

Coffee Moisture Classifier Using Convolutional Neural Networks

Yurley Tovar Martínez*, Andrés Calvo Salcedo, and Arley Bejarano Martínez

¹ Technological University of Pereira, Colombia - Risaralda

Abstract

One of the most important agricultural sectors in Colombia is the coffee industry. The coffee provides important income to the country and is highly recognized for its flavor and quality. To ensure quality production, the drying process is one of the most important factors. This process requires the humidity of the production be in the range of 10% to 12%. Due to the low impact of technification process in the country, many coffee growers measure empirically this characteristic because specialized equipment tends to be highly expensive or unknown. In addition, generating inaccurate estimation of coffee moisture could produce economic losses, due to the cost overrun for hiring an experts, or the loss of weight because of the coffee dehydration, which is fundamental in the sale of coffee. In this work, we developed an algorithm that detects coffee moisture using digital images. Our method uses deep learning from a convolutional network with LeNet CNN topology to classify the image and obtain the measurement. This proposal was validated with images captured by the coffee growers themselves and finally, the results shows the implemented method is robust to outliers and changes in the capture of the photographs.

Keywords: Convolutional neural network; machine learning; Support Vector Machine (SVM).

1. Introduction

One of the most important sectors in Colombian agriculture is caficulture, its product represents one of the most important elements of consumption around the country (Portafolio,2019, Federación Nacional de Cafeteros de Colombia, 2019). Where the quality and taste of coffee depend on a good process of drying the beans. To meet these standards, humidity must be between 10% and 12%, due to 2 reasons: First the storage process is taken into account, First the storage process is taken into account, since the grains must be stored in long periods of time, to prevent the proliferation of fungi, molds and fermentations that affect their characteristics (Rodríguez et al., 2019). Second, the largest sales of coffee are obtained with low-humid coffee, where shoppers verify that the value of moisture has the correct percentage. Additionally, coffee is sold by how much weight, so over-drying represents monetary losses for growers and losses of attributes that make a good coffee (Rodríguez et al., 2019: Centro de comercio internacional, 2018). However, despite the relevance of this product, research and technologies that seek to technify this crop is scarce.

A big problem in the sales process is that coffee growers use their empirical knowledge to measure moisture. Since the grains have some coloration in their almond when the humