

# Classification of Immature and Mature Coffee Beans Using Texture Features and Medium K-Nearest Neighbor

Edwin R. Arboleda

Department of Computer and Electronics Engineering, College of Engineering and Information Technology,  
Cavite State University, Indang, Cavite, Philippines

(Received 06 March 2023; Revised 30 April 2023; Accepted 30 April 2023; Published online 11 May 2023)

**Abstract:** In this study, texture features namely entropy, contrast, energy, and homogeneity are extracted from mature and immature coffee beans using image processing, and the values are inputted into MATLAB's Classification Learner App for discrimination. Among the 23 machine-learning algorithms, the best performance was achieved by medium K-nearest neighbor which has 97% accuracy and 0.14574 seconds in speed. When compared with previous studies that used RGB and HSV color features to differentiate mature and immature coffee beans, it can be concluded that texture features are far superior in distinguishing the two coffee bean groups.

**Keywords:** coffee; energy; entropy; homogeneity; machine learning; texture

## I. INTRODUCTION

Coffee is one of the world's most important food commodities [1]. Coffee ranks second in global importance only behind crude oil, with an annual turnover of \$10 billion [2]. Nearly 60 tropical and subtropical countries export coffee as their main agricultural product. Over 70 African, Latin American, and Southeast Asian nations considered coffee as their most important agricultural product [3]. In most Asian countries, the coffee industry provides employment for many families and allows them to meet their daily needs [4]. The quality of coffee and its taste is affected by a variety of environmental factors. In different countries, higher altitudes and more shade are better for the quality [5]. Climate and soil conditions are perfect for four different coffee species to grow in the Philippines: Arabica, Robusta, Liberica, and Excelsa.

It took over a century before Philippine coffee reached its peak. A Spanish friar from Lipa, Batangas, brought the country's first coffee plant to the country in 1740. After the coffee rust destroyed Brazil, Africa, and Java in the year 1880, the Philippines became the largest producer of coffee beans in the world. By 1889, coffee rust reached the Philippine shores. In 1889, coffee rust and bug invasion nearly destroyed the Philippines's coffee crops. In the 1950s, the United States developed pest-resistant varieties, and in the 1960s, many farmers returned to coffee farming. Brazil experienced a coffee frost in the 1970s, which led to a spike in coffee prices. The Philippines became a member of International Coffee Organization in 1980 [6]. The Philippines is among the 43 countries that export coffee in the world, but its production pales in comparison with Brazil, Vietnam, Columbia, Indonesia, Ethiopia, Honduras, India, and Peru. The Philippines ranks 31st among 36 countries that export coffee, with 12,000,000 kg exported [7].

The presence of flaws, size, color, shape, and roasting potential are all factors that can be used to evaluate the quality of coffee. The quality of coffee is affected by defects, which are related to the harvesting and preprocessing methods. Only pick coffee cherries that are fully red on the branch. The coffee cherries do not all ripen simultaneously. Some cherries are completely red, others yellow, and still others green. Farmers should first pick the red cherries, and then wait until the yellow and green beans have fully developed and turned red. To save money, because the red, yellow, and green coffee beans are all identical when dried, farmers collect them regardless of their color. Although mature and immature coffee beans look almost identical, their taste is completely different. Immature coffee lowers the price of beans and reduces their quality [8]. A typical method for separating coffee beans involves a manual examination, followed by the selection and separation of the defective ones. This method is time-consuming and subjective, as it depends on the condition of the person who selects the defects [9].

Manual inspection methods rely on human judgment and can differ between people even though they are both doing the same thing. This study builds on previous research in immature or mature coffee beans [10–12] to address the research question, "Which attribute, is best for distinguishing mature and immature coffee beans—the color features (RGB and HSV) or the texture attributes?" The study used image processing techniques to extract texture characteristics, which were subjected to machine learning (ML) to determine which feature was best.

