

# Computer Vision for Identifying and Classifying Green Coffee Beans: A Review

Bonang Waspadadi Ligar<sup>1\*</sup>

<sup>1</sup>Departemen of Informatics Engineering, Faculty of Industrial Technology, Gunadarma University, Indonesia  
<sup>1</sup>bonangligar@staff.gunadarma.ac.id\*

**Abstract.** Coffee is widely consumed around the world, also considered one of the most important beverages today. Factors contributing to the quality of coffee beans such as color, texture, size, aroma, etc. and other processes along the production chain such as plant, roasting, and grinding. Those processes will be worthless if the quality of the coffee beans are low. It is important to only use the best quality coffee beans. Therefore, the challenge is to develop a system that uses computer vision to either identify high quality beans or classify them by their species to ease the effort needed by all actors in the supply chain. Providing information for end customers is a defining factor to push forward the coffee industry. This paper aims to review literatures within the topic of using computer vision for green coffee beans. After reviewing a selected number of studies which corresponds with the topic chosen in our paper, computer vision techniques were used for two main reasons, identification and classification. Research on this topic are still limited. Hence, it can be concluded that there are still plenty of room for study on this topic. This study also aims to help provide research material for future researchers.

**Keywords:** Classification, Coffee Beans, Computer Vision, Identification, Image Processing.

Received June 2022 / Revised June 2022 / Accepted June 2022

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



## INTRODUCTION

Coffee is widely consumed around the world, and is considered one of the most important beverages that exist today [1]. Although coffee was not founded until the 17<sup>th</sup> century, it has now developed not only into a daily routine drink, but also as an expression of lifestyle [2]. There are a number of factors in which a coffee bean is considered as good quality, such as color, texture, size, aroma, etc. [1], [3], and there are also other processes along the production chain such as plant, roasting, and grinding processes [4], [5]. But all of those processes will be worthless if the quality of the coffee bean is low. So, it is important to only use the best quality coffee beans. The geographical origin of a coffee bean also differentiates coffee bean quality [6]. The morphology of a coffee bean such as the surface area of the bean, perimeter, equivalent diameter, and percentage of roundness are several important features of a coffee bean, which differs between different species [6]. It is important to able to identify the quality levels of coffee beans, and also to identify its origin for traceability purposes.

Proper methods for identifying and classifying coffee beans are lacking, which has troubled the coffee industry [6]. The process of identifying coffee defects still use old methods such as generally inspecting only by seeing using the naked eye. This is time consuming and also low in accuracy [1]. In order to find a solution for those problems, many researchers have proposed other methods, one of them is using image processing techniques or in a broader sense, computer vision. Computer vision aims to establish algorithms which extracts and analyses useful knowledge about an object or scene from a studied image, image sets or image sequences [4]. Computer vision has been used by researchers to identify quality factors of coffee beans such as color [1], [3], [4], texture [1], size [1], and shape [1]. Also to classify coffee beans by their species [6], [7].

Failure in identifying which coffee bean is high quality or low quality could affect the whole coffee industry. This is caused by increasing concerns on food safety and quality of consumers, industry players, and also the government as a policy maker [8]. They care about the origin of the product, raw materials, production methods, labor standards applied, and the impact of the production process on the environment [9]. This awareness is formed as a result of many incidents of food counterfeiting, fraud, transmission of diseases through food, harmful substances found in food, the use of genetically modified plants, and others [10]. Therefore, the challenge is to develop a system that uses computer vision to either identify high quality

beans or classify them by their species to ease the effort needed by all actors in the supply chain. Providing information for end customers could also be a defining factor to push forward the coffee industry.

This study focuses on reviewing the identification and classification of green coffee beans and their quality level. Quality is a top priority and one of the parameters that determines the price of coffee beans during negotiations in the coffee trade. However, checking each sample during the coffee trade is a huge task because quality assessors have limitations in examining many samples of green coffee beans. Thus it is not possible to check the quality of each coffee bean sample manually during the sale of coffee beans. A tool that can quickly check the quality of coffee beans based on defined parameters is very much needed in the coffee bean trade, especially at the negotiation stage. Conduct a needs assessment of the quality assessment model. Green coffee bean quality assessment is focused on visual parameters, especially the assessment of defective beans. The contribution of the article lies in the position of the latest research and future research proposals regarding the application of Computer Vision technology in agriculture, especially in coffee commodities.

