

# Coffee Bean Recognition Using Shape Features Using Decision Trees And Ensemble Classifiers

Diego L. Bersabal, Jumelyn L. Usa, Edwin R. Arboleda, Elbert M. Galas

**Abstract:** Artificial intelligence as a medium for coffee bean extraction and differentiation can be used by farmers to ease their works. In this given study, Decision Trees and Ensemble Classifiers were used to identify the given data by using the laid patterns that have been developed. The test was implemented on 120 coffee beans belonging to 3 coffee species, Robusta, Excelsa, and Liberica. The system was able to identify the coffee species based on the following values area, perimeter, equivalent diameter and roundness that could be extracted in order to deduce their equivalent morphological values. Based on the results, the Bagged Tree Classifier got the highest accuracy rate of 95.6%.

**Index Terms:** Decision Trees, Ensemble Classifiers, Excelsa, Liberica, Robusta

## 1. INTRODUCTION

### 1.1 Coffee

Coffee categorized to the genus *Coffea* L. covering a hundred species. It is also considered second to petroleum as the most important product in the world trade.[1] In this generation, coffee is one of the most commercialized type of food product and vastly consumed beverage in the earth.[2], [3] It is also one of the most produced products in developing countries such as Latin America, Africa, and Asia, averaging for more than fifty countries in the said places. It serves as a main source of income for some.[4] The production of coffee beans consists of farming, collecting, and processing – which is labor-intensive, thus making it an exceptional job that requires a lot of strength and guts, which why it is perfectly suited for abundant developing countries.[5] On the other hand, the process of roasting and branding of coffee is suited for northern industrialized countries.[6] Throughout the Earth, the top five contenders for being the most consumers of coffee are the United States of America, Brazil, Germany, Japan, and France.[7] The framework of the production is identical which isn't dependent on whether the country is a producing or consuming country. The framework consists of different phases such as Cultivation, Processing, Roasting, and Consumption.[8] The diverse areas of the framework are each related to environmental, social, economic, and governance issues which are significant in affecting the future sustainability of extracting the cultured coffee bean.[9], [10], [11].

The plantations in the region of Central highlands Arabica grew, but a case of severe leaf rust outbreaks were plagued. Eventually, after the long run almost of these plantations were transformed to Robusta and Liberica. Coffee was also initially cultivated by the French in Quang Tri, coffee. And this type of coffee was mostly of the Liberica type.[12]

The scenario of extracting and identifying coffee types like robusta and arabica became interesting for the target consumers, raising industries and agencies.[13] Brazilian authorities had considered the chances of mandating the extracted information regarding the composition of blends and roast degree on the cover or shell of the roasted and ground commercial coffee.[14]

### 1.2 Decision Tree Classifiers

Decision trees are known for being an efficient classifier where it represents their classification through a tree form. Each of its interior points or node of the tree form is a test on classification.[15] The succession of the test results in the parameter being taken to one branch of that node. On the other hand, failing results in the parameters being taken out on another branch. This type of classifier is mainly used for classifying an instantaneous scenario by starting from the root node of the decision. The second type of tree is a class probability type. This contains a vector class probability in which each of its leaf decides on its own.[16] The algorithm basically builds a tree top-down by having the standard greedy research principle. It includes stopping, splitting and pruning rules.[17]

### 1.3 Ensemble Classifiers

In this type of classifier, it is visible that it is useful for each of the component/part classifier is well train and flexible even on a different region of feature space. It is also known as mixture-of-expert models, sometimes called as modular classifiers, or occasionally pooled classifiers. The idea of this classifier is the unique resampling technique which suggests the building of multiple component classifiers.[18] One of the convenient technique and most used software where processing of image can be executed in MATLAB. With the use of image processing, some objects can be classified and determined as well. In using image processing, objects can be determined and classified, diseases can be spotted and many other applications. [19]

## 2 METHODOLOGY

### 2.1 Coffee Beans

The coffee beans gathered and used in the study came from the National Coffee Research Development and Extension Center (NCRDEC), at Cavite State University, Indang, Cavite. The coffee bean varieties were Robusta, Excelsa, and Liberica. One hundred twenty (120) samples per variety were

