Quality Control Process of Cocoa Beans Through Computer Vision: Concept

Rodrigo Espinal[1,2], Jóse Luis Ordoñez-Avila[1,2], Angie Carolina Santos Fernández[1,2] and Maria Elena Perdomo[1,2]

[1]Faculty of Engineering, Central American University of Technology, (Unitec), San Pedro Sula 21102, Honduras
[2]Faculty of Engineering, Central American University of Technology, (Unitec), San Pedro Sula 21102, Honduras
[3]Faculty of Engineering, Central American University of Technology, (Unitec), San Pedro Sula 21102, Honduras
[4]Faculty of Engineering, Central American University of Technology, (Unitec), San Pedro Sula 21102, Honduras

Abstract
Cocoa beans are often manually classified according to their quality status, a process that can be time-consuming and prone to human error. The aim of this research was the development of a data acquisition system using artificial vision for the elaboration of a diffuse neural network. This research includes a review of cocoa processes, quality tests for beans, open-source computer vision libraries, and adaptive neuro-fuzzy systems. The algorithm was tested using an adaptive neuro-fuzzy inference system (ANFIS) with a fuzzy interface in the MATLAB mathematical application. The Gaussian membership function was used and the network training consisted of 500 epochs. In the test, 24 beans were evaluated and 22 were correctly classified, resulting in an accuracy rate and an F1 score of 92%. These results suggests that our approach using computer vision is a viable method for classifying cocoa beans their physical defects or deformities.

Keywords
computer vision, algorithm, neuro-fuzzy networks, cocoa beans

1. Introduction

There are several procedures currently used to assess the quality of cocoa beans. One of the methods is known as the cut-test, which is based on color changes registered during fermentation. This test is used to determine if the bean is properly fermented and to ensure its quality [1]. Another method is a visual test where the person in charge observes the outside of the bean for physical defects and determines its quality as good or bad. However, these methods are quite subjective as they depend on the farmer's experience and judgment. Cocoa processing involves several stages, including harvesting, fermentation, drying, and storage. During harvesting, ripe cobs are removed from trees and opened to extract the moist cocoa beans [2]. Fermentation helps eliminate slime or mucilage and is the stage where biochemical transformations occur that reduce bitterness and trigger internal reactions, which modifies the composition of the

rodriespinal@gmail.com (R. Espinal); jlordoñez@unitec.edu (J. L. Ordoñez-Avila); angfernandez13@unitec.edu (A. C. S. Fernández); maria_perdomo@unitec.edu (M. E. Perdomo)
0000-0002-4964-8604 (R. Espinal); 0000-0002-3537-1739 (J. L. Ordoñez-Avila); 0000-0002-6059-0650 (M. E. Perdomo)

cocoa beans and promotes the formation of aroma and flavor precursors [3]. Finally, during the drying stage, humidity is reduced, and the formation of flavor and aroma is completed [4].

In this study, we proposed a computer vision algorithm for feature extraction of cacao bean images and an artificial neural fuzzy inference system to classify the cocoa beans according to their quality. Fuzzy logic, proposed by Zadeh in 1965 [5], is a popular computing framework that uses fuzzy set theory, fuzzy if-then rules, and fuzzy reasoning. By integrating the fuzzy systems with ANN models, an effective tool is obtained that takes advantage of the learning characteristics of the ANN models and performs equally well as an inference fuzzy model. As proposed by Jang [6], an artificial neural fuzzy inference system takes inputs and fuzzifies them using membership functions. The objective of this research was to develop a computer vision algorithm and an artificial neural fuzzy inference system that helps in the quality control process of cocoa beans by classifying them based on their status. The first section of this experiment is the Method, where the methodology of this experiment is explained. The second section describes the feature extraction of the cocoa bean's with the algorithm, including data collection and information. The third section explains the development and training of the artificial neural fuzzy inference system model. In the final section, the model is evaluated, and the results are analyzed.

