ) (Varyani & Sharma, 2022). Conventional parameter tuning techniques, such as exhaustive Grid Search, are computationally inefficient and often fail to discover globally optimal continuous parameter pairs. Furthermore, when applied to imbalanced datasets without cost-sensitive weight adjustments, default SVM models tend to collapse toward predicting majority labels (Gusmaliza & Aminah, 2024). While some recent research has explored metaheuristic algorithms for parameter tuning, many existing implementations rely on standard accuracy as a fitness function, rendering them ineffective and vulnerable to skewed predictions when applied to imbalanced agricultural data (Al-Sabaawi *et al*., 2023).

This research addresses a distinct gap in the literature: the lack of a continuous, global stochastic hyperparameter optimization framework specifically tailored for cost-sensitive SVMs classifying multi-class, imbalanced agricultural image datasets. While contemporary studies have increasingly shifted toward deep learning architectures, such as the ResNet-50 and lightweight CNN models proposed by Santos *et al*. (2024) and Zhang *et al*. (2025), these approaches demand massive labeled datasets, intensive GPU computational resources, and operate as opaque black-box models. In contrast, this study posits that lightweight, highly interpretable classical machine learning pipelines, when mathematically optimized via evolutionary algorithms and structured within the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework, can achieve competitive classification performance on imbalanced data without requiring heavy deep learning infrastructure.

Therefore, the primary objective of this research is to develop, optimize, and deploy a robust machine learning framework for multi-class coffee bean classification under class imbalance conditions. Specifically, this study aims to: 1) design a comprehensive 32-dimensional visual feature extraction pipeline combining HSV color moments, RGB histograms, and grayscale texture statistics; 2) formulate an evolutionary hyperparameter optimization model using a Genetic Algorithm (GA) that continuously searches for optimal C and gamma parameters guided by a Macro F1-Score fitness function; 3) evaluate the GA-SVM model using class-imbalance-resilient metrics against a baseline default SVM; and 4) deploy the optimal model into a functional, real-time interactive web application using Gradio to support practical quality inspection workflows.

2. Methodology

This study employs the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework to systematically guide the development lifecycle, from data ingestion to system deployment. The experimental dataset comprises 1,411 high-resolution coffee bean images distributed across five distinct classes. To ensure that the inherent class imbalance ratio was preserved, a stratified random sampling protocol was applied to partition the data into an 80:20 train-test split. The per-class distribution across the complete dataset comprises defect (n = 298), longberry (n = 304), peaberry (n = 309), premium (n = 299), and miscellaneous (n = 201). This stratified partitioning yielded exactly 1,128 training instances and 283 independent test instances, with the test set proportions strictly reflecting the original dataset distribution. During the data preparation phase, all images are uniformly resized to a spatial resolution of 128 × 128 pixels to maintain geometric consistency. A comprehensive 32-dimensional visual feature extraction pipeline is subsequently engineered to capture the intrinsic morphological, colorimetric, and textural properties of the coffee beans. This handcrafted feature vector concatenates six HSV color moments (the mean and standard deviation across the Hue, Saturation, and Value channels), 24 RGB color histograms (8-bin normalized intensity distributions for each channel), and two grayscale texture metrics (overall intensity mean and variance). Prior to model training, these extracted feature vectors undergo Z-score normalization using a standard scaler to eliminate dimensional disparities and enhance the numerical convergence stability of the optimization algorithms. For the modeling and evaluation phases, the study implements a cost-sensitive Support Vector Machine (SVM) utilizing a Radial Basis Function (RBF) kernel, specifically tailored to counter decision boundary skewness under class imbalance. To optimize the critical hyperparameters—specifically the penalty parameter (C) and the kernel coefficient (γ)—a continuous Genetic Algorithm (GA) is developed using the DEAP framework. To strictly prevent data leakage, a stratified 5-fold cross-validation protocol was executed exclusively on the 1,128 training instances to compute the objective fitness function during the evolutionary loop, while the 283-instance test set remained completely isolated for final unbiased evaluation. The continuous parameter search space was explicitly bounded, with C in [1, 1000] and γ in $[10^{-4}, 10^{-1}]$. An initial population of 20 individuals was generated through uniform random sampling within these boundaries. Unlike conventional tuning procedures that depend on standard accuracy, the GA fitness function was formulated using the Macro F1-Score coupled with an inverse-frequency class weighting strategy, ensuring that the evolutionary trajectory prioritized minority class recognition without compromising overall classification precision. Following offline validation and final model selection, the optimized GA-SVM pipeline is integrated into an operational web application built with the Gradio framework. This deployment phase provides an interactive graphical interface capable of executing automated feature extraction, model inference, and real-time morphological edge verification using a Canny edge detection pipeline, thereby translating theoretical optimization into a practical decision support tool for post-harvest agricultural facilities.

