Preprints.org

Machine Learning Techniques for Coffee Classification: A Comprehensive Review of Scientific Research

Isabela V. de C. Motta [*] , Nicolas Vuillerme , Huy-Hieu Pham , Felipe Augusto Pereira de Figueiredo

Article

Machine Learning Techniques for Coffee Classification: A Comprehensive Review of Scientific Research

Isabela V. de C. Motta [1],*, Nicolas Vuillerme [2], Huy-Hieu Pham [3], Felipe A. P. de Figueiredo [1]

[1]   National Institute of Telecommunications, Av. João de Camargo, 510, Santa Rita do Sapucaí 37540-000, Minas Gerais, Brazil; felipe.figueiredo@inatel.br

[2]   AGEIS, Université Grenoble Alpes, Grenoble 38000, France; nicolas.vuillerme@univ-grenoble-alpes.fr

[3]   College of Engineering & Computer Science and VinUni-Illinois Smart Health Center, VinUniversity, Hanoi,Vietnam; hieu.ph@vinuni.edu.vn

*   Correspondence: isabela.motta@dtel.inatel.br

Abstract: In the realm of agribusiness, transformative shifts are underway, propelled by the growing demands and expanding scales of grain production. This evolution calls for a critical reevaluation of the existing paradigms in coffee production and marketing paradigms, with a specific focus on integrating Artificial Intelligence (AI). This work aims to review, synthesize, and summarize the available data regarding how Machine Learning (ML) has been used to detect and classify characteristics in coffee beans and leaves. For this purpose, a comprehensive literature review of the most significant research contributions describing the application of AI for advanced classification techniques in coffee agriculture has been carried out. Our analysis suggests that implementing AI technologies allows the classification of coffee, encompassing various attributes such as maturity, roast intensity, disease identification, flavor profiles, and overall quality. More largely, this technological advancement holds the potential to revolutionize coffee farming by providing producers and agricultural specialists with sophisticated tools to enhance production efficiency, minimize costs, and improve the accuracy and confidence of their decision-making processes. This article allows us to learn about the latest studies in ML in the coffee area, observe the methodology used, and allow researchers to develop new research that covers gaps in the literature, bringing a real contribution to the scientific field. Furthermore, this article listed the databases used in the studies and which may be useful for future ML projects.

Keywords: artificial intelligence; machine-learning; computer vision; coffee bean and leaves classification

1. Introduction

Coffee stands as a cornerstone crop globally and its consumption has been increasing in recent years. [1]. This surge underscores coffee's pivotal role in the global market, given its status as a crucial economic food commodity [2].

Current methodologies for managing agricultural data in coffee farming are increasingly considered outdated—labor-intensive, time-consuming, and, most importantly, prone to inaccuracies and errors. In this context, the advent of Machine Learning (ML) and intelligent data analytics emerges as a transformative force, propelling sustainable agriculture by enhancing food production and addressing the pressing challenges in this sector [3].

Although a considerable volume of work addresses the use of Artificial Intelligence (AI) in coffee culture, it should be noted that research has followed different paths, depending on its various aspects. There is a considerable volume of work on the use of AI in agriculture, especially in the category of predictive analysis [4]. Indeed, there are numerous areas in coffee production where ML techniques can be applied. Intelligent systems can provide advances and improvements, from grain harvesting to commercialization. In a scenario where specialized systems surpass human diagnoses in terms of trust, speed, and accuracy, it will be of fundamental importance for coffee growers to make their use widely applicable and essential to increase productivity in agriculture.

Therefore, the present work aims to review and summarize the available data regarding how ML has been used in coffee production and marketing in classification activities, which encompass the classification of defects, roasting, maturation, and sensory characteristics of coffee beans and also leaf

diseases that affect the coffee plantation. This research is important and justified by understanding the relationship between AI and coffee production, identifying trends in AI applications in coffee classification, and, mainly, identifying research gaps that can be eliminated to improve processes throughout the coffee production chain.

Considering that global coffee consumption is increasing, and its importance of participation in revenue in more than 70 developing countries economies [1], combined with the capacity of AI techniques for problem-solving and decision-making, this paper presents a detailed summary of research covering coffee and AI.

Table 1 compares our literature and related literature reviews. The first column corresponds to the reference of the literature review; the second column describes the paper; the third column lists the number of papers analyzed in the work; and the last column informs the time range of the analyzed works. The review's authors [5] analyzed seven studies that classify the quality and type of coffee beans based on images and using ML. The reviews [6–8] analyze research works and list challenges of disease detection of various plants. The literature review [6] focuses on leaf diseases. The study of [9] analyzes technologies for detecting roast-level coffee beans, but not all studies use ML techniques. Many studies related to smart agriculture [10,11] address machine learning techniques for classifications of other types of crops, such as rice [12–15], corn [16,17] and soybean [18].

Table 1. Related Literature Reviews

Reference	Brief Description	Quantity	Year
[5]	Machine learning techniques for classifying the quality and types of coffee beans on image-based.	7	2018-2022
[6]	Machine learning techniques for detecting leaf diseases in various plants.	118	2010-2022
[7]	Machine learning techniques for detecting diseases in various plants.	160	2017-2022
[9]	Techniques for the detection of Coffee Bean Roasting Levels.	31	2014-2022
[8]	Machine learning techniques for detecting disease in various plants.	135	2017-2023

A Comprehensive Review was used to achieve the proposed objective of synthesizing and understanding how ML techniques for coffee classification are presented in scientific research. The review was carried out in the IEEE, Science Direct, and Springer databases, with terms compatible with the object of this research.

The main goal of this review is to survey current research in the coffee area that uses machine and deep learning techniques. This research lists the works related to the classification of coffee beans and leaves in a comprehensive way. In this context, our main contributions include:

- A more up-to-date and comprehensive literature review of current ML research in the classification and detection of coffee beans and coffee leaves, covering 72 papers and different ML algorithms, including comparisons and discussions among them;
- An analysis of the evaluation metrics used in ML for classification and detection of coffee beans and coffee leaves for understanding results and reproducing them, benefiting future research;
- A presentation of architectures used based on ML for detecting and classifying coffee beans and coffee leaves;
- A detailed exploration of the main limitations, challenges, and future directions related to the use of ML techniques for the classification and detection of coffee beans and coffee leaves;
- A summary of the databases used in research to detect and classify coffee beans and coffee leaves.

This work was divided into five more sections to contextualize the use of ML in coffee farming, specifically its use on coffee beans and leaves. Following this introduction, the second section describes

the research strategy and details the important points of each selected study. The third section introduces the main ML techniques used for classification tasks. The fourth section summarizes and analyzes the findings, presents the importance of coffee classification in the coffee industry, and demonstrates how agribusiness uses ML techniques. The fifth section discusses the studies' challenges, future trends, and limitations and summarizes the dataset for classification tasks. Finally, the sixth section presents the final considerations.

2. Methodology

This paper is a descriptive research of the comprehensive review type carried out through online access to the IEEE, Science Direct, and Springer databases. In the browsing for research articles, the descriptors "machine learning" and "coffee." For the inclusion of works in the present study, the following criteria were established: articles published in the last five years before the consultation, articles in which the topic of "machine learning" is related to coffee classification techniques, and articles available in full format in the databases above. Theoretical studies and systematic literature reviews were not included in the overview of synthesized research data.

The article analysis process followed these steps: 1) search for descriptors in the databases above; 2) exclusion of articles published more than five years ago; 3) exclusion of works with low relevance; 4) critical reading of the articles and checking whether they meet the proposed theme; 5) exclusion of theoretical articles and literature reviews; 6) final analysis. After selecting the articles using inclusion and exclusion criteria, tables containing the main information about the objectives and scenario of application, the methodology used for coffee classification analysis, and the results were created.

In the IEEE database, 114 articles were found with the words "machine learning" and "coffee". Fifty-eight (58) articles were disregarded because they were unrelated to the coffee area; for example, the author's last name is Coffee or the three searched words do not appear simultaneously in the text. Only "machine learning", or only "coffee". Ten (10) articles were disregarded because they were not related to classifying coffee beans, such as Predicting Coffee Crop Yield, detecting coffee trees, coffee price prediction, and agriculture 4.0. Finally, two (2) articles were disregarded in the analysis because it was a literature review. In the same way, in the Science database, 2,828 articles were found; 2,770 were disregarded because it was not related to the general theme; 34 were disregarded because it was not related to the specific theme; and two (2) articles were disregarded in the analysis because it was a literature review. In summary, 7,630 articles were found in the Springer database, but 7,617 were disregarded; 6 were disregarded because they were unrelated to the specific theme; and 1 article was disregarded in the analysis because it was a literature review. Figure 1 shows the research methodology used in this study. Figure 1 shows the total number of publications per database and the steps used to choose the selected publications.
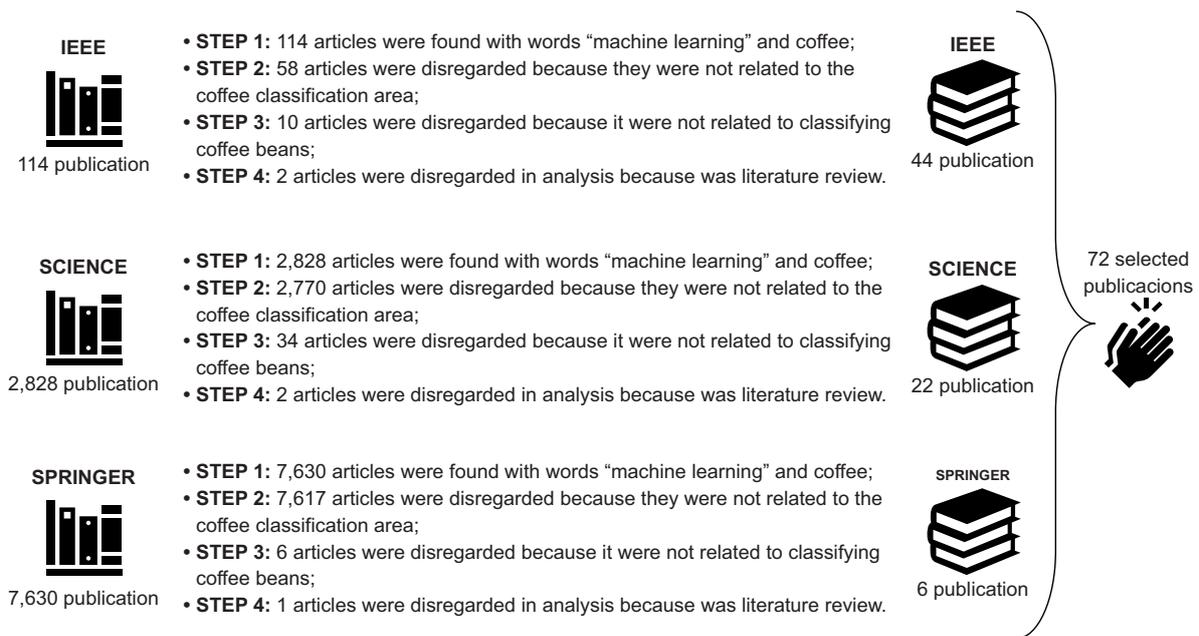
IEEE

114 publication

- **STEP 1:** 114 articles were found with words "machine learning" and coffee;
- **STEP 2:** 58 articles were disregarded because they were not related to the coffee classification area;
- **STEP 3:** 10 articles were disregarded because it were not related to classifying coffee beans;
- **STEP 4:** 2 articles were disregarded in analysis because was literature review.

IEEE

44 publication

SCIENCE

2,828 publication

- **STEP 1:** 2,828 articles were found with words "machine learning" and coffee;
- **STEP 2:** 2,770 articles were disregarded because they were not related to the coffee classification area;
- **STEP 3:** 34 articles were disregarded because it were not related to classifying coffee beans;
- **STEP 4:** 2 articles were disregarded in analysis because was literature review.

SCIENCE

22 publication

SPRINGER

7,630 publication

- **STEP 1:** 7,630 articles were found with words "machine learning" and coffee;
- **STEP 2:** 7,617 articles were disregarded because they were not related to the coffee classification area;
- **STEP 3:** 6 articles were disregarded because it were not related to classifying coffee beans;
- **STEP 4:** 1 articles were disregarded in analysis because was literature review.

SPRINGER

6 publication

72 selected publicacions

Figure 1. Research Methodology

In Figure 2, we can see a fluctuation in the number of publications over the last five (5) years, which various factors may have influenced. In general, the number of publications has increased over the years.
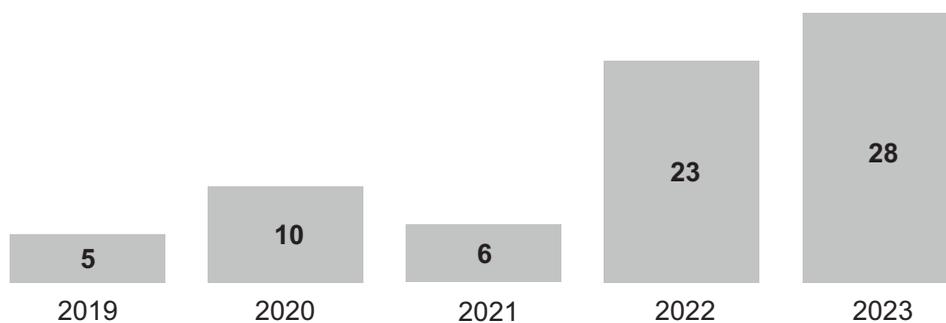
Publication Year

Figure 2. Number of Articles Selected

Additionally, Table 2 below shows the number of publications per database searched. As illustrated, the first column corresponds to the database, and the second column describes the number of publications.