1.1. State of the Art

In the cacao sector, there are numerous investigations regarding the quality of cacao beans. Bueno [7] stated that the demand for high-quality cacao products is expected to rise due to advances in technology. Computer vision has made great strides [8] in the agricultural sector. Computer vision solutions can help sort produce by weight, color, size, maturity, and identify defects, among other factors [9]. Combining multivariate statistics with image analysis has become a dominant tool to deal with several problems in the food sector. Multiple techniques are used for classifying cocoa beans by their quality. One such technique involves using a multiclass ensemble and least-squares support vector machine based on color features, as shown in the work of [1]. Feature extraction is an example of artificial intelligence that can be used to reduce the amount of data under processing while still maintaining the fundamental data [10]. In a study by [11], feature extraction was used for cocoa bean digital image classification prediction for smart farming applications. Additionally, [12] extracted and analyzed physical features of cocoa beans using image processing and a pre-trained neural network. In [13], an SVM-classifier was used to classify cocoa beans by their fermentation degree. Artificial neural networks (ANNs) have been used in various works, such as those of [14] and [15], as instruments to model non-linear trends within the data where there are complicated relationships to be modeled. ANNs are useful when the theoretical relationship between input and output variables is lacking, such as in the case of fermentation index and color of cocoa beans in [14]. Backpropagation and Principal Component Analysis (PCA) Artificial Neural Networks were used in the work of [15] to classify cocoa beans by their quality. Moreover, in [9], image analysis combined with a random forest algorithm was used to classify cocoa beans by the grade of fermentation.

Cocoa Bean Quality	Images
Good	
Flat	
Broken	
United	
Sprouted	

Figure 1: Cocoa Bean

2. Method and Data

This experiment was conducted in three stages: data collection and feature extraction, algorithm and adaptive neuro-fuzzy inference system development, and evaluation. In the data collection stage, we gathered photographs of cocoa beans with and without physical deformations from the Association of Agro-forestry Producers of the Choloma River basin. A total of 198 photographs were collected and grouped based on the bean's defects or deformations. While most of the beans were of good quality, some had defects or deformations such as flat, united, broken, or sprouted cocoa beans.

During the algorithm development stage, we used a statistical sample of 174 cocoa beans to collect, analyze, and train an adaptive neuro-fuzzy inference system. The aim was to predict the quality status of the cocoa beans based on their contour area, surface area, and perimeter.

Specifically, the adaptive neuro-fuzzy inference system was designed to predict whether the beans were of good or bad quality, meaning they had any of the deformations mentioned earlier. We used the information gathered from the statistical sample to make the prediction.

In the evaluation phase, we analyzed the results of the predictions using four categories: true positives (correctly evaluated beans), true negatives (correctly evaluated beans), false positives (incorrectly evaluated beans), and false negatives (incorrectly evaluated beans). We used these results to calculate several metrics, including the model's accuracy (1), sensitivity (2), specificity (3), false positive rate (4), and F1-Score (5-7), using the method described by [16].

$$Accuracy = \frac{TP + TN}{TotalSample} \tag{1}$$

$$TruePositiveRate(Sensitivity) = \frac{TP}{FN + TP} \tag{2}$$

$$TrueNegativeRate(Specificity) = \frac{TN}{TN + FP} \tag{3}$$

$$FalsePositiveRate = \frac{FP}{TN + FP} \tag{4}$$

To obtain the F1 score, we calculated precision and recall and passed them to the formula described in [16] to evaluate the model's performance. Finally, we evaluated the performance and accuracy of the neuro-fuzzy inference system using 24 out-of-sample observations.

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$F1\ \text{Score} = \frac{2 * Precision * Recall}{Precision + Recall} \tag{7}$$

2.1. Data Collection

The data collection process for this experiment was a crucial step in ensuring the accuracy and reliability of the results. To gather a representative sample of cocoa beans, we visited the Association of Agroforestry Producers of the Choloma River basin, a group of farmers who specialize in producing high-quality cacao beans. With the cooperation of the farmers, we were able to collect a total of 198 photographs of cocoa beans, carefully selected to represent both good quality and poor quality beans. To ensure consistency in our data collection process, all photographs were taken at the same distance and under the same lighting conditions, minimizing any potential sources of variability that could affect the quality of the images. Furthermore, the poor-quality beans were classified based on their physical deformations, which included flat cocoa beans, multiple united cocoa beans, broken cocoa beans, and sprouted cocoa beans, as shown in Figure 1. By classifying the poor-quality beans in this way, we were able to capture a range of common defects that can affect the quality of cocoa beans.