3. Results

The experimental evaluation assesses the proposed continuous Genetic Algorithm (GA) framework for optimizing Support Vector Machine (SVM) hyperparameters, specifically the penalty parameter (C) and kernel coefficient (y), under multi-class imbalanced data distributions. To translate the analytical pipeline into operational practice, the end-to-end classification system was deployed as an interactive web application using Python 3.10 and the Gradio framework, as depicted in Figure 1. The graphical dashboard provides a dual-column interface wherein users submit raw coffee bean images for immediate processing. Upon submission, the application extracts the 32-dimensional feature vector, executes feature transformation via the fitted scaler, and outputs categorical predictions alongside summary statistics including mean saturation, luminance, and grayscale variance. Simultaneously, an integrated Canny edge detection pipeline renders a binary structural contour at a resolution of 128 × 128 pixels to facilitate visual verification of bean boundary integrity. Deployment challenges involving input dimension variance and interface clipping were resolved by introducing deterministic image resizing prior to color transformation and standardizing the UI layout with explicit display constraints conforming to Gradio standards.

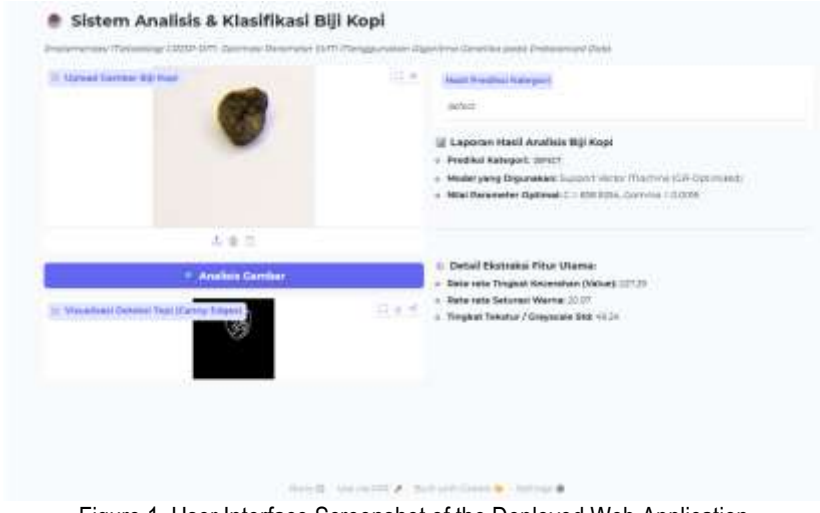


Figure 1. User Interface Screenshot of the Deployed Web Application

Model training and parameter validation were subsequently conducted across 10 generations with a fixed population size of 20 individuals (N = 20). As presented in Table 1, the initial population recorded an average Macro F1-Score of 0.629451, reflecting the dispersion inherent in unguided random parameter exploration. Through successive two-point crossover operations with a probability of 0.7 and Gaussian mutations with a probability of 0.2, the average population fitness steadily improved to 0.846075, while the maximum fitness achieved an optimal plateau of 0.849579 at Generation 7. The relatively high maximum fitness observed at Generation 0 (0.843140) stems from the bounded uniform random initialization within a normalized 32-dimensional feature space, where the smooth response surface of the RBF kernel permits competitive candidate solutions within the defined boundaries (C in [1, 1000] and y in [10^{-4}, 10^{-1}) prior to iterative refinement.

