

Diseases Detection in Blueberry Leaves using Computer Vision and Machine Learning Techniques

Cecilia Sullca, Carlos Molina, Carlos Rodríguez, and Thais Fernández

Abstract—This paper explains how image processing techniques and Machine Learning algorithms were used, such as Support Vector Machine (SVM), Artificial Neural Networks (ANN) and Random Forest; and Deep Learning's technique Convolutional Neural Network (CNN) was also used so we can determine which is the best algorithm for the construction of a recognition model that detects whether a blueberry plant is being affected by a disease or pest, or if it is healthy. The images were processed with different filters such as medianBlur and gaussianblur for the elimination of noise, the add Weighted filter was used for the enhancement of details in the images.

The images were compiled by the authors of this work, since there was no accessible database of this specific kind of fruit, for which we visited Valle and Pampa farm so we could take pictures of different blueberry leaves, labeled in three different tags: diseased, plagued and healthy. The extraction of characteristics was done with algorithms such as HOG (Histogram of oriented gradients) and LBP (Local binary patterns), both normalized and not normalized. The results of the model showed an 84% accuracy index using Deep Learning, this model was able to classify whether the blueberry plant was being affected or not. The result of this work provides a solution to a constant problem in the agricultural sector that affects the production of blueberries, because pests as well as diseases are constant problems in this sector.

Index Terms—Artificial vision, deep learning, machine learning, random forest, support vector machine.

I. INTRODUCTION

The investment in blueberries is a matter of great care because the cultivation of these fruits is highly complex, in addition to needing as much agronomic knowledge as possible. The companies that venture for the first time in this field can end into failure if they do not have all the important knowledge to start. The approximate investment per hectare is 29 thousand dollars and this doesn't include any irrigation system so in the end the investment ends being higher. Considering this, it is very important to invest in all the necessary care so the investment in this crop is not lost, but blueberries like other fruits are susceptible to certain risks, such as diseases and pests that can affect the blueberry plantations, which can end in loss of hectares of crops that they were ready to harvest [1].

The pests and diseases that attack blueberries not only affect them during the growth of these, but can also affect

them just weeks before starting the harvest, in some cases the plants reduce the amount of produced fruit, in others the quality of the fruits, harvesting very small fruits that were expected, in others the plant does not produce anything until it is free of the disease or plague and in other cases, the plant dies thus losing the fruits it produced and the investment that was done in the purchase of each plant. This happens many times because the control that is made to combat these pests or diseases is not enough, in addition sometimes the disease is not detected in its initial phase or the type of pest or disease is not determined correctly causing the plant not to heal and to infect those that are nearby, causing a spread of the disease or plague, and as a result financial losses to the company.

According to Mr. César Torres, head of crops at Bayer, who exhibited at the Berries seminar organized by Sierra Exportadora, he explained the main pests and diseases that attack blueberries in Peru, as follows: One of the important problems that is being presented in the blueberry is the anomalous insect sp that is a white beetle whose larvae mainly affect the roots, they eat rootlets and can even cause the death of the plant. It has come to report up to 30% of death of the plant, this pest attacks all year and there is still no effective control. It usually happens when organic material that is not well decomposed is used [2].

A plague that has bothered the blueberry in the north of Peru, accurately in Trujillo, is the *Prodiplosis longifila* an insect that mainly affects the asparagus and has a very high reproduction capacity, therefore the ability to do harm is constant. Another plague that is occurring particularly in the blueberry production areas is the *Heliothis*, a fruit-boring worm belonging to the *Noctuidae* family, the larvae perforate the fruits, the damaged fruits rot and fall, causing defoliation of shoots, terminals and fruits, produces up to 40% damage on the total population of plants [3].

II. THEORETICAL ANALYSIS

For this work, other research was taken as a reference with topics related to those used in this research. In the thesis "Assessment of Internal and External Quality of blueberries using images", addresses situations related to how to improve the quality of blueberries in order to increase performance and improve the marketing of these, this thesis uses machine learning techniques with images.[4]

They were used for training different classifiers in order to determine which of them gives a better classification, some classifiers we used were: discriminant linear analysis, vector support machine and probabilistic neural network being these also reported with better performance after training and pattern recognition.