2.2. Data Preprocessing

During the data preprocessing step, we aimed to ensure that the extracted features were accurate and reliable. To achieve this, we performed several operations on the collected images. Firstly, we resized the images to a uniform size to facilitate their analysis. Then, we applied several filters, including gray-scale, threshold, and median blur, to enhance the image quality and facilitate feature extraction. We also used the Canny edge detection algorithm to accurately detect the contour of the cocoa beans. This allowed us to extract the contour area, surface area, and perimeter of each bean using the OpenCV functions cv.contourArea, cv.arcLength, and np.histogram, as shown in Figure 2. These features were used as inputs to the neuro-fuzzy interface, which was designed to classify the beans as either good or bad based on their physical deformations. To train the neuro-fuzzy interface, we used a statistical finite sample of 174 beans, which were representative of the population of cocoa beans in the Choloma River basin. This dataset was used to calibrate the interface and to ensure that it could accurately predict the quality status of the beans.

2.3. Artificial Neural Fuzzy Inference System

After the data preprocessing step, the extracted features were utilized to develop an artificial neural fuzzy inference system. This system was designed to classify the cocoa beans based on their physical deformations as either good quality or bad quality. The fuzzy plugin in Matlab was used to implement this system. To train the ANFIS system, a set of input and output vectors were utilized. These vectors were used to find the premise parameters for the membership functions. In an artificial neural fuzzy inference system, the membership function is used to represent the degree to which an input value belongs to a particular class. The Gaussian member function was used for this system. It has a smooth curve and utilizes only two parameters: c for locating center and σ for determining the width of the curve, as expressed mathematically in (8). The fuzzy neural network was trained for 500 epochs to obtain the network shown in Figure 3. Each input variable has its own set of membership functions, one for each class. These membership functions are often graphically represented, with the x-axis representing the input value and the y-axis representing the membership degree.

$$f(x;\sigma,c) = \frac{e^{-(x-c)^2}}{2\sigma^2} \tag{8}$$

The ANFIS system generated 27 rules for handling quantitative formulation between contour area, surface area, and bean perimeter in predicting the quality of cocoa beans. The relation between the inputs and the outputs is shown in Figure 4. Overall, the artificial neural fuzzy inference system was trained to predict whether the beans were of good quality or bad quality, using the information gathered from the statistical sample.

3. Results

To evaluate the performance of the artificial neural fuzzy inference system models, a preliminary test was conducted using 24 cocoa beans that were not included in the training sample. The

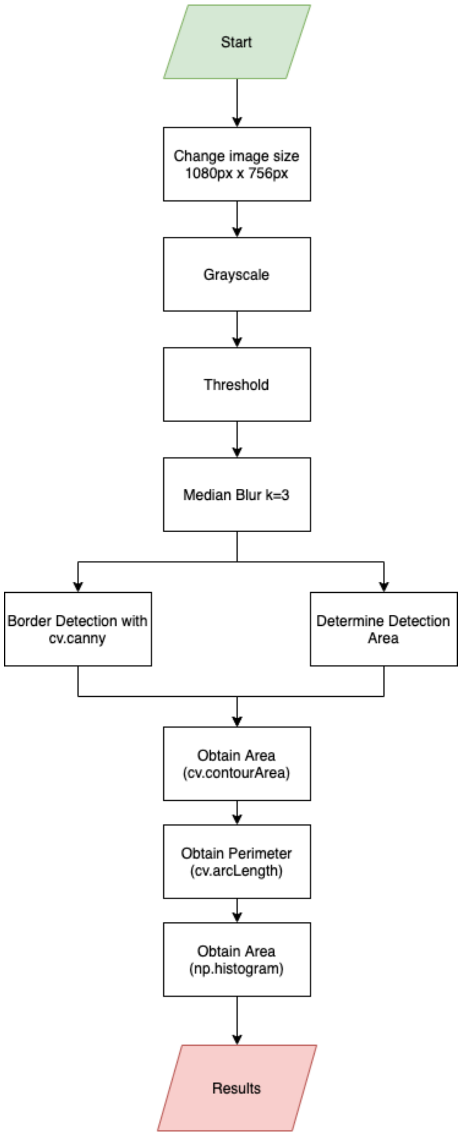

Figure 2: Algorithm Flowchart

Gaussian membership function was found to perform optimally, correctly classifying 14 beans as positive (good) and 10 as negative (bad). Out of the 14 good beans, 12 were correctly identified as true positives (TP), while all 10 bad beans were correctly identified as true negatives (TN). The model produced 2 false positives and 0 false negatives. A summary of these results can be found in Table I and Figure 5.