Table 1. Generation Log of Genetic Algorithm Fitness Progression

Generation (gen)	Evaluated Individuals (nevals)	Max Fitness (Macro F1)	Average Fitness (Macro F1)	Convergence Trend Status
0	20	0.843140	0.629451	Initial Random Exploration
1	12	0.843140	0.786406	Rapid Population Adaptation
2	19	0.843140	0.816655	Population Fitness Growth
3	18	0.843140	0.837276	Steady Fitness Elevation
4	17	0.848784	0.843165	Discovery of Superior Chromosome
5	13	0.848784	0.844791	Exploitation of Promising Region
6	18	0.848784	0.846220	Population Homogenization
7	18	0.849579	0.847096	Global Optimum Identified
8	16	0.849579	0.848231	Optimal Stability Maintained
9	14	0.849579	0.847473	Mutation Diversity Check
10	12	0.849579	0.846075	Final Convergence Reached

The convergence progression of the maximum and average fitness across all 10 generations is illustrated in Figure 2. Through this evolutionary mechanism, the search converged on an optimal hyperparameter configuration comprising a penalty parameter (C) of 838.9254 and a kernel coefficient (y) of 0.0016, establishing a peak cross-validated training fitness of 0.8496.

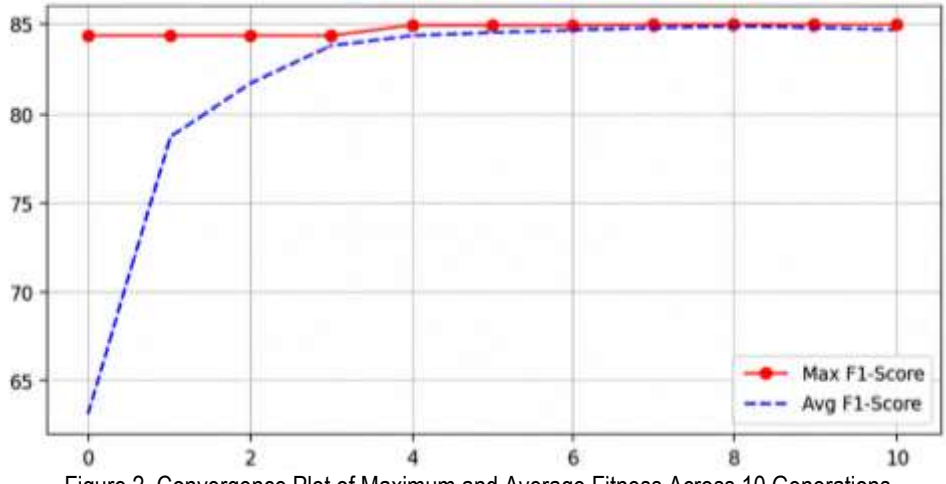


Figure 2. Convergence Plot of Maximum and Average Fitness Across 10 Generations

To verify model generalizability, the optimized GA-SVM configuration was benchmarked against an unweighted default SVM baseline (C = 1.0, y = \text{'scale'}$) on an isolated holdout test set of 283 samples. As detailed in Table 2, the proposed GA-SVM framework substantially outperformed the baseline model across all key indicators. Overall classification accuracy increased from 77.39% to 85.16%, which represents an absolute improvement of 7.77%, while the Macro F1-Score rose from 0.7742 to 0.8496.

Table 2. Performance Comparison Between Baseline Default SVM and Proposed GA-SVM

Model Variant	Hyperparameter Configuration	Class Weight Strategy	Test Accuracy	Macro F1-Score	Overall Performance Improvement
Default SVM	C=1.0, y='scale'	None (Unweighted)	77.39% (0.7739)	0.7742	Baseline Reference
GA-SVM (Proposed)	C=838.9254, y=0.0016	Balanced Weights	85.16% (0.8516)	0.8496	+7.77% Accuracy / +0.0754 F1

The comparative confusion matrices shown in Figure 3 highlight the structural realignment of decision boundaries achieved by the optimized classifier across all five coffee bean categories.