Manuscript received May 13, 2019; revised August 12, 2019. This work was supported in part by ESAN University.

The authors are with ESAN University, Peru (e-mail: 14100970@ue.edu.pe, 14100968@ue.edu.pe, 13100138@ue.edu.pe, 10200324@ue.edu.pe).

Another investigation taken into account was the thesis "Evaluation of Classifiers for Automatic Disease Detection in Citrus leaves using machine vision", this investigation was intended to evaluate an adequate control for the diseases presented in the citrus industry because this industry is important in the agricultural economy of Florida. In this thesis, the study that was implemented investigated the use of artificial vision and image processing techniques in the classification of diseased leaves of citrus fruits. In addition, algorithms based on image processing techniques were used to extract characteristics. The classifiers used in this study were: statistical classifier using the Mahalanobis minimum distance method, neural network based on the use of the back propagation algorithm and neural network using radial basis functions. His study determined that such classification methods are suitable for the classification of citrus leaves [5].

Finally, in the research "Deep Neural Networks based recognition of plant diseases by leaf image classification", the results of the research indicate that climate change can alter the stages and the rates of pathogen development. The situation is further complicated by the fact that, nowadays, diseases are transferred globally more easily than ever. New diseases can occur in places where previously they were not identified and, inherently, where there is no local experience to fight them [6]. The use of inexperienced pesticides can lead to the development of long-term resistance of pathogens, which drastically reduces the ability to fight. The timely and accurate diagnosis of plant diseases is one of the pillars of precision agriculture. It is crucial to avoid unnecessary waste of financial and other resources, thus achieving healthier production, addressing the problem of developing long-term resistance to pathogens and mitigating the negative effects of climate change. The images of this research were downloaded from the Internet in different formats along with different resolutions and quality. In order to obtain better feature extraction, the final images intended

to be used as a data set for the deep neural network classifier were preprocessed to gain consistency.

Training a deep convolutional neuronal network was proposed to make a classification model of images from a set of described data. In this work, a new approach to the use of the deep learning method was explored to automatically classify and detect plant diseases from leaf images. The developed model was able to detect the presence of leaves and distinguish between healthy leaves and 13 different diseases, which can be diagnosed visually.

III. WORK METHODOLOGY

The proposal is to implement a recognition model of the status of a blueberry plant and identify if it is being affected by a disease or pest, or if it's healthy.

This proposal aims to solve a number of problems in blueberry crops because reducing the time of analysis or recognition that was normally done in laboratories, would make it immediately with only get an instant capture or a short video of the plant. At the end of the investigation, a disease and pest recognition system will be obtained and will be able to detect the disease. In addition people can use this system in a mobile device or a mechanical device that monitors the sowings of blueberries all the time so the accuracy of the system is greater; if it is used in a mechanical device it would have to be go around the fields for several hours and send alerts to the operations center when finding certain anomalies in the blueberry plants.

The work methodology used in this investigation was the one shown in Fig. 1.

As shown in Fig. 1, the first phase that we are going to go thru is the input of the database of the photographs; pictures of different varieties of blueberry plants were collected, we took the pictures in Valle and Pampa farm, located in Pampa California, in the Humay desert in Pisco, 240 kilometers to the south of Lima.

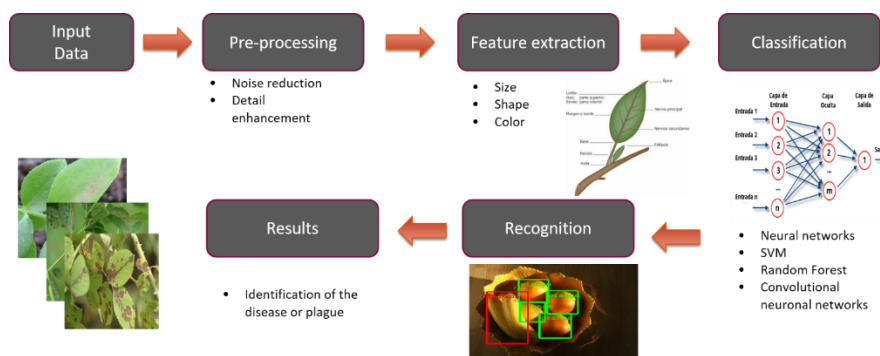


Fig. 1. Work methodology that we follow in this research: input data, pre-processing, extraction, classification, recognition and results.



