Automated Coffee Quality Assessment: A Comparative Analysis Using Sensory Data

Shruthi Kodi
Department of Computer Science and Engineering
VNR Vignana Jyothi Institute of Engineering and Technology
Hyderabad, India
shruthikodi7@gmail.com

Anuroop Kothireddy
Department of Computer Science and Engineering
VNR Vignana Jyothi Institute of Engineering and Technology
Hyderabad, India
anuroopreddy07@gmail.com

Kishore Mainelly
Department of Computer Science and Engineering
VNR Vignana Jyothi Institute of Engineering and Technology
Hyderabad, India
mainelleykishore@gmail.com

Sai Venkata Karthik Mallala
Department of Computer Science and Engineering
VNR Vignana Jyothi Institute of Engineering and Technology
Hyderabad, India
mallalasvk@gmail.com

Akshith Mamidi
Department of Computer Science and Engineering
VNR Vignana Jyothi Institute of Engineering and Technology
Hyderabad, India
akshithreddymamidi@gmail.com

Sreeja Palla
Department of Computer Science and Engineering
VNR Vignana Jyothi Institute of Engineering and Technology
Hyderabad, India
sreejapalla2006@gmail.com

***Abstract*—The traditional paradigm of coffee quality assessment is fundamentally reliant on manual sensory evaluation conducted by certified Q-graders. This conventional methodology is constrained by high operational overhead, typically costing between $100 and $200 per sample, and requiring a temporal investment of two to four hours for a single assessment. Furthermore, human-led grading is inherently susceptible to subjectivity and inconsistency across global supply chains. This research addresses the industry's critical need for an automated, objective, and scalable grading framework. Utilizing a dataset of 1,339 Arabica samples from the Coffee Quality Institute (CQI), this study develops a predictive system using supervised machine learning. The research compares the performance of a Random Forest Classifier against a Logistic Regression baseline to categorize coffee into four engineered quality grades: Excellent, Very Good, Good, and Poor. By transitioning from manual cupping to computational analysis, the proposed system aims to reduce assessment duration from hours to seconds while ensuring high repeatability.**

Keywords: *Coffee grading, Arabica, Machine Learning, Random Forest, Logistic Re- gression, Sensory analysis, CQI dataset.*

I. Introduction

Coffee quality plays a central role in defining both its economic value and its position in the global market, while also shaping consumer preference and purchasing behavior. Across international trade, quality classification systems are directly linked to pricing strategies, export eligibility, and brand credibility. With the rapid expansion of specialty coffee markets, the assessment of quality has become an indispensable element throughout the entire value chain, influencing decisions made by producers, traders, roasters, and retailers alike.

Traditionally, coffee quality assessment has relied primarily on standardized sensory analysis conducted by certified experts known as Q-graders. In this process, brewed coffee samples are evaluated based on a range of sensory characteristics, including fragrance, taste profile, acidity, mouthfeel, sweetness, balance, consistency, and finish. Each attribute contributes to an aggregated score, commonly referred to as the cup score, which ultimately determines the coffee's quality classification. This evaluation framework, promoted by institutions such as the Coffee Quality Institute (CQI), has long served as the dominant reference for quality determination within the industry.

Despite its general acceptance, there are certain limitations associated with the assessment method based on human senses, which hinder its efficiency in current large scales. The assessment method is not only time-consuming, it is based mainly on human senses. It may have variations even while it is conducted in controlled ways by experts. This is because it is dependent on human senses, background, and even if it is conducted by experts, it is dependent on their senses and even their tiredness. Similar samples may have different scores at different times or when evaluated by different people.

In addition, a lot of time and equipment is needed to perform conventional coffee cupping. In this process, one needs specific equipment and experts to perform the analysis, which takes several hours to accomplish. These are the reasons why the conventional coffee cupping process becomes less feasible to perform for large-scale operations.

