International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Supervised Model for the Detection of Coffee Leaf Diseases by Image Analysis

Camilo-Enrique Rocha-Calderón*[1], Julio Barón-Velandia[2], Sebastian-Camilo Vanegas-Ayala[3]

Submitted: 26/04/2023 **Revised**: 27/06/2023 **Accepted**: 07/07/2023

Abstract: An alternative model is presented for the early detection of coffee leaf rust disease, since this disease usually causes yield losses of up to 30% during the pandemic season, considering that the traditional detection method known as direct observation requires monetary resources that the farmer does not have. There are several technological alternatives to detect plant diseases in a short time, which are accurate but not very interpretable, therefore a supervised image analysis model is generated at pixel level according to the EG and GCC color indices to detect coffee leaf rust based on a fuzzy inference system, being this developed under experimental and prototype-based methodologies. The model obtained an accuracy of 93.75%, being considered effective and ready to be tested in uncontrolled environments, where the GCC color index presents a better discrimination of the state of a plant against the EG.

Keywords: *Coffee, coffee rust, fuzzy inference system, image analysis, machine learning*

1. Introduction

The identification of plant diseases is one of the most important activities in the agricultural sector, since this action, developed early, allows the application of a treatment adapted to the crop, the disease and its stage [1]. The importance of this activity is related to the economic loss that can occur if the necessary measures are not taken to prevent and treat the diseases that affect the sector worldwide, considering that, according to the FAO, US$ 220 billion are lost annually [2].

One way to reduce the costs associated with crop yield loss is to use mechanisms or techniques that allow early detection of diseases that occur in a given crop [3]. The most common mechanism for disease detection is direct observation, which consists of a crop expert being able to identify the appearance of typical characteristics of the disease through his visual sense; however, this technique is not very efficient because the verification of all the plants of the crop, depending on the extent of the sown field, is delayed and usually not performed [4][5]. In addition, for the farmer, the use of a crop expert represents a high monetary cost that most of them cannot assume, even more so in developing regions or countries where the presence and support of the State to farmers is almost nonexistent [6], making them rely on past experiences or rumors to identify and treat diseases with diagnoses that may be wrong, generating an environmental cost framed by the deterioration of arable land in the future, of nearby water sources and high probability of not being able to eradicate the disease [7].

1 Faculty of Engineering, Universidad Distrital Francisco José de Caldas, Bogotá D.C. - 110110, Colombia ORCID ID : 0000-0002-6395-822X
2 Faculty of Engineering, Universidad Distrital Francisco José de Caldas, Bogotá D.C. - 110110, Colombia ORCID ID : 0000-0002-9491-5564
3 Faculty of Engineering, Universidad Distrital Francisco José de Caldas, Bogotá D.C. - 110110, Colombia ORCID ID : 0000-0002-8610-9765
** Corresponding Author Email: cerochac@correo.udistrital.edu.co*

The agricultural sector has been involved in the application of technological advances in its various fields of activity, among which the detection of plant diseases through machine learning and deep learning algorithms stands out [8], supported by the estimated figure corresponding to $5098 million in the smart agriculture market for the year 2016 [1].

The automated detection of diseases has increased its research and use in recent years [9] for different reasons and consequences, one of the main ones being the depression of crop yields and economic losses suffered by farmers, regardless of the scale of the affection of the same, ranging from 5% to 80% depending on the plant species, type of disease, level of infection, climate and soil [10][11], for example in the strawberry crop some diseases affect the photosynthetic activity of the plant, which are directly involved in the quality of the fruit, representing a yield loss of 30% to 70% [12].

One of the challenges of precision agriculture is related to the lack of data, the pre-processing and processing required in research using digital image processing and artificial vision techniques, and the comparability of results by an expert [13].

Artificial intelligence models and research results focused on disease detection are usually implemented in machine learning and deep learning-type image analysis techniques, such as Support Vector Machines [14], KNN [15], Neural Networks [16], Convolutional Neural Networks [17], ANFIS [18], Evolutionary Algorithms [19], Correlations [20], Generative Neural Networks [21], among others [2][14].