The model achieved an F1-Score of 92%, demonstrating its ability to accurately classify cocoa beans as either free of defects or with defects. Table II provides a summary of the evaluation results.

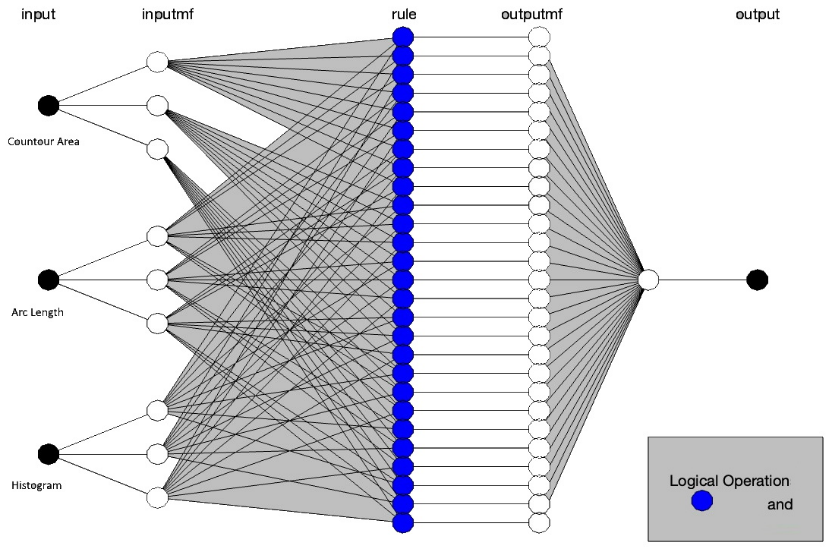

Figure 3: Artificial Neural Fuzzy Inference System

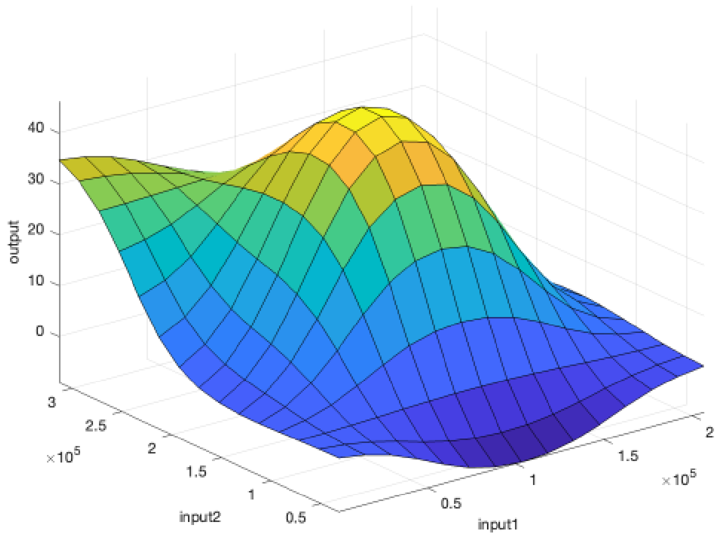

Figure 4: Surface graph showing the inputs and outputs based con the membership function and fuzzy rules. a) Arc Length and Histogram. b) Histogram and Contour c) Arc Length and Contour

4. Conclusions

A computer vision algorithm was developed to classify cocoa beans according to their quality based on physical defects. The algorithm uses a classification artificial neural fuzzy infer-

Table 1
Prediction Results

Metric	Amount of Cocoa Beans
True Positive	12
True Negative	10
False Positive	2
False Negative	0
Found Positive	14
Found Negative	10
Total Beans in Eval	24