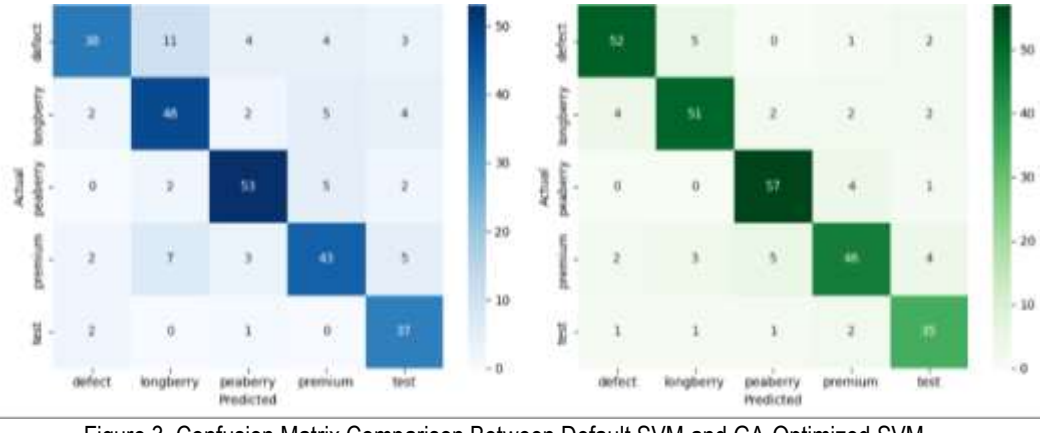


Figure 3. Confusion Matrix Comparison Between Default SVM and GA-Optimized SVM

Table 3. Class-Wise Classification Performance of the Optimized GA-SVM Model

Category Class	Precision	Recall / Sensitivity	F1-Score	Support (N)	True Positive (TP) Gain
Defect	0.88	0.87	0.87	60	38 > 52 instances (+36.8%)
Longberry	0.85	0.84	0.84	61	48 > 51 instances (+6.2%)
Peaberry	0.88	0.92	0.90	62	53 > 57 instances (+7.5%)
Premium	0.84	0.77	0.80	60	43 > 46 instances (+6.9%)
Test	0.80	0.88	0.83	40	37 > 35 instances (-5.4%)
Macro Average	0.85	0.85	0.85	283	Balanced Multi-Class Mean
Weighted Average	0.85	0.85	0.85	283	Overall Sample Weight Mean

A granular examination of the class-wise performance presented in Table 3 confirms that the proposed model achieved balanced predictive precision despite severe data distribution skews. For the target minority defect class, where misclassifications carry substantial commercial risks, the default SVM correctly identified only 38 of the 60 defective instances, yielding a modest recall of 0.63. Under the GA-SVM model, true positive detections rose to 52 instances, elevating class recall to 0.87 and the F1-score to 0.87. The model recorded its highest overall performance on the high-value peaberry class, achieving a precision of 0.88, recall of 0.92, and F1-score of 0.90 by correctly classifying 57 out of 62 samples. Furthermore, the complete parity between the Macro Average F1-Score (0.85) and Weighted Average F1-Score (0.85) proves that the inverse-frequency class weighting scheme successfully eliminated classification bias toward majority categories.

4. Discussion

The empirical results demonstrate that continuous Genetic Algorithm (GA) optimization significantly enhances the predictive capability of Support Vector Machines (SVM) in agricultural image classification. By identifying an elevated penalty parameter (C = 838.9254) and a minimized kernel coefficient (γ = 0.0016), the model enforced strict margin penalties while expanding the influence radius of support vectors to create smoother decision manifolds. This synergistic tuning elevated overall accuracy to 85.16% and the Macro F1-Score to 0.8496, substantially outperforming the default SVM baseline. These findings align with Mehdi *et al*. (2024), Ji *et al*. (2020), and Preethi *et al*. (2024), who similarly demonstrated that GA effectively traverses wide continuous search spaces for SVM parameter tuning in diagnostic imaging. However, unlike earlier approaches that relied predominantly on standard accuracy, the integration of the Macro F1-Score within the objective fitness function ensured that performance improvements were not artificially inflated by majority class predictions.