## II. RELATED WORK

Image processing is the process of performing specific operations on an image to refine, enhance, or extract valuable characteristics [13]. Image processing is a process that extracts desirable information from images using various algorithms and modification techniques such as morphology and color [14]. The values that can be retrieved by image processing represent unique traits to one

Corresponding author: Edwin R. Arboleda (e-mail: [edwin.r.arboleda@cvsu.edu.ph](mailto:edwin.r.arboleda@cvsu.edu.ph)).

class [15]. There are several studies that employ image processing in extracting features of coffee. The morphological features retrieved by image processing were 96.66% accurate and K-nearest neighbor (KNN) was 84.12% accurate in identifying three coffee bean species [16]. Image processing was used to extract RGB values that were then used to distinguish black beans from regular beans. The texture of an image is determined by the roughness, granulation, and regularity of its pixel structure. Textural aspects of an image can be used for segmentation, classification, and interpretation. In 1973, energy, entropy, and contrast were among the texture characteristics that were presented. The Texture Feature Extraction Algorithm is based on binarized images' qualities according to texture [17].

ML, combined with machine vision algorithms, has been used in agricultural applications, leading to increased crop production through automation [18–20]. A classifier is needed to distinguish between classes, as feature values are often overlapping. Various classifiers were developed over the years, and their performance and use depends on the application. The MATLAB Classification Learner App contains 23 machine-learning classifiers. The input data can be extracted from image processing, or any data that consists of numeric values. The app validates and trains the data to improve accuracy and speed.

In recent years, a combination of image processing, features extraction, and ML has been used to discriminate crop classes based on their characteristics.

The color values (R, G, B, H, S, and I) of the grape skin were used to predict grape maturation. The color values of the grape skin were extracted using image processing. The features were subjected to a Back Propagation Neural Network (BPNN) algorithm. The results showed that BPNN could classify Drunk Incense Grapes with recognition accuracy as high as 79.3%. Muscat Hamburg Grapes were classified with 78.2% accuracy, and Xiang Yue with 79.4% [21].

This paper uses image processing to extract the shape parameters of tobacco leaves, such as their length, width, and area, as well as their circumference and roundness. Then, a BPNNs with Levenberg Marquardt algorithm is used as a classifier for tobacco leaves grade. The BPNN LM algorithm provided a 99.495% accuracy in predicting the tobacco leaves grades [22].

This study used the color of the Bluecrop Northern Highbush Blueberries (green, green-red, red, blue, and red-blue), to identify five different maturation stages. Image processing is used to extract color features, which are then classified using discriminant analysis. The combination of image processing and discriminant analysis accurately classified the maturity stages with 98.3% accuracy [23].

In this study, image processing and machine-learning algorithms were used to classify the following fruit varieties: Starkrimson Delicious and Golden Delicious apple varieties, Washington Navel and Valencia Midnight orange varieties, and Ekmek and Esme quince varieties. Image processing techniques were used to extract color and size features. The success rate for classifiers was 93.6% with K-Nearest Neighbor, 90.3% with decision trees, 88.3% with Naive Bayes, and 92.6% with multi-level perceptrons [24].

In this study, levels of melons are classified into mature and immature. This research shows that the most accurate method to detect melon maturation levels is a combination (mean filtering  $[5 \times 5]$  + gamma-contrast stretching 2.0) + (mean adaptive threshold) + (histogram with oriented gradients) + (support vector machines) with an accuracy rate of 78.67% [25].

### III. MATERIALS AND METHODS

#### A. COFFEE BEAN SAMPLES

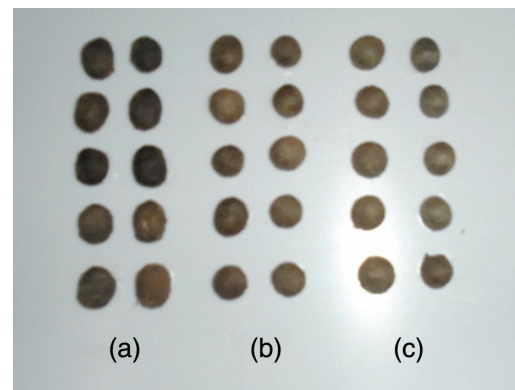
The dataset for this study was created using prototype hardware that the author developed to extract green coffee bean characteristics [26]. The coffee beans were collected during the Philippines' 2019 coffee harvest. The coffee beans were kept in their husks for 1 year, before they were removed in March 2020 and used as samples. Two groups are formed from mature and immature Robusta coffee beans. In this experiment, 100 coffee beans, mature and immature, were used. The coffee samples used in the study are shown in Fig. 1. In Fig. 1, you can see that the first two columns in the table are dried green beans. The third and fourth columns show dried yellow beans. And the fifth and sixth columns show dried red beans. The dried yellow and green beans are immature coffee beans. The mature coffee beans are red-dried beans. In Fig. 1, the texture difference between immature coffee beans and mature ones is visible to the naked eye. Immature coffee beans have a rougher texture than mature coffee beans. The mature coffee bean is smoother than the immature beans.

