

Using Artificial Neural Networks for Detecting Damage on Tobacco Leaves Caused by Blue Mold

Himer Avila-George*, Topacio Valdez-Morones†, Humberto Pérez-Espinosa‡, Brenda Acevedo-Juárez*, and Wilson Castro§¶

*Centro Universitario de los Valles, Universidad de Guadalajara. Ameca, Jalisco 46600, México

†Unidad de Transferencia Tecnológica Tepic. Tepic, Nayarit 63173, México

‡CONACYT - CICESE, Unidad de Transferencia Tecnológica Tepic. Tepic, Nayarit 63173, México

§Facultad de Ingeniería, Universidad Privada del Norte. Cajamarca, Cajamarca 06002, Perú

¶Centro de Investigaciones e Innovaciones de la Agroindustria Peruana. Amazonas 1061, Perú

Abstract—Worldwide, the monitoring of pests and diseases plays a fundamental role in the agricultural sustainability; making necessary the development of new tools for early pest detection. In this sense, we present a software application for detecting damage in tobacco (*Nicotiana tabacum* L.) leaves caused by the fungus of blue mold (*Peronospora tabacina* Adam). This software application processes tobacco leaves images using a pattern recognition technique known as Artificial Neural Network. For the training and testing stages, a total of 40 images of tobacco leaves were used. The experimentation carried out shows that the developed model has accuracy higher than 97% and there is no significant difference with a visual analysis carried out by experts in tobacco crop.

Keywords—*Nicotiana tabacum* L.; *Peronospora tabacina* Adam, image processing; artificial neural networks

I. INTRODUCTION

Worldwide, there is considerable interest in working towards agricultural sustainability. Therefore, the pests and diseases that attack different crops are some of the enormous challenges that must be faced. For this reason, integrated pest control is usually one of the most used approaches, where the monitoring of pests and diseases plays a fundamental role.

Tobacco (*Nicotiana tabacum* L.) is an economically important crop worldwide, which is grown in different agricultural areas around the world [1]. According to the SIAP¹, in 2015, 7,936 ha were planted in Mexico, which 1,264 ha were reported as damaged.

Blue mold (*Peronospora tabacina* Adam) is considered the main phytosanitary problem of tobacco cultivation in many countries worldwide [2]. The humid and cold climates favor the development of the fungus, which can be germinated in a period from 2 to 4 hours. This fungus can infect the tobacco plant in all its phenological stages. Initially, several localized lesions appear in different areas of the infected leaves, as the disease progresses, these lesions gather forming necrotic zones of light brown color. Fig. 1 shows a leaf of tobacco damaged by blue mold in its initial stage.

In general, farmers and agriculture experts detect diseases of their crops through visual inspection, which is based on their experience. This type of work requires continuous monitoring



Fig. 1. Tobacco leaf damaged by blue mold.

of expert people, which is often prohibitively expensive for many farmers due to the vast extensions of the plots. In several countries, there is government support for the monitoring of crops in order to guarantee agrifood sustainability. In Mexico, there is economic and technical support from some government agencies to help farmers monitor some pests and diseases^{2,3}. However, this evaluation process is tedious, time-consuming and moreover very much subjective. For this reason, the detection of diseases based on the processing of images is a very relevant research topic [3]. This technique allows non-expert people, helped by a software application, automatically detect the signals of diseases that occur in the crops.

The process of plant diseases recognition using images consists of extracting the feature information of the diseased regions in the obtained images by the use of image processing techniques, and then achieve the identification of diseases by employing some pattern recognition technique such as Neural Networks (NNs) [4], [5], Support Vector Machine (SVM) [6], [7], Genetic Algorithms (GA) [8], among others.

In particular, artificial neural networks (ANNs) stand out as a popular machine learning technique due to their relative ease of use and understanding compared to statistical methods, as well as their excellent performance in various machine learning tasks. In recent years, several studies have shown that ANNs have unique abilities such as learning and generalization which facilitates a reliable diagnosis of plant diseases [9], see Table I; in general, ANNs have a higher degree of diagnostic than other machine learning techniques.