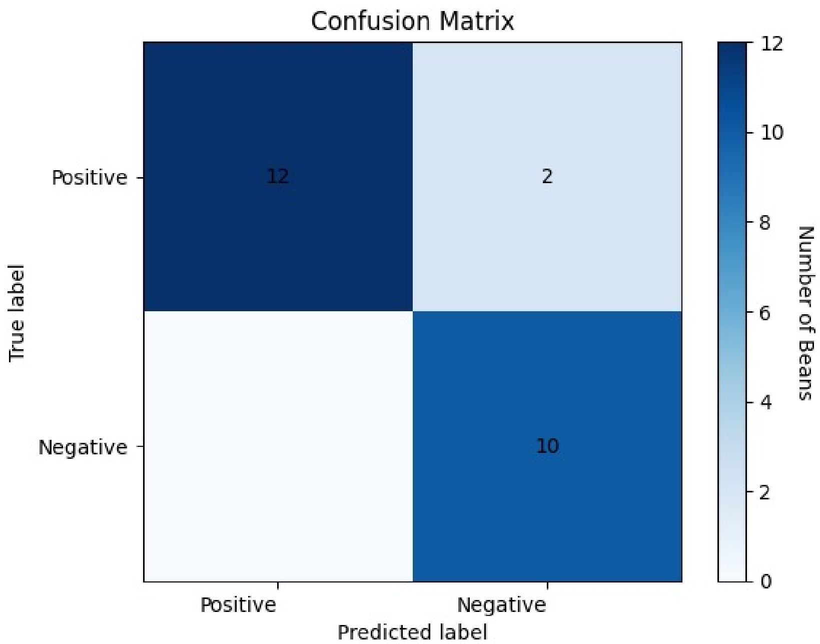

Figure 5: Confusion matrix summarizing the prediction results for the classification of cocoa beans based on their physical defects or deformities. The x-axis represents the predicted values while the y-axis represents the true values.

ence system to calculate bean physical characteristics such as contour area, surface area, and perimeter, and then uses this information to classify the beans as either free of defects or with defects. The algorithm was found to have an accuracy level of 92%, a sensitivity score of 100%, a specificity score of 83% and an F1-Score of 92%. These results compare very well to previous studies on defect detection in agro-products in Honduras, including the use of a neural network for detecting red ring pest in oil palm [17], which had an accuracy of 98%; the identification of coral beef disease using computer vision [18], which had an accuracy of 94%; and the detection of coffee rust [19], which had an accuracy of 96%. This difference is because the authors used a neural network to make their applications.

One advantage of this study is that it can be implemented at a low cost, as it only requires

Table 2
Evaluation Results

Evaluation	Membership Function Gaussmf Results
Accuracy	92%
Sensibility	100%
Specificity	83%
False Positive Rate n	17%
Presicion	86%
Recall	100%
F1 Test	92%

a phone camera to take pictures of the cocoa beans. Another advantage is that it eliminates subjectivity in the classification of cocoa beans, as the decision is based on concrete and enumerable data rather than human perception. Future work could focus on improving the accuracy of the algorithm by incorporating more training data and fine-tuning the parameters of the fuzzy neural network, or by exploring other machine learning techniques and considering additional factors that may affect cocoa bean quality. By automating this process, we aim to optimize time and enhance the quality of cocoa bean classification.

References

[1] Y. Adhitya, A. Lawi, H. Hartono, M. Mursalin, Cocoa Beans Digital Image Classification Based On Color Features using Multiclass Ensemble Least-Squares Support Vector Machine, ???? doi:10.4108/eai.20-1-2018.2281929.

[2] G. A. R. Wood, R. A. Lass, Cocoa, Tropical agriculture series, 4th ed., repr. with corrections ed., Longman Scientific & Technical ; Wiley, ????

[3] M. C. López-Navarrete, El proceso de fermentación del CACAO (theobroma cacao l.) 4 (????). URL: https://revista-agroproductividad.org/index.php/agroproductividad/article/view/572, number: 1.

[4] J. Nogales, L. Graziani de Fariñas, L. Ortiz de Bertorelli, Cambios físicos y químicos durante el secado al sol del grano de cacao fermentado en dos diseños de cajones de madera[1] 56 (????) 5–20. URL: http://ve.scielo.org/scielo.php?script=sci_abstract&pid=S0002-192X2006000100001&lng=es&nrm=iso&tlng=es, number: 1 Publisher: Instituto Nacional de Investigaciones Agrícolas (INIA).

[5] L. A. Zadeh, Fuzzy sets 8 (????) 338–353. URL: https://www.sciencedirect.com/science/article/pii/S001999586590241X. doi:10.1016/S0019-9958(65)90241-X.