**1) TEXTURE FEATURE EXTRACTION.** The texture of the beans can be used to distinguish mature from immature beans. When the two groups are combined, the problem arises. The two groups are difficult to differentiate. In previous research, it was found that the RGB and HSV values of mature coffee beans and immature ones differ. This study goes further to determine if the texture values of mature and immature beans can be distinguished.

The following information was used to calculate the texture properties, which can be described as a co-occurrence matrix.

- i. Entropy can be used to calculate the degree of predictability in the intensity distribution.
- ii. In a matrix of co-occurrences, energy is the feature that measures the concentration of intensity pairs.
- iii. Contrast measures the contrast between images.
- iv. Homogeneity, the inverse to contrast, determines how homogenous an image in terms of intensity variations is.

The author [26] developed a machine vision prototype to extract the texture features from the coffee beans. Figure 2 illustrates how texture features are extracted. The prototype was used to photograph the coffee beans at a camera height of 13.5 cm. Python software was used to retrieve the texture properties from the prototype. The hardware prototype saves the extracted features



**Fig.1.** Immature Coffee Beans - (a) Dried Green Beans (b) Dried Yellow Beans and Mature Coffee Beans - (c) Dried Red Beans.

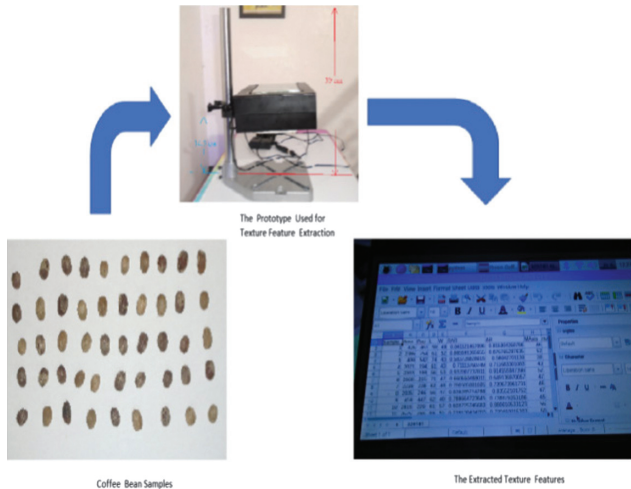


Fig. 2. The process of extracting texture features.

in its USB drive. The data saved in the USB drive of the hardware prototype is transferred to the computer. The texture characteristics of mature and immature coffee beans are then arranged into an MS Excel file with four columns: Entropy, Contrast, Energy, and Homogeneity. Finally, the fifth column will contain the classification, whether the coffee bean belongs to the mature bean or the immature bean. The first 100 rows of the MS Excel file's 5th column are labeled as "Mature Coffee Bean", and rows 101–200 are labeled as "Immature Coffee Bean". The approach used in this paper is the supervised learning because the dataset already contains the classes, "Mature and Immature Coffee Beans" corresponding to the texture features of every sample. The coffee bean classes are labeled manually before it is inputted into the machine-learning algorithms of the MATLAB Classification learner app.

### IV. RESULTS AND DISCUSSION

Table I shows the texture values that can be recovered through image processing. In Table I, the texture values of mature and immature beans are overlapping. The author used classification algorithms to determine whether a coffee bean was immature or mature.

As a classifier, the 23 machine-learning algorithms from the MATLAB Classification Learner App were used. As inputs, 200 beans (100 mature coffee beans and 100 immature coffee beans) were used at once. The settings were set on 5 folds. This means that 80% of the samples were used for training. Divide the total number of samples by five to get five times as many, or divide 100 by five to get 20. The remaining 20 samples served as test samples. A total of 80 were used as samples for training. The results of the classification are shown in Table II.