The models currently implemented usually have a low level of significance or interpretability and high accuracy, so this research seeks to have an initial model oriented to present high accuracy and interpretability [22][23]. The supervised model resulting from this research is focused on the use of fuzzy inference systems to detect the rust disease in the coffee crop, which represents up to 30% of the yield loss of the crops that suffer from it [24], where through digital image processing of the crop leaf is analyzed at the pixel level using color indices that determine whether the state of the leaf is healthy or diseased. The color indices chosen are EG (Extreme Green) and GCC (Green Chromatic Coordinates), which are proposed to differentiate the state of a plant using visual techniques [25].

The paper consists of the methodological technique implemented, the design and implementation of the model, the results obtained, and the conclusions.

2. Methodology

The methodological technique used for the development of the research has two components, as developed in [26]: the one related to the development of the software required for the functional model and the one related to its operability. In the case of software development, taking into account that it is an artificial intelligence model and the adaptations to be made, a methodology based on prototypes is established, which is empirical, experimental and iterative-incremental, so that the functionalities can be analyzed, designed, tested and evaluated on the basis of preset acceptance parameters, following the phases of communication, rapid plan, modeling and rapid design, construction of the prototype and, deployment and feedback of the prototype [27].

On the other hand, the operational methodology is proposed as experimental, which allows to adapt the case study environment to obtain an expected result, looking for the relevant characteristics and properties in the experiment [28]. Therefore, the basis of the proposed methodology is based on image analysis at the pixel level, showing the state of a coffee leaf.

As shown in Figure 1, the procedure to be followed consists of a flow that starts with the respective loading of the images, which are then analyzed at the pixel level under the EG and GCC color indices, from which the characteristic values are extracted to identify a pixel as healthy or diseased, allowing to configure the input and output sets of a fuzzy logic system designed for early disease detection. The following stages are available:

2.1. Data Loading

The first step corresponds to the loading of images of healthy and rust affected coffee leaves, focused directly on the area of interest. The data set used corresponds to "BRACOL" [29], from which 96 examples were taken for this research.

2.2. Pixel-level analysis

After loading the data, 48 images are used in this phase: 24 representing the healthy state and 24 representing the diseased state, in which a pixel representative of its state is selectively selected, from which the RGB color components are extracted in order to subsequently calculate the characteristic values of the EG and GCC color indices at pixel level.

2.3. Establish independent variables and ranges

Once the values of the indices per pixel representative of the 48 images have been calculated, ranges are defined to define whether a pixel corresponds to a representation of the diseased leaf or to an optimal state. Defining the ranges selects which of the color indices best describes the state of a pixel.

2.4. Parameterize the model

A model based on a fuzzy inference system is parameterized from the intervals of the selected color index determined in the previous step, defined in the input set. In contrast, the output set corresponds to the resulting pixel state.

2.5. Testing the model

Once the structure of the fuzzy model is established, it is tested on the 48 examples not selected for the "pixel-level analysis" phase; that is, the validation is performed with observations the system has not seen. The test consists of selecting one characteristic pixel per image, feeding it to the model, and verifying that the resulting state matches the state represented by the image.

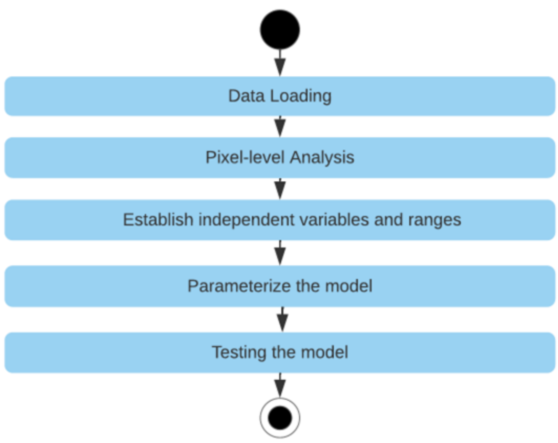

Fig. 1. Operational methodology.

3. Design and implementation

Taking into account the operational methodology described in the previous section, the images of the "BRACOL" dataset are first loaded into the scientific development environment Matlab; in Figure 2, two observations of the dataset are shown; on the left a leaf infected by rust and on the right a healthy leaf.

Fig. 2. Leaf affected by rust vs. healthy leaf. Taken from: [29].

Then, from 48 images, one pixel characterizing the image type (healthy or diseased leaf) is selected to extract its R, G, and B color components, values necessary to calculate the EG and GCC color indices, which are proposed to differentiate a healthy leaf from a diseased one [25].

$$EG = 2G - (R + B) \quad (1)$$

$$GCC = G/(R + G + B) \quad (2)$$

It should be noted that although the stages of a coffee leaf in the study of rust disease mentioned so far correspond to healthy or diseased, there are different stages of the disease and the state of the leaf that present color variations, so 4 subclasses are defined to verify the state of the pixels, as shown in Table 1.