- *Diego Bersaba is from the Department of Computer and Electronics Engineering, College of Engineering and Information Technology, Cavite State University, Indang, Cavite.*
- *Jumelyn Usa is from the Department of Computer and Electronics Engineering, College of Engineering and Information Technology, Cavite State University, Indang, Cavite.*
- *Edwin Arboleda is from the Department of Computer and Electronics Engineering, College of Engineering and Information Technology, Cavite State University, Indang, Cavite. E-mail: [edwin.r.arboleda@cvsu.edu.ph](mailto:edwin.r.arboleda@cvsu.edu.ph)*
- *Elbert M. Galas is from the Information Technology Department, College of Computing, Pangasinan State University-Urdaneta Campus*

gathered. This consisted of 60 that was for training and also 60 for data testing.[20]

**2.2 Image Processing**

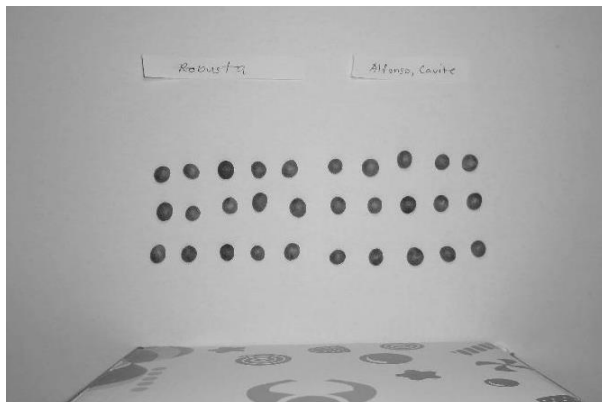
Images gathered per variety were processed using Matlab version 2016a software. The images were first converted to grayscale in rgb2gray function and then converted to black and white by typing im2bw.(15) The converted images were then labeled using bwlabel and vislabels function. [19], [21], [22] With the use of image processing, some objects can be classified.



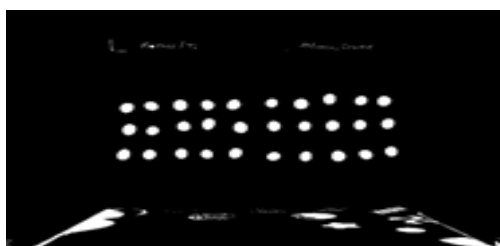
**Fig. 1. Gathered Coffee Bean Image**

**2.3 Data Extraction**

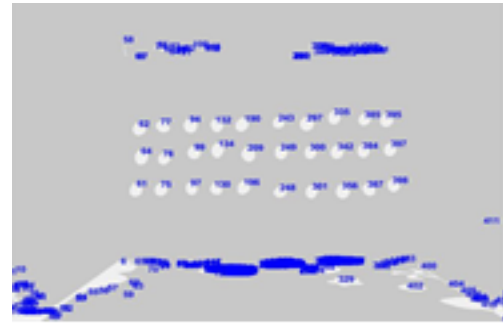
In the labeled images, only the white circles which correspond to the coffee beans were considered. Morphological features to be used for recognition were gathered using the region props function.[23], [24]



**Fig. 2. Grayscale Coffee Bean Image**



**Fig. 3. Black and White Coffee Bean Image**



**Fig. 4. Labeled Coffee Bean Image**

Here are some of the morphological features extracted:

1. Area – it is the number of pixels in the region
2. Perimeter – simply the distance on the boundary of the region
3. Equivalent Diameter – it is the circle's diameter having the same area as the region
4. Roundness - In a non-circular object, it is computed as