In Fig. 3, the scatter plot for the medium KNN texture features of mature and immature beans is shown. Immature coffee beans are

Table II. Classification results

Classifier type	Accuracy (%)	Training time (Seconds)
Fine Tree	95.5	0.41367
Medium Tree	95.5	0.14974
Coarse Tree	95.5	0.12787
Linear Discriminant	97	0.1813
Quadratic Discriminant	97	0.14574
Logistic Regression	96	0.40414
Support Vector Machine (SVM)	97	0.16595
Quadratic SVM	97	0.16595
Cubic SVM	97	0.15512
Fine Gaussian SVM	96.5	0.14866
Medium Gaussian SVM	96.5	0.16383
Coarse Gaussian SVM	96.5	0.13939
Fine KNN	96.5	0.14219
Medium KNN	97	0.14219
Coarse KNN	92	0.16051
Cosine KNN	97	0.15295
Cubic KNN	96.5	0.1448
Weighted KNN	96	0.15377
Boosted Trees	50	0.21365
Bagged Trees	96.5	1.4138
Subspace Discriminant	96.5	1.4176
Subspace KNN	95.5	1.6234
RUS Boosted Trees	50	0.22916

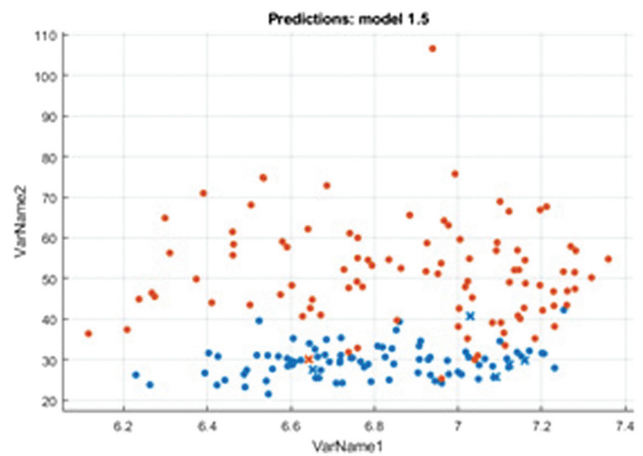


Fig. 3. Scatterplot of medium KNN.

represented in blue, and mature coffee beans in orange. Figure 3 shows the texture values between the two groups from the perspective of medium KNN.

Table I. Texture values extracted from mature and immature coffee beans

Coffee bean type	Entropy	Contrast	Energy	Homogeneity
Mature Coffee Beans	6.229132 to 7.420944	21.52108 to 72.77092	12.9994 to 49.95692	2613 to 11894
Immature Coffee Beans	0.153334 to 0.405179	6.558494 to 7.479654	36.73359 to 101.0826	12.9994 to 42.95692

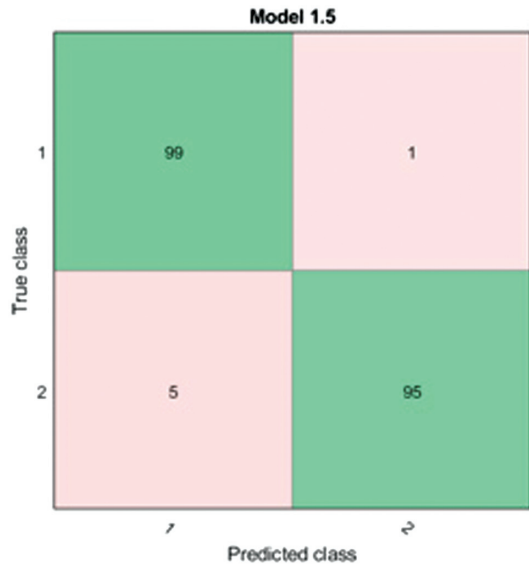


Fig. 4. Confusion matrix of medium KNN.