Table 2. Number of Publication per Database.

Database	Number of Publication
IEEE	44
Science	22
Springer	6
TOTAL	72

2.1. Research Questions

This review was designed to answer twelve Research Questions (RQ). The RQ are as follows.

RQ1: What are the most used architectures based on ML for detecting and classifying defects in coffee beans?

RQ2: What are the most used architectures based on ML for detecting and classifying coffee beans roasting?

RQ3: What are the most used architectures based on ML for the sensory classification of coffee?

RQ4: What are the most used architectures based on ML for detecting and classifying the maturity of coffee beans?

RQ5: What are the most used architectures based on ML for detecting and classifying coffee diseases?

RQ6: What defects are considered when classifying the coffee bean in the studies?

RQ7: What stages are considered to classify coffee roasting in the works studied?

RQ8: What stages are considered to classify coffee maturity in the works studied?

RQ9: What are the different coffee diseases addressed in the works studied?

RQ10: What challenges exist, and how can research be improved in detecting and classifying coffee beans and leaves based on the studies reviewed?

RQ11: How diverse is the research on coffee bean and coffee leaves detection or classification, and for what classification purpose are ML applications the most?

RQ12: Which databases are available for research into detecting and classifying coffee beans and coffee leaves?

The questions RQ1, RQ2, RQ3, RQ4, RQ5, RQ6, RQ7, RQ8, and RQ9 aim to map the most used architectures for ML in coffee bean classification and, consequently, classify commonly used ML approaches to resolve them. Based on this review, these five questions identify the main ML algorithms and their main metrics used to achieve the classification objective. In these questions, it is possible to carry out a quantitative and qualitative analysis of the more effective ML algorithms to be applied to coffee bean classification. Through mapping, it is possible to identify how researchers are addressing the techniques that involve ML and consequently analyzing potential gaps related to the topic.

By answering RQ10, RQ11, and RQ12, we can understand the challenges, future trends, and final observations when implementing ML techniques in various coffee bean and coffee leaves classification scenarios.

Finally, RQ1 and RQ6 are answered in Subsection 4.1, RQ2 and RQ7 in Subsection 4.2, RQ3 in Subsection 4.3, RQ4 and RQ8 in Subsection 4.4, RQ5 and RQ9 in Subsection 4.5, and RQ10, RQ11 and RQ12 in Section 5.

3. Machine Learning Techniques for Coffee Classification

ML is an area of AI capable of generalizing a problem from a set of training data to correctly classify or predict data that has not been previously observed [19]. It is a tool that can be applied to any field that needs data analysis.

There are three ML techniques: unsupervised, supervised, and reinforcement learning. A training set is observed in supervised learning and comprises examples of input and output pairs. The machine learns a function that maps the input to the output, and this function is called the hypothesis function [20]. The supervised learning problem can be classified when the output value has finite values or regression when the output value has infinite values. The idea of classification involves grouping items into classes or groups. The difference between supervised and unsupervised learning can be seen in Figures 3 and 4 below.
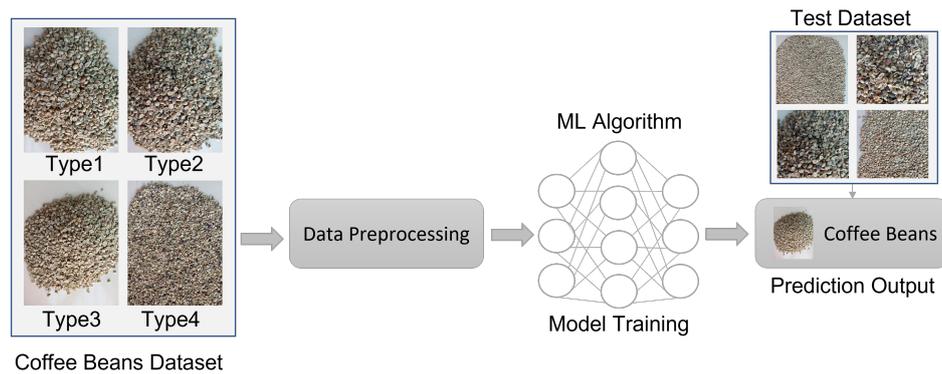
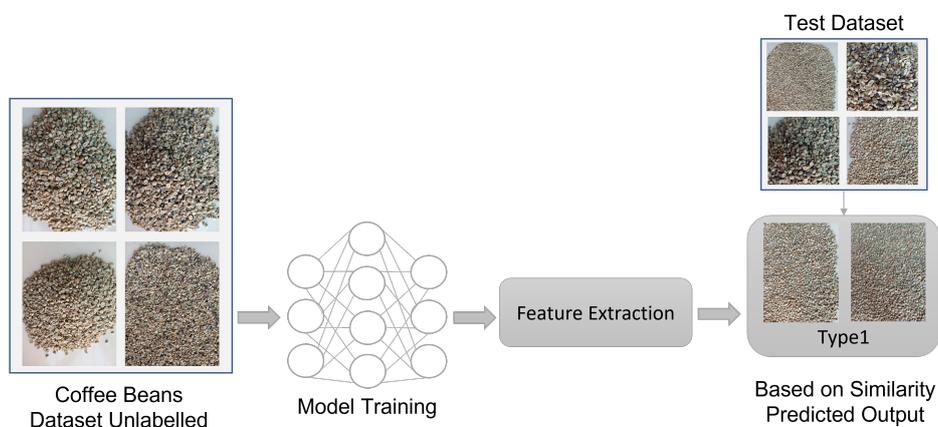**Figure 3.** Supervised Learning. Adapted from [21]

Figure 4. Unsupervised Learning. Adapted from [21]

Computer vision is an area of AI that seeks to analyze, interpret, and extract relevant information from images so that decisions can be made. Object recognition in its complete generality is difficult [20]. Convolutional Neural Network (CNN) is a ML model that extracts features in various computer vision tasks like image classifications. Several CNN architecture models, such as AlexNet, VGGNet, ResNet, MobileNet, EfficientNet, and DenseNet, are used for image classifications. The Figure 5 shows a typical CNN.
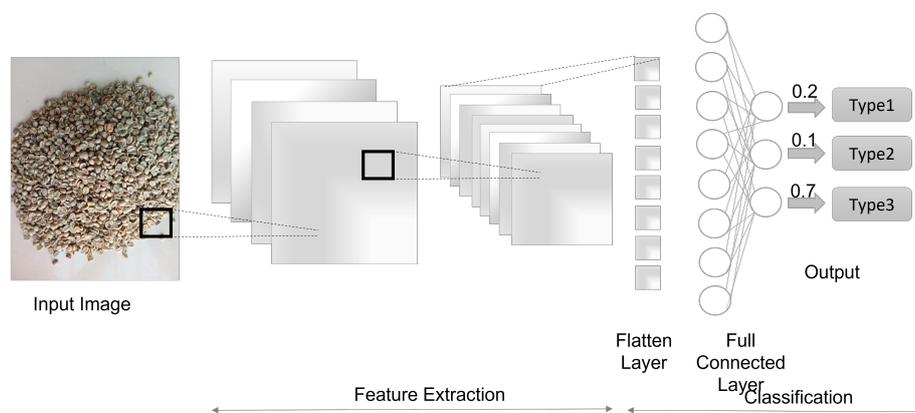

Figure 5. A Typical CNN. Adapted from []

Vision Transformer is a neural network with the potential for extensive use in image classification tasks [22]. The Figure 6 shows a illustration of the Vision Transformer Model.
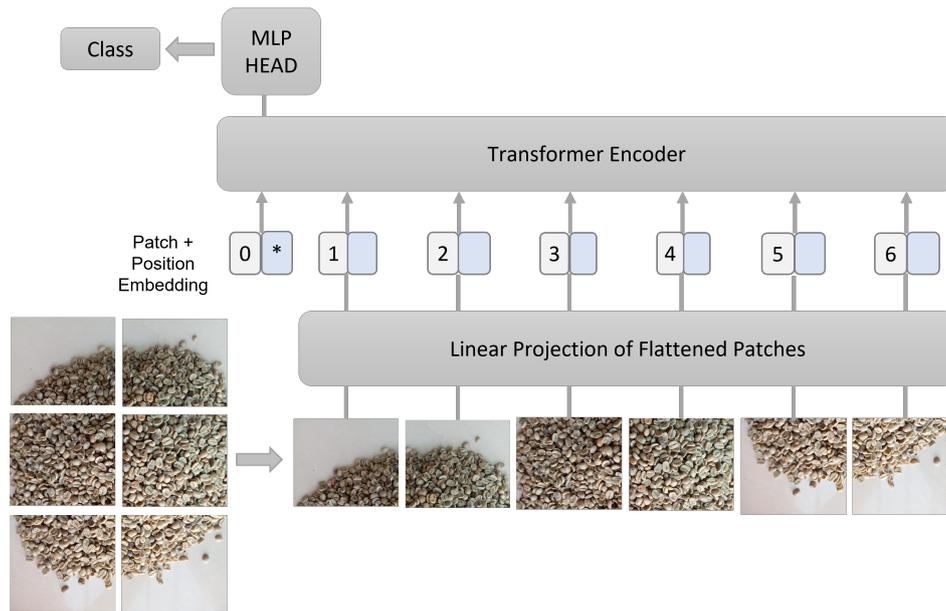
Figure 6. Vision Transformer Model. Adapted from [23]

ML techniques combined with computer vision have become decisive in developing agricultural efficiency and have been widely applied in research in various areas of agriculture [24], including coffee farming. Due to different algorithms, this research seeks to identify which models achieve greater accuracy in image classification.

4. Overview of Synthesized Research Data

This section summarizes the most significant studies on ML in coffee classification. It details each study, including the reference (authors and publication year), objectives and application scenarios, methodology (architectures and dataset), and main results.

We identified 72 selected studies that could be grouped into 5 common themes: classification of (1) coffee defects, (2) coffee roasting, (3) coffee aroma and flavor, (4) coffee maturity, and (5) coffee diseases. The 5 categories are listed and commented below.

4.1. Coffee Defects Classification

Defects and impurities in coffee beans reduce their quality. Defective beans affect the flavor of the coffee and, therefore, devalue it and affect the producer's profitability. Several factors can cause these defects, such as problems in production and storage and physiological and genetic changes. Identifying defects makes it possible to improve management and prevent the occurrence of these defects, which depreciate the coffee.

As illustrated in Tables 3 and 4, the first column corresponds to the reference of the literature review; the second column describes the main objectives of the paper; the third column lists the models and algorithms used in the paper; and the last column informs the results achieved.

Table 3. Coffee Defects Classification.

Reference	Objectives and Scenario of Application	Methodology	Results
[25]	The study created an inspection machine to classify defects in coffee beans.	Three deep learning models (Enhanced, ResNet-50, and AlexNet) are used to analyze the images.	The most efficient model was ResNet-50, with an accuracy of 93.33%.
[26]	The study presents a system for inspecting coffee beans, identifying good and bad ones.	The model uses DenseNet201 architecture.	The proposed system achieved 98.97%.
[27]	The study explores automated detection of the quality of coffee beans based on their color and texture characteristics.	The model uses a Support Vector Machine (SVM), Deep Neural Network (DNN), and Random Forest (RF).	The method was evaluated using cross-validation with K-Fold of 5 and 10. The highest accuracy value achieved was 96.11%.
[28]	The study proposes a scheme for the automated inspection of defects in dense Arabica coffee beans.	The model is based on a generative adversarial network (GAN) and can generate synthetic training images with defects at multiple locations.	More than 90% of the labeling time is done by the proposed prototype, which uses less time than a human for the same task.
[29]	The study presents a new dataset containing 8,000 images of green Arabica coffee beans divided into 4 classes: Peaberry, Longberry, Premium, and Defect.	The model was implemented with two CNN architectures: ResNet-18 and MobileNetV2.	The final average test accuracy was 81.12% for ResNet-18 and 81.31% for MobileNetV2.
[30]	The study categorizes coffee beans into seven classes: sour, black, broken, moldy, shell, insect damage, and good beans.	The proposed CNN model has 12 layers. The dataset has about 1,700 images in each folder.	The proposed model achieved an accuracy rate greater than 90% for all categories except shell beans (88%).
[31]	The study classifies images of coffee beans as defective or normal.	The proposed CNN model has three convolution layers with three max-pooling layers and three fully connected layers. A dataset of 1,813 images and two classes was used.	The model achieved an accuracy of 90.44%.
[32]	The study proposes a system called Hough circle assisting deep-network inspection scheme (HCADIS) which identifies defects in coffee beans.	The proposed method uses the Hough Circle Transform algorithm, an image processing method that detects circles in an image.	The proposed scheme achieved half of the testing images greater than 80% in defect inspection accuracy values.
[33]	The study classifies defects in coffee beans into four classes: black, broken, holey, and normal.	The proposed method uses the Mask R-CNN (Region-based CNN) algorithm. A dataset of 480 images with two forms of images was used: images with individual coffee beans and images with multiple objects.	The proposed method obtained an accuracy of 93.3% for tests with individual objects and 75% for tests with multiple objects.