²Mexican Service for Agroalimentary Public Health, Safety and Quality (Servicio Nacional de Sanidad, Inocuidad y Calidad Agroalimentaria). Available: <https://www.gob.mx/senasica>

³Nayarit Plant Health Committee (Comité Estatal de Sanidad Vegetal de Nayarit). Available: <http://cesavenay.org.mx/>

¹Mexican Agrifood and Fisheries Information System (Servicio de Información Agroalimentaria y Pesquera). Available: <http://infosiap.siap.gob.mx>

TABLE I. STRATEGIES FOR THE RECOGNITION OF DISEASES IN PLANTS BASED ON IMAGE PROCESSING AND NEURAL NETWORKS TECHNIQUES

Description	Technique	Accuracy	Reference
It is introduced that an algorithm for extracting lesion area and application of Probabilistic Neural Network (PNN) to classify seedling diseases such as anthracnose and frog-eye spots on tobacco leaves.	PNN	88.59%	[10]
It is presented a method to recognize wheat and grape diseases which is based on the principal component analysis (PCA) and PNN techniques. PCA is used to reduce the dimensions of the feature data to reduce the number of neurons in the input layer and to increase the speed of PNN.	PCA and PNN	95%	[4]
The authors introduced an image processing method to detect diseases which attack the pomegranate crop. The method is based on the artificial neural network (ANN) technique.	ANN	91%	[11]
Sannakki et al. presented a model to classify diseases that attack the leaves of the grape, in specific, downy mildew and powdery mildew. The model was based on the ANN technique.	ANN	100%	[5]
A method for identifying fungal diseases that affect vegetable crops is presented, the approach is based on the ANN technique.	ANN	84.11%	[12]

In this paper, with the intention of promoting the early identification of pests and the levels of damage in tobacco crops, a mobile application it is introduced to achieve the detection of damage in tobacco leaves caused by blue mold. A classification model was trained to boost this application, which is based on the multilayer perceptron ANN technique.

II. MATERIALS AND METHODS

A. Tobacco Leaves

In Mexico, more than 80% of tobacco crop is grown in the state of Nayarit, mainly in its coastal municipalities; being *Santiago Ixcuintla* the municipality where more than 70% of Nayarita tobacco is sown [13]. Fig. 2 shows the Nayarit state in solid green color, the leading tobacco region is highlighted in yellow tones, and the municipality of *Santiago Ixcuintla* is in intense solid yellow.

The tobacco leaves used in this study came from tobacco plantations located in the municipality of *Santiago Ixcuintla*; There, we identified healthy plants and plants with signs of damage by blue mold.

B. Protocol for the Acquisition of Images

The acquisition of the images was made in the following way: activating the focus option of the mobile device, taking a photograph at around 30 cm distance, making sure that there is proper solar illumination and absence of shadows and non-uniformities in light distribution over the leaf.

C. Samples

The data used in this research work consists of 40 images of tobacco leaves with different levels of damage. The photos

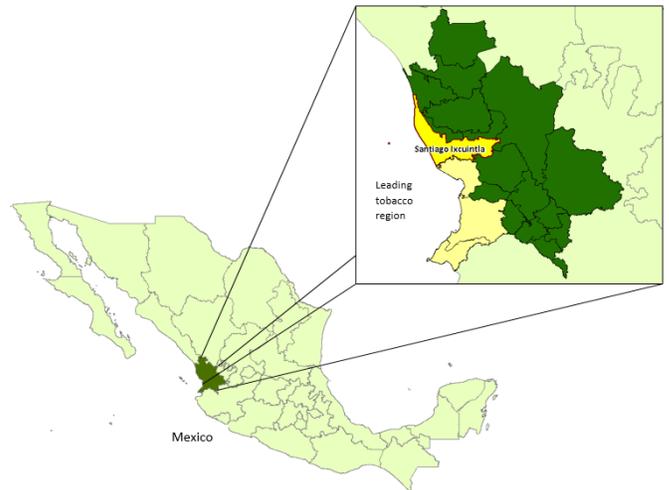


Fig. 2. Santiago Ixcuintla (highlighted in intense yellow color), municipality of the state of Nayarit (solid green color), the leader in the production of tobacco in Mexico.

were taken in different tobacco plantations located in Nayarit, Mexico.

D. Proposed Software System

This section outlines the overall scheme of the developed software application and the underlying technique. Fig. 3 shows the architecture of the proposed software system, which follows the client-server principle. On the client side, there are two components (1) a mobile application for the user interaction and (2) a local database. On the server side, there are three components, (3) a web service to address the requests of clients, (4) a database, and (5) a classification model based on the ANN technique. Next, each of the components is described.