The seven classifiers that achieve the highest accuracy of 97% are linear discriminant (LD), quadratic discriminant (QD), linear support vector machine (SVM), cubic SVM, and quadratic SVM. The medium KNN is the fastest classifier, with a time to train of 0.14219 seconds. The medium KNN will take only 0.14219 seconds to differentiate between the two classes of 40 mixed samples of mature and immature beans. Even the most experienced human classifiers cannot match that level of accuracy or speed.

Figure 4 shows the confusion matrix of the medium KNN.

As shown in Fig. 4, the medium KNN has one incorrect classification for mature coffee beans and five incorrect classifications for immature coffee beans. This indicates that medium KNN can distinguish between the texture values of mature and immature coffee beans with relatively few incorrect classifications in both groups.

Fig. 5 shows the receiver operating characteristics (ROCs) of the medium KNN.

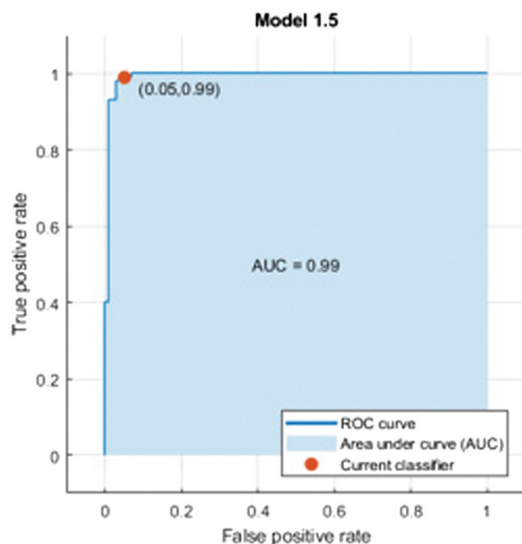


Fig. 5. The ROC curve of medium Gaussian SVM.

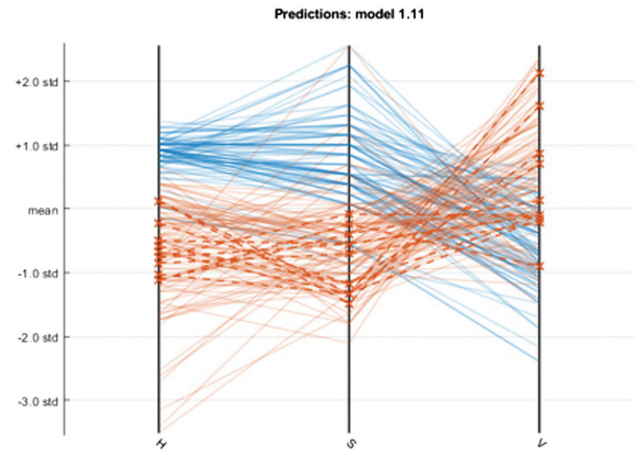


Fig. 6. Parallel coordinates plot of the medium Gaussian SVM.

Table III. Comparison of results

Study	Features	Classifier	Accuracy (%)	Training time
Eustaquio <i>et al.</i> [12]	RGB	Quadratic SVM	94	0.62 seconds
Eustaquio <i>et al.</i> [10]	RGB	ANN	95	300 seconds
De Guzman [11]	HSV	Medium Gaussian SVM	94	0.14624 seconds
This Study	Texture	Medium KNN	97	0.14219 seconds

The curve of the ROC shows the relationship between the features input and the classifier. The area under the curve in Fig. 3 is 0.99, which indicates that the features given are used more often for classification.

Figure 6 shows the plot of parallel coordinates for the medium KNN. Figure 4 shows that mature coffee bean (orange) has lower texture values than immature beans (blue). The medium KNN could differentiate between two classes even though the data overlapped.

Table III shows the comparison between the results of the current study and the previous studies on the classification of mature and immature beans.

Table III compares the results of previous studies to those of this study. In a study conducted previously, quadratic SVM was able to achieve a classification accuracy rate of 94% when RGB was used as a feature. In another study, using RGB color features with a back propagation neural network as a classification yielded 95% accuracy. A study that used HSV data found the medium Gaussian SVM to be 94% accurate. This study used texture features and achieved 97% accuracy with a medium KNN classification. Comparing previous studies which used RGB and HSV colors to distinguish mature and immature beans, texture features give the highest accuracy. The shortest training time could be achieved by combining texture features with medium KNN.