Table 1. Categorization of pixels by color

Class	Color	Category
1	Dark green	Health
2	Green Medium light	Health
3	Orange	Sick
4	Coffee	Sick

Table 2 shows a fragment of the values corresponding to the pixels identified in each of the 48 reference images, consisting of the R, G, B color components and the EG and GCC color indices.

Table 2. Characterization of pixels according to their category and EG and GCC color indices

Píxel	Pixel 1	Pixel 2	Pixel 3	Pixel 4
Class	1	2	3	4
R	127	153	255	45
G	144	162	207	28
B	12	35	53	10
EG	149	136	106	1
GCC	0,5088	0,4629	0,4019	0,3373

Based on the characteristic values found in the 48 pixels parameterized in the EG and GCC color indices, the intervals necessary to identify a pixel as healthy or diseased are determined, which are necessary to define the membership functions in the input and output sets of the fuzzy inference system.

4. Results and Discussion

The results presented in this paper are segmented into three, presenting the characteristic values found to define the intervals of the fuzzy sets, the supervised model based on a fuzzy inference system obtained, and the comparison of the classification obtained in the test pixels resulting from the model versus the expected.

4.1. Characteristic values

From the analysis of the 48 initial pixels under the EG and GCC color indices, intervals of values were extracted that determine when a pixel represents a coffee leaf with rust disease or in a healthy state:

- For the EG color index, the intervals for characterizing the state of a pixel were established, as shown in Table 3.

Table 3. Determination of the categorization intervals of a pixel according to the EG color index

Observation /class	1	2	3	4
1	149	136	106	10
2	107	112	101	1
3	99	109	87	11
4	226	215	160	9
5	106	131	71	2
6	74	91	78	7
7	96	96	118	18
8	91	105	72	47
9	64	75	63	-4
10	107	129	76	4
11	97	108	66	2
12	96	109	116	-6
Minimum	64	75	63	-6
Maximum	226	215	160	47

Therefore, the healthy state of a pixel falling into categories 1 and 2 has a defined interval between 64 and 226, while the sick state is set between -6 and 160.

- For the GCC color index, the characterization intervals were defined in Table 4.

Table 4. Determination of pixel categorization intervals according to the GCC color index

Observation /class	1	2	3	4
1	0,5088	0,4629	0,4019	0,3373
2	0,5272	0,4957	0,4498	0,3615
3	0,6016	0,4977	0,434	0,3604
4	0,7216	0,5518	0,4588	0,3391
5	0,5857	0,4728	0,4006	0,3605
6	0,566	0,4744	0,4209	0,369
7	0,5814	0,4795	0,4477	0,4158
8	0,6018	0,5816	0,4346	0,3279
9	0,5446	0,4677	0,38	0,3415
10	0,5605	0,4636	0,3913	0,339
11	0,4975	0,4833	0,4019	0,2821
12	0,4836	0,4479	0,4174	0,3244
Minimum	0,4836	0,4479	0,38	0,2821
Maximum	0,7216	0,5816	0,4588	0,4158

The healthy state of a pixel falling into categories 1 and 2 has a defined interval between 0.4479 and 0.7216, while the sick state is set between 0.2821 and 0.4588.

4.2. Model Obtained

Considering the results presented above, it can be inferred that the EG color index does not adequately discriminate whether a pixel is in a healthy or diseased state, while the GCC color index can clearly discriminate between these states. Therefore, the fuzzy inference system was designed and implemented based on the GCC index, which was taken as the input set of the system, with the output set defined as the pixel state. The fuzzy inference system is a Mamdani-type set, as shown in Figure 3.

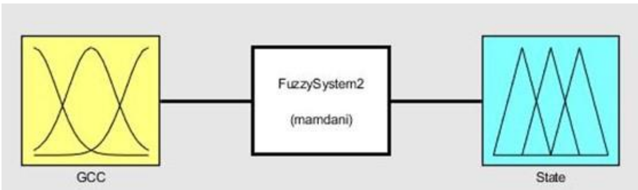

Fig. 3. Mamdani-type fuzzy inference system.

The input set shown in Figure 4 was defined with two triangular activation functions representing a high or low value of the GCC index, as follows.