The main purpose of this research is to find out the latest developments in Computer Vision technology for the identification and classification of coffee beans, as well as to see opportunities for the application of Computer Vision technology in the agricultural sector, especially in the coffee bean commodity. The benefits that can be drawn from this research are to provide alternative solutions in helping research actors and coffee commodity businesses to facilitate the identification and classification of coffee beans, and also as a basis for conducting further research activities in various fields.

**METHODS**

This paper is an idea whose main purpose is to provide up-to-date information on the chosen topic. The steps used in this study were adapted to the steps described in the Kitchenham [11] and Cruz-Benito [12] papers. In general, the literature review process in this paper can be seen in Figure 1.

**A. Inclusion and Elimination Criteria**

In order for the paper to focus more on the research topic, the selection criteria determined can be seen in table 1. The initial search stage is carried out manually to find papers that both in the title and abstract contain the words "Coffee Identification" and/or "Coffee Classification". At this stage, 14 papers were found that met the criteria.

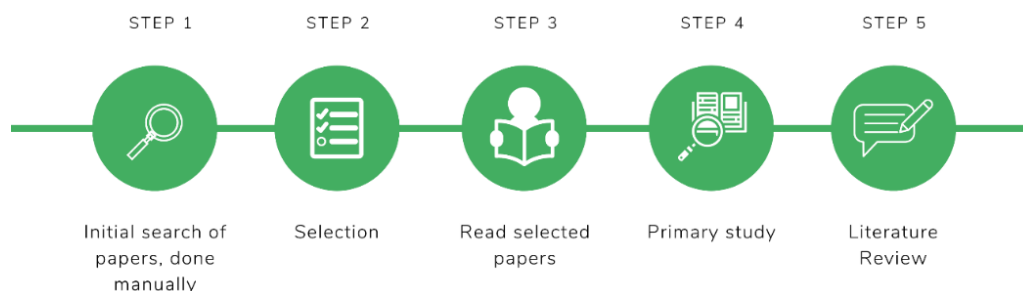


Figure 1. Literature review process in this paper.

In the second stage, namely the selection stage, all the papers that have been obtained in the first stage are then read their contents to find out whether they are related to the research question or not. If not, it will be eliminated. At this stage the selection results get 6 papers that match the research criteria. The last stage is the primary study, where an in-depth literature review is carried out on each paper to find answers to research questions.

Table 1. Inclusion and Exclusion Criteria

<b>Inclusion Criteria</b>
Papers that focuses on the use of computer vision for coffee identification or classification
Papers published between 2016 and 2021
Papers in English or Indonesian language
<b>Exclusion Criteria</b>
Papers in the form of books or research reports
Papers for which full-text articles are not available

## RESULTS AND DISCUSSIONS

After reviewing a selected number of studies which corresponds with the topic chosen in our paper, computer vision techniques were used for two main reasons

1. Identification of Coffee Bean Features
2. Classification of Coffee Beans

A detailed survey on the methods for identification and classification of coffee beans will be explained in the subsequent chapters.

### A. Coffee Bean Quality Assessment

The quality of green coffee beans comes from many variables such as variety, soil, climate, processing methods, and many others. However, green coffee bean quality determination is mainly observed from three categories: green coffee assessment, sensory evaluation, and chemical measurement [13]. Among these three categories, green coffee assessment is the only visual-based evaluation, which consists of bean size distribution, damaged bean count, and bean color assessment [7], [13]. However, the assessment of defective coffee beans in green coffee grading is one of the most influencing quality assessments that correlates with off-flavors in the cup after brewing. Usually, defective beans are assessed by self-selecting and counting the number of defective beans based on the type of defect and calculating the defect point using a predetermined coffee bean standard [13], [14]. This is done during quality assessment sessions in green coffee processors, traders, and certification processes. It is worth mentioning that the green coffee bean quality assessment process is to ensure the green coffee bean screening process during the production process runs correctly for green coffee bean processors and to ensure the green coffee quality is properly labeled in the coffee trade for traders and the certification process [15].