## V. CONCLUSION

The research question of the study is whether texture variables are better for identifying mature and immature coffee beans than RGB color variables or HSV color variations. In terms of accuracy and time required to train, it can be concluded texture features are superior in identifying immature and mature coffee beans. The results show that texture features are better at identifying maturity in coffee beans than color features RGB or HSV. Color photos can easily identify RGB and HSV features, but they have a limited range of variation. This method is also slower to train than texture features.

## References

- [1] K. P. P. Nair, *The Agronomy and Economy of Important Tree Crops of the Developing World*: Elsevier, Amsterdam, Netherlands, 2010. DOI: [10.1016/B978-0-12-384677-8.00006-0](https://doi.org/10.1016/B978-0-12-384677-8.00006-0).
- [2] M. Sanders, "Breeding for coffee leaf rust resilience in *Coffea* sp." *Nat. Sci. Educ.*, vol. 48, no. 1, p. 190102, 2019. DOI: [10.4195/nse2019.01.0102](https://doi.org/10.4195/nse2019.01.0102).
- [3] M. K. Mishra and A. Slater, "Recent advances in the genetic transformation of coffee." *Biotechnol. Res. Int.*, vol. 2012, pp. 1–17, 2012. DOI: [10.1155/2012/580857](https://doi.org/10.1155/2012/580857).
- [4] F. J. P. Montalbo and A. A. Hernandez, "Classifying Barako coffee leaf diseases using deep convolutional models." *Int. J. Adv. Intell. Inf.*, vol. 6, no. 2, pp. 197–209, 2020. DOI: [10.26555/ijain.v6i2.495](https://doi.org/10.26555/ijain.v6i2.495).
- [5] B. Cheng, A. Furtado, H. E. Smyth, and R. J. Henry, "Trends in food science & technology influence of genotype and environment on coffee quality," vol. 57, pp. 20–30, 2016. DOI: [10.1016/j.tifs.2016.09.003](https://doi.org/10.1016/j.tifs.2016.09.003).
- [6] Philippine Coffee Board, "OUR COFFEE HERITAGE: coffee's rich history in the Philippines," pp. 2–3, 2017, [Online]. Available: <http://philcoffeeboard.com/philippine-coffee/>.
- [7] International Coffee Organization, "Coffee prices increase as concerns about Brazil's 2017/18 crop emerge," 2017, [Online]. Available: [https://www.ico.org/show\\_news.asp?id=612](https://www.ico.org/show_news.asp?id=612).
- [8] C. Pinto, J. Furukawa, H. Fukai, and S. Tamura, "Classification of green coffee bean images based on defect types using convolutional neural network (CNN)," in *Proceedings - 2017 International Conference on Advanced Informatics: Concepts, Theory and Applications, ICAICTA*, 2017, pp. 1–5. DOI: [10.1109/ICAICTA.2017.8090980](https://doi.org/10.1109/ICAICTA.2017.8090980).
- [9] D. Talukdar et al., "Coffee bean grade determination based on image parameter," *Telkomnika*, vol. 9, no. 3, pp. 547–554, 2011. DOI: [10.11591/telkomnika.v16i1.8686](https://doi.org/10.11591/telkomnika.v16i1.8686).
- [10] W. R. Eustaquio and J. L. Dioses, "Artificial neural network for classification of immature and mature coffee beans using RGB values," *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 8, pp. 4301–4305, 2020. DOI: [10.30534/ijeter/2020/41882020](https://doi.org/10.30534/ijeter/2020/41882020).
- [11] L. B. de Guzman, "Classification of immature and mature coffee beans using HSV features and machine learning algorithms," *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 8, pp. 4350–4356, 2020. DOI: [10.30534/ijeter/2020/49882020](https://doi.org/10.30534/ijeter/2020/49882020).