Fig. 2. Blueberry plants affected by alternaria.



Fig. 3. Blueberry leaves affected by whitefly.



Fig. 4. Blueberry plant affected by whitefly.

The pictures were taken with different types of devices, this was in order to have captures with different resolutions and images of different qualities because the objective for the system is to recognize the blueberry disease in any type of image that the user captures.

The main diseases and pests that we have been able to identify with the help of the farm specialist will be detailed below:

- *Alternaria* sp: It is a fungus that reproduces quickly in dry leaves. See Fig. 2. [7]
- Whitefly: They hide in the back of the leaves, and they eat the leaves from there. See Fig. 3 and 4. [8]
- Fruit borer (*Heliothis virescens*): They perforate the fruits, which are contaminated by their feces and pathogens. See Fig. 5. [9]
- Trips (*Thrips tabaci*): It feeds on the floral parts and interferes with the pollination and mooring of the fruit. See Fig. 6. [10]
- *Fumagina*: Initially presented as a thin black layer. See Fig. 7.



Fig. 5. Blueberry plant with worm (left) and fruit borer (right).



Fig. 6. Trip.



Fig. 7. Blueberry plant with sooty mold (left) and blueberry leaf with sooty mold (right).

Approximately 400 captures of different plants were collected, both diseased, with plagues and healthy, the diseased plants were classified by type of disease that we found in the farm, from a complete image of a plant we proceeded to make trims of all the leaves that could contain the image, this in order to better segmentation, then we obtained a total of 800 images obtained from the trims made. See Fig. 8 to see the process of trimming.

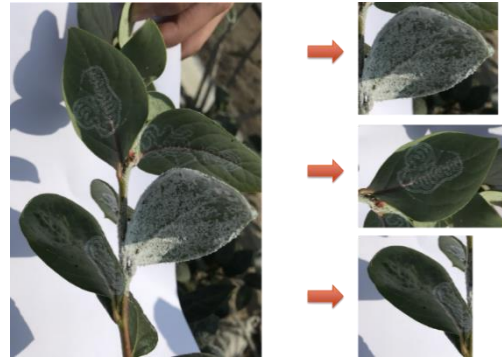


Fig. 8. Process of trimming the images of whole blueberry plants, obtaining in this case, 3 images of leaves affected by whitefly.

The second phase and the third is the pre-processing and segmentation respectively, where we will convert the RGB image (by its acronym Red, Green and Blue) to gray (as seen in Fig. 9) to later be able to perform binarization, this step is very important because it converts the black and white image to a binary image. The binary image, as its name says, will consist of two values, of 0 and 1, where 0 indicates black and 1 indicates white. This step is important because it will improve the extraction quality of the object. [11]

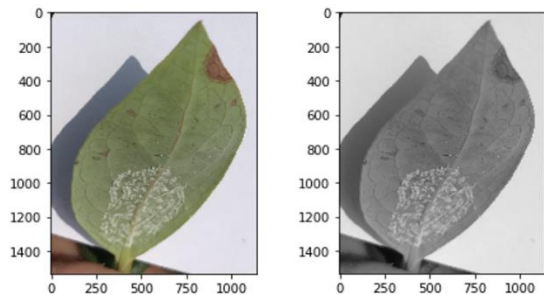


Fig. 9. Process of converting an RGB image to a grayscale image.