At the processing stage, the quality of coffee beans ranging from cherry coffee to green coffee beans is shown in Figure 2. The quality of green coffee beans in processing plants is generally classified as defective. Green coffee bean defects are created during harvesting, pulping, mucilage removal, drying, grinding, polishing, and separation by size. However, separation, cleaning, density separation, and color sorting reduce defects in production batches by eliminating them [16]. Because of this, coffee bean quality standards were developed in many countries. This ensures the initial parameters for the quality of the resulting drink.

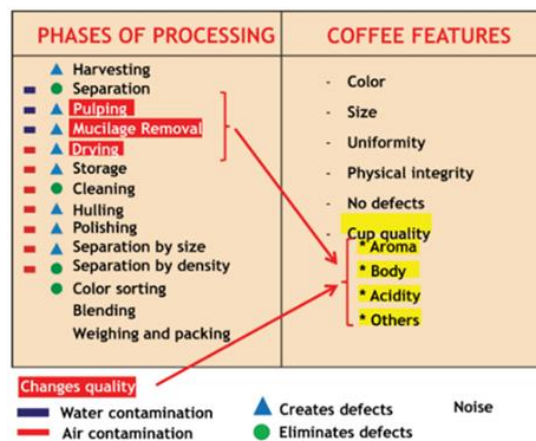


Figure 2. Defects and contamination during green coffee bean processing [16].

### B. Green Coffee Bean Quality Standard

Green coffee bean quality assessment is regulated and standardized by specialized coffee concentrate organizations such as the Specialty Coffee Association (SCA) with its SCA green coffee standards, government standards of coffee importing countries, or standards by large coffee producing countries such as Indonesia, which regulates levels of quality standards. coffee in the Indonesian National Standard (SNI) [17]. This study uses the SCA green coffee standard because it is used internationally in the coffee market [7], [18]. The quality of green coffee beans is determined by several aspects based on SCA standards. The first is the water content. The moisture content of the coffee beans should be 10-12% when imported to the buyer. Second, the grain size must be uniform with a cut tolerance of 5% of the contracted specifications. This is measured using a traditional perforated grading screen. In addition, it is necessary to have different

taste attributes in a cupping session. And the last one is the point of the seed defect which is calculated and below the specified point while the more significant defect point will be chosen as the main defect if there are two defects in one seed [19].

SCA grades green coffee beans into five grades within their standards: specialty grade, premium grade, exchange grade, under grade, and out grade. A special grade coffee requirement is that coffee has five full points of defects or less in 1 sample without category one defects. The screen size below and above the maximum is 5% and has a moisture content of 10-12%. Premium grade allows nine full handicap points with category one defects permitted. The maximum screen size and moisture content requirements are the same as for the special class. Exchange rate copy allows defect points up to 23 full defects, with all other requirements being the same as above standard. However, substandard grades and off grade grades only have a defect point value requirement, with grades below up to 86 full defects and off grade green coffee grades above 86 defective points [20].

### C. Image Acquisition

By reviewing the methods used in several studies, image acquisition is the first step in identifying coffee bean features. A study done by Arboleda et al [1], proposed a method for managing the quality of coffee beans using digital images. The acquired images of 180 Robusta coffee beans were taken using Sony DCS-800 20.1 Megapixels Camera digital camera. The coffee beans were placed on a white background (to contrast between coffee beans and the background), intentionally placed well spread so that the beans do not touch each other in order to ease the process of bean segmenting, and also improve the accuracy of morphology features extraction. Hendrawan [21] opted to use a black background rather than a white one. Given that there is constant fluorescent lighting. The use of digital cameras to capture images of coffee is not mandatory, others studies opted with different devices such as flatbed scanners [22], and smartphones [23].