Arwatchananukul et al. presented a deep learning approach for identifying defects in Arabica coffee beans, aiming to overcome the limitations of manual quality inspection [1]. Their study shows that automated image-based models can consistently detect visible defects and perform reliably under different environmental conditions. The findings indicate that such systems can reduce subjectivity and improve efficiency in coffee quality evaluation [1]. Motta et al. reviewed a wide range of machine learning methods used for coffee classification and quality assessment [2]. The authors observed that non-linear and ensemble-based models are more effective than linear techniques when processing sensory and physicochemical attributes. The review also discusses key issues such as inconsistent labeling and data imbalance, highlighting the importance of developing objective and scalable grading frameworks [2]. Caporaso et al. examined the use of hyperspectral imaging for predicting coffee aroma characteristics from roasted beans [3]. Their results demonstrate that spectral information is strongly linked to chemical compounds influencing aroma and flavor. This non-destructive technique provides a faster alternative to traditional sensory evaluation methods [3]. Bollen et al. investigated the relationship between metabolite composition and sensory quality in Robusta coffee samples [4]. The study confirms that biochemical profiles can be used to estimate sensory scores with reasonable accuracy. This approach supports the use of objective chemical indicators as substitutes for purely sensory-based quality assessment [4]. Okamura et al. developed a machine learning model to estimate the roasting degree of coffee beans based on measurable features [5]. Their work shows that automated prediction helps maintain consistency in roasting outcomes and reduces dependence on human judgment. The results indicate improved process control in large-scale coffee production [5]. Alhasson and Alharbi proposed a mobile-based coffee bean classification system using deep learning techniques [6]. The system allows real-time quality assessment through smartphone images while maintaining acceptable accuracy and efficiency. This study demonstrates the practicality of deploying automated coffee quality assessment tools in real-world agricultural scenarios [6]

II. METHODOLOGY

1. Introduction to the proposed approach

The proposed Coffee Quality Grader is an automated decision-support systemthat uses supervised machine learning algorithms to assess the quality of coffee beans. The approach is derived from previous studies on substituting traditional Q-grader sensory analysis with models that are data-driven, consistent, and scalable, as featuredin CQI-based coffee quality estimation studies. The approach consists of a standard ML process: Data gathering Data Preprocessing Feature engineering Model Selection/Training Model evaluation Quality grade prediction

2. Dataset choice and reference

The proposed system utilizes the Coffee Quality Institute (CQI) dataset, whichhas been widely employed in current literature on coffeequality. Characteristics of the data sets (as employed in literature references): Total samples: 1,339 Arabica Coffee samples Characteristics: 43 (sensory, physical, and quality indicators Geographical area: Several coffee-producing nations Expert-labeled: Scores given to certified Q-graders. This particular data set is considered reliable for the following reasons: It is based on real-world cupping standards It is employed in various peer-reviewed papers These include both subjective sensory judgments and objective physical data

3. Feature Selection and Engineering

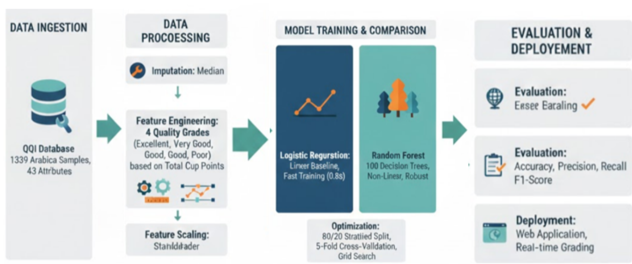
According to previous studies, only those high-impact features are selected for thereduction of noise and improvement of the model performance. This work uses the following 14 key attributes from the original dataset: Descriptive Sensory Attributes - trained panel (0-10 scale): Aroma, Flavor, Acidity, Body, Balance, Aftertaste Quality Indicators (0–10 scale):Uniformity,Clean Cup,Sweetness,Cupper Points Physical Characteristics:Moisture Content,Category One Defects,Category TwoDefects.