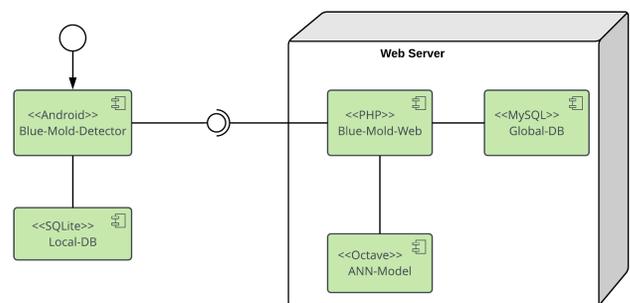


Fig. 3. The architecture of the proposed software application.

1) *Blue-Mold-Detector*: This software component is a mobile application based on the Android system. The user interacts with this component to take photographs of tobacco leaves; the photographs are sent to the central server for analysis; finally, the results are displayed on the mobile device.

2) *Local-DB*: In the fieldwork, there is often no Internet signal, for this reason, it is necessary to store the photographs taken on the mobile device and then when there is phone signal send them for analysis. The Local-DB component, based on SQLite software, is responsible for doing these tasks.

3) *Blue-Mold-Web*: This component is responsible for addressing the client's requests to analyze the images in search of blue mold. For this, this component interacts with the Global-DB and ANN-Model components. This component is a web service implemented in PHP.

4) *GlobalDB*: This component has the task of storing the images taken with the mobile application so that they can be processed later. The images are stored in a database, based on MySQL software.

5) *ANN classification model*: This module is responsible for analyzing the images sent by the Mold-Blue-Web component. This component is the classification model which is based on a multilayer perceptron neural network. The following section describes the method used to create the classification model based on the ANN technique.

E. Proposed Method for Developing the ANN Model

The following steps were followed to develop the proposed classification model: preprocessing, feature extraction, creation, and validation.

1) *Preprocessing*: The images were enhanced by a Gaussian filter; the function is shown in (1) was used for this purpose.

$$g(x, y) = \frac{1}{2\pi\sigma} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \tag{1}$$

where

g = Filtered image

x, y = Position of pixel

σ = Standard deviation of Gaussian filter.

2) *Feature extraction*: Similar to that presented by Castro et al. [14], a visual evaluation scale was established and three levels of tobacco leaf damage caused by blue mold are proposed. In Table II, the three levels of damage are described, from a healthy stage, through initial damage, to advanced damage showing large necrotic areas.

TABLE II. VISUAL SCALE FOR EACH LEVEL OF DAMAGE IN TOBACCO LEAVES CAUSED BY BLUE MOLD

Level	Stage	Visual aspect
1	Healthy	No visible stain.
2	Initial	Few yellow spots without defined borders.
3	Advance	The spots turn dark brown.

Subsequently, an application was developed in GNU Octave⁴ to extract characteristics of each level of damage. This application allows it to select manually rectangular regions of interest (ROIs) and associate the value of the RGB combinations of each pixel to a level of damage. In Fig. 4, the three histograms of blue mold damage levels are represented.

⁴GNU Octave. A Scientific Programming Language. Available: <https://www.gnu.org/software/octave/>