### D. Color Feature Extraction

Color extraction is an important step to identify coffee bean quality levels. It is the most widely used method to retrieve images and index them [24]. The ability to represent visual content of images, uncomplicated in extracting color information, efficient in distinguishing on image to another, durable to background complications, and not dependent to the size and orientation of an image [25]. Basic colors that are combined together is called a color image. Arboleda et al collected the RGB values of a selected pixel using a function called `impixel` [1]. MATLAB<sup>®</sup> software was used to divide every individual pixel in a color image into individual Red, Green and Blue (RGB) values. They further determine the minimum surface area that a normal coffee bean would have, eliminate images that have lesser surface area, lastly use white color to show which are the normal coffee beans.

Hendrawan [21] converted RGB spaces to grey, Hue Saturation Lightness (HSL), Hue Saturation Value (HSV) and L\*a\*b\* colour spaces. It resulted with the color co-occurrence matrix (CCM) in each color group (Red(RGB), Green(RGB), Blue(RGB), grey, Hue, Saturation(HSL), Lightness(HSL), Saturation(HSV), Value(HSV), L\*, a\*, and b\*). Nasution [26] further extracted Hue Saturation Intensity (HSI) values from RGB images. They normalized RGB values to align with the HIS values (0-1). Lastly they conduct the thresholding process which is done by establishing a threshold value. RGB values that are equal to or more than the threshold value are converted to white (1), whereas RGB values that are less than the threshold value are converted to black (0). Figure 3 shows (a) the thresholding image and (b) RGB image on the thresholding (c) HSI image in the Thresholding image.

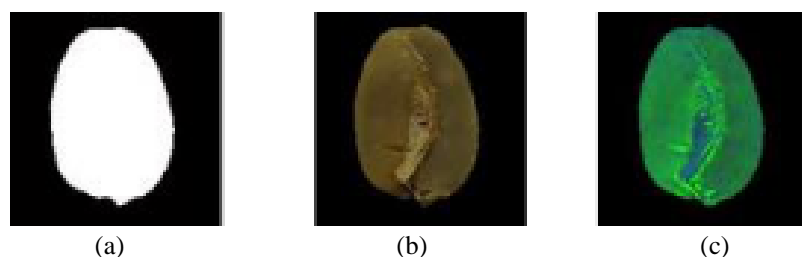


Figure 3. Image thresholding results [21].

**E. Identification of Features**

Arboleda et al extracted the parameter values needed in the testing process by doing training processes [1]. As many of 180 Robusta coffee beans used in the research were divided for different purposes. Training processes used 70 normal beans and 50 black beans, testing processes used 35 normal beans and 25 black beans. Training was intended to derive the RGB value range of normal beans and the black beans. The values obtained from the training procedure then were utilized in the testing procedure. The RGB values of normal beans were plotted as the higher and lower limits as values that are out of range were excluded. An accuracy of 100% was achieved in classifying and excluding the black beans in an image using the developed technique.

**F. Classification of Coffee Beans**

Pinto et al conducted a study to construct an automatic coffee bean sorting system for coffee bean growers in Timor-Leste [7]. Pictures of green coffee beans were taken using a digital camera. Pictures of both sides of the coffee bean were taken. A total of about 13,000 colored images were gained from 6,500 beans. All images were divided in different groups for different purposes such as training, validating and testing processes. The training data were utilized for neural networks learning process. The validation data was used to confirm the accuracy of classification in the learning part of neural networks. Testing data were used to evaluate the accomplishment of the neural networks sorting ability with final parameters. Images were labeled manually according to the type of defect a coffee bean has i.e. fade, black, sour, broken, peaberry, and no defect. These types of defects decrease the quality of a coffee bean, and type of defect has its own deduction point value.

Wang et al proposed a system which could intelligently inspect coffee bean quality using deep learning and computer vision [27]. They used deep neural network (DNN), knowledge distillation (KD), and residual neural network (ResNet) to achieve their goals. The image dataset was open-sourced, and consists of 4626 images of green coffee beans, where there were 2150 of good coffee images and 2476 of bad coffee images. The dataset then was further divided into a set of 4000 images for training purposes and 626 images for testing purposes.