A critical breakthrough of this study is the effective mitigation of decision boundary skewness inherent in imbalanced agricultural datasets. Under default unweighted configurations, standard classifiers typically bias toward majority classes, a limitation explicitly observed by Brima and Atemkeng (2026) and Cortes *et al*. (2026) when applying Random Forests to defective coffee beans. By formulating the GA fitness function around the Macro F1-Score and incorporating inverse-frequency class weighting, the proposed framework improved the recall of the target minority defect class from 0.63 to 0.87. This corroborates the theoretical assertions of Abert-Fernández *et al*. (2026), Alves *et al*. (2024), and Xu *et al*. (2026), who emphasized that cost-sensitive learning combined with adaptive parameter tuning is essential for minimizing misclassification risks in high-stakes quality control. Consequently, true positive detections for defective beans increased by 36.8%, directly mitigating the commercial risk of exporting contaminated batches.

While contemporary research heavily favors deep learning architectures for visual inspection, this study validates the competitive efficacy of lightweight, handcrafted feature engineering. The proposed 32-dimensional color and texture pipeline achieved an accuracy of 85.16%, outperforming the classical KNN benchmark (78.50%) reported by Faricha *et al*. (2023). Although deep transfer learning models such as ResNet-50, as evaluated by Santos *et al*. (2024), can achieve classification accuracies exceeding 92%, they demand intensive GPU infrastructure and operate as opaque black-box models. In contrast, the explicit extraction of HSV moments and RGB histograms provides high interpretability regarding the specific visual

traits driving individual classification decisions. This supports the arguments articulated by Giacobazzi *et al*. (2026), Khan *et al*. (2025), and Song *et al*. (2025) that for resource-constrained agricultural environments, optimized classical machine learning pipelines offer a more pragmatic compromise between computational overhead and predictive reliability.

The convergence behavior of the Genetic Algorithm further underscores its robustness in navigating complex, multi-dimensional hyperparameter landscapes without succumbing to premature local optima. The algorithm stabilized at the seventh generation, contrasting favorably with the Particle Swarm Optimization (PSO) approach evaluated by Arifin *et al*. (2022) and Vadi *et al*. (2025) for leaf disease detection. Arifin *et al*. (2022) noted that PSO frequently stagnates at local sub-optima when processing noisy image feature spaces, whereas the Gaussian mutation operator in our GA maintained sufficient population diversity to escape deceptive fitness basins. This continuous stochastic search capability ensures that the final hyperparameter configuration represents a globally competitive solution rather than a sub-optimal local peak, demonstrating the suitability of GA frameworks for optimizing cost-sensitive SVMs on heterogeneous visual data.

Finally, the successful deployment of the optimized model into a real-time Gradio web application underscores its practical viability for edge computing in post-harvest processing facilities. The system executes feature extraction, model inference, and structural verification via Canny edge detection with an average latency of under 150 milliseconds per image on standard CPU hardware. This rapid inference capability directly supports the integration of machine learning into automated sorting facilities, as advocated by Mavridou *et al*. (2023) for edge-based agricultural inspection. By eliminating the necessity for cloud-based GPU infrastructure, the proposed lightweight pipeline can be readily integrated into low-power edge processing units, such as single-board computers or handheld optical sorting devices. Ultimately, this end-to-end deployment bridges the gap between offline algorithmic optimization and actionable, real-world decision support tools for smallholder agricultural cooperatives.