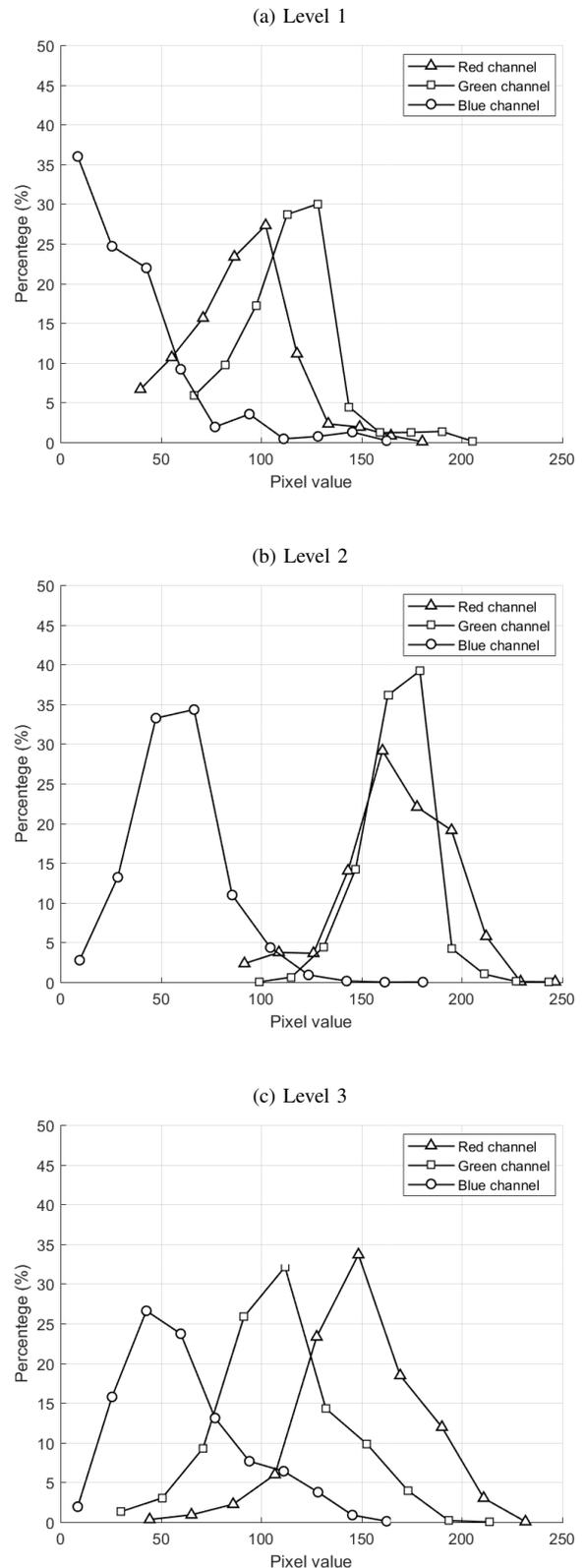


Fig. 4. Histograms of the different levels of damage.

3) *Creation and validation of the classification model*: In order to create and validate a classification model based on the ANN technique, the dataset was divided as follows: 50% for the creation and 50% for validation.

- **Creation.** In this step, we used the methodology suggested by Castro et al. [15], which were proposed for hyperspectral images; and we modified it to use tri-band images (RGB) and create a multilayer perceptron neural network, see Fig. 5.

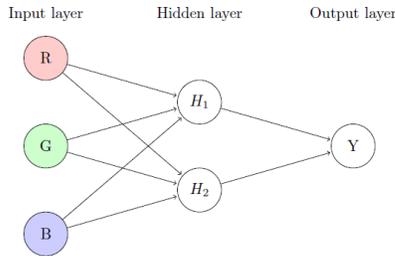


Fig. 5. The multilayer perceptron neural network architecture proposed.

In the proposed model, the number of neurons in the hidden layer was determined by the pyramidal principle given in (2). Three layers with three neurons in the input layer (one for each RGB band) and one neuron in the output layer with three possible values, one for each class (healthy, initial and advanced).

$$H_n = \frac{X_n + Y_n}{2} \quad (2)$$

where

H_n = hidden neurons

X_n = input neurons

Y_n = output neurons.

Once the algorithm based on the ANN technique was developed, 50% of the dataset was used as follows: training (70%), validation (15%), and tests (15%).

- **Validation.** In the same way, as in the previous step, the other 50% of the data set was used to evaluate the accuracy of the classification model created, the results are projected in a confusion matrix.

4) *Testing:* All the leaves were evaluated visually by a panel of four trained judges who determined the total area damaged. These values were compared with those obtained through the computational model, using descriptive statistics techniques.

III. RESULTS

A. Android APP

A prototype of a mobile application based on the Android system was developed, it offers an option to analyze the photographs captured on the secure digital card of the mobile device, see Fig. 6(a). Once the image is analyzed the results are shown in Fig. 6(b), it includes detailed information about the analysis, the date and time of the capture, the device that took the photograph, the GPS location of the leaf that was sent to be analyzed, and a graph of the levels of damage due to blue mold.

B. Obtaining Information by Level of Damage

The ROI selector system described in Section II-E2 was used to obtain the information by level of damage of tobacco

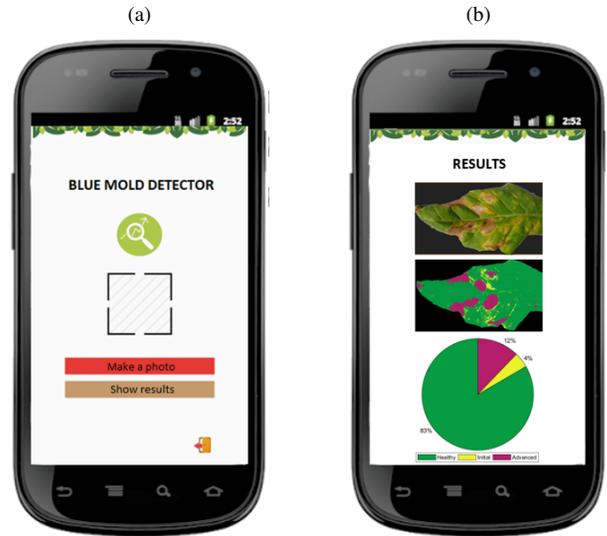


Fig. 6. The mobile software application proposed. (a) Main screen to take a photo of tobacco leaf and analyze it for damage caused by blue mold. (b) It shows the results of the analysis of a tobacco leaf.