Table 4. Coffee Defects Classification.

Reference	Objectives and Scenario of Application	Methodology	Results
[34]	The study presents a method to detect defects in green coffee beans in four categories: healthy, black, insects damaged, and shell.	Spectral and spatial characteristics were extracted from the images, and ML algorithms were applied to classify the coffee beans.	The model presented an overall accuracy of 98.6% using SVM with dimensionality reduction and band selection.
[35]	The study classifies green coffee beans as special or traditional.	Coffee samples were collected from different regions of Brazil, and a multispectral camera captured images of the beans at different wavelengths. Four ML models were used: SVM, K-Nearest Neighbors (KNN), RF, and Multilayer Perceptron (MLP).	The SVM model obtained the best performance, with an accuracy of 97.5%, followed by MLP with 96.9%, RF with 95.6%, and KNN with 94.4%.
[36]	The study classifies two types of coffee bad beans: insect bite and broken.	The model was based on NFNet-F3, which combines semi-supervised learning.	The proposed model achieved an F1-score of 97.21% and a precision of 97.38%.
[37]	The study classifies three types of coffee bean species: espresso, Kenya, and Starbucks Pike Place.	The transfer learning models used were: SqueezeNet, Inception V3, VGG16, and VGG19. A dataset of 1,554 images was used.	The best model was Squeezenet, with an average classification success of 87.3%.
[38]	The study identifies Arabica and Robusta coffee types through the images of the leaves.	The proposed model used four convolution layers, four pooling layers, and a fully connected layer. It was compared with LeNet, AlexNet, ResNet-50, and GoogleNet. A dataset of 19,980 images was used.	The CNN developed achieved an accuracy rate of 97.67% better than other nets.
[39]	The study presents a system that classifies different kinds of coffee beans: Normal beans, Peaberries, insect-infested, black beans, shell beans, and sour beans.	The proposed model adjusted a Slim-CNN using the least parameters. A dataset of 5,435 images was used.	The lightweight model achieved an accuracy of 92%.
[40]	The study presents a new dataset that classifies 17 different defects of coffee beans.	Two CNN architectures were tested: MobileNet and InceptionResnetV2. Two datasets were used.	with InceptionResnetV2, the classification for 17 classes achieved an accuracy of 53.35%, and for 3 classes, it achieved 92.52%.
[41]	The study classifies palm civet coffee beans based on their mass spectra.	The neural network model has an input layer with 301 neurons, and the hidden layer has 50 neurons.	The proposed model achieved an accuracy of 99.58%.
[42]	The study classifies good-quality raw coffee beans and subpar-quality raw coffee beans	The model compares VGG16 and Inception V3. A dataset of 100 images was used.	The best model was Inception, which achieved an accuracy of 99%.
[43]	The study classifies 4 Arabica coffee bean varieties: Aceh Gayo, Bali Kintamani, Lintong, and Surya Sabana Selo.	Five classification methods was tested: Artificial Neural Network (ANN), Decision Tree (DT), KNN, Naive Bayes, and SVM. A dataset of 400 images was used.	The best model was ANN, wich achieved an accuracy of 99.75%.
[]			
[]			
[]			
[]			
[]			

Most coffee defect classification research consists of categorizing coffee according to effective beans of an intrinsic nature (imperfect beans), which are beans damaged by the imperfect application of agricultural processes such as improper drying or picking of overripe cherries [44]. These coffee beans are black, broken, and damaged by insects, including the coffee borer beetle.

The challenge in future research in defect classification is to cover all defects originating from coffee cultivation. Few studies have identified external defects in the coffee bean, such as the presence of sticks, stones, clods, bark, and small insects. These defects originate during harvesting due to coffee falling on the ground and poor fanning. In the same way, few studies have identified acidic coffees derived from prolonged fermentation.

The paper [30] classified 6 types of coffee intrinsic and extrinsic defects. The authors created a screening system with a user interface. The authors of [34] created a high-speed near-infrared (NIR) camera to capture images of coffee beans moving on a conveyor belt. [25] also proposed a conveyor for capturing and uploading images.

[26] used an automated system to classify good and bad coffee beans. The bad coffee beans are divided into three categories: broken, insect-infested, and mold. The system recognizes the coffee beans; if all classification results are negative, the coffee bean is considered good and kept. If at least one output indicates a defect, the coffee bean is labeled with the corresponding type of defect. [28] classified all defects categorized by the Specialty Coffee Association of America (SCAA).

Meanwhile, [27] developed a Python algorithm to extract texture, shape, and color characteristics from images of coffee beans. Specialty coffees go through a more laborious production process, require special care from planting to roasting, and are made of pure, unmixed, high-quality coffees. At the same time, traditional coffee has inferior and defective beans. This mixture reduces the quality of the coffee as it changes its flavor and reduces its cost. [35] classified coffee beans in different regions of Brazil as special or traditional. The papers [38] and [37] classify coffee bean species. [38] classified arabica and robusta coffee plants, and [37] classified espresso, Kenya, and Starbucks pike place species.

The networks tested for classifying coffee defects were Resnet-50, Densenet201, SVM, DNN, RF, GAN, MobilenetV2, Hough Circle Transformer, Mask R-CNN, NFNet-F3, SqueezeNet, Slim-CNN, InceptionResnetV2, ANN , InceptionV3 and CNN networks proposed by the researches authors.

Enhanced, AlexNet, ResNet-18, KNN, MLP, VGG-16, VGG-19, MobileNet were also tested, but obtained inferior results when compared to other networks.

The highest accuracy achieved was 99.58% for the ANN network proposed by the work (tsai, 2023). Tsai's study combines ANN with Mass Spectrometry and analyzes 2 classes of single beans: civet coffee beans and regular beans. It is not possible to say which network model presents the best performance among all the researches analyzed because the databases tested were different. The works differ in the defects they seek to classify and consequently in the number of classes presented in each model. The researches used datasets from 100 images to around 20,000 images. The size of the dataset also influences the model result.

4.2. Coffee Roast Classification

Coffee roasting is one of the final stages of production and a determining factor in producing a good drink. Good roasting highlights the characteristics of each type of bean, which are flavor, aroma, and acidity. The same coffee bean may have different characteristics depending on the type of roast. The color of the coffee bean and the aroma can differentiate roast levels. The coffee beans change color as the temperature increases, and the roast can be light, medium, or dark. Each of them will have different flavor characteristics.

Roasting is defined by the time the coffee bean remains at a high temperature under the careful supervision of a specialized professional. Work that classifies the roasting level and analyzes the process until the coffee reaches the final carbonization stage contributes to automating the process without professional supervision.

As illustrated in Table 5, the first column corresponds to the reference of the literature review; the second column describes the paper's main objectives; the third column lists the models and algorithms used in the paper; and the last column informs the results achieved.

Table 5. Coffee Roast Classification.

Reference	Objectives and Scenario of Application	Methodology	Results
[45]	The study classifies Arabica coffee beans based on their light, medium, and dark roast levels.	The work proposes two main procedures: feature extraction and classification. Four kernel types are used in the SVM method for classification: linear, polynomial, radial basis, and sigmoidal.	The polynomial kernel achieved a maximum accuracy of 100% using k-fold values of 5 and 10.
[46]	The study presents a method to classify the quality of coffee beans based on their roast level: good, medium, and bad.	A dataset of 160 images was used. The proposed model was tested with ResNet-152 and VGG16.	The best model was ResNet-152, wich achieved an accuracy of 73.3%.
[47]	The study presents a coffee roasting process to classify three classes of roast coffee beans: accepted, rejected, and not yet.	The model used Android smartphones and a dataset of 10,944 images. The model tested MobileNetv1, MobileNetV2, NasNetMobile, and DenseNet121 architectures.	The best model was MobileNetV2, which achieved an accuracy of 97.75%.
[48]	The study presents a method that recognizes the brightness of the beans before and after grinding.	The model was tested with 5 algorithms: linear regression, DT, RF, support vector regression, and fully coupled neural network.	The fully connected neural network performed the best, with 2.52 of color numerical difference.
[49]	The study presents a method that searches for optimal coffee bean roasting conditions.	Starling particle swarm optimization (SPSO) and other swarm intelligence and gradient-based algorithms were used.	SPSO achieved performance superior to other algorithms with average errors of 1.2–8.5%.
[50]	The study presents a method to recognize the different grades of coffee beans based on their features and patterns.	The model used the Densenet121 architecture. A dataset of 363 images was used.	The proposed model achieved an accuracy of 81.89%.
[51]	The study presents a method to detect the roast level of coffee beans.	The model developed the board using CNN architecture and NVIDIA Jetson Nano. A dataset of 2,489 images was used with three classes: under-roasting, optimum-roasting, and over-roasting.	The proposed model achieved an accuracy of 91.33%.
[52]	The study presents a method to detect the roasting level of coffee beans into 4 classes.	The study proposed a CNN to classifies 4 roasting levels: green, light roast, medium roast, and dark roast. A dataset of 1,200 images was used.	The CNN proposed achieved an accuracy of 97.5%.

The authors of papers [46,47] developed a mobile application that automatically classifies coffee bean quality based on its roast degree.

Additional work has been carried out on the automated diagnosis of coffee roasting level, achieving excellent results. [45] achieved a maximum 100% accuracy in classifying Arabica coffee into light, medium, and dark. To predict the roasting degree of coffee beans, the authors of [48] created a model that consists of 4 input data: temperature, humidity, place of origin, and temperature of the roasting curve sampled every 15 seconds.

The networks tested for coffee roast classification were SVM, ResNet-152, MobileNetV2, Fully Connected Neural Network, SPSO, DenseNet121, and CNN proposed by the research authors.

VGG16, MobileNetV1, NasNetMobile, Linear Regression, DT, RF, Support Vector Regression, and Fully Couped Neural Network were also tested, but obtained inferior results when compared to other network models.

The research by (Septiarini, 2022) achieved maximum accuracy to classify 3 roast levels using SVM. It is not possible to say which network presents the best performance among all the works analyzed because the databases tested were different. The works differ in the roast levels they seek to classify and consequently in the number of classes presented in each model. The studies studied used datasets from 160 images to around 11,000 images. The size of the dataset also influences the model result. More studies are needed to consider different coffee roasting levels.

4.3. Coffee Sensory Classification

During coffee production, the beans can go through different processes. While traditional coffee has undesirable characteristics such as firm bitterness, special coffees have a diverse range of flavors and aromatic notes and are better balanced. Sensory analysis of coffee involves identifying its properties in the drink. The method takes place when the taster separates, inhales, and ingests a small amount of roasted and ground coffee and evaluates the following sample factors: Fragrance, Aroma, Uniformity, Sweetness, Flavor, Acidity, Body, Finish, and Balance.

As illustrated in Table 6, the first column corresponds to the literature review's reference; the second column describes the paper's main objectives; the third column lists the models and algorithms used in the paper; and the last column informs the results achieved.

Table 6. Coffee Sensory Classification.

Reference	Objectives and Scenario of Application	Methodology	Results
[53]	The study explores how to classify and characterize coffee beans based on their aroma.	The SVM method was used with an electronic nose system and a dataset with 100 coffee bean samples.	The proposed model achieved an accuracy of 70%.
[54]	The study proposes the use of an electronic nose (E-nose) to recognize original Arabica civet coffee (authentic and non-authentic).	Nine different mixture combinations were used for a total of 90 samples. Three classification methods were compared: Logistic Regression (LR), Linear Discriminant Analysis (LDA), and KNN.	The best model was KNN, which achieved an accuracy of 97.77%.
[55]	The study presented a new technique to analyze and characterize the flavor of 21 varieties of coffee.	A miniaturized potentiometric electronic tongue based on low-selectivity polymeric sensors was used.	The proposed model achieved an accuracy of 91.3 %.
[56]	The study used an e-nose device to estimate the caffeine content of samples.	The methods used were PCA (principal component analysis), LDA, PLSR (partial least squares regression), and ANN. Seven coffee bean classes and a total of 147 samples were made.	PLSR and ANN models achieved an R2 of 0.9576 and 0.9634, respectively, and the LDA model achieved an R2 of 0.9714.
[57]	This study presents a model for predicting cup coffee quality.	Two algorithms were implemented: SVM and ANN. Fifty-six samples were analyzed.	The most efficient model was ANN, which had an average accuracy of 81%.
[58]	This study presents a model for coffee aroma classification.	The model used an electronic nose that combined the separability indicator and the support vector machine margin. It collected data on the coffee aromas of two coffee brands.	The proposed model achieved an accuracy of 100%.
[59]	The study compares coffee quality to recommend the best coffee combination based on various features related to aroma.	The paper compares various regression models using cross-validation. The models included Linear Regression, Ridge, Lasso, ElasticNet, DT Regressor, Random Forest Regressor, Gradient Boosting Regressor, Support Vector Regressor (SVR), and MLP Regressor.	The best model achieved an MAE of 0.2567.