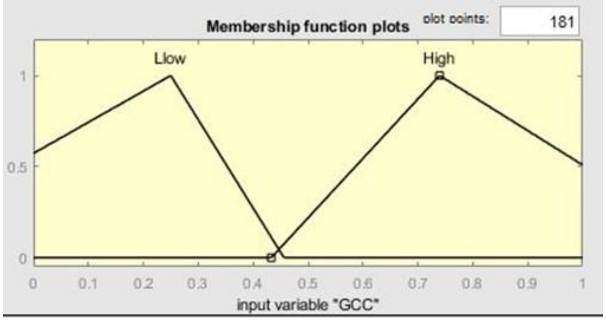

Fig. 4. Input set of the fuzzy inference system - GCC.

The output set of the fuzzy inference system, shown in Figure 5, represents the state of a pixel and presents two triangular activation functions set as healthy and sick.

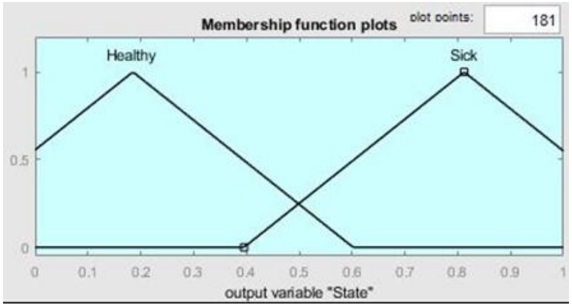

Fig. 5. Output setting of the fuzzy inference system - state.

The fuzzy inference system is associated with rules that allow establishing the relationship between the GCC variable and the pixel state.

1. If (GCC is low) then (State is sick)
2. If (GCC is high) then (State is healthy)

4.3. Results from the model

The resulting supervised model was tested on 48 previously unobserved pixels (24 with healthy and 24 with diseased state). Therefore, the state was defined for each pixel: 0 for healthy and 1 for disease, then the pixels were evaluated under the GCC color index. Once the color index values for the 48 pixels were obtained, they were passed through the model, which provided values between 0 and 1, characterized based on the fuzzy output set. The model evaluation results show an accuracy of 93.75%, with 45 correct and 3 incorrect classifications, as shown in the confusion matrix in Figure 6.

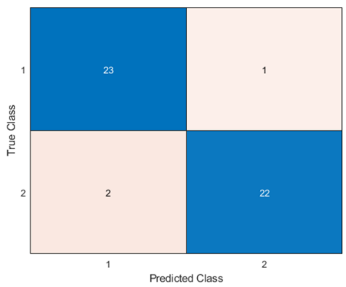

Fig. 6. Confusion matrix.

5. Conclusion

The GCC color index allows a better discrimination of the state of the pixels defined as healthy or diseased concerning the EG color index, since the latter presents a significant overlap of values between the two states.

The supervised model obtained represents an alternative to the research results found in the literature since it presents a high degree of interpretability based on fuzzy inference systems of the Mamdani type.

The model resulting from the present research provides an accuracy of 93.75% in the test phase using examples not used in previous stages, representing a high effectiveness in discriminating pixels of images of coffee leaves considered healthy or diseased due to rust.

Acknowledgements

This research has been financed by the “Centro de Investigaciones y Desarrollo Científico” of the “Universidad Distrital Francisco José de Caldas”, within the framework of the “Convocatoria 05-2022”.

We thank the faculty of engineering and the master's program in “Ciencias de la Información y las Comunicaciones” of the “Universidad Distrital Francisco José de Caldas” for their academic support.

Author contributions

Camilo-Enrique Rocha-Calderón: Conceptualization, Data curation, Writing-Original draft preparation, Methodology **Julio Barón-Velandia:** Investigation, Validation, Writing-Reviewing and Editing **Sebastian-Camilo Vanegas-Ayala:** Visualization, Investigation, Writing-Reviewing and Editing, Methodology.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] J. G. A. Barbedo, “A new automatic method for disease symptom segmentation in digital photographs of plant leaves,” Eur. J. Plant Pathol., vol. 147, no. 2, pp. 349–364, 2017, doi: 10.1007/s10658-016-1007-6.

[2] M. A. Khan et al., “An Optimized Method for Segmentation and Classification of Apple Diseases Based on Strong Correlation and Genetic Algorithm Based Feature Selection,” IEEE Access, vol. 7, no. c, pp. 46261–46277, 2019, doi: 10.1109/ACCESS.2019.2908040.