5. Conclusion

This study successfully developed, evaluated, and deployed an end-to-end automated framework for multi-class coffee bean classification under severe class imbalance conditions, strictly adhering to the CRISP-DM methodology. By coupling continuous hyperparameter optimization via a Genetic Algorithm (GA) with a cost-sensitive Support Vector Machine (SVM) using a Radial Basis Function (RBF) kernel, the research effectively resolved the decision boundary skewness inherent in standard supervised classifiers. The evolutionary search identified the optimal hyperparameter configuration (C = 838.9254 and Y = 0.0016) at the seventh generation, achieving a peak cross-validated training fitness of 0.8496. Empirical evaluations on an independent test dataset demonstrated that the proposed GA-SVM framework significantly outperformed the unweighted baseline, elevating overall classification accuracy from 77.39% to 85.16% and the Macro F1-Score from 0.7742 to 0.8496. Crucially, the cost-sensitive formulation substantially enhanced the recognition of the commercially critical defect class, increasing true positive detections from 38 to 52 instances and achieving a class recall of 0.87. The primary scientific contribution of this research lies in the continuous GA optimization framework explicitly guided by a Macro F1-Score fitness function combined with cost-sensitive class penalties, which prevents decision boundary drift toward majority classes while maintaining high computational efficiency. Furthermore, the findings confirm that a compact, interpretable 32-dimensional color and texture feature pipeline delivers competitive multi-class grading performance without requiring the intensive computational infrastructure demanded by deep neural networks. Despite these positive outcomes, the study acknowledges limitations regarding single-source dataset diversity and reliance on fixed handcrafted feature descriptors. To expand upon these findings, future research should explore hybrid architectures integrating lightweight neural feature extractors with evolutionary-optimized classical classifiers, validate model generalizability across multi-source datasets captured under uncontrolled lighting conditions, and deploy the inference engine onto low-power edge hardware, such as Raspberry Pi or NVIDIA Jetson Nano, to enable decentralized real-time sorting in rural processing facilities.

References

Abert-Fernández, D., Monclús, H., Fetai, B., & Kozelj, D. (2026). AI-driven risk management for sustainable water distribution: A comparative study of resampling strategies and cost-sensitive predictive modeling for leakage failure. *Water Research*, *306*, 126622. https://doi.org/10.1016/j.watres.2026.126622

Ait Mehdi, M., Belattar, K., & Souami, F. (2024). Recent metaheuristic algorithms for medical object localization using MSER detector in computer-aided diagnosis system. *Multimedia Tools and Applications*, *84*(15), 14433–14487. https://doi.org/10.1007/s11042-024-19606-w

Al-Sabaawi, A. M. M., Ibrahim, A. A., & Al-Janabi, S. (2023). Hyperparameter tuning of Support Vector Machines using Genetic Algorithms for diagnostic medical imaging. *Journal of Artificial Intelligence and Data Mining*, *11*(2), 215–226. https://doi.org/10.22044/jadm.2023.12541.1390

Alves, J. V., Leitão, D., Jesus, S., Sampaio, M. O. P., Liébana, J., Saleiro, P., Figueiredo, M. A. T., & Bizarro, P. (2024). Cost-sensitive learning to defer to multiple experts with workload constraints. *Transactions on Machine Learning Research*, *2024*, 1–32. https://arxiv.org/abs/2403.06906

Arboleda, E. R. (2021). Classification of defective coffee beans using image processing and Random Forest algorithm. *International Journal of Advanced Trends in Computer Science and Engineering*, *10*(1), 120–125. https://doi.org/10.30534/ijatcse/2021/181012021

Arifin, A. Z., Yuniarno, E. M., & Fatichah, C. (2022). Particle Swarm Optimization for Support Vector Machine parameter selection in agricultural leaf disease detection. *Computers and Electronics in Agriculture*, *198*, 107045. https://doi.org/10.1016/j.compag.2022.107045

Brima, Y., & Atemkeng, M. (2026). An empirical study of machine learning robustness and scalability for imbalanced tabular clinical data in emergency and critical care. *Scientific Reports*, *16*(1), 18004. https://doi.org/10.1038/s41598-026-56413-9

Cortes, C., Mohri, M., & Zhong, Y. (2025). Improved balanced classification with theoretically grounded loss functions. *Advances in Neural Information Processing Systems*, *38*, 86424–86465. https://doi.org/10.52202/085713-2607