The models for finding the sensory characteristics of coffee (aroma and flavor) do not use images, these studies employ a dataset with coffee samples and can use an electronic tongue or nose. The electronic nose (e-nose) has gas sensors that can identify patterns when combined with MA algorithms. The dataset of these studies is composed of samples of coffee odors collected by the gas sensors present in the e-nose. The samples are submitted to machine learning algorithms to recognize and classify the coffee brand [58] and coffee bean types [56]. [53,54,56,58] classified coffee beans based on their aroma using an e-nose. The method proposed by [58] achieved 100% accuracy for classifying two distinct coffee brands using odor samples collected by e-nose gas sensors. The electronic tongue (e-tongue) has low-selective potentiometric sensors that respond to a wide variety of flavors. [55] classified the taste of coffee using an e-tongue and achieved excellent results. The dataset [55]'s research is composed of coffee cup samples prepared by a coffee machine, and the model identifies 21 varieties of coffee and achieves an average accuracy of 91.3%. The study [57] classified the quality of coffee drinks based on measurements made to almond and roasted coffee beans. [59] used a sensory evaluation of the coffee quality scores dataset with features related to aroma, flavor, acidity, body, balance, uniformity, and others using various linear regression models to predict coffee quality scores.

The networks tested for coffee sensory analysis were SVM, KNN, PLSR, LDA, ANN, SVM and various Regressions Models. It is not possible to say which network presents the best performance among all the works analyzed because the techniques used were different (e-nose and e-tongue) and the databases tested were also different.

4.4. Coffee Maturity Classification

Coffee maturation begins with the fruits turning light green, evolving into cherry fruits ready to be harvested. The previous year's flowering influences fruit maturation and is directly associated with the quality of the drink. Identifying the maturity of the coffee is very important because the coffee must be harvested with the highest percentage of ripe "cherry" fruits to provide the best quality drink.

Research in coffee classification must consider the ability to generalize to other coffee-producing regions and different coffee varieties. The characteristics extracted from the images may not capture all the nuances of coffee maturity.

As illustrated in Table 7, the first column corresponds to the literature review's reference; the second describes the papers's main objectives; the third lists the models and algorithms used in the paper; and the last column informs the results achieved.

Table 7. Coffee Maturity Classification.

Reference	Objectives and Scenario of Application	Methodology	Results
[60]	The study presents a framework for classifying coffee cherry fruits into 5 ripening stages.	Five CNN architectures were validated: VGG16, VGG19, Inception-ResNet-V2, Inception-V3, and DenseNet201 on a dataset with 600 images.	DenseNet201 achieved an accuracy over 98%.
[61]	The study explores the classification of coffee cherries as green, half broken, nosimetrics, and red.	The model used the KNN algorithm and ANN in a dataset of 1,159 images.	The best model achieved with KNN an accuracy of 72.12%.
[62]	The study classifies the type and maturity of the coffee.	The article compares the performance of two models, the Deep-CNN and ResNet50.	The ResNet50 model performed best and achieved an accuracy of 99.01%.
[63]	The study presents a method to monitor the ripeness of coffee fruits using multispectral images obtained by Unmanned Aerial Vehicles (UAV).	The model used a UAV with a modified multispectral camera to capture images of five coffee fields with distinct characteristics.	The NDVI (Normalized Vegetation Index) correlated well with the percentage of ripe fruits observed in the field ($R^2 = 0.81$).
[64]	The study describes a method to classify coffee fruits according to their ripening stage: Unripe, Semiripe, Ripe-Overripe.	The model uses high-frequency vibrations, and electrical impedance measurements were planned to be conducted and correlated with the ripening stage. The model used a Naive Bayes classifier.	The article evidenced an alternative for classifying coffee fruit differently from traditional operations.
[65]	The study presents a dataset for classifying coffee berries.	The dataset has 5,048 images in six classes: overripe, immature, semi-mature, mature, small, and elephant. It was tested on the CNN GoogLeNet and ShuffleNet.	The best model was GoogLeNet, which had a precision of better than 96.9%.

CNN networks are used to classify the ripening stage of coffee beans by [60,62,65]. The paper's authors [60] classified coffee cherry fruit into immature, semi-mature, mature, overripe, and dry. [61] classified coffee cherries based on skin color and shape characteristics. [63] classified coffee cherries based on spectral and textural characteristics.

The networks tested to identify coffee maturity were DenseNet201, KNN, ResNet50, Naive Bayes, and GoogLeNet. VGG16, VGG19, InceptionResNet-V2, Inception-v3, ANN, Deep-CNN, ShuffleNet were also tested, but did not achieve satisfactory results when compared with other networks. It is not possible to say which network presents the best performance among all the works analyzed because the databases tested were different. The researches differ in the maturity levels they seek to classify and consequently in the number of classes presented in each model. The studies studied used datasets from 600 images to around 5,000 images. The size of the dataset also influences the model result.

4.5. Coffee Disease Classification

Several diseases affect coffee farming. These diseases cause serious damage and reduce or nullify coffee productivity when poorly controlled. Studies on many coffee classifications aim to accurately and quickly identify diseases in coffee leaves for effective management and control.

As illustrated in Tables 8, 9, 10, and 11, the first column corresponds to the literature review's reference; the second column describes the paper's main objectives; the third column lists the models and algorithms used in the paper; and the last column informs the results achieved.

Table 8. Coffee Disease Classification.

Reference	Objectives and Scenario of Application	Methodology	Results
[66]	The study presented an approach for classifying disease severity in coffee leaves.	The model extracts low-level features in coffee leaves, such as color, to reduce memory and computational cost.	The model achieved an accuracy of over 93.8%.
[67]	The study addresses the detection of rust in coffee leaves.	The proposed method is based on a CNN and uses the transfer learning technique.	The best model was SVM, which achieved an accuracy of 96%.
[68]	The study proposes applying methods to classify 4 diseases in Barako coffee: spots, insect infestation, rust, and health.	The method used 3 configurations of the VGG16 model to classify a dataset of 3,958 coffee leaf images.	VGG16 model achieved 100% accuracy with setting: 100 epochs, 512 neurons, 0.5 dropouts, 0.0001 learning rate, adam, and 32 batch size.
[69]	The study identified and categorized coffee leaf diseases: rust, wilt, and brown eye.	Resnet50 and Mobilenet models were used, and a dataset of 1,120 images of coffee leaves with 3,360 images after data augmentation techniques.	Resnet50 model presented the best performance with 99.89% accuracy while MobileNet presented 97.01%.
[70]	The study describes a method for finding regions damaged by coffee rust.	The model uses a Deep Belief Network and a dataset of 624 images of coffee leaves.	The proposed model achieved an accuracy of 99.75%.
[71]	The article proposed a model for detecting the rate of coffee rust infection.	The model applies three deep learning algorithms: Backpropagation Neural Network (BPNN), CNN, and Recurrent Neural Network (RNN).	The best model was modified BPNN, which achieved a minimum MAE of 1,2462.
[72]	The study identifies diseases in Arabica coffee leaves such as Healthy, Miner, Rust, Phoma, and Cescospora.	The proposed model was trained using max pooling, dropout layer, two dense layers, and Adam optimizer.	The proposed model achieved an accuracy of 99.84% for 5 classes.
[73]	The study presents a method to detect Coffee Leaf Rust.	The model was based on CNN with few layers.	The proposed model achieved an accuracy of 95%.
[74]	The study implements a model to identify biotic agents present in Robusta coffee leaves.	The model used YOLOv3 and the MobileNetV2 algorithm.	The model achieved an accuracy of 90%.
[75]	The study presents a method for detecting white stem borer disease in coffee plants using an autonomous multi-terrain robot.	The model used YOLOv5 and a dataset of 492 images after data augmentation.	The model achieved a precision of 89.7%.
[76]	The study presents a methodology to detect Coffee Leaf Rust disease.	The model used twenty-five CNNs.	The ResNet101V2 model achieved the highest test with an accuracy of 95.56%.
[77]	The study presents a hybrid approach to identify various diseases in coffee leaves, including Red Spider Mite and Rust.	The proposed method uses MobileNetV3, Swin Transformer, and variational autoencoder (VAE). The model used the Robusta Coffee Leaf (RoCoLe) dataset.	The hybrid proposed model achieved an accuracy of 84.29%.

Table 9. Coffee Disease Classification.

Reference	Objectives and Scenario of Application	Methodology	Results
[78]	The study proposed a model for detecting diseases and pests in coffee leaves in Panama. Classified diseases include Cercospora, Rust, Miner, and Phoma.	The study proposed two CNN models: the detection of coffee leaves and the detection of coffee leaf diseases.	The proposed model achieved an overall accuracy of 90%.
[79]	The study presents a smartphone application classifier of coffee leaf diseases, containing the classes healthy, Rust, Sooty Molds, Cercospora, Phoma, and Leaf Miner.	The model used MobileNet on a low-cost microcontroller board in two architectures (cascaded and single-stage). Two datasets were used: BRACOL (1,747 samples) and LiCoLe (4,667 samples).	The embedded cascade model presented an accuracy of around 90%.
[80]	The study presents a few-shot learning model to classify and estimate the severity of biotic stresses on coffee leaves such as Rust, Miner, Brown Leaf Spot, and Cercospora.	The proposed model used two models, Triplet-Net and ProtoNet. Two datasets were used: the Leaf Dataset with 1,685 images and the Symptoms Dataset with 2,722 images.	The best result was using the Symptoms Dataset for biotic stress classification, which achieved an accuracy of 96.72%. The severity estimation task achieved an accuracy of 93.25%.
[81]	The study presents a method for detecting healthy and diseased coffee leaves.	An algorithm was developed by modifying VGG16 architecture. A dataset was collected with 4,000 images.	The proposed model achieved an accuracy of 97.9%.
[82]	The study presents a dataset of images of Peruvian coffee leaves, called CoLeaf, with 10 different nutritional deficiencies.	The dataset contains 1,006 leaf images divided into three subsets: Catimor, Caturra, and Borbon. ResNet50 network was used for classification.	The proposed model achieved an accuracy of 87.75%.
[83]	The study presents a method for detecting coffee diseases: Wilt, Rust, Cercospora, Sooty Moldy, Phoma, and Phoma Costaricensis.	The model consists of two algorithms, GoogLeNet and RESNET, to extract high-level features. MLP was used to classify coffee leaf and berry diseases. A dataset of 3,288 images was used.	The model achieved an accuracy of 99.08%.
[84]	The study describes an app to identify diseases and pests of coffee leaves: Miner, Rust, Brown leaf spot, Cercospora, and Healthy.	The CNN model used two datasets, Segmentation Dataset and Symptoms Dataset, and two architectures, UNet and PSPNet.	For semantic segmentation, the UNet model presented the best performance with 99.53% accuracy. The best result for the symptom classification was achieved with ResNet50 with an accuracy of 97.07%.
[85]	The study describes a dataset of images of Arabica coffee leaves, which have five classes: Rust, Cescospora, Phoma, Miner, and Healthy.	The dataset is called JMuBEN and JMuBEN2 and contains 58,555 images.	The dataset can evaluate ML models for coffee disease classification.

Table 10. Coffee Disease Classification.