[3] V. Gonzalez-Huitron, J. A. León-Borges, A. E. Rodriguez-Mata, L. E. Amabilis-Sosa, B. Ramírez-Pereda, and H. Rodriguez, “Disease detection in tomato leaves via CNN with lightweight architectures implemented in Raspberry Pi 4,” Comput. Electron.

Agric., vol. 181, no. January, 2021, doi: 10.1016/j.compag.2020.105951.

[4] K. R., H. M., S. Anand, P. Mathikshara, A. Johnson, and M. R., "Attention embedded residual CNN for disease detection in tomato leaves," Appl. Soft Comput., vol. 86, p. 105933, 2020, doi: https://doi.org/10.1016/j.asoc.2019.105933.

[5] J. Zhu, M. Cheng, Q. Wang, H. Yuan, and Z. Cai, "Grape Leaf Black Rot Detection Based on Super-Resolution Image Enhancement and Deep Learning," Front. Plant Sci., vol. 12, no. June, pp. 1–16, 2021, doi: 10.3389/fpls.2021.695749.

[6] J. Sun, Y. Yang, X. He, and X. Wu, "Northern Maize Leaf Blight Detection under Complex Field Environment Based on Deep Learning," IEEE Access, vol. 8, pp. 33679–33688, 2020, doi: 10.1109/ACCESS.2020.2973658.

[7] D. Bienkowski, M. J. Aitkenhead, A. K. Lees, C. Gallagher, and R. Neilson, "Detection and differentiation between potato (Solanum tuberosum) diseases using calibration models trained with non-imaging spectrometry data," Comput. Electron. Agric., vol. 167, p. 105056, 2019, doi: https://doi.org/10.1016/j.compag.2019.105056.

[8] A. V. Paéz Perilla, A. Patiño Villarraga, S. Lozada Garcés, and T. F. Tabares Carreño, "Calidad agrícola," Quindío, 2018.

[9] D.-D. Leal-Lara, J. Barón-Velandia, and C.-E. Rocha-Calderón, "Interpretabilidad en el campo de la detección de enfermedades en las plantas: Una revisión," Rev. Fac. Ing., vol. 30, no. 58, pp. e13495–e13495, Nov. 2021, doi: 10.19053/01211129.V30.N58.2021.13495.

[10] A. Ramcharan, K. Baranowski, P. McCloskey, B. Ahmed, J. Legg, and D. P. Hughes, "Deep learning for image-based cassava disease detection," Front. Plant Sci., vol. 8, no. October, pp. 1–7, 2017, doi: 10.3389/fpls.2017.01852.

[11] J. Morel et al., "Exploring the potential of PROCOSINE and close-range hyperspectral imaging to study the effects of fungal diseases on leaf physiology," Sci. Rep., vol. 8, no. 1, pp. 1–14, 2018, doi: 10.1038/s41598-018-34429-0.

[12] L. Yuan et al., "Detection of anthracnose in tea plants based on hyperspectral imaging," Comput. Electron. Agric., vol. 167, p. 105039, 2019, doi: https://doi.org/10.1016/j.compag.2019.105039.

[13] P. Jiang, Y. Chen, B. Liu, D. He, and C. Liang, "Real-Time Detection of Apple Leaf Diseases Using Deep Learning Approach Based on Improved Convolutional Neural Networks," IEEE Access, vol. 7, pp. 59069–59080, 2019, doi: 10.1109/ACCESS.2019.2914929.

[14] L. M. Tassis, J. E. Tozzi de Souza, and R. A. Krohling, "A deep learning approach combining instance and semantic segmentation to identify diseases and pests of coffee leaves from in-field images," Comput. Electron. Agric., vol. 186, no. September 2020, 2021, doi: 10.1016/j.compag.2021.106191.

[15] Q. Zeng, X. Ma, B. Cheng, E. Zhou, and W. Pang, "GANS-based data augmentation for citrus disease severity detection using deep learning," IEEE Access, vol. 8, pp. 172882–172891, 2020, doi: 10.1109/ACCESS.2020.3025196.

[16] X. E. Pantazi, D. Moshou, and A. A. Tamouridou, "Automated leaf disease detection in different crop species through image features analysis and One Class Classifiers," Comput. Electron. Agric., vol. 156, pp. 96–104, 2019, doi: https://doi.org/10.1016/j.compag.2018.11.005.