Another filter that was used was the Gaussian filter where a box filter consisting of equal coefficients is applied, a Gaussian core is used. This is done with the function, `cv2.GaussianBlur` (image, (5,5), 0) of the openCV library. The width and height of the kernel must be given as input parameters, which must be positive and odd. In addition, the standard deviation must be specified in the X and Y directions, `sigmaX` and `sigmaY`, respectively. This type of filtering is very effective to eliminate the Gaussian noise of the image [12].

It can be seen in Fig. 10 an original image (left) and one with the Gaussian filter (right).

The Medium filter was also applied, this filter calculates the median of all the pixels under the kernel window, and the central pixel is replaced with this median value. This is very effective in eliminating the noise known as salt and pepper noise. OpenCV has the function `cv2.medianBlur`

(img, 5) to apply this type of filter to an image. As in the Gaussian filter, the size of the kernel in the median filter must be a positive odd integer. [13]

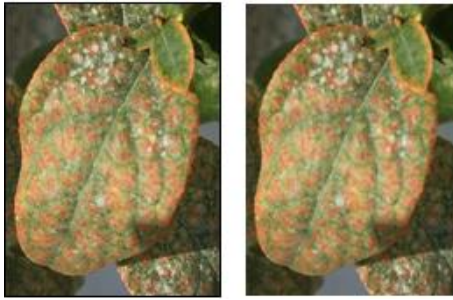


Fig. 10. Image of blueberry leaf processed with the Gaussian filter, original photo (left) and processed photo (right).

This can be seen in Fig. 11 an original image and one with the Medium filter.

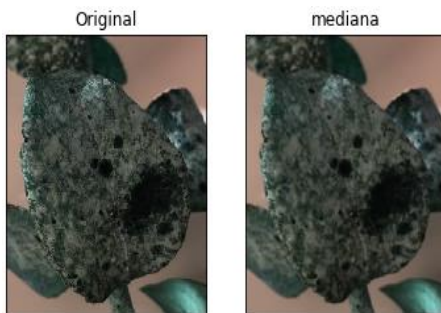


Fig. 11. Image of blueberry leaf processed with the medium filter, original photo (left) and processed photo (right).

Then we proceeded to pass the photos through a filter to improve details, this in order to see even the smallest details, such as the nerves of the leaves. The filter that was used is the following: `cv2.addWeighted (image_EG, 1.5, image_EG2, -0.5, 0, image_EG).`[14]

This can be seen in Fig. 12 an original image (left) and one with the details filter (right).



Fig. 12. Detail enhancement process. Original healthy leaf (left) and processed healthy leaf (right).

According to the methodology used, the fourth phase is the extraction of characteristics, for this different algorithms were used to extract characteristic vectors from the images, the extraction algorithms were the following and they were implemented in python language with OpenCV libraries.

- HoG (Histograms of Oriented Gradient) - Robert K. McConnell: Used to detect objects in computer vision and image processing. Counts occurrences of the orientation gradient in localized portions of an image - detection window, or region of interest (ROI) [15].

There are libraries that already implement this algorithm, for example: HOG OpenCV Descriptor. The implementation of the HoG algorithm is the following and can be seen in Fig 13:

A detection window inside the box.

Color normalization is done.

Gradients of the sub-picture are obtained to calculate the histogram of candidates on uniform cells.

The cells are grouped in blocks with certain overlap.

The cell blocks are normalized independently.

The feature vector is made up by the set of descriptors of the blocks.

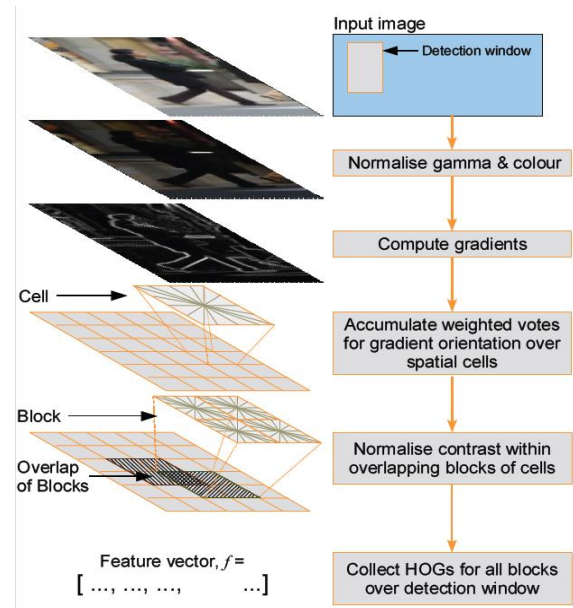


Fig. 13. HoG technique process.