Arboleda et al conducted a research with the objective to classify coffee bean species using image processing, ANN and K nearest neighbors (KNN) [6]. The coffee beans samples were from the National Coffee Research Development and Extension Center (NCRDEC), Cavite State University, Indang, Cavite. The beans were samples taken from distinct coffee producing towns in Cavite (Mendez, Amadeo, Silang and Indang, General Emilio Aguinaldo). Digital coffee bean sample images were shot using a 20.1 Megapixels Camera (Sony DCS-800). Coffee bean samples were placed on a white background. The MATLAB software was utilized to construct a computer routine pseudo code to preprocess and extract features of coffee samples images. Then the images were pre-processed, and an extraction of morphological features were done for classification purposes.

**G. Morphological Feature Extraction**

Arboleda [6] identified and classified coffee beans origins within Cavite using morphology features. Morphology is the geometric property of objects. Surface area, Roundness, Equivalent Diameter, and Perimeter are the features that were extracted. Their average values were used for classification and every property were calculated from their binary images (Table 2).

Table 1. Shape Features of Coffee Beans Samples [6].

Features	Robusta	Excelsa	Liberica
Surface Area	12,551 - 21,542	12,033 - 21,779	20,380 - 28,876
Perimeter	425.45 - 624.28	521 - 1756.10	549.29 - 1375.90
Equivalent Diameter	126.41 - 165.61	123.78 - 166.52	161.09 - 191.74
Roundness Percentage	22% - 28%	3% - 18%	6% - 27%

Liberica had the largest surface area between the three, while Robusta and Excelsa are almost same in size and average range. Perimeter wise, Excelsa has the largest while Robusta has the smallest. The highest equivalent diameter was Liberica. Robusta is the most spherical of the three species.

Wang et al [27] used saliency maps to segment important features of a coffee bean. Green spots in the saliency maps refer to the defective coffee bean parts, then the model computes the difference of pixel

values in the images between defective and good coffee beans. A linear regression model was also used to compute the coffee bean features with the highest positive weight which are later on used to classify coffee beans. The process flow can be seen in Figure 4.

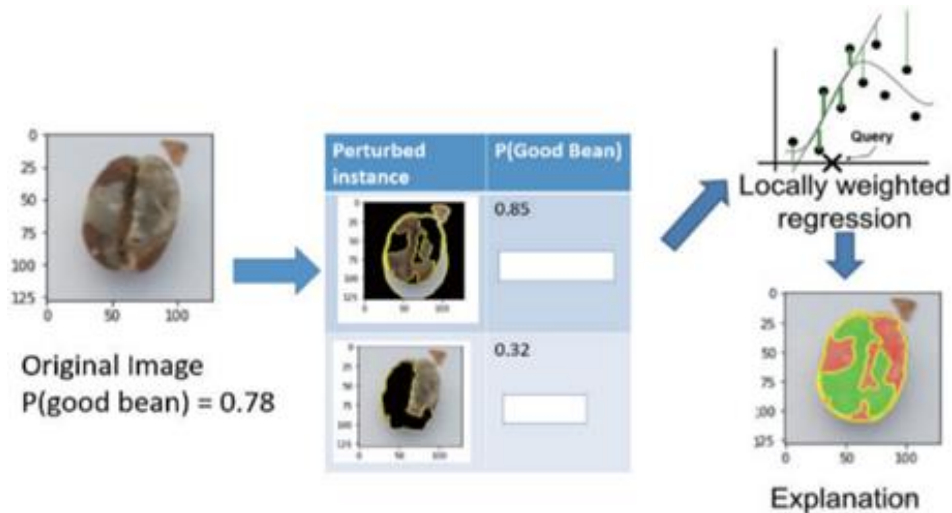


Figure 4. Computing local weights of coffee beans images [27].