Reference	Objectives and Scenario of Application	Methodology	Results
[86]	The study presents a method to recognize Cercospora, Rust, and Healthy coffee leaves based on Deep Learning and texture attributes	Two models were created. The first used a modification of AlexNet, and the second used a texture attribute extraction as input to the ANN classifier. A dataset of about 500 images for each class was used after augmentation.	Using the texture model, the kappa rate was 0.900, and the sensitivity was 0.933. Using DLDR, the best result for Kappa was 0.970, and the sensitivity was 0.980.
[87]	The study presents a model to identify diseases and pests in coffee leaves based on field images.	The Mask R-CNN was used for instance segmentation. For semantic segmentation, two networks were used: UNet and PSPNet. BRACOT and BRACOL datasets were used.	The model achieved an average precision AP of 73.90% for instance segmentation and AP of 71.90% for object detection.
[88]	The study presents a model to identify and estimate the stress severity caused by biotic agents on coffee leaves: Leaf Miner, Rust, Brown Leaf Spot, and Cercospora Leaf Spot.	The model used different CNN architectures: AlexNet, GoogLeNet, VGG16, ResNet50, and MobileNetV2. A dataset BRACOL was used.	The best model achieved an accuracy of 95.24% for the biotic stress classification and 86.51% for severity estimation.
[89]	The study aims to predict the incidence of Phoma leaf spot disease in coffee plantations considering the climatic variables in the coffee-producing regions of Brazil.	The model was tested with KNN, MLP, SVN, RF, Extreme Gradient Boosting (XGBoost), and Gradient Boosting (GradBOOSTING) algorithms.	The best model was XGBoost, with a Root Mean Square Error (RMSE) of 3.45% for the high-yielding trees.
[90]	The study aims to predict the incidence of Rust, Cercospora, Miner, and Coffee Borer in coffee plantations considering the climatic variables in the coffee-producing regions of Brazil.	The model was tested with Multiple Linear Regression (RLM), KNN, Random Forest Regressor (RFT), and MLP.	The best model was RFT, with RMSE values ranging from 0.227 to 0.853 for high yield.
[91]	The study classifies healthy and unhealthy coffee cherries in various stages of maturity.	The paper presents a model that uses the Restricted Boltzmann Machine Algorithm with Grey Wolves Optimization (GWO). A dataset with 475 coffee cherries.	GWO achieved an accuracy of 98%.
[92]	The study presents a method to detect the presence of Coffee Rust.	The study trained two neural network architectures: with OpenCV and with NVIDIA Digits.	The neural network trained with NVIDIA achieved the best performance with an accuracy of 98%.
[93]	The study presents a method to classify the biotic stress on coffee leaves.	The study presents three CNN-based methods: ECNN, HLGGM, and HLGCM. A combination of datasets was used: Bracol, JMUBEN, and PDCMD.	The HLGCM achieved the best performance with an accuracy of 99.49%.
[94]	The study presents a dataset with robusta coffee leaf images called RoCoLe.	RoCoLe has 1,560 coffee leaf images with six classes: healthy, Red Spider Mite, Rust level 1, Rust level 2, Rust level 3, and Rust level 4.	RoCoLe can be used to train and validate the performance of ML algorithms.
[95]	The study identifies and categorizes Cucurbit leaf diseases in four severity levels.	The model used federated learning with CNN and DT. A dataset with 4,585 images was used with 14 classes.	The proposed model achieved an average accuracy of 89.54%.

Table 11. Coffee Disease Classification.

Reference	Objectives and Scenario of Application	Methodology	Results
[96]	The study detects four classes of the severity of Coffee Leave Rust vegetation indices extracted from UAV imagery.	The paper compares different decision tree models.	The best model was the Logistic Model Tree (LTM), which achieved a precision of 0.672.
[97]	The study detects contamination by Aspergillus ochraceous in Robusta green coffee beans.	The model used 6 ML algorithms: linear discriminant analysis (LDA), SVM, KNN, DT, Naive Bayes (NB), and quadratic discriminant analysis (QDA).	The proposed model achieved an accuracy of 97.5%.
[98]	The study classifies coffee plant diseases as Cescospora, Miner, Phoma, Rust, and Healthy.	The paper compares DenseNet121 with Enhanced EfficientNetV2-5. The Kaggle coffee plant dataset and the JMuBEN Mendeley dataset were used with 5000 images.	The best model was Enhanced EfficientNetV2-S with an accuracy of 98.1%.
[99]	The study classifies coffee plant diseases as Cescospora, Miner, Phoma, and Rust.	The model used a Simple Linear Iterative Clustering (SLIC) segmentation algorithm and the Densenet-264. The JMuBEN Mendeley and Kaggle coffee plant datasets were used with 7,044 images.	The proposed model achieved an accuracy of 99.17 %.
[100]			The proposed model achieved Kappa coefficient of 0.96 for N and P deficiency, and 0.92 for B.

The white stem borer disease is a major pest of Arabica [75,101] proposed a method to identify them. The beetle pierces the coffee stem and remains buried until the plant falls completely. As the disease is difficult to diagnose in the early stages of its infestation, it causes major problems for coffee producers. The authors of the paper [75] used the YoloV5 deep learning model to identify white stem borer disease using color, texture, and shape characteristics. The Jackal UGV robot navigated and mapped the simulated environment of an Arabica coffee plantation using a depth camera and state and location estimation algorithms.

Coffee Leaf Rust is a serious disease that affects many coffee-producing regions. It is caused by a pathogenic fungus that attacks the underside of coffee leaves and is described by yellow-orange, powdery dots. If left untreated, rust can cause a drop in coffee production. [76] carried out a comparative study designed to identify Coffee Leaf Rust with twenty-five models and achieved the best performance of 90% with models ResNet101V2, InceptionV3, ResNet50V2, Xception, and DenseNet169. [67] used a CNN architecture to spot rust infestation and identify rust in coffee leaves of different varieties and stages of development.

The paper [84] presented an application for identifying and quantifying diseases and pests in coffee leaves using smartphone images. The application consists of two main modules: semantic segmentation for severity and symptom classification.

The authors of the paper [66] proposed a method to capture the discolored distribution of coffee leaves, allowing her to identify the severity of the disease more easily.

In [68], the authors achieved 100% accuracy in identifying disease with no errors using VGG16 architecture, hence demonstrating that the appropriate configuration and optimization can generate efficient results. [81] also used VGG16 to classify diseases in coffee leaves. In this paper, the authors modified VGG16 and achieved an accuracy of 97.9% to classify healthy and infected coffee leaves.

The work in [77] proposed a hybrid feature fusion with MobileNetV3 that extracts local features, while Swin Transformer extracts high-level features classification performance and rust. [78] used two models. The first model identifies coffee leaves. This model used a CNN architecture with 10 convolutional layers, Adam optimization, softmax, and 100 epochs and achieved an accuracy of 95%. The second model was used to detect diseases from coffee leaves, and the CNN architecture has 20 convolutional layers, Adamax optimization, and 100 epochs, achieving an accuracy of 90%. In the same way, [79] proposed two architectures to classify diseases of coffee leaves. The first identifies the diseases of coffee in six classes. The second architecture is used to improve the classification performance made by the first stage.

On the other hand, [87] proposed three experiments that can learn to classify using just a few examples. Each experiment was tested with two architectures: Tripletnet and Protonet. The first two experiments were tested to classify the biotic stress of coffee leaves. The third experiment sought to automatically estimate the severity of the disease: low, very low, high, very high.

The paper [82] introduced a dataset with images captured in a controlled environment showing nutritional deficiencies in coffee leaves. The nutritional deficiencies are Nitrogen, Phosphorus, Potassium, Magnesium, Boron, Manganese, Calcium, and Iron. The nutritional deficiencies are detected and classified by analyzing the color and shape characteristics. Leaves with more than one deficiency can also be detected.

Moreover, [83] proposed a model with GoogLeNet and RestNet in the feature extraction phase and MLP with traditional ML classifiers such as KNN, DT, SVM, and RF for classification tasks.

Three CNN-based methods were proposed in [93]. ECNN focuses on the effective concatenation of five CNNs; HLGGM combines dimension-reduced Mobile-Net v3 features with handcrafted features; and HLGCM has reduced dimensionality and uses a DT.

The networks tested to identify coffee diseases were SVM, VGG16, ResNet50, Deep Belief Network, BPNN, MobileNetV2, ResNet101V2, TripletNet, ProtoNet, GooLeNet, ResNet, UNet, AlexNet, ANN, Mask R-CNN, XGBoost, RFT, NVIDIA Digits, Decision Tree, Enhanced EfficientNetV2-5, and DenseNet264. MobileNet, MobileNetV3, RNN, CNN, Swin Transformer, PSPNet, KNN, MLP, Grad-

Boosting, RLM, OpenCV, LDA, NB, QDA, DenseNet121, were also tested, but did not achieve satisfactory results when compared with other networks. It is not possible to say which network presents the best performance among all the works analyzed because the databases tested were different. The works differ in the diseases they seek to classify and consequently in the number of classes presented in each model. The studies studied used datasets from 400 images to around 58,000 images. The size of the dataset also influences the model result.

The proposed methodologies for classifying coffee diseases can be applied to assist farmers in early detection and making decisions to protect coffee plantations. It is necessary to include more classifications for coffee diseases and expand the size of the dataset. Future studies can also compare other deep learning algorithms with other pre-trained models to achieve higher accuracy.

5. Challenges and Future Trends

ML, along with complementary technologies, such as Computer Vision, are revolutionizing coffee farming, propelling agribusiness towards unprecedented efficiency, speed, and precision in decision-making while reducing operational costs. Our review evidenced the existence of several computational models devised for coffee bean classification, addressing various attributes such as defects, roasting, diseases, aroma and flavor, and maturation of coffee, hence demonstrating the potential for continuous advancements. However, despite this progress, most of the existing AI-based models and tools predominantly reside in experimental phases, lacking widespread practical deployment.

They further often exhibit a restricted focus, tailored to particular regions [82,89] or coffee types [35,45,54,64,68,74,85,90,94,97] and further demonstrate a limited scope in identifying nuanced characteristics like roast degrees [47] and maturity stages [60,61]. The need for more universally applicable models is evident, especially in areas like aromatic profiling, where ML remains underutilized, revealing significant untapped potential for research and industry applications.

Furthermore, the scarcity of extensive image datasets is a critical bottleneck, underscoring the necessity for enriched coffee bean data to enable robust model generalization across diverse environmental and cultural contexts. Tables 12, 13 and 14 provide a concise summary of the datasets used in the analyzed articles and are available for download. Additionally, the Tables list datasets that could be valuable for future coffee classification tasks using ML.

Table 12. Summary Datasets.

Reference	Number of Images	Number of Categories	Types	Download Link
[29]	8,000	4 (peaberry, long berry, premium, defect.)	Single bean	https://comvis.unsyiah.ac.id/usk-coffee/
[31]	1,813	2 (defective and normal)	Single bean	https://github.com/Tauranis/deep_coffee/
[36]	4,617	3 (good and bad - insect bite and broken)	Single bean	https://github.com/tanius/smallopticalsorter/tree/master/classifier-trainingdata
[37]	1554	3 (espresso, Kenya and Starbucks pike place)	Single bean	https://muratkoklu.com/dataset/Coffee_Image_Dataset.zip
[57]	56	.csv archive		https://github.com/Javiersuing/GitHub/blob/master/AlmacafeDataBase_CrossV_v4.ipynb
[68]	3958	4 (Healthy, Rusted, Infested, and Spotted Barako)	Leaves	https://github.com/francismontalbo/swatdcnn/tree/main
[79]	6000	6 (Healthy, Rust, Sooty Molds, Cercospora, Phoma and Leaf Miner)	Leaves	Licole F.J.P. Montalbo, A.A. Hernandez, Classifying Barako coffee leaf diseases using deep convolutional models (2020).
[82]	1006	10 (Boron, Iron, Potassium, Calcium, Magnesium, Manganese, Nitrogen, Phosphorus, calcium and healthy)	Leaves	CoLeaf https://data.mendeley.com/datasets/brfgw46wzb/1
[80], [84]	142	6 (Leaf Miner, Cercospora Leaf Spot, Rust, Bacterial Blight, Blister Spot, Brown Leaf Spot)	Leaves	BARBEDO, Jayme Garcia Arnal. Plant disease identification from individual lesions and spots using deep learning. Biosystems Engineering, v. 180, p. 96-107, 2019.
[85]	58,555	5 (Phoma, Cerscospora, Rust, Healthy, and Miner)	Leaves	JMuBEN2 https://data.mendeley.com/datasets/tgv3zb82nd/1
[87]	300	5 (Leaf Miner, Rust, Brown Leaf Spot, and Cercospora Leaf Spot)	Leaves	BRACOT - A Brazilian Arabica Coffee Tree images dataset, for instance segmentation of coffee leaves https://data.mendeley.com/datasets/pmkbyjpf6k/1

Table 13. Summary Datasets.

Reference	Number of Images	Number of Categories	Types	Download Link
[66], [79], [80], [84], [87], [88], [93]	1747	4 (Leaf Miner, Leaf Rust, Brown Leaf Spot, and Cercospora Leaf Spot)	Leaves	BRACOL - A Brazilian Arabica Coffee Leaf images dataset to identification and quantification of coffee diseases and pests https://data.mendeley.com/datasets/yy2k5y8mxg/1
[66], [77], [94]	1,560	6 (Healthy, Red Spider Mite Presence, Rust level 1, Rust level 2, Rust level 3 and Rust level 4)	Leaves	RoCoLe: A dataset with robusta coffee leaf images https://data.mendeley.com/datasets/c5yvn32dzg/2
[93], [98], [99]	22,591	3 (Coffee Rust, Cescospora and Phoma)	Leaves	JMuBEN https://data.mendeley.com/datasets/t2r6rszp5c/1
[93], [99]	76,000 (for coffee disease are 1,103 images)	88 classes, for the coffee disease are 4 classes (Cercospora, Healthy, Red Spider Mite and Rust)	Leaves	Kaggle Plant Disease Classification Merged Dataset https://www.kaggle.com/datasets/alinedobrovsky/plant-disease-classification-merged-dataset

Table 14. Summary Datasets.