- LBP (Local Binary Pattern)

It is a method used in extraction of texture characteristics with classification reasons.

It has a characteristic that makes it invariant to changes of illumination in the levels of grays.

It is ideal for applications that require fast extraction of characteristics and classification of textures [16].

The process of the extracting algorithms is shown in Fig 14.

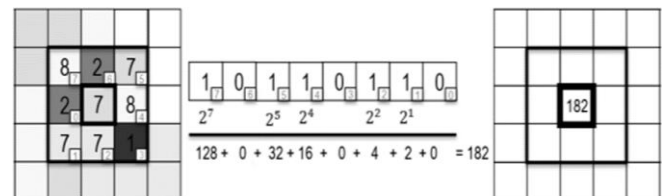


Fig. 14. LBP process.

Finally, after extracting the characteristics of all the images, we proceeded to separate them into two classes: train and test, the class "train" to train the model and the class "test" for validation. After this, these vectors will be processed with learning algorithms, different accuracies will be obtained, and the best of these will be chosen.

The algorithms that were used for learning the recognition model were: Support Vector Machine, Random Forest,

Neural Networks and Deep Learning using Neural Networks Convolutional which will be described below.

Support Vector Machines (created by Vladimir Vapnik) is a learning-based method for solving classification and regression problems. In both cases, this resolution is based on a first phase of training (where they are informed with multiple examples already solved, in pairs (problem, solution)) and a second phase of use for solving problems. In it, the SVMs become a "black box" that provides a response (output) to a given problem (input) [17].

Random Forest is a supervised algorithm, this algorithm creates a kind of "forest" which is a set of decision trees and it is random, most of the time is trained with the "bagging" method, this method is a combination of learning models that increases the overall result. One advantage of the Forest is that it can be used for classification and regression problems. Random Forest adds additional randomness to the model, since it looks for the best feature among a random subset of characteristics, this results in a better model. [18]

When talking about Deep Learning refers to a class of machine learning algorithms based on neural networks, these networks are characterized by a cascade process, the input data is sequentially passed through different "layers" in which rules of application are applied, learning modulated according to a weight, when they pass through the last layer the results are compared with the "correct" result, and the parameters are adjusted (obtained by the "weight" functions) [19].

Convolutional neural networks are used to process images, they can learn input-output, where the input is an image, they work by sequentially modeling small pieces of information, and then combining this information into the deeper layers of the network. One way to understand them is that the first layer will try to detect the edges and establish patterns of edge detection. Then, the subsequent layers try to combine them in simpler forms and, finally, in patterns of the different positions of the objects, lighting scales, etc. The final layers will try to match an input image with all the patterns and arrive at a final prediction as a weighted sum of all of them [20].

IV. RESULTS

The process of this project consisted first in dividing the database into two classes: train and test. We used the class "train" to train the model and that can learn and the class "test" to validate if the model correctly classified.

These results are shown in Table I.

We worked with two extraction algorithms: HOG and LBP, with these we obtained characteristic vectors for each image. These vectors were processed with 4 learning algorithms: Support Vector Machine, Random Forest, Neural Networks and Convolutional Neural Networks.

When SVM and characteristic vectors extracted with HOG were used, the following was obtained: with standardized characteristic vectors, an accuracy of 74% was obtained and 72% with unstandardized characteristic vectors.

When SVM and characteristic vectors extracted with LBP were used, the following was obtained: with standardized characteristic vectors, an accuracy of 83% was obtained and 82.4% with unstandardized characteristic vectors.

When Random Forest and characteristic vectors extracted with HOG were used, the following was obtained: with standardized characteristic vectors, an accuracy of 58% was obtained and 56% with unstandardized characteristic vectors.