#### H. Classification Process

Pinto et al [7] used a classifier based on Convolutional neural network (CNN). The convolution layer is intended to extract features from images using spatial filters, the pooling layer intends to decrease the position sensitivity of the extracted features from the convolution layer, so even if the target feature moves slightly, the output of the pooling layer would not be affected. Coffee beans were classified Using CNN into two distinct classes. The accuracy of classification process was derived by dividing the total data correctly classified by the total of test data.

Results show that the accuracy of classifying for the Black bean was the highest (98.75%) and 2<sup>nd</sup> place was the Sour bean (92.93%). The lowest accuracy was held by the Broken bean (gray image) (67.50%). These results indicate that the developed CNN model constructed showed great performance in extracting features for those labels (Black Bean and Sour Bean). This was due to the typical color that were easy to identify. Palebean (color) and Distortedbean (color) have 72.41% and 72.50%, this due to their typical color that varies between semi-transparent and amber to yellow or crystallized, not so easy to identify. Generally, CNN has advantages for image feature identification spacial filters. Although, in the classifying process, the color characteristics of coffee bean has strong influences.

Arboleda [6] used a classifier based on ANNs and KNN. They used two models to compare which results the higher accuracy and explain the factors. Research results using ANN indicated that the proposed algorithm is feasible achieving accuracy greater than 96.67% on all coffee species. Using KNN the accuracy of 82.56% was achieved when the coffee species were classified. ANN performed better in classifying coffee beans over the KNN method using the output dataset. This is because the morphological features of excelsa and robusta are quite resemblant so that they are very similar. The results shows that there is high probability that robusta is classified as excelsa and vice versa using the KNN classifier.

The local explainable model-agnostic explanation method (LIME) was used by Wang et al [27]. LIME classifies coffee beans by converting pixels with green colors into an explainable model by using the quality analysis of green beans produced by linear regression in the previous step. The classification focused on the shape, and colors of beans. The model can classify sour and moldy beans, also crushed beans with broken shapes. The accuracy of the model reached 95% using ResNet optimized by the adaptive moment estimation (ADAM), 93% for ResNet optimized by the stochastic gradient descent (SGD) model, and lastly ResNet with KD reached 91%. The model they proposed were lightweight and could accurately classify the quality of beans, which also allow interpretation and further understanding of image features to make information more open and transparent.



## CONCLUSION

Computer Vision has been used in many research in order to extract information from coffee beans. Mainly the goal of these studies are to identify features of coffee beans and classify them into an intended class. Varying methods were used to achieve these goals. Different computer intelligence and image processing techniques were used. These methods and techniques have already been provided in this paper sequentially. We believe that further research will improve the quality in identifying and classifying coffee beans. We hope that this paper could be a sufficient guide for further researchers to conduct studies on this topic for various reasons and in various fields.