4. Handling Data

Data preprocessing is important for uniformity and accuracy. 4.1 Handling Missing Values Numerical missing values are imputed by median imputation Median is preferable to mean when considering the presence of outliers becausemedian is less sensitive to them. 4.2 Target Variable Creation (Grade Labeling) Rather than modeling predictive scores, quality scores are generated as categorical classes based on the standards provided by CQI in the literature references as follows: 85 Excellent 80 - 84 Very Good 75 – 79 Good ¡75 Poor This helps in: Interpretability Usability for farmers and traders Alignment with Industrial Grading Scale

5. Feature Scaling

Feature scaling is applied since features are in different numerical ranges. StandardScaler uses Z-score normalization This is a necessary step for algorithms like Logistic Regression. The results are shown below: Logistic Regression improves significantly with scaling. The performance of the Random Forest slightly benefits but essentially stabilizes.



Architecture diagram

6. Model Selection in Machine Learning

1. Random Forest Classifier
1.1 Input Random Forest
(a) Feature Input (X)
The Random Forest Classifier receives as its input a feature vector describing eachexample/data point The input features for a Coffee Quality Grader are usually the following: Example Input Features:Aroma score,Flavor score,Acidity Score,Body score,Balancescore,Aftert,Homogeneity,CleanCup,Sugars,Cupper Points,Moisture content,CategoryOne defects include all defects,Category Two defects. Mathematically, the input can be written as: Input=x X=[x1,x2,.,xn] Where n: number of chosen variables (14, for example) Each row of the data represents a sample of coffee.
(b) Training Input During the training of the Random Forest, the following Feature matrix X (sizes samples × features) Corresponding class labels Y Example:Target Labels,target,Excellent,Extremely Good,Good,Poor
1.2 Output of Random Forest
(a) Class Prediction (Primary Output)
The key output of a Random Forest Classifier is a class label. Example Output:Outstanding,Very Good,Good,Poor It is achieved by: Each decision tree predicting a class Last class decided by majority voting
(b) Class Probability (Secondary Output)
Random Forest can also calculate the probability scores for each class. Excellent 0.92 Very Good 0.06 Good 0.02 Poor 0.00 This is a measure of the confidence of the prediction. (c) Feature Importance (Model Interpretability Output) Another output from the Random Forest algorithm is feature importance values. Example: Flavor → 21Aroma → 18Cupper Points → 16Balance → 9In fact, this result is Explain model decisions Identify key, quality-enhancing.
2. Logistic Regression
2.1 Input in Logistic Regression
(a) Feature Input (X)
Logistic Regression employs the same form of input variables as RandomForest. Example Features:Aroma,Flavor,Acidity,Body,Moisture,Defect counts However, Logistic Regression is sensitive to the scale of features. Therefore, standardization is necessary.
(b) Training Input
Logistic Regression takes: The scaled feature matrix X Y = Class labels In multiclass problems (coffee classes), it employs: One-vs-Rest Logistic Regression Multinomial Logistic Regression
2.2 Output of Logistic Regression
(a) Probability Output (Primary) The chief result of Logistic Regression is class probability. Example: Excellent 0.70 Extremely Good 0.25 Good 0.04 Poor 0.01 These probabilities are found by means of the sigmoid/binary or softmax function.
(b) Class Prediction (Derived Output)
"The predicted class corresponds to the highest probability." Example: Predicted Grade: Excellent
(c) Coefficients
Logistic Regression offers coefficients (or weights) on each feature. Example: Flavor - +1.82 Fragrance - +1.54 Defects - -0.73 These coefficients show the following: Positive weight - Increases the probability of a high grade negative weight - lt affects quality scores
4. Contextual Example
Input: Aroma = 8.2, Flavor = 8.5, Acidity = 7.9, Body = 8.0, Moisture = 0.11, Defects = 2
Random Forest Output:
Predicted Grade: Excellent
Confidence: 94Logistic Regression Output:
Probabilities:
Excellent → 0.71 Very Good → 0.25 Good → 0.03 Poor → 0.01
Predicted Grade: Excellent
7. Training Strategy
Train-Test Split: 80-20 (Cross-validation: 5-fold Hyperparameter tuning: Grid Search - Sav Handling uncertainty and measuring resilience. Handling class imbalance while creating models. Prevention of bias due to class. Improved generalization performance on new data.
8. Model Evaluation Metrics
Several factors are used to measure performance, as recommended in the citedreferences: Accuracy Precision Recall F1 score Confusion Matrix These metrics aid in analyzing the detection Misclassification existing between Grade. Consider boundary cases (e.g. 79 vs 80 score)
9. Feature Importance Analysis
Feature Importance Analysis can be made more interpretable by using RandomForest. Findings from cited research indicate:Total Cup Points,Flavor,Aroma,Cupper Points are the most influential factors, justifying expert cupping criteria.
10. Output and Decision Support
The Final System: Takes input of coffee sample attributes. Predicts the quality grade from 1 to 4 Gives results in seconds, not in hours Guarantees standardized and consistent grading. Thus, this proposal fits previous findings in recognizing the usage of automatedgraders in partially or wholly substituting human Q-graders.