Reference	Number of Images	Number of Categories	Types	Download Link
–	2,000	2 (good and bad)	Single bean	https://github.com/tanius/smallopticalsorter/tree/master/classifier-trainingdata
–	1,600	4 (Dark, green, light and medium)	Single bean	https://www.kaggle.com/datasets/gpiosenka/coffee-bean-dataset-resized-224-x-224
–	542	2(Miner and Rust)	Leaves	https://www.kaggle.com/datasets/alvarole/coffee-leaves-disease
–	1,000	3 (Healthy, Red Spider Mite and Rust)	Leaves	https://www.kaggle.com/datasets/coffeedisease/coffee-plant-disease
–	1,664	4 (Miner, Nodisease, Phoma and Rust)	Leaves	https://www.kaggle.com/datasets/gauravduttakiit/coffee-leaf-diseases https://www.kaggle.com/datasets/stunningvisionai/coffee-leaf-diseases-yolo https://www.kaggle.com/datasets/badasstechie/coffee-leaf-diseases
–	600	1 (green coffee beans)	Multiple beans	https://www.kaggle.com/datasets/danielmejiar/green-coffee-beans
–	1,800	3 (Cercospora, Healthy and Rust)	Leaves	https://www.kaggle.com/datasets/miladatulmuharromah/coffee-leaves-disease
–	6,000	4 (Dark, green, light, and medium)	Multiple beans	https://www.kaggle.com/datasets/dewanakretarta/resize-224x224-coffee-bean-dataset-v2
–	175	3 (half ripe, ripe, unripe)	Multiple cherries	https://www.kaggle.com/datasets/maulanaimanpratama/coffee-fruit-images
–	–	Is a merge of BrACoL , JMuBEN, JMuBEN2 , RoCoL , and LiCoLe	–	CoffeeNet https://drive.google.com/drive/folders/1dDEw-VhuGx37wSVqhYQ-m1zKFug2SpDz

Although little explored, agribusiness professionals should prepare themselves for the future tools that will emerge from the research results presented in this article. These agribusiness professionals will encounter new challenges and changes in the way they work. They must adapt and relearn their roles, focusing on their analytical skills in data-driven agriculture.

Recognizing the urgent demand for refined classification tools to aid coffee growers, this work underscores the importance of machine learning-based image classification in supporting decision-making and enhancing overall industry efficiency.

Finally, the transformative power of ML and Computer Vision offers promising prospects for coffee bean diagnostics, promising to enhance productivity and management confidence significantly. Future research should, therefore, expand upon these findings, focusing on refining classification techniques for a broader and more accurate application in coffee bean classification.

As a result, Figure 7 lists the roadmap of future trends for resources in ML coffee classification. To maximize production, reduce costs, and increase precision in decision-making in coffee farming, future research needs to increase the capacity to implement higher-performance ML algorithms, increase the size of image datasets, and, finally, develop fewer generalist methods and more classes considered.
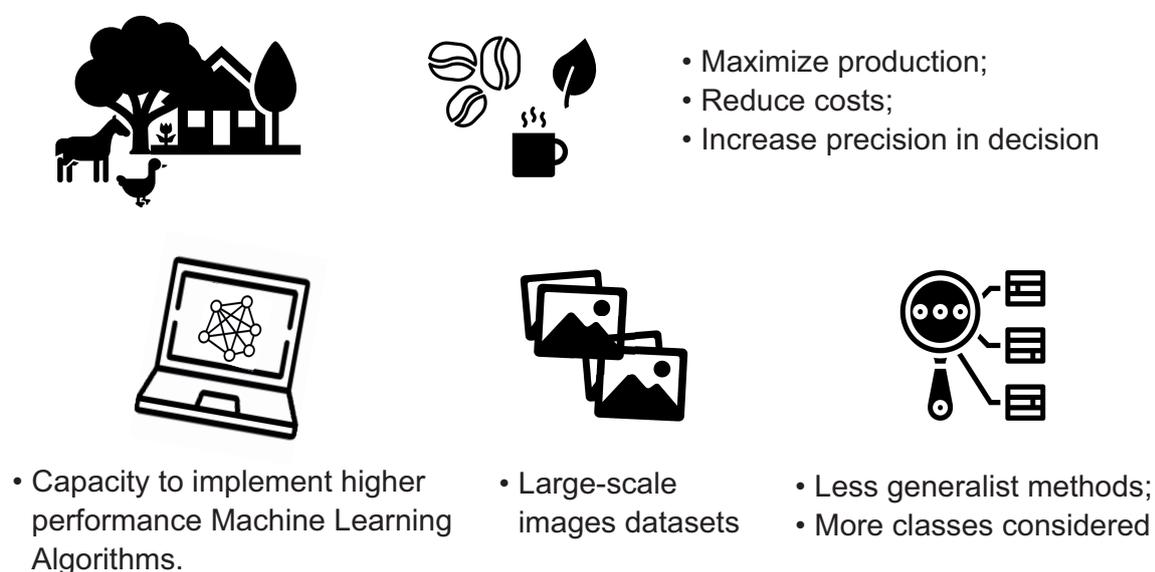

- Maximize production;
- Reduce costs;
- Increase precision in decision

- Capacity to implement higher performance Machine Learning Algorithms.

- Large-scale images datasets

- Less generalist methods;
- More classes considered

Figure 7. Roadmap of future trends for resource in Machine Learning coffee classification

6. Conclusions

This comprehensive literature review examined the application of ML techniques to coffee bean and leaves classification, focusing on the last five years of research related to coffee defects, roasting, maturity, sensing, and diseases.

The synthesis of data extracted from the reviewed works offers valuable insights, presenting a comprehensive overview of contemporary day research, the most frequently used algorithms, and model comparisons.

Some of the research directions for the future include creating new data sets that are larger, less general, and with more classes, and, above all, the search for increasing the capacity to implement higher-performance ML algorithms.

Notably, there has been a significant increase in relevant scientific publications, reflecting the growing impact of machine learning-based image classification in advancing coffee agriculture. This surge in scholarly activity has catalyzed increased investments in developing new products and services, underscoring the promising trajectory for future ML research within the coffee domain.

Author Contributions: I.V.C.M. contributed to this study's conceptualization, data curation, resources, writing, review, and editing. F.A.P.F., N.V., and H.P. contributed to methodology, supervision, review, validation, and funding. All authors read and approved the final manuscript.

Data Availability Statement: Data declarations do not apply to this paper.

Conflicts of Interest: The authors declare that they have no competing interests.

References

1. Gois, T.C.; Thomé, K.M.; Balogh, J.M. Behind a cup of coffee: International market structure and competitiveness. *Competitiveness Review: An International Business Journal* **2023**, *33*, 993–1009. https://doi.org/10.1108/CR-10-2021-0141.

2. Bosso, H.; Barbalho, S.M.; de Alvares Goulart, R.; Otoboni, A.M.M.B. Green coffee: economic relevance and a systematic review of the effects on human health. *Critical Reviews in Food Science and Nutrition* **2023**, *63*, 394–410. https://doi.org/10.1080/10408398.2021.1948817.

3. Dhanya, V.; Subeesh, A.; Kushwaha, N.; Vishwakarma, D.K.; Kumar, T.N.; Ritika, G.; Singh, A. Deep learning based computer vision approaches for smart agricultural applications. *Artificial Intelligence in Agriculture* **2022**. https://doi.org/10.1016/j.aiia.2022.09.007.

4. Elbasi, E.; Mostafa, N.; AlArnaout, Z.; Zreikat, A.I.; Cina, E.; Varghese, G.; Shdefat, A.; Topcu, A.E.; Abdelbaki, W.; Mathew, S.; others. Artificial intelligence technology in the agricultural sector: a systematic literature review. *Ieee Access* **2022**, *11*, 171–202. https://doi.org/10.1109/ACCESS.2022.3232485.

5. Pragathi, S.; Jacob, L. Review On Image based Coffee Bean Quality Classification: Machine Learning Approach. 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N). IEEE, 2022, pp. 706–711. https://doi.org/10.1109/ICAC3N56670.2022.10074124.

6. Sarkar, C.; Gupta, D.; Gupta, U.; Hazarika, B.B. Leaf disease detection using machine learning and deep learning: Review and challenges. *Applied Soft Computing* **2023**, p. 110534. https://doi.org/10.1016/j.asoc.2023.110534.

7. Sunil, C.; Jaidhar, C.; Patil, N. Systematic study on deep learning-based plant disease detection or classification. *Artificial Intelligence Review* **2023**, pp. 1–98. https://doi.org/10.1007/s10462-023-10517-0.

8. Sajitha, P.; Andrushia, A.D.; Anand, N.; Naser, M. A Review on Machine Learning and Deep Learning Image-based Plant Disease Classification for Industrial Farming Systems. *Journal of Industrial Information Integration* **2024**, p. 100572. https://doi.org/10.1016/j.jii.2024.100572.

9. Anto, I.A.F.; Munandar, A.; Wibowo, J.W.; Salim, T.I.; Mahendra, O. Coffee Bean Roasting Levels Detection: A Systematic Review. 2023 IEEE 7th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE). IEEE, 2023, pp. 146–151. https://doi.org/10.1109/ICITISEE58992.2023.10404775.

10. Sharma, K.; Sharma, C.; Sharma, S.; Asenso, E. Broadening the research pathways in smart agriculture: predictive analysis using semiautomatic information modeling. *Journal of Sensors* **2022**, *2022*, 1–19. https://doi.org/10.1155/2022/5442865.

11. Sharma, S.; Sharma, C.; Asenso, E.; Sharma, K.; others. Research Constituents and Trends in Smart Farming: An Analytical Retrospection from the Lens of Text Mining. *Journal of Sensors* **2023**, *2023*. https://doi.org/10.1155/2023/6916213.

12. Sharma, K.; Sethi, G.K.; Bawa, R.K. A comparative analysis of deep learning and deep transfer learning approaches for identification of rice varieties. *Multimedia Tools and Applications* **2024**, pp. 1–18. https://doi.org/10.1007/s11042-024-19126-7.

13. Komal.; Sethi, G.K.; Bawa, R.K. Automatic Rice Variety Identification System: state-of-the-art review, issues, challenges and future directions. *Multimedia Tools and Applications* **2023**, *82*, 27305–27336. https://doi.org/10.1007/s11042-023-14487-x.

14. Komal.; Sethi, G.K.; Bawa, R.K. A prototype of automatic rice variety identification system using artificial intelligence techniques. *AIP Conference Proceedings* **2022**, *2455*, 040004.

15. Sethi, G.; Bawa, R.; others. A Hybrid Approach of Preprocessing and Segmentation Techniques in Automatic Rice Variety Identification System. *Journal of Scientific Research* **2022**, *14*. https://doi.org/10.3329/jsr.v14i1.54811.

16. Javanmardi, S.; Ashtiani, S.H.M.; Verbeek, F.J.; Martynenko, A. Computer-vision classification of corn seed varieties using deep convolutional neural network. *Journal of Stored Products Research* **2021**, *92*, 101800. https://doi.org/10.1016/j.jspr.2021.101800.

17. Hu, R.; Zhang, S.; Wang, P.; Xu, G.; Wang, D.; Qian, Y. The identification of corn leaf diseases based on transfer learning and data augmentation. Proceedings of the 3rd International Conference on Computer Science and Software Engineering, 2020, pp. 58–65. https://doi.org/10.6041/j.issn.1000-1298.2020.02.025.

18. de Medeiros, A.D.; Capobiango, N.P.; da Silva, J.M.; da Silva, L.J.; da Silva, C.B.; dos Santos Dias, D.C.F. Interactive machine learning for soybean seed and seedling quality classification. *Scientific reports* **2020**, *10*, 11267. https://doi.org/10.1038/s41598-020-68273-y.

19. Janiesch, C.; Zschech, P.; Heinrich, K. Machine learning and deep learning. *Electronic Markets* **2021**, *31*, 685–695. https://doi.org/10.1007/s12525-021-00475-2.

20. Russell, S.J.; Norvig, P. *Artificial intelligence: A modern approach*, 4 ed.; Pearson, 2021.

21. Mahadevkar, S.V.; Khemani, B.; Patil, S.; Kotecha, K.; Vora, D.R.; Abraham, A.; Gabralla, L.A. A review on machine learning styles in computer vision—techniques and future directions. *Ieee Access* **2022**, *10*, 107293–107329.

22. Han, K.; Wang, Y.; Chen, H.; Chen, X.; Guo, J.; Liu, Z.; Tang, Y.; Xiao, A.; Xu, C.; Xu, Y.; others. A survey on vision transformer. *IEEE transactions on pattern analysis and machine intelligence* **2022**, *45*, 87–110. https://doi.org/10.1109/TPAMI.2022.3152247.

23. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; others. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* **2020**.