- [12] W. R. Eustaquio and J. L. Dioses, "Classification of immature and mature coffee beans using RGB values and machine learning algorithms," *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 8, pp. 4301–4305, 2020. DOI: [10.30534/ijeter/2020/41882020](https://doi.org/10.30534/ijeter/2020/41882020).
- [13] F. Faridah, G. O. F. Parikesit, "Coffee bean grade determination based on image parameter," *Telkomnika*, vol. 9, no. 3, pp. 547–554, 2011.
- [14] J. Tian, Y. Zhang, J. Lei, C. Sun, and G. Hu, "Lightweight classification network for pulmonary tuberculosis based on CT images," *J. Artif. Intell. Technol.*, vol. 3, pp. 25–31, 2023.
- [15] H. Ponce, C. Cevallos, R. Espinosa, and S. Gutiérrez, "Estimation of low nutrients in tomato crops through the analysis of leaf images using machine learning," *J. Artif. Intell. Technol.*, vol. 1, pp. 131–137, 2021.
- [16] E. R. Arboleda, "Classification of coffee bean species using image processing, artificial neural network and K nearest neighbors," in 2018 IEEE Int. Conf. Innov. Res. Dev., no. May, 2018, pp. 1–5, DOI: [10.1109/ICIRD.2018](https://doi.org/10.1109/ICIRD.2018).
- [17] R. M. Haralick, I. Dinstein, and K. Shanmugam, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-3, no. 6, pp. 610–621, 1973. DOI: [10.1109/TSMC.1973.4309314](https://doi.org/10.1109/TSMC.1973.4309314).
- [18] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: a review," *Sensors*, vol. 18, no. 8, p. 2674, Aug. 2018. DOI: [10.3390/S18082674](https://doi.org/10.3390/S18082674).
- [19] S. C. Lauguico, R. I. S. Concepcion, J. D. Alejandrino, R. R. Tobias, and E. P. Dadios, "Lettuce life stage classification from texture attributes using machine learning estimators and feature selection processes," *Int. J. Adv. Intell. Inf.*, vol. 6, no. 2, pp. 173–184, 2020. DOI: [10.26555/ijain.v6i2.466](https://doi.org/10.26555/ijain.v6i2.466).
- [20] K. Kiratiratanapruk et al., "Development of paddy rice seed classification process using machine learning techniques for automatic grading machine," *J. Sens.*, vol. 2020, p. 2020. DOI: [10.1155/2020/7041310](https://doi.org/10.1155/2020/7041310).
- [21] X. Wei, L. Wu, D. Ge, M. Yao, and Y. Bai, "Prediction of the maturity of greenhouse grapes based on imaging technology," *Plant Pnemon.*, vol. 2022, pp. 1–14, 2022.
- [22] H. Su, "Data research on tobacco leaf image collection based on computer vision sensor," *J. Sens.*, vol. 2021, pp. 1–11, 2021.
- [23] A. M. Zahra, T. Chosa, and S. Tojo, "Blueberries quality evaluation during maturation using image processing," *Indones. Food Nutr. Prog.*, vol. 18, no. 2, pp. 41–49, 2021. DOI: [10.22146/ifnp.63897](https://doi.org/10.22146/ifnp.63897).
- [24] D. G. Koc and M. Vatandas, "Turkish journal of agriculture – food science and technology," *Turkish J. Agric. – Food Sci. Technol.*, vol. 9, no. 12, pp. 2189–2196, 2021. DOI: [10.24925/turjaf.v9i12.2189-2196.4445](https://doi.org/10.24925/turjaf.v9i12.2189-2196.4445).
- [25] A. Tawakal and A. Prayoga, "The development of methods for detecting Melon maturity level based on fruit skin texture using the histogram of oriented gradients and the support vector machine," in *2019 Fourth Int. Conf. Inf. Comput. (ICIC)*, Semarang, Indonesia, 2019, pp. 1–4. DOI: [10.1109/ICIC47613.2019.8985953](https://doi.org/10.1109/ICIC47613.2019.8985953).
- [26] E. R. Arboleda, A. C. Fajardo, and R. P. Medina, "Green coffee beans feature extractor using image processing," *TELKOMNIKA (Telecommunication Comput. Electron. Control)*, vol. 18, no. 4, p. 2027, 2020. DOI: [10.12928/telkomnika.v18i4.13968](https://doi.org/10.12928/telkomnika.v18i4.13968).