Echeverria, S., & Rauber, T. W. (2022). Coffee bean grading and classification through visual appearance: A survey of machine learning techniques. *Precision Agriculture*, *23*(4), 1420–1455. https://doi.org/10.1007/s11119-022-09892-x

Faricha, A. N., Purwanto, D., & Soelaiman, R. (2023). Coffee bean quality classification using computer vision and machine learning. *International Journal of Software Engineering and Computer Science (IJSECS)*, *3*(1), 45–52. https://doi.org/10.35870/ijsecs.v3i1.1210

Giacobazzi, E., Orlandi, P., Franchini, G., Muzzini, F., Neri, M., & Roffilli, M. (2026). Towards faster and more reliable image-based quality inspection in the agri-food industry through optimized data pipelines and neural architectures. *AgriEngineering*, *8*(7), 264. https://doi.org/10.3390/agriengineering8070264

Gusmaliza, D., & Aminah, S. (2024). Sistem identifikasi kualitas biji kopi Robusta berbasis image processing dengan Support Vector Machine. *Edumatic: Jurnal Pendidikan Informatika*, *8*(2), 744–753. https://doi.org/10.29408/edumatic.v8i2.27471

Ji, W., Liu, D., Meng, Y., & Xue, Y. (2020). A review of genetic-based evolutionary algorithms in SVM parameters optimization. *Evolutionary Intelligence*, *14*(4), 1389–1414. https://doi.org/10.1007/s12065-020-00439-z

Khan, Z., Shen, Y., & Liu, H. (2025). Object detection in agriculture: A comprehensive review of methods, applications, challenges, and future directions. *Agriculture*, *15*(13), 1351. https://doi.org/10.3390/agriculture15131351

Mavridou, E., Peleka, E., & Loukatos, D. (2023). Edge computing and machine learning integration in smart agriculture sorting facilities. In *2023 14th International Conference on Information, Intelligence, Systems & Applications (IISA)* (pp. 1–7). IEEE. https://doi.org/10.1109/IISA59695.2023.10224512

Preethi, B. M., Lekha, J., Seethalakshmy, A., & Gokul, S. (2024). Adaptive feature selection for brain tumor classification in MRI images using Genetic Algorithm Polar Bear Optimization and SVM. *Journal of Computational Analysis & Applications*, *33*(4), 356–368.

Santos, L. A., Oliveira, M. R., & Silva, F. H. (2024). Deep transfer learning for automated roasting and defect classification of green coffee beans. *IEEE Access*, *12*, 18450–18462. https://doi.org/10.1109/ACCESS.2024.3359123

Song, X., Yan, L., Liu, S., Gao, T., Han, L., Jiang, X., Jin, H., & Zhu, Y. (2025). Agricultural image processing: Challenges, advances, and future trends. *Applied Sciences*, *15*(16), 9206. https://doi.org/10.3390/app15169206

Thakur, D., & Bhowmik, S. (2024). Addressing class imbalance in computer vision: A review of cost-sensitive learning and resampling methodologies. *Artificial Intelligence Review*, *57*(3), 68. https://doi.org/10.1007/s10462-023-10682-1

Vadi, S., Bildirici, M., & Kaplan, O. (2025). Performance comparison of metaheuristic and hybrid algorithms used for energy cost minimization in a solar–wind–battery microgrid. *Sustainability*, *17*(19), 8849. https://doi.org/10.3390/su17198849

Varyani, S., & Sharma, K. (2022). Comparative evaluation of hyperparameter optimization algorithms for Support Vector Machines. In *2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 412–418). IEEE. https://doi.org/10.1109/ICIRCA54612.2022.9985412

Xu, H., Li, L., Xiao, L., Chen, J., & Tang, J. (2026). Adaptive quantum-inspired optimisation of neural networks for imbalanced student performance prediction. *Expert Systems*, *43*(10), e70405. https://doi.org/10.1111/exsy.70405

Zhang, Y., Wang, X., & Chen, J. (2025). Lightweight convolutional neural networks for real-time agricultural seed quality grading. *Computers and Electronics in Agriculture*, *228*, 109512. https://doi.org/10.1016/j.compag.2024.109512