$$Roundness = \frac{(4)(Area)(\pi)}{Perimeter^2}$$

**TABLE 1**

**SAMPLE TABLE OF MORPHOLOGICAL FEATURES EXTRACTED (ROBUSTA)**

| Coffee Bean No. | Area | Perimeter | Diameter    | Roundness   |
|-----------------|------|-----------|-------------|-------------|
| 62              | 2592 | 178.752   | 57.44768837 | 101.9396274 |
| 77              | 2299 | 172.401   | 54.10339835 | 97.2006735  |
| 94              | 2904 | 190.074   | 60.8069703  | 101.009227  |
| 132             | 2386 | 173.046   | 55.11759749 | 100.1283763 |
| 190             | 2619 | 184.392   | 57.74611993 | 96.79684736 |
| 243             | 2032 | 158.341   | 50.86474963 | 101.8466623 |
| 297             | 2864 | 188.875   | 60.38673742 | 100.8867015 |
| 335             | 2517 | 177.867   | 56.61045781 | 99.97751075 |
| 385             | 2184 | 165.164   | 52.73286608 | 100.6078142 |
| 395             | 2621 | 181.746   | 57.76816464 | 99.71193915 |
| 64              | 2890 | 190.727   | 60.66021995 | 99.83512165 |
| 78              | 2101 | 160.897   | 51.72113962 | 101.9858743 |
| 99              | 2426 | 174.46    | 55.57768559 | 100.1633692 |
| 134             | 2745 | 187.998   | 59.11888488 | 97.59908692 |
| 209             | 2994 | 194.102   | 61.74203751 | 99.86231726 |
| 249             | 2509 | 178.033   | 56.52042124 | 99.47398242 |
| 300             | 2240 | 167.215   | 53.40464942 | 100.6717032 |
| 342             | 2636 | 181.134   | 57.9332326  | 100.9613884 |
| 384             | 2379 | 172.31    | 55.03668664 | 100.6893046 |
| 397             | 2504 | 177.23    | 56.46407548 | 100.1773903 |
| 61              | 2752 | 187.915   | 59.19421616 | 97.93442933 |
| 75              | 2462 | 175.403   | 55.98853239 | 100.5596791 |
| 97              | 2320 | 170.697   | 54.34993784 | 100.0566762 |
| 130             | 2048 | 159.392   | 51.06461189 | 101.2993774 |
| 196             | 2527 | 178.016   | 56.72280255 | 100.2067621 |
| 248             | 2178 | 164.288   | 52.66038101 | 101.4042259 |
| 301             | 2290 | 169.837   | 53.99739398 | 99.76557885 |
| 356             | 2900 | 191.149   | 60.76507779 | 99.73872251 |
| 387             | 2383 | 171.896   | 55.08293597 | 101.3450088 |
| 398             | 2350 | 171.83    | 54.7002096  | 100.0183639 |

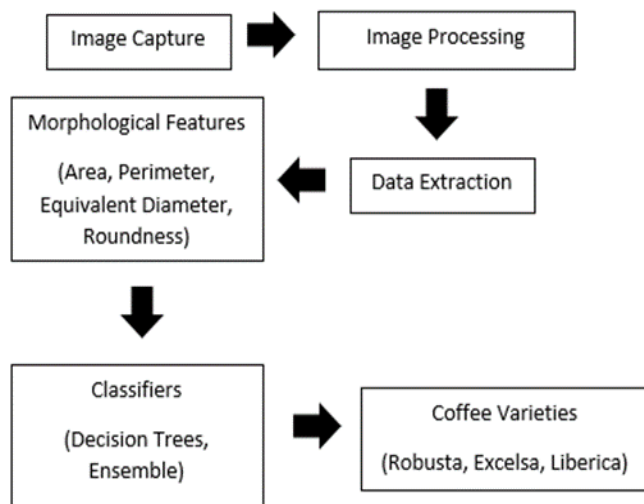
**2.4 Data Classifier**

Decision trees and ensemble classifiers from the Matlab

classifier learner were used in this study.[25], [26] These classifiers were further classified into several types. Table 2 shows the classifiers and their subtypes. (Turcanu, 2016) (Maria & Ribeiro, 2015)

**TABLE 2**  
DATA CLASSIFIERS USED IN THE STUDY

| Decision Tree | Ensemble              |
|---------------|-----------------------|
| Complex Tree  | Boosted Trees         |
| Medium Tree   | Bagged Trees          |
| Simple Tree   | Subspace Discriminant |
|               | Subspace KNN          |
|               | RUSBoosted Trees      |



**Fig. 5.** Coffee Recognition Flow

## 3 RESULTS AND ANALYSIS

### 3.1 Morphological Features

Four morphological features such as area, perimeter, equivalent diameter, and roundness were taken for recognition. Table 3 shows the range of values extracted per feature in each coffee variety.