leaves. From the obtained RGB combinations, we uniformly select 30,000 data (pixels) per damage level. As a result of the training process, an accuracy of 97.2% was obtained, see the confusion matrix shown in Table III.

TABLE III. CONFUSION MATRIX RESULTING FROM THE TRAINING PROCESS

	Healthy	Initial	Advanced	Accuracy
Healthy	33.2%	1.8%	0.8%	92.7%
Initial	0.1%	31.5%	0.0%	99.6%
Advanced	0.0%	0.0%	32.5%	100%
				97.2%

An example of the use of the developed classification model is shown in Fig. 7; (a) shows a classified leaf with different levels of damage, and (b) shows a graph which represents the percentage of damage for each level.

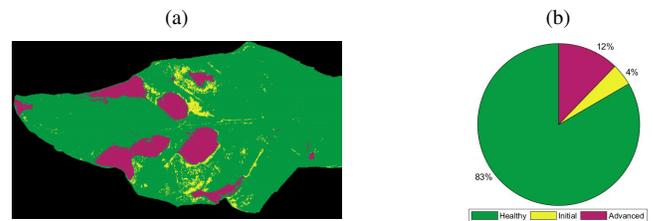


Fig. 7. An example of the classification model developed. (a) A leaf with different levels of damage. (b) A corresponding graph with the damage percentages for each level.

In order to assess the quality of the results obtained by the classifying model, we proceeded to compare the results against the reported in the literature. Table I shows the accuracy achieved by other classifiers based on neural networks techniques.

The only previous study that we found recorded in the literature regarding the use of automatic learning techniques applied to the detection of diseases and pests in tobacco crops

was reported by Guru et al. [10] in 2011. However, Guru et al. worked with tobacco seedlings and the damage caused by anthracnose. They proposed an algorithm to recognize the damaged areas, which was based on the technique of the probabilistic neural network (PNN). The developed classification model was able to detect and classify damages in leaves of tobacco seedlings with an accuracy of 88.58%. Therefore, the results we obtained in the classification of damage levels in tobacco leaves caused by blue mold using the ANN technique are very competitive (accuracy = 97.2%).

Wang et al. [4] used the PNN technique to detect diseases in wheat and grapes, obtaining accuracies of 95% and 94.29% respectively. Concerning the sample size, they used an average of 25 images for each disease. The results that we report in this paper using a sample of 40 images are superior to those reported by Wang et al.

Regarding the specific use of the ANN technique in the classification of diseases in crops, we find in the literature the works of Kulkarni and Patil [11] (pomegranate), Sannakki et al. [5] (grape), and Pujari et al. [12] (beans, soybean, sunflower, and tomato). In summary, our results are very competitive compared to those reported in the literature and that only the results reported by Sannakki et al. are superior, see Table I.

Analyzing the work reported by Sannakki et al. [5], we found that the addressed problem could be considered more accessible to tackle compared to the problem proposed in this paper. Sannakki et al. addressed the problem of identifying between two types of diseases that attack the leaves of grapes (downy mildew and powdery mildew); instead, we address the problem of determining levels of damage caused by the blue mold in tobacco leaves. A detail that caught our attention is that Sannakki et al. mentioned that they used 94% of their data for training and only 6% for validation, which could have positively influenced their results.

C. System Testing

First, four judges specialized in tobacco pests carried out a visual analysis (VA). They made a visual examination of the tobacco leaf samples using the scale showed in Table II.

Due to the differences in the results of the applied methods (ANN and VA), it is necessary to determine if such differences are significant. For this purpose, a nonparametric statistical analysis was performed using the Kolmogorov-Smirnov test, which we calculated using (3).

$$D = \max |VA_i - ANN_i| = 0.2023 \quad (3)$$

For $\alpha = 0.05$, we obtain a $D_\alpha = 0.2101$, since $D < D_\alpha$ it is concluded that there is no significant difference in the results.