24. Uddin, M.S.; Bansal, J.C. *Computer Vision and Machine Learning in Agriculture, Volume 2*; 2022; pp. 1–8.

25. Micaraseth, T.; Pornpipatsakul, K.; Chancharoen, R.; Phanomchoeng, G. Coffee Bean Inspection Machine with Deep Learning Classification. 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME). IEEE, 2022, pp. 1–5. https://doi.org/10.1109/ICECCME55909.2022.9987835.

26. Liang, C.S.; Xu, Z.Y.; Zhou, J.Y.; Yang, C.M.; Chen, J.Y. Automated Detection of Coffee Bean Defects using Multi-Deep Learning Models. 2023 VTS Asia Pacific Wireless Communications Symposium (APWCS). IEEE, 2023, pp. 1–5. https://doi.org/10.1109/APWCS60142.2023.10234059.

27. Septiarini, A.; Hamdani, H.; Burhandenny, A.E.; Nur, S.; Winarno, E. The color-texture features and machine learning approach for quality detection of coffee beans. 2023 9th International Conference on Computer and Communication Engineering (ICCCE). IEEE, 2023, pp. 217–222. https://doi.org/10.1109/ICCCE58854.2023.10246091.

28. Kuo, C.J.; Chen, C.C.; Chen, T.T.; Tsai, Z.; Hung, M.H.; Lin, Y.C.; Chen, Y.C.; Wang, D.C.; Homg, G.J.; Su, W.T. A labor-efficient gan-based model generation scheme for deep-learning defect inspection among dense beans in coffee industry. 2019 IEEE 15th International Conference on Automation Science and Engineering (CASE). IEEE, 2019, pp. 263–270. https://doi.org/10.1109/COASE.2019.8843259.

29. Febriana, A.; Muchtar, K.; Dawood, R.; Lin, C.Y. USK-COFFEE Dataset: A Multi-class Green Arabica Coffee Bean Dataset for Deep Learning. 2022 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom). IEEE, 2022, pp. 469–473. https://doi.org/10.1109/CyberneticsCom55287.2022.9865489.

30. Shao, B.; Hou, Y.; Huang, N.; Wang, W.; Lu, X.; Jing, Y. Deep Learning based Coffee Beans Quality Screening. 2022 IEEE International Conference on e-Business Engineering (ICEBE). IEEE, 2022, pp. 271–275. https://doi.org/10.1109/ICEBE55470.2022.00054.

31. Lee, J.Y.; Jeong, Y.S. Prediction of defect coffee beans using CNN. 2022 IEEE International Conference on Big Data and Smart Computing (BigComp). IEEE, 2022, pp. 202–205. https://doi.org/10.1109/BigComp54360.2022.00046.

32. Kuo, C.J.; Wang, D.C.; Chen, T.T.; Chou, Y.C.; Pai, M.Y.; Horng, G.J.; Hung, M.H.; Lin, Y.C.; Hsu, T.H.; Chen, C.C. Improving defect inspection quality of deep-learning network in dense beans by using hough circle

transform for coffee industry. 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). IEEE, 2019, pp. 798–805. https://doi.org/10.1109/SMC.2019.8914175.

33. Heryanto, T.A.; Nugraha, I.G.B.B. Classification of Coffee Beans Defect Using Mask Region-based Convolutional Neural Network. 2022 International Conference on Information Technology Systems and Innovation (ICITSI). IEEE, 2022, pp. 333–339. https://doi.org/10.1109/ICITSI56531.2022.9970890.

34. Chen, S.Y.; Chiu, M.F.; Zou, X.W. Real-time defect inspection of green coffee beans using NIR snapshot hyperspectral imaging. *Computers and Electronics in Agriculture* **2022**, *197*, 106970. https://doi.org/10.1016/j.compag.2022.106970.

35. Gomes, W.P.C.; Gonçalves, L.; da Silva, C.B.; Melchert, W.R. Application of multispectral imaging combined with machine learning models to discriminate special and traditional green coffee. *Computers and Electronics in Agriculture* **2022**, *198*, 107097. https://doi.org/10.1016/j.compag.2022.107097.

36. Chen, P.H.; Jhong, S.Y.; Hsia, C.H. Semi-supervised learning with attention-based CNN for classification of coffee beans defect. 2022 IEEE International Conference on Consumer Electronics-Taiwan. IEEE, 2022, pp. 411–412. https://doi.org/10.1109/ICCE-Taiwan55306.2022.9869187.

37. Unal, Y.; Taspinar, Y.S.; Cinar, I.; Kursun, R.; Koklu, M. Application of pre-trained deep convolutional neural networks for coffee beans species detection. *Food Analytical Methods* **2022**, *15*, 3232–3243. https://doi.org/10.1007/s12161-022-02362-8.

38. Putra, B.T.W.; Amirudin, R.; Marhaenanto, B. The evaluation of deep learning using convolutional neural network (CNN) approach for identifying Arabica and Robusta coffee plants. *Journal of Biosystems Engineering* **2022**, *47*, 118–129. https://doi.org/10.1007/s42853-022-00136-y.

39. Wang, Y.F.; Cheng, C.C.; Tsai, J.K. Implementation of Green Coffee Bean Quality Classification Using Slim-CNN in Edge Computing. 2022 IEEE 5th International Conference on Knowledge Innovation and Invention (ICKII). IEEE, 2022, pp. 133–135. https://doi.org/10.1109/ICKII55100.2022.9983596.

40. Kesiman, M.W.A.; Sulaiman, I.; Maysanjaya, I.M.D.; Dermawan, K.T. Benchmarking A New Dataset for Coffee Bean Defects Classification Based on SNI 01-2907-2008. 2023 International Conference on Information Technology Research and Innovation (ICITRI). IEEE, 2023, pp. 75–80. https://doi.org/10.1109/ICITRI59340.2023.10249345.

41. Tsai, J.J.; Chang, C.C.; Huang, D.Y.; Lin, T.S.; Chen, Y.C. Analysis and classification of coffee beans using single coffee bean mass spectrometry with machine learning strategy. *Food Chemistry* **2023**, *426*, 136610. https://doi.org/10.1016/j.foodchem.2023.136610.

42. Pratondo, A.; Zani, T.; Novianty, A.; Pudjoatmodjo, B. Raw Coffee Bean Classification for Roasting Suitability Assessment Using Transfer Learning. 2023 IEEE 11th Conference on Systems, Process & Control (ICSPC), 2023, pp. 1–6. https://doi.org/10.1109/ICSPC59664.2023.10419990.

43. Hamdani, H.; Septiarini, A.; Akbar, F.; Saputra, R.; Nurmadewi, D.; Priyatna, S.E. Classification of Arabica Coffee Beans Based on Multi-Features Using Artificial Neural Networks. 2023 1st International Conference on Advanced Engineering and Technologies (ICONNIC), 2023, pp. 85–90. https://doi.org/10.1109/ICONNIC59854.2023.10467549.

44. Agresti, P.D.M.; Franca, A.S.; Oliveira, L.S.; Augusti, R. Discrimination between defective and non-defective Brazilian coffee beans by their volatile profile. *Food chemistry* **2008**, *106*, 787–796. https://doi.org/10.1016/j.foodchem.2007.06.019.

45. Septiarini, A.; Hamdani, H.; Rifani, A.; Arifin, Z.; Hidayat, N.; Ismanto, H. Multi-Class Support Vector Machine for Arabica Coffee Bean Roasting Grade Classification. 2022 5th International Conference on Information and Communications Technology (ICOIACT). IEEE, 2022, pp. 407–411. https://doi.org/10.1109/ICOIACT55506.2022.9971897.

46. Janandi, R.; Cenggoro, T.W. An Implementation of convolutional neural network for coffee beans quality classification in a mobile information system. 2020 International Conference on Information Management and Technology (ICIMTech). IEEE, 2020, pp. 218–222. https://doi.org/10.1109/ICIMTech50083.2020.9211257.

47. Hakim, M.; Djatna, T.; Yuliasih, I. Deep learning for roasting coffee bean quality assessment using computer vision in mobile environment. 2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS). IEEE, 2020, pp. 363–370. https://doi.org/10.1109/ICACSIS51025.2020.9263224.

48. Okamura, M.; Soga, M.; Yamada, Y.; Kobata, K.; Kaneda, D. Development and evaluation of roasting degree prediction model of coffee beans by machine learning. *Procedia Computer Science* **2021**, *192*, 4602–4608. https://doi.org/10.1016/j.procs.2021.09.238.

49. Ratanasanya, S.; Chindapan, N.; Polvichai, J.; Sirinaovakul, B.; Devahastin, S. Model-based optimization of coffee roasting process: Model development, prediction, optimization and application to upgrading of Robusta coffee beans. *Journal of food engineering* **2022**, *318*, 110888. https://doi.org/10.1016/j.jfoodeng.2021.110888.

50. J, B.N.B.; M, A.N.K.; S, S.A.; Mohethe G, L.R.; Raghavendra, V. Coffee Bean Grading Based on Weight Estimation Using Densenet121 Model. 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA). IEEE, 2023, pp. 1–6. https://doi.org/10.1109/ICCUBEA58933.2023.10392243.

51. Vilcamiza, G.; Trelles, N.; Vinces, L.; Oliden, J. A coffee bean classifier system by roast quality using convolutional neural networks and computer vision implemented in an NVIDIA Jetson Nano. 2022 Congreso Internacional de Innovación y Tendencias en Ingeniería (CONIITI). IEEE, 2022, pp. 1–6. https://doi.org/10.1109/CONIITI57704.2022.9953636.

52. Naik, N.K.; Sethy, P.K. Roasted Coffee beans Classification based on Convolutional Neural Network. 2022 International Conference on Futuristic Technologies (INCOFT), 2022, pp. 1–3. https://doi.org/10.1109/INCOFT55651.2022.10094378.

53. Caya, M.V.C.; Maramba, R.G.; Mendoza, J.S.D.; Suman, P.S. Characterization and Classification of Coffee Bean Types using Support Vector Machine. 2020 IEEE 12th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM). IEEE, 2020, pp. 1–6. https://doi.org/10.1109/HNICEM51456.2020.9400144.

54. Harsono, W.; Sarno, R.; Sabilla, S.I. Recognition of original arabica civet coffee based on odor using electronic nose and machine learning. 2020 International Seminar on Application for Technology of Information and Communication (iSemantic). IEEE, 2020, pp. 333–339. https://doi.org/10.1109/iSemantic50169.2020.9234234.

55. Gabrieli, G.; Muszynski, M.; Thomas, E.; Labbe, D.; Ruch, P.W. Accelerated estimation of coffee sensory profiles using an AI-assisted electronic tongue. *Innovative Food Science & Emerging Technologies* **2022**, *82*, 103205. https://doi.org/10.1016/j.ifset.2022.103205.

56. Aghdamifar, E.; Sharabiani, V.R.; Taghinezhad, E.; Szymanek, M.; Dziwulska-Hunek, A. E-nose as a non-destructive and fast method for identification and classification of coffee beans based on soft computing models. *Sensors and Actuators B: Chemical* **2023**, *393*, 134229. https://doi.org/10.1016/j.snb.2023.134229.

57. Suarez-Peña, J.A.; Lobaton-García, H.F.; Rodríguez-Molano, J.I.; Rodriguez-Vazquez, W.C. Machine learning for cup coffee quality prediction from green and roasted coffee beans features. Workshop on Engineering Applications. Springer, 2020, pp. 48–59. https://doi.org/10.1007/978-3-030-61834-6_5.

58. Wu, J.C.; Chou, T.I.; Chiu, S.W.; Shihabudeen, P.; Chen, P.A.; Tang, K.T. Development of Coffee Classification by Feature Selection and Classifier Optimization Based on An Electronic Nose. 2023 IEEE Conference on AgriFood Electronics (CAFE). IEEE, 2023, pp. 104–107. https://doi.org/10.1109/CAFE58535.2023.10291414.

59. Rajbharath, R.; Vijayamalaiya, S.; Aabid, M.; Salilan, A. Cofee Quality Prediction Using Machine Learning. 2023 International Conference on System, Computation, Automation and Networking (ICSCAN). IEEE, 2023, pp. 1–6. https://doi.org/10.1109/ICSCAN58655.2023.10395728.

60. Tamayo-Monsalve, M.A.; Mercado-Ruiz, E.; Villa-Pulgarin, J.P.; Bravo-Ortíz, M.A.; Arteaga-Arteaga, H.B.; Mora-Rubio, A.; Alzate-Grisales, J.A.; Arias-Garzon, D.; Romero-Cano, V.; Orozco-Arias, S.; others. Coffee maturity classification using convolutional neural networks and transfer learning. *IEEE Access* **2022**, *10*, 42971–42982. https://doi.org/10.1109/ACCESS.2022.3166515.

61. Anita, S.; others. Classification Cherry's Coffee using k-Nearest Neighbor (KNN) and Artificial Neural Network (ANN). 2020 International Conference on Information Technology Systems and Innovation (ICITSI). IEEE, 2020, pp. 117–122. https://doi.org/10.1109/ICITSI50517.2020.9264927.