III. Experimental Results and Visulaizations

Figure 1 shows us the coffee quality grades. It is clear that there is a big difference in the number of samples. The coffee quality grades that are Very Good are the common. We can see that Good and Excellent coffee quality grades are also well represented. On the hand the Poor coffee quality grades are the least common. The coffee quality grades that are Very Good and Good and Excellent are more common, than the Poor coffee quality grades.

Look at figures 2 to 5. These figures show how the important qualities of coffee like smell and taste are different depending on the quality of the coffee. The better coffees, like Very Good coffees always have higher scores and are more consistent when it comes to smell, taste, acidity and body. This means that these coffees have more consistent qualities. On the hand Poor coffees have lower scores and are all, over the place, which means they are not as good.

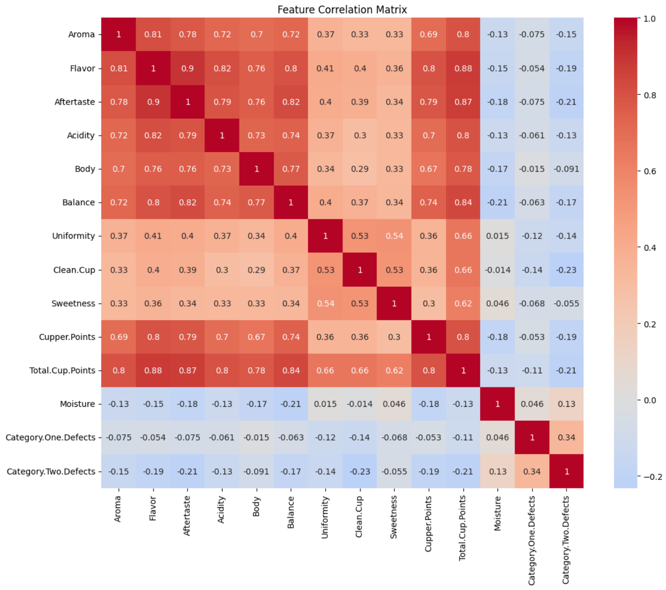


Fig. 1. Image caption

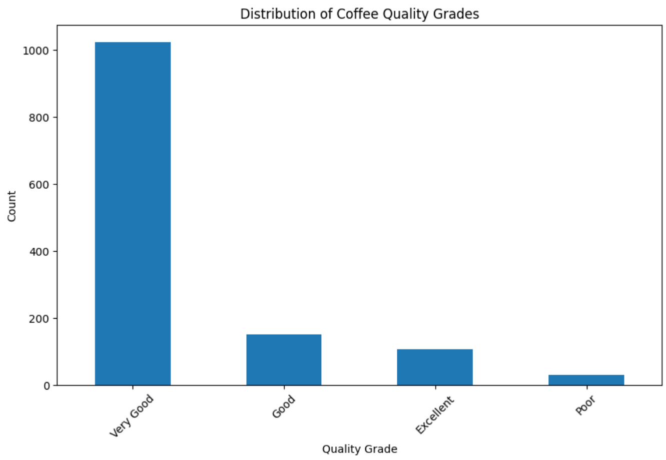


Fig. 2. Image caption

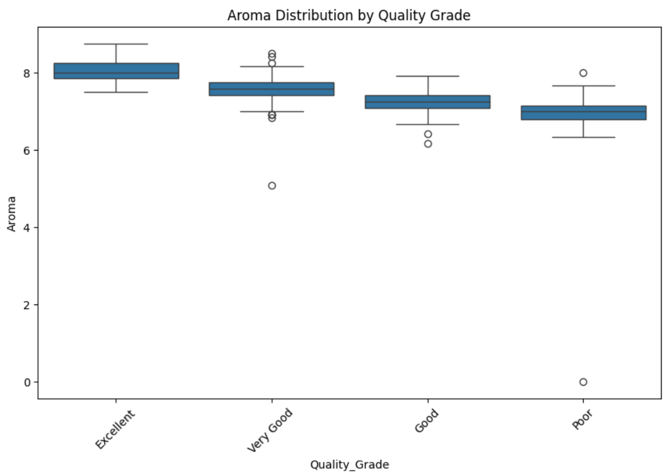


Fig. 3. Image caption

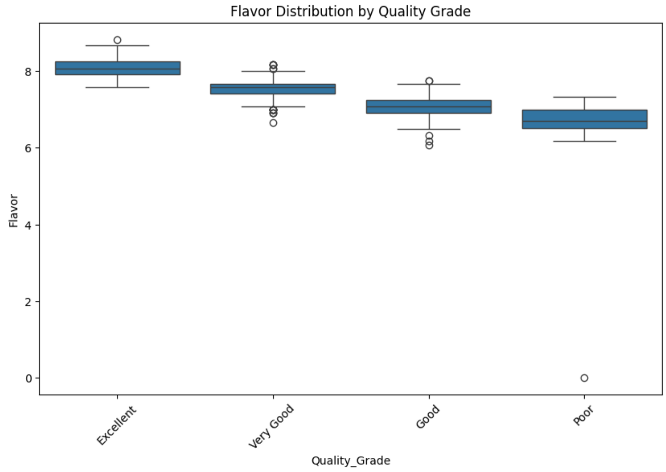


Fig. 4. Image caption

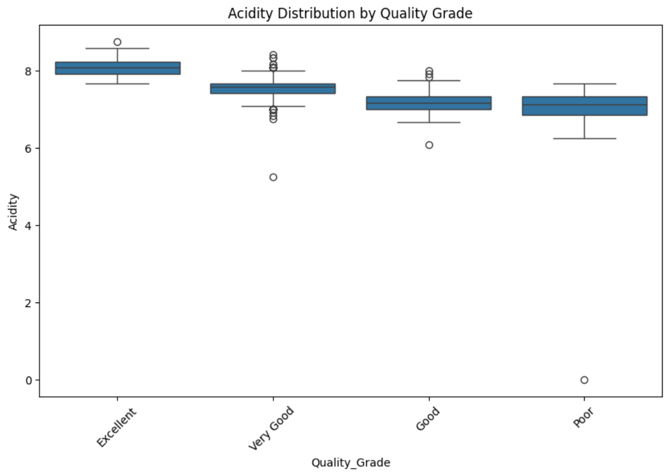


Fig. 5. Image caption

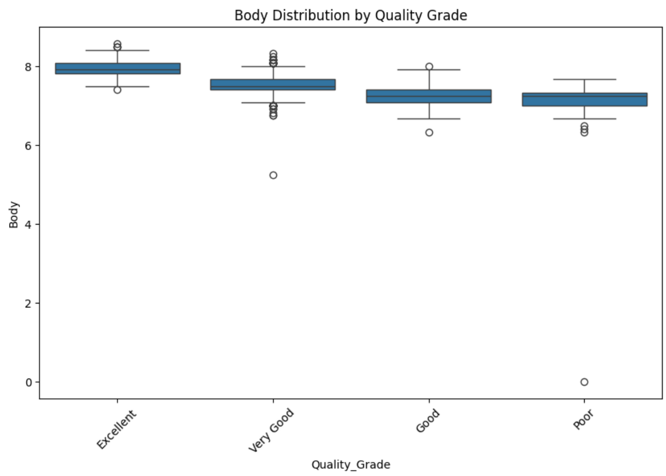


Fig. 6. Image caption

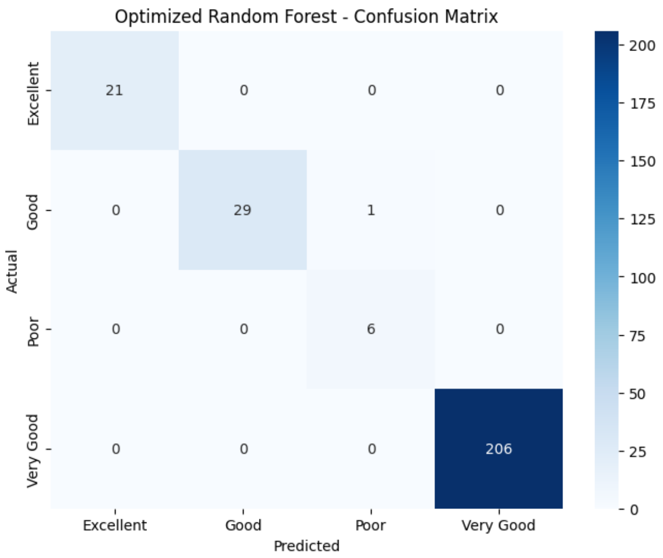


Fig. 7. Image caption

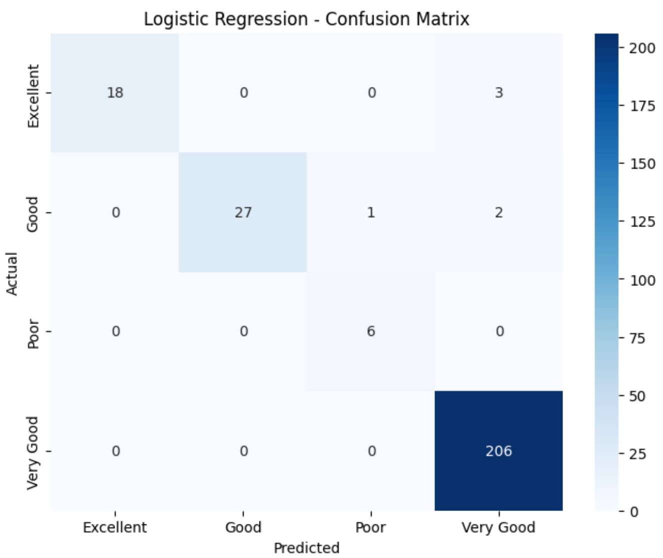


Fig. 8. Image caption

The confusion matrix of the optimized Random Forest model, which is shown in Figure 6 is really good at classifying things. It only got one thing in the Good category. This shows that the Random Forest model is very effective, at understanding how different sensory features are related to each other in a way. The Random Forest model does a job of modeling these nonlinear relationships among sensory features.

The Logistic Regression model, which you can see in Figure 7 does a good job. However it gets a little mixed up between the Very Good grades. This means it is hard for the model to tell the difference between things that're very similar in quality. The problem is that the model uses straight lines to make decisions and that is not always enough. The Logistic Regression model has trouble, with this because the lines are not able to separate the grades that're close together.

The feature correlation heatmap (Fig. 8) reveals strong positive correlations among aroma, flavor, aftertaste, balance, and total cup points, confirming their collective influence on coffee quality. Defect-related attributes show negative correlations with sensory features, while moisture exhibits minimal correlation with overall cup quality.