[17] N. R. Deepa and N. Nagarajan, "Kuan noise filter with Hough transformation based reweighted linear program boost classification for plant leaf disease detection," J. Ambient Intell. Humaniz. Comput., vol. 12, no. 6, pp. 5979–5992, 2021, doi: 10.1007/s12652-020-02149-x.

[18] M. Sultan Mahmud, Q. U. Zaman, T. J. Esau, G. W. Price, and B. Prithiviraj, "Development of an artificial cloud lighting condition system using machine vision for strawberry powdery mildew disease detection," Comput. Electron. Agric., vol. 158, no. January, pp. 219–225, 2019, doi: 10.1016/j.compag.2019.02.007.

[19] W. Zhang, M. F. Hansen, M. Smith, L. Smith, and B. Grieve, "Photometric stereo for three-dimensional leaf venation extraction," Comput. Ind., vol. 98, pp. 56–67, 2018, doi: 10.1016/j.compind.2018.02.006.

[20] S. Kumar, B. Sharma, V. K. Sharma, H. Sharma, and J. C. Bansal, "Plant leaf disease identification using exponential spider monkey optimization," Sustain. Comput. Informatics Syst., vol. 28, 2020, doi: 10.1016/j.suscom.2018.10.004.

[21] I. Navrozidis, T. K. Alexandridis, A. Dimitrakos, A. L. Lagopodi, D. Moshou, and G. Zalidis, "Identification of purple spot disease on asparagus crops across spatial and spectral scales," Comput. Electron. Agric., vol. 148, no. October 2017, pp. 322–329, 2018, doi: 10.1016/j.compag.2018.03.035.

[22] M. Toseef and M. J. Khan, "An intelligent mobile application for diagnosis of crop diseases in Pakistan using fuzzy inference system," Comput. Electron.

Agric., vol. 153, pp. 1–11, Oct. 2018, doi: 10.1016/j.compag.2018.07.034.

[23] T. M. Matsunaga, D. Ogawa, F. Taguchi-Shiobara, M. Ishimoto, S. Matsunaga, and Y. Habu, "Direct quantitative evaluation of disease symptoms on living plant leaves growing under natural light," Breed. Sci., vol. 67, no. 3, pp. 316–319, 2017, doi: 10.1270/jsbbs.16169.

[24] J. Avelino et al., "The coffee rust crises in Colombia and Central America (2008–2013): impacts, plausible causes and proposed solutions," Food Secur., vol. 7, no. 2, pp. 303–321, Apr. 2015, doi: 10.1007/s12571-015-0446-9.

[25] A. M. Reid, W. K. Chapman, C. E. Prescott, and W. Nijland, "Using excess greenness and green chromatic coordinate colour indices from aerial images to assess lodgepole pine vigour, mortality and disease occurrence," For. Ecol. Manage., vol. 374, pp. 146–153, Aug. 2016, doi: 10.1016/j.foreco.2016.05.006.

[26] C. E. R. Calderón, O. J. S. Parra, and S. C. V. Ayala, "Water Quality Prediction and Detection of the Vibrio Cholerae Bacteria," Int. J. Adv. Sci. Eng. Inf. Technol., vol. 11, no. 6, pp. 2369–2374, 2021, doi: 10.18517/IJASEIT.11.6.13598.

[27] R. S. Pressman, Ingenieria del Software. Un Enfoque Practico. 2010.

[28] F. A. Coutelieris and A. Kanavouras, Experimentation Methodology for Engineers, 1st ed. Springer Cham, 2018.

[29] R. Krohling, J. Esgario, and J. A. Ventura, "BRACOL - A Brazilian Arabica Coffee Leaf images dataset to identification and quantification of coffee diseases and pests," vol. 1, 2019, doi: 10.17632/YY2K5Y8MXG.1.

[30] Saxena, K. ., & Gupta, Y. K. . (2023). Analysis of Image Processing Strategies Dedicated to Underwater Scenarios. International Journal on Recent and Innovation Trends in Computing and Communication, 11(3s), 253–258. https://doi.org/10.17762/ijritcc.v11i3s.6232

[31] Hernandez, A., Hughes, W., Silva, D., Pérez, C., & Rodríguez, C. Machine Learning for Predictive Analytics in Engineering Procurement. Kuwait Journal of Machine Learning, 1(2). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/124