**TABLE 3**  
RANGES OF MORPHOLOGICAL FEATURES EXTRACTED

|                     | Robusta             | Excelsa             | Liberica            |
|---------------------|---------------------|---------------------|---------------------|
| Area                | 1816 – 2994         | 2409 – 3554         | 3093 – 5385         |
| Perimeter           | 157.076 – 194.102   | 184.193 – 222.576   | 202.52 – 275.554    |
| Equivalent Diameter | 48.08537 – 61.74204 | 55.38262 – 67.26881 | 62.75452 – 82.80335 |
| Roundness           | 64.48868 – 102.0404 | 67.25677 – 102.3105 | 80.84326 – 97.84629 |

### 3.2 Coffee Recognition using Morphological Values

The Matlab classification learner was trained in decision tree and ensemble classifiers using 60 samples. To test the performance of each of the proposed classifiers, 60 samples were also used. The classifier that garnered the highest accuracy rate in coffee variety recognition was the bagged tree with 95.6% which is a type of ensemble. The lowest of them

all were Boosted Tree and the RUSBoosted Tree with 33.3%. Table 4 shows the coffee recognition results in the study.

**TABLE 4**  
COFFEE RECOGNITION PERCENTAGE OF ACCURACY

| Classifier    |                       | %           |
|---------------|-----------------------|-------------|
| Decision Tree | Complex Tree          | 91.1        |
|               | Medium Tree           | 91.1        |
|               | Simple Tree           | 86.7        |
| Ensemble      | Boosted Trees         | 33.3        |
|               | Bagged Trees          | <b>95.6</b> |
|               | Subspace Discriminant | 82.2        |
|               | Subspace KNN          | 86.7        |
|               | RUSBoosted Trees      | 33.3        |

## 4 CONCLUSION

This paper presented the coffee variety recognition by image processing and by the use of Decision Tree and Ensemble classifiers. Each classifier was tasked to identify what type of coffee was each of the test samples. The results showed that the best classifier fit for the job is the Bagged Tree classifier that has the highest accuracy with 95.6% compared to the 7 other classifiers.

## REFERENCES

- [1] E. R. Arboleda, "Comparing Performances of Data Mining Algorithms for Classification of Green Coffee Beans," International Journal of Engineering and Advanced Technology vol. 8, no. 5, pp. 1563–1567, 2019.
- [2] S. N. Kane, A. Mishra, and A. K. Dutta, "Preface: International Conference on Recent Trends in Physics (ICRTP 2016)," J. Phys. Conf. Ser., vol. 755, no. 1, 2016.
- [3] E. R. Arboleda, A. C. Fajardo, and R. P. Medina, "Classification of coffee bean species using image processing, artificial neural network and K nearest neighbors," 2018 IEEE Int. Conf. Innov. Res. Dev. ICIRD 2018, no. May, pp. 1–5, 2018.
- [4] L. Lu, L. Di, and Y. Ye, "A decision-tree classifier for extracting transparent plastic-mulched Landcover from landsat-5 TM images," IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 7, no. 11, pp. 4548–4558, 2014.
- [5] M. A. Istiadi, W. A. Kusuma, and I. M. Tasma, "Application of decision tree classifier for single nucleotide polymorphism discovery from next-generation sequencing data," Proc. - ICACIS 2014 2014 Int. Conf. Adv. Comput. Sci. Inf. Syst., pp. 339–343, 2014.
- [6] J. T. Diaz et al., "Methodology for decisions making in transportation logistics function For Supply Chain (SC) colombian green coffee study," 2015 Int. Conf. Ind. Eng. Oper. Manag., p. 8, 2015.
- [7] S. Tadjudin and D. A. Landgrebe, "Decision tree classifier design for high-dimensional data with limited training samples," Int. Geosci. Remote Sens. Symp.,