When Random Forest and characteristic vectors extracted with LBP were used, the following was obtained: with standardized characteristic vectors, an accuracy of 68% was obtained and 66% with unstandardized characteristic vectors.

When Neural Network and characteristic vectors extracted with HOG were used, the following was obtained: with standardized characteristic vectors, an accuracy of 67% was obtained and 63% with unstandardized characteristic vectors.

When Neural Network and characteristic vectors extracted with LBP were used, the following was obtained: with standardized characteristic vectors, an accuracy of 78% was obtained and 75% with unstandardized characteristic vectors.

When Deep learning (Convolutional Neural Networks) and characteristic vectors extracted with HOG were used, the following was obtained: with standardized characteristic vectors, an accuracy of 72% was obtained and 70% with unstandardized characteristic vectors.

When Deep learning (Convolutional Neural Networks) and characteristic vectors extracted with LBP were used, the following was obtained: with standardized characteristic vectors, an accuracy of 84% was obtained and 68% with unstandardized characteristic vectors.

Then, the best result was an accuracy of 0.84, using the algorithm of Convolutional Neural Networks using Deep Learning and using the standardized characteristic vectors extracted with LBP. It is suitable to predict the state of the blueberry plant; this can improve if the model is trained with more images.

TABLE I: TABLE OF RESULTS

	Channels	SVM	Random Forest	Neural Network	Deep Learning
HOG	Normalized	0.74	0.58	0.67	0.72
	Not normalized	0.72	0.56	0.63	0.7
	Normalized	0.83	0.68	0.78	0.84
LBP	Not Normalized	0.824	0.66	0.75	0.68

V. CONCLUSIONS

This work concludes that the recognition model has a good level of prediction, with an 84% prediction it can correctly classify a healthy leaf and one sick or affected by a pest.

The advantage of developing a database of images of blueberry plants was that we had the segmented images ready to work, but one disadvantage was the time of collection of these images because we had to take pictures of different blueberry leaves to classify them.

The work in the future is to continue collecting more images of blueberry leaves to increase the database. It is also expected to build a model that classifies diseased blueberry leaves by type of disease or plague.

In the future we will work with other features extraction algorithms such as sift (Scale-invariant feature transform) or surf(Speeded-Up Robust Features).

The idea is to develop a mobile application and that it can be used by anyone in this sector, the person would take a photo of the affected plant and the system will tell you what the disease is and what the solution would be. Another idea would be to build a small robot that monitors cranberry fields and if it finds an affected plant, then it will send an alarm to the supervisor.

Finally after finishing this project, we want to support other people who have an interest in the area of artificial vision by providing them with the image database that was built in this project.

ACKNOWLEDGMENT

We thank the following entities and people for their collaboration and guidance in the process of this research topic.

- Agroinversiones Valle y Pampa, for their collaboration in allowing us to visit their facilities, their willingness to provide us with guides during visits, and their interest in technology projects.
- Fabián Arteaga Junior (Professor of Advanced Artificial Intelligence at the esan University), for his guidance in the research process.
- Carlos Molina Mendoza (Mg in Information Technology and Systems), for his collaboration in the management during the research project.

REFERENCES

[1] N. R. Marroquin. (December 10, 2016). How to succeed in a blueberry project? gro Negocios Perú. [Online]. Available: <https://agronegociosperu.org/2016/12/10/como-tener-exito-en-un-proyecto-de-arandanos/>

[2] N. R. Marroquin. (November 20, 2015). Blueberries Peru. BlueberriesPeru.pe. [Online]. Available: <https://arandanosperu.pe/2015/11/20/principales-plagas-y-enfermedades-en-el-arandano-en-el-peru/>

[3] R. Agrícola. (September 2017). Arándanos en Perú: Situación actual y perspectivas. RedAgrícola. [Online]. Available: <http://www.redagricola.com/cl/arandanos-en-peru-situacion-actual-y-perspectivas/>

[4] G. Leiva, "Assessment of internal and external quality of blueberries using images," doctoral theses, Santiago of Chile: Pontifical Catholic University of Chile, 2013.