IV. CONCLUSIONS

In this work, a mobile application was presented for the detection of damage levels in tobacco leaves due to blue mold. Internally, this mobile application has a damage classification model based on the ANN technique. The experimentation carried out, showed that this model has an accuracy superior to 97% and there is no significant difference with a visual

analysis made by experts in pests and diseases of the tobacco. As future work, we will extend the use of the proposed APP to identify other diseases and pests in tobacco crops.

ACKNOWLEDGMENT

The authors would like to thank the technical staff of the *CONTMIPEBIO S.P.R. de R.L.* company for all the help given, both professional and logistical when it comes to collecting the samples.

REFERENCES

- [1] S. Sukanya and O. Spring, "Influence of temperature and ultra-violet light on viability and infectivity of peronospora tabacina sporangia," *Crop Protection*, vol. 51, pp. 14 – 18, 2013.
- [2] O. Spring, T. Hammer, R. Zipper, and N. Billenkamp, "Population dynamics in tobacco blue mold incidences as a consequence of pathogen control and virulence performance of peronospora tabacina phenotypes," *Crop Protection*, vol. 45, pp. 76 – 82, 2013.
- [3] V. Singh and A. Misra, "Detection of plant leaf diseases using image segmentation and soft computing techniques," *Information Processing in Agriculture*, vol. 4, no. 1, pp. 41 – 49, 2017.
- [4] H. Wang, G. Li, Z. Ma, and X. Li, "Image recognition of plant diseases based on principal component analysis and neural networks," in *Natural Computation, 2012 Eighth International Conference on*. IEEE, 2012, pp. 246–251.
- [5] S. S. Sannakki, V. S. Rajpurohit, V. B. Nargund, and P. Kulkarni, "Diagnosis and classification of grape leaf diseases using neural networks," in *2013 Fourth International Conference on Computing, Communications and Networking Technologies*. IEEE, 2013, pp. 1–5.
- [6] A. Camargo and J. S. Smith, "Image pattern classification for the identification of disease causing agents in plants," *Computers and Electronics in Agriculture*, vol. 66, no. 2, pp. 121–125, 2009.
- [7] T. Rumpf, A.-K. Mahlein, U. Steiner, E.-C. Oerke, H.-W. Dehne, and L. Plümer, "Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance," *Computers and Electronics in Agriculture*, vol. 74, no. 1, pp. 91–99, 2010.
- [8] M. Arshad, S. Ullah, K. Khurshid, and A. Ali, "Estimation of leaf water content from mid-and thermal-infrared spectra by coupling genetic algorithm and partial least squares regression," *Journal of Applied Remote Sensing*, vol. 12, no. 2, p. 022203, 2018.
- [9] K. Golhani, S. K. Balasundram, G. Vadamalai, and B. Pradhan, "A review of neural networks in plant disease detection using hyperspectral data," *Information Processing in Agriculture*, 2018.
- [10] D. S. Guru, P. B. Mallikarjuna, and S. Manjunath, "Segmentation and classification of tobacco seedling diseases," in *Proceedings of the Fourth Annual ACM Bangalore Conference*. ACM, 2011, p. 32.
- [11] A. H. Kulkarni and A. Patil, "Applying image processing technique to detect plant diseases," *International Journal of Modern Engineering Research*, vol. 2, no. 5, pp. 3661–3664, 2012.
- [12] J. D. Pujari, R. Yakkundimath, and A. S. Byadgi, "Image processing based detection of fungal diseases in plants," *Procedia Computer Science*, vol. 46, pp. 1802–1808, 2015.
- [13] SIAP, "Atlas agroalimentario 2016," Mexican Agrifood and Fisheries Information System (Servicio de Información Agroalimentaria y Pesquera), 2016. [Online]. Available: https://nube.siap.gob.mx/gobmx_publicaciones_siap/pag/2016/Atlas-Agroalimentario-2016
- [14] W. Castro, J. Oblitas, J. Maicelo, and H. Avila-George, "Evaluation of expert systems techniques for classifying different stages of coffee rust infection in hyperspectral images," *International Journal of Computational Intelligence Systems*, vol. 11, no. 1, pp. 86–100, 2018.
- [15] W. Castro, J. Oblitas, R. Santa-Cruz, and H. Avila-George, "Multilayer perceptron architecture optimization using parallel computing techniques," *PLoS ONE*, vol. 12, no. 12, p. e0189369, 2017.