- vol. 1, pp. 790–792, 1996.
- [8] D. V. V. V. Jaya and R. Krishniah, "Hemorrhage Detection Using Decision Tree Classifier."
- [9] M. Murphy and T. J. Dowding, "The Coffee Bean: A Value Chain and Sustainability Initiatives Analysis," pp. 1–19, 2017.
- [10] V. V. Satyanarayana Tallapragada and E. G. Rajan, "Morphology based non ideal iris recognition using decision tree classifier," 2015 Int. Conf. Pervasive Comput. Adv. Commun. Technol. Appl. Soc. ICPC 2015, vol. 00, no. c, 2015.
- [11] V. Kadappa, S. Guggari, and A. Negi, "Decision Tree classifier using theme based partitioning," 2015 Int. Conf. Comput. Netw. Commun. CoCoNet 2015, no. 2, pp. 540–546, 2016.
- [12] R. Dias and M. Benassi, "Discrimination between Arabica and Robusta Coffees Using Hydrosoluble Compounds: Is the Efficiency of the Parameters Dependent on the Roast Degree?," *Beverages*, vol. 1, no. 3, pp. 127–139, 2015.
- [13] M. Pal and P. M. Mather, "A comparison of decision tree and backpropagation neural network classifiers for land use classification," *Int. Geosci. Remote Sens. Symp.*, vol. 1, no. C, pp. 503–505, 2002.
- [14] M. Kuit, D. M. Jasen, and N. Van Thiet, "Manual for Arabica cultivation," p. 213, 2004.
- [15] Pooja Kamavisdar, Sonam Saluja, and Sonu Agrawal, "A Survey on Image Classification Approaches and Techniques," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 2, no. 1, pp. 1005–1009, 2013.
- [16] Anuradha and G. Gupta, "A self explanatory review of decision tree classifiers," *Int. Conf. Recent Adv. Innov. Eng. ICRAIE 2014*, no. 3, 2014.
- [17] M. Cerňak, "A comparison of decision tree classifiers for automatic diagnosis of speech recognition errors," *Comput. Informatics*, 2010.
- [18] B. Ma, H. Li, and R. Tong, "Spoken language recognition using ensemble classifiers," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 15, no. 7, pp. 2053–2062, 2007.
- [19] E. R. Arboleda, A. C. Fajardo, and R. P. Medina, "An image processing technique for coffee black beans identification," 2018 IEEE Int. Conf. Innov. Res. Dev. ICIRD 2018, no. May, pp. 1–5, 2018.
- [20] R. H. M. Condori, J. H. C. Humari, C. E. Portugal-Zambrano, J. C. Gutiérrez-Cáceres, and C. A. Beltrán-Castañón, "Automatic classification of physical defects in green coffee beans using CGLCM and SVM," *Proc. 2014 Lat. Am. Comput. Conf. CLEI 2014*, 2014.
- [21] N. P. Lakkakula, M. M. Naidu, and K. K. Reddy, "An entropy based elegant decision tree classifier to predict precipitation," in *Proceedings - UKSim-AMSS 8th European Modelling Symposium on Computer Modelling and Simulation, EMS 2014*, 2014.
- [22] B. M. Ayitenfsu, "Method of Coffee Bean Defect Detection," *Int. J. Eng. Res. Technol.*, 2014.
- [23] C. Kelly and J. Scott, "Microwave aquametry of roasting coffee beans," *Asia-Pacific Microw. Conf. Proceedings, APMC*, vol. 3, pp. 1–3, 2015.
- [24] M. Wang, A. Perera, and R. Gutierrez-Osuna, "Principal Discriminants Analysis for small-sample-size problems: Application to chemical sensing," *Proc. IEEE Sensors*, vol. 2, pp. 591–594, 2004.
- [25] J. T. Díaz, J. D. V. Cubillos, and H. Bolivar, "Application for decision-making in transportation logistics function: Supply chain colombian green coffee," *Proc. - 2nd Int. Symp. Stoch. Model. Reliab. Eng. Life Sci. Oper. Manag. SMRLO 2016*, pp. 575–582, 2016.
- [26] L. De Gonzaga Ferreira, F. M. Da Siva, and D. D. Ferreira, "Displacement Tracking of Harvester Rods of a Coffee Harvester," *IEEE Lat. Am. Trans.*, vol. 14, no. 12, pp. 4674–4680, 2016.