[5] R. Pydipati, "Evaluation of classifiers for automatic disease detection in citrus leaves using machine vision," Thesis, University of Florida, 2014.

[6] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep neural networks based recognition of plant diseases by leaf image classification," *Computational Intelligence and Neuroscience*, vol. 11, 2016.

[7] M. Naresh, G. S. Saharan, and P. D. Meena, *Alternaria Diseases of Crucifers: Biology, Ecology and Disease Management*, Springer, 2015.

[8] W. M. Thompson, *The Whitefly, Bemisia Tabaci (Homoptera: Aleyrodidae) Interaction with Geminivirus-Infected Host Plants: Bemisia Tabaci, Host Plants and Geminiviruses*, Springer, 2011.

[9] H. C. Sharma, *Heliothis/ Helicoverpa Management: The Emerging Trends and Need for Future Research*, CRC Press, 2005.

[10] T. Lewis, "Thrips as Crop Pests," *CABI*, 1997.

[11] A. T. Amin, *Computer Vision with OpenCV 3 and Qt5*, Packt Publishing, 2018.

[12] E. R. Davies, *Computer Vision: Principles, Algorithms, Applications, Learning*, Academic Press, 2017.

[13] A. Kaehler and G. Bradski, *Learning OpenCV 3: Computer Vision in C++ with the OpenCV Library*, O'Reilly Media, 2017.

[14] A. Spizhevoi and A. Rybnikov, *OpenCV 3 Computer Vision with Python*, Packt Publishing, 2018.

[15] M. Nixon, *Feature Extraction and Image Processing for Computer Vision*, Academic Press, 2012.

[16] E. A. Gutiérrez, G. P. Martinsanz and A. D. la E. Hueso, *Concepts and Methods in Computer Vision*, España: CEA, 2016.

[17] P. Harrington, *Machine Learning in Action*, New York: Manning Publications Co. All rights reserved, 2012.

[18] L. Rokach, O. Maimon and O. Maimon, "Data mining WITH decision trees: Theory and applications," *WSPC*, 2014.

[19] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, Granada: The MIT Press, 2016.

[20] R. Venkatesan and B. Li, *Convolutional Neural Networks in Visual Computing*, CRC Press, 2017.



Cecilia V. Sulca was born in Lima, Peru on March 3, 1993. She is currently studying information technology and systems engineering at ESAN University in Peru.

She has competed in congresses related to the application of technologies in her country as "open innovation applied to the Peruvian energy mining sector" held in Peru by the National Society of Petroleum and Energy Mining in May 2017. She also attended the 4th Business Forum Analysis and Its Impact on Businesses conducted in October 2014 in Peru. Currently, she participates proactively in research projects related to the field of data science, machine learning, artificial vision, etc.



Carlos J. Molina was born in Lima, Peru on April 30, 1996, he is currently studying Information technology and systems engineering at the ESAN University in Peru.

He has attended technology-related fairs such as "INTEROP New York 2014" held in the United States in 2014. He has competed in conferences related to the application of technologies in his country as "open innovation applied to the Peruvian energy mining sector" in Peru by the National Society of Petroleum and Energy Mining in May 2017. It currently participates proactively in research projects related to the field of data science, machine learning, artificial vision, etc.



Carlos E. Rodríguez was born in Lima, Peru on April 30, 1996, he studied information technology and systems engineering at the ESAN University in Peru.

He has competed in competitions related to data science in Peru as "data datafest" organized by Banco de Credito del Peru and Universidad ESAN. He currently participates proactively in research projects related to the field of technological innovation and data science.



Thais Fernández was born in Lima, Peru on September 18, 1993, she is studying information technology and systems engineering at the ESAN University in Peru.

She attended the 4th forum on Business Analysis and Its Impact on Businesses held in October 2014 in Peru. He works for the petroleum sector in Petroperú as a systems implementation practitioner.

Currently she participates proactively in research projects related to the field of systems implementation projects.