62. Raveena, S.; Surendran, R. ResNet50-based Classification of Coffee Cherry Maturity using Deep-CNN. 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT). IEEE, 2023, pp. 1275–1281. https://doi.org/10.1109/ICSSIT55814.2023.10061006.

63. Martins, R.N.; de Carvalho Pinto, F.d.A.; de Queiroz, D.M.; Valente, D.S.M.; Rosas, J.T.F.; Portes, M.F.; Cerqueira, E.S.A. Digital mapping of coffee ripeness using UAV-based multispectral imagery. *Computers and Electronics in Agriculture* **2023**, *204*, 107499. https://doi.org/10.1016/j.compag.2022.107499.

64. Pineda, M.F.; Tinoco, H.A.; Lopez-Guzman, J.; Perdomo-Hurtado, L.; Cardona, C.I.; Rincon-Jimenez, A.; Betancur-Herrera, N. Ripening stage classification of Coffea arabica L. var. Castillo using a machine learning approach with the electromechanical impedance measurements of a contact device. *Materials Today: Proceedings* **2022**, *62*, 6671–6678. https://doi.org/10.1016/j.matpr.2022.04.669.

65. Ngocho, B.M.; Mwangi, E.; Kamucha, G.; Jeon, G. An Image-Set of Coffee Berries for CNN Classification. 2023 IEEE AFRICON. IEEE, 2023, pp. 1–6. https://doi.org/10.1109/AFRICON55910.2023.10293226.

66. Okada, T.; Huang, Y.; Hao, G.; Iizuka, S.; Fukui, K. Low-Level Feature Aggregation Networks for Disease Severity Estimation of Coffee Leaves. 2023 18th International Conference on Machine Vision and Applications (MVA). IEEE, 2023, pp. 1–5. https://doi.org/10.23919/MVA57639.2023.10215626.

67. Prabhu, A.; Isiri, K. A Deep Learning Approach to Identify Defects in Coffee Leaves Using Convoluional Neural Network. 2022 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON). IEEE, 2022, pp. 1–8. https://doi.org/10.1109/SMARTGENCON56628.2022.10084091.

68. Montalbo, F.J.P.; Hernandez, A.A. An Optimized Classification Model for Coffea Liberica Disease using Deep Convolutional Neural Networks. 2020 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA). IEEE, 2020, pp. 213–218. https://doi.org/10.1109/CSPA48992.2020.9068683.

69. Paulos, E.B.; Woldeyohannis, M.M. Detection and Classification of Coffee Leaf Disease using Deep Learning. 2022 International Conference on Information and Communication Technology for Development for Africa (ICT4DA). IEEE, 2022, pp. 1–6. https://doi.org/10.1109/ICT4DA56482.2022.9971300.

70. Raveena, S.; Surendran, R. Clustering-based Hemileia Vastatrix Disease Prediction in Coffee Leaf using Deep Belief Network. 2023 8th International Conference on Communication and Electronics Systems (ICCES). IEEE, 2023, pp. 1094–1100. https://doi.org/10.1109/ICCES57224.2023.10192835.

71. Lyimo, D.A.; Narasimhan, V.L.; Mbero, Z.A. Sensitivity Analysis of Coffee Leaf Rust Disease using Three Deep Learning Algorithms. 2021 IEEE AFRICON. IEEE, 2021, pp. 1–6. https://doi.org/10.1109/AFRICON51333.2021.9571007.

72. Madhukar, R.K.; Chaurasiya, A.; Chaturvedi, P. A Systematized Chronicity based Disease Classification in Coffee Leaves using Deep Learning. 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC). IEEE, 2022, pp. 1336–1342. https://doi.org/10.1109/ICOSEC54921.2022.9951915.

73. Marcos, A.P.; Rodovalho, N.L.S.; Backes, A.R. Coffee leaf rust detection using convolutional neural network. 2019 XV Workshop de Visão Computacional (WVC). IEEE, 2019, pp. 38–42. https://doi.org/10.1109/WVC.2019.8876931.

74. Javierto, D.P.P.; Martin, J.D.Z.; Villaverde, J.F. Robusta Coffee Leaf Detection based on YOLOv3-MobileNetv2 model. 2021 IEEE 13th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM). IEEE, 2021, pp. 1–6. https://doi.org/10.1109/HNICEM54116.2021.9731899.

75. Geddam, L.S.; Mungara, A.; Kapavari, K.; Jayarama, K.; Tripathi, S. Detection of White Stem Borer Disease in Coffee Plantation using Autonomous Multi Terrain Robot. 2023 19th IEEE International Colloquium on Signal Processing & Its Applications (CSPA). IEEE, 2023, pp. 230–235. https://doi.org/10.1109/CSPA57446.2023.10087558.

76. Lelis, A.K.; Ferriols, E.G.I.; Vallesteros, K.M.A.; Delmo, J.A.B. A Comparative Analysis of Convolutional Neural Network Architectures for Coffee Leaf Rust Detection. 2023 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS). IEEE, 2023, pp. 213–218. https://doi.org/10.1109/I2CACIS57635.2023.10193074.

77. Faisal, M.; Leu, J.S.; Darmawan, J.T. Model Selection of Hybrid Feature Fusion for Coffee Leaf Disease Classification. *IEEE Access* **2023**. https://doi.org/10.1109/ACCESS.2023.3286935.

78. Grimaldo, G.; Rodriguez, H.; Cabrera, V.L. Convolutional Neural Network Model for the Detection of Diseases and Pests in Coffee Crops. 2022 8th International Engineering, Sciences and Technology Conference (IESTEC). IEEE, 2022, pp. 684–690. https://doi.org/10.1109/IESTEC54539.2022.00113.

79. Yamashita, J.V.Y.B.; Leite, J.P.R. Coffee disease classification at the edge using deep learning. *Smart Agricultural Technology* **2023**, *4*, 100183. https://doi.org/10.1016/j.atech.2023.100183.

80. Tassis, L.M.; Krohling, R.A. Few-shot learning for biotic stress classification of coffee leaves. *Artificial Intelligence in Agriculture* **2022**, *6*, 55–67. https://doi.org/10.1016/j.aiia.2022.04.001.

81. Milke, E.B.; Gebiremariam, M.T.; Salau, A.O. Development of a coffee wilt disease identification model using deep learning. *Informatics in Medicine Unlocked* **2023**, *42*, 101344. https://doi.org/10.1016/j.imu.2023.101344.

82. Tuesta-Monteza, V.A.; Mejia-Cabrera, H.I.; Arcila-Diaz, J. CoLeaf-DB: Peruvian coffee leaf images dataset for coffee leaf nutritional deficiencies detection and classification. *Data in Brief* **2023**, *48*, 109226. https://doi.org/10.1016/j.dib.2023.109226.

83. Abuhayi, B.M.; Mossa, A.A. Coffee disease classification using Convolutional Neural Network based on feature concatenation. *Informatics in Medicine Unlocked* **2023**, *39*, 101245. https://doi.org/10.1016/j.imu.2023.101245.

84. Esgario, J.G.; de Castro, P.B.; Tassis, L.M.; Krohling, R.A. An app to assist farmers in the identification of diseases and pests of coffee leaves using deep learning. *Information Processing in Agriculture* **2022**, *9*, 38–47. https://doi.org/10.1016/j.inpa.2021.01.004.

85. Jepkoech, J.; Mugo, D.M.; Kenduiywo, B.K.; Too, E.C. Arabica coffee leaf images dataset for coffee leaf disease detection and classification. *Data in brief* **2021**, *36*, 107142. https://doi.org/10.1016/j.dib.2021.107142.

86. Sorte, L.X.B.; Ferraz, C.T.; Fambrini, F.; dos Reis Goulart, R.; Saito, J.H. Coffee leaf disease recognition based on deep learning and texture attributes. *Procedia Computer Science* **2019**, *159*, 135–144. https://doi.org/10.1016/j.procs.2019.09.168.

87. Tassis, L.M.; de Souza, J.E.T.; Krohling, R.A. A deep learning approach combining instance and semantic segmentation to identify diseases and pests of coffee leaves from in-field images. *Computers and Electronics in Agriculture* **2021**, *186*, 106191. https://doi.org/10.1016/j.compag.2021.106191.

88. Ventura, J.; Esgario, J.; Krohling, R. Deep learning for classification and severity estimation of coffee leaf biotic stress. *Computers and Electronics in Agriculture* **2020**, *169*. https://doi.org/10.1016/j.compag.2019.105162.

89. de Oliveira Aparecido, L.E.; Lorençone, P.A.; Lorençone, J.A.; Torsoni, G.B.; de Lima, R.F.; Padilha, F.; de Souza, P.S.; de Souza Rolim, G. Addressing coffee crop diseases: forecasting Phoma leaf spot with machine learning. *Theoretical and Applied Climatology* **2023**, pp. 1–22. https://doi.org/10.1007/s00704-023-04739-z.

90. de Oliveira Aparecido, L.E.; de Souza Rolim, G.; da Silva Cabral De Moraes, J.R.; Costa, C.T.S.; de Souza, P.S. Machine learning algorithms for forecasting the incidence of Coffea arabica pests and diseases. *International Journal of Biometeorology* **2020**, *64*, 671–688. https://doi.org/10.1007/s00484-019-01856-1.

91. Selvanarayanan, R.; Rajendran, S. Roaming the Coffee Plantations Using Grey Wolves Optimisation and the Restricted Boltzmann Machine to Predict Coffee Berry Disease. 2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS). IEEE, 2023, pp. 681–689. https://doi.org/10.1109/ICSSAS57918.2023.10331629.

92. Caballero, E.M.T.; Duke, A.M.R. Implementation of artificial neural networks using nvidia digits and opencv for coffee rust detection. 2020 5th International Conference on Control and Robotics Engineering (ICCRE). IEEE, 2020, pp. 246–251. https://doi.org/10.1109/ICCRE49379.2020.9096435.

93. Latif, M.A.; Afshan, N.; Mushtaq, Z.; Khan, N.A.; Irfan, M.; Nowakowski, G.; Alqhtani, S.M.; Mursal, S.; Telenyk, S. Enhanced classification of coffee leaf biotic stress by synergizing feature concatenation and dimensionality reduction. *IEEE Access* **2023**. https://doi.org/10.1109/ACCESS.2023.3314590.

94. Parraga-Alava, J.; Cusme, K.; Loor, A.; Santander, E. RoCoLe: A robusta coffee leaf images dataset for evaluation of machine learning based methods in plant diseases recognition. *Data in brief* **2019**, *25*, 104414. https://doi.org/10.1016/j.dib.2019.104414.

95. Jindal, V.; Kukreja, V.; Bhattacherjee, A.; Rana, S.; Mehta, S. Agricultural Innovation: Unleashing Federated Learning CNNs on Coffee Leaf Disease Severity Analysis. *Communication, and Intelligent Systems (ICCCIS)* **2023**, pp. 782–787. https://doi.org/10.1109/ICCCIS60361.2023.10425426.

96. Marin, D.B.; Santana, L.S.; Barbosa, B.D.S.; Barata, R.A.P.; Osco, L.P.; Ramos, A.P.M.; Guimarães, P.H.S.; others. Detecting coffee leaf rust with UAV-based vegetation indices and decision tree machine learning models. *Computers and Electronics in Agriculture* **2021**, *190*, 106476. https://doi.org/10.1016/j.compag.2021.106476.

97. Ruttanadech, N.; Phetpan, K.; Srisang, N.; Srisang, S.; Chungcharoen, T.; Limmun, W.; Youryon, P.; Kongtragoul, P. Rapid and accurate classification of Aspergillus ochraceous contamination in Robusta green coffee bean through near-infrared spectral analysis using machine learning. *Food Control* **2023**, *145*, 109446. https://doi.org/10.1016/j.foodcont.2022.109446.

98. Kulkarni, S.; N C, K.; C K, S.; Pal, S.; Dash, S.; Shenoy, P.D.; K R, V. Coffee Plant Disease Identification using Enhanced Short Learning EfficientNetV2. 2023 IEEE 20th India Council International Conference (INDICON), 2023. https://doi.org/10.1109/INDICON59947.2023.10440883.

99. Kulkarni, S.; H N, S.; V, V.M.; Shenoy, P.D.; R, V.K. Detection of Coffee Leaf Diseases using DenseNet-264 and SLIC Segmentation. 2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), 2023, Vol. 10, pp. 769–774. https://doi.org/10.1109/UPCON59197.2023.10434875.

100. Sosa, J.; Ramírez, J.; Vives, L.; Kemper, G. An Algorithm For Detection of Nutritional Deficiencies from Digital Images of Coffee Leaves Based on Descriptors and Neural Networks. 2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA), 2019, pp. 1–5. https://doi.org/10.1109/STSIVA.2019.8730286.

101. Rajus, S.; Bhagavan, S.G.; Kharva, H.; Rao, S.; Olsson, S.B. Behavioral Ecology of the Coffee White Stem Borer: Toward Ecology-Based Pest Management of India's Coffee Plantations. *Frontiers in ecology and evolution* **2021**, *9*, 607555. https://doi.org/10.3389/fevo.2021.607555.