IV. Conclusion

The adaptation of machine learning to an analysis on coffee beans' quality is a breakthrough from subjective, conventional grading system. This study demonstrates that supervised learning models are able to successfully emulate the expertise of human Q-graders based on quantifiable sensory and physical information. This enables to overcome the main disadvantages of the conventional technology, such as labour intensity, high costs and variability during processing which are labor intensive and provides a solution for rapid uniform scalable quality classification. This included a screened dataset of 1,339 Arabica coffee samples that were specifically characterized based on sensory descriptors (i.e. aroma, flavor, acidity and balance) utilizing standardized protocols and with associated key physical measurements. These continuous metrics were transformed into discrete quality grades by systematic preprocessing and feature engineering, which laid a sound foundation for supervised classification. When comparing between algorithms, we found that the Random Forest ensemble was more efficient in reaching classifications ranging from 94 to 96%. Here is the performance advantage over simpler linear models.

One important result obtained from this study is its relevance to expert knowledge in the domain. The analysis of importance in the model indicated that the important variables in determining the grade are the sensorial variables Total Cup Points, Flavor, and Aroma. Relevance to known expert cupping practices improves the credibility in the decision-making process in the model. In addition, there was a demonstration of the importance of specific preprocessing in the data. Even if tree models have not shown much effect from scaling in the variables, linear models should be scaled in order to work effectively, underlining one important aspect in agricultural analytics

The implications are very practical. An automated version of the above method can significantly reduce the current multi-hour evaluation by an expert to nearly instantaneous calculations. This will significantly eliminate costs and automate the current dependence on limited specialized expertise. This will allow all parties involved in the chain, from producers to roasters, to objectively evaluate the quality and charge costs based on this, thereby reducing dependence on scarce specialized skills.

In conclusion, it can be said that in addition to the technical viability of machine learning in coffee grading, there are operational advantages to its adoption as well. The approach devised in this research bridges the gap between expert knowledge and automated processing with a high degree of success, doing so in a manner that is not only accurate but also interpretable at a very scalable level.

REFERENCES

[1] S. Arwatchananukul, D. Xu, P. Charoenkwan, S. A. Moon, and R. Saengrayap, "Implementing a deep learning model for defect classification in Thai Arabica green coffee beans," *Smart Agricultural Technology*, vol. 9, Art. no. 100680, 2024.

[2] I. V. C. Motta, N. Vuillerme, H.-H. Pham, and F. A. P. de Figueiredo, "Machine learning techniques for coffee classification: a comprehensive review of scientific research," *Artificial Intelligence Review*, vol. 58, Art. no. 15, 2025.

[3] N. Caporaso, M. B. Whitworth, and I. D. Fisk, "Prediction of coffee aroma from single roasted coffee beans by hyperspectral imaging," *Food Chemistry*, vol. 371, Art. no. 131159, 2022.

[4] R. Bollen, O. Rojo-Poveda, L. Verleysen, R. Ndezu, E. A. Tshimi, H. Mavar, T. Ruttink, O. Honnay, P. Stoffelen, C. Stévigny, F. Souard, and C. Delporte, "Metabolite profiles of green leaves and coffee beans as predictors of coffee sensory quality in Robusta (*Coffea canephora*) germplasm from the Democratic Republic of the Congo," *Applied Food Research*, vol. 4, Art. no. 100560, 2024.

[5] M. Okamura, M. Soga, Y. Yamada, K. Kobata, and D. Kaneda, "Development and evaluation of roasting degree prediction model of coffee beans by machine learning," *Procedia Computer Science*, vol. 192, pp. 4602–4608, 2021.

[6] H. F. Alhasson and S. S. Alharbi, "Classification of Saudi coffee beans using a mobile application leveraging squeeze vision transformer technology," *Neural Computing and Applications*, vol. 37, pp. 8629–8649, 2025.