## REFERENCES

- [1] E. R. Arboleda, A. C. Fajardo, and R. P. Medina, "An image processing technique for coffee black beans identification," in *2018 IEEE International Conference on Innovative Research and Development, ICIRD 2018*, 2018.
- [2] C. Yeretian, "Coffee," in *Springer Handbooks*, 2017.
- [3] E. M. De Oliveira, D. S. Leme, B. H. G. Barbosa, M. P. Rodarte, and R. G. F. Alvarenga Pereira, "A computer vision system for coffee beans classification based on computational intelligence techniques," *J. Food Eng.*, 2016.
- [4] D. S. Leme, S. A. da Silva, B. H. G. Barbosa, F. M. Borém, and R. G. F. A. Pereira, "Recognition of coffee roasting degree using a computer vision system," *Comput. Electron. Agric.*, 2019.
- [5] T. H. Nasution and U. Andayani, "Recognition of roasted coffee bean levels using image processing and neural network," in *IOP Conference Series: Materials Science and Engineering*, 2017.
- [6] 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," in *2018 IEEE International Conference on Innovative Research and Development, ICIRD 2018*, 2018.
- [7] 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*, 2017.
- [8] U. Lehtinen, "Traceability in Agrifood Chains," *Intell. Agrifood Chain. Networks*, pp. 151–166, 2011.
- [9] S. Khan, A. Haleem, M. I. Khan, M. H. Abidi, and A. Al-Ahmari, "Implementing traceability systems in specific supply chain management (SCM) through critical success factors (CSFs)," *Sustain.*, 2018.
- [10] M. M. Aung and Y. S. Chang, "Traceability in a food supply chain: Safety and quality perspectives," *Food Control*, vol. 39, no. 1, pp. 172–184, 2014.
- [11] B. Kitchenham and P. Brereton, "A systematic review of systematic review process research in software engineering," *Information and Software Technology*. 2013.
- [12] J. Cruz-Benito, "Systematic Literature Review & Mapping," 2016, p. 67.
- [13] H. Steinhart, "Coffee: growing, processing, sustainable production—a guidebook for growers, processors, traders, and researchers. Jean Nicolas Wintgens (Ed). Wiley-VCH Verlag, Weinheim, 2004. 976 pp, ISBN 3-527-30731-1," *J. Sci. Food Agric.*, vol. 85, no. 11, 2005.
- [14] S. Schenker and T. Rothgeb, "The roast—Creating the Beans' signature," *Cr. Sci. coffee*, 2017.
- [15] C. Lambot, J. C. Herrera, B. Bertrand, S. Sadeghian, P. Benavides, and A. Gaitán, "Cultivating Coffee Quality-Terroir and Agro-Ecosystem," in *The Craft and Science of Coffee*, 2017.
- [16] M. D. del Castillo, B. Fernandez-Gomez, N. Martinez-Saez, A. Iriondo, and M. D. Mesa, *Coffee By-Products*. 2018.
- [17] M. Castillo, B. Fernandez-Gomez, N. M. Sáez, and ..., *Coffee by-products*. digital.csic.es, 2019.
- [18] Y. C. Chou *et al.*, "Deep-learning-based defective bean inspection with GAN-structured automated labeled data augmentation in coffee industry," *Appl. Sci.*, vol. 9, no. 19, 2019.
- [19] J. Kosalos, R. Stephen, S. Diaz, P. Songer, and M. Alves, "SCAA Arabica Green Coffee Defect Handbook," *Spec. Coffee Assoc. Am. SCAA*, 2006.
- [20] C. R. Institute, "SCAA Coffee Beans Classification," 2006. [Online]. Available: <http://www.coffeeresearch.org/coffee/scaaclass.htm>. [Accessed: 22-Feb-2022].
- [21] Y. Hendrawan, S. Widyaningtyas, and S. Sucipto, "Computer vision for purity, phenol, and pH detection of Luwak coffee green bean," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 17, no. 6, pp. 3073–3085, 2019.
- [22] F. F. L. Dos Santos, J. T. F. Rosas, R. N. Martins, G. de M. Araújo, L. de A. Viana, and J. de P. Gonçalves, "Quality assessment of coffee beans through computer vision and machine learning algorithms," *Coffee Sci.*, vol. 15, no. 1, pp. 1–9, 2020.

- [23] J. N. C. Sarino, M. M. Bayas, E. R. Arboleda, E. C. Guevarra, and R. M. Dellosa, "Classification of coffee bean degree of roast using image processing and neural network," *Int. J. Sci. Technol. Res.*, vol. 8, no. 10, pp. 3231–3233, 2019.
- [24] D. S. Kalel, P. M. Pisal, and R. P. Bagawade, "A Study of Color, Shape and Texture Feature Extraction for Content Based Image Retrieval System," *Int. J. Adv. Res. Comput. Commun.*, 2016.
- [25] F. Alamdar and M. R. Keyvanpour, "A new color feature extraction method based on quad histogram," in *Procedia Environmental Sciences*, 2011.
- [26] T. H. Nasution, M. Rumansa, and H. Lukman Adlin, "Designing the quality of coffee bean detection application using Hue Saturation Intensity," in *IOP Conference Series: Materials Science and Engineering*, 2019, vol. 648, no. 1.
- [27] P. Wang, H. W. Tseng, T. C. Chen, and C. H. Hsia, "Deep convolutional neural network for coffee bean inspection," *Sensors Mater.*, vol. 33, no. 7, pp. 2299–2310, 2021.