[6] J.-S. Jang, ANFIS: adaptive-network-based fuzzy inference system 23 (????) 665–685. doi:10.1109/21.256541, conference Name: IEEE Transactions on Systems, Man, and Cybernetics.

[7] G. E. Bueno, K. A. Valenzuela, E. R. Arboleda, Maturity classification of cacao through spectrogram and convolutional neural network 8 (????) 228–233. URL: https://jtsiskom.undip.ac.id/index.php/jtsiskom/article/view/13733. doi:10.14710/jtsiskom.2020.13733.

[8] D. K. Shetty, U. Dinesh Acharya, N. Malarout, R. Gopakumar, P. P.J., A review of application

of computer-vision for quality grading of food products, in: 2019 International Conference on Automation, Computational and Technology Management (ICACTM), IEEE, ????, pp. 297–303. URL: https://ieeexplore.ieee.org/document/8776763/. doi:10.1109/ICACTM.2019.8776763.

[9] M. M. Oliveira, B. V. Cerqueira, S. Barbon, D. F. Barbin, Classification of fermented cocoa beans (cut test) using computer vision 97 (????) 103771. URL: https://www.sciencedirect.com/science/article/pii/S0889157520314769. doi:10.1016/j.jfca.2020.103771.

[10] R. Essah, D. Anand, S. Singh, An intelligent cocoa quality testing framework based on deep learning techniques 24 (????) 100466. URL: https://linkinghub.elsevier.com/retrieve/pii/S2665917422001003. doi:10.1016/j.measen.2022.100466.

[11] Y. Adhitya, S. W. Prakosa, M. Köppen, J.-S. Leu, Feature extraction for cocoa bean digital image classification prediction for smart farming application 10 (????) 1642. URL: https://www.mdpi.com/2073-4395/10/11/1642. doi:10.3390/agronomy10111642, number: 11 Publisher: Multidisciplinary Digital Publishing Institute.

[12] C. C. Hortinela, K. J. R. Tupas, Classification of cacao beans based on their external physical features using convolutional neural network, in: 2022 IEEE Region 10 Symposium (TENSYMP), IEEE, ????, pp. 1–5. URL: https://ieeexplore.ieee.org/document/9864337/. doi:10.1109/TENSYMP54529.2022.9864337.

[13] Y. Aubain, C. N'Zi, K. Kpalma, Cocoa beans fermentation degree assessment for quality control using machine vision and multiclass SVM classifier (????).

[14] N. León-Roque, M. Abderrahim, L. Nuñez-Alejos, S. M. Arribas, L. Condezo-Hoyos, Prediction of fermentation index of cocoa beans (theobroma cacao l.) based on color measurement and artificial neural networks 161 (????) 31–39. URL: https://www.sciencedirect.com/science/article/pii/S0039914016305914. doi:10.1016/j.talanta.2016.08.022.

[15] U. Lestari, R. A. Kumalasanti, E. S. Wulandari, Identifying the quality system of cocoa beans to increase productivity using backpropagation neural network algorithm: A case study at sumber rejeki farmers group, patuk gunung kidul 1413 (????) 012033. URL: https://iopscience.iop.org/article/10.1088/1742-6596/1413/1/012033. doi:10.1088/1742-6596/1413/1/012033.

[16] A. Géron, Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems, second edition ed., O'Reilly Media, Inc, ????

[17] O. A. Fernandez, J. L. Ordóñez-Ávila, I. A. Magomedov, Evaluation of parameters in a neural network for detection of red ring pest in oil palm 2442 (????) 030015. URL: https://aip.scitation.org/doi/abs/10.1063/5.0076481. doi:10.1063/5.0076481, publisher: American Institute of Physics.

[18] G. A. Bautista-Hernández, D. A. Jimenez-Nixon, A. M. Reyes-Duke, Coral reef disease and bleaching indentification through computational vision algorithm, in: 2022 IEEE Central America and Panama Student Conference (CONESCAPAN), ????, pp. 1–6. doi:10.1109/CONESCAPAN56456.2022.9959584.

[19] E. M. T. Caballero, A. M. R. Duke, Implementation of artificial neural networks using NVIDIA digits and OpenCV for coffee rust detection, in: 2020 5th International Conference on Control and Robotics Engineering (ICCRE), ????, pp. 246–251. doi:10.1109/ICCRE49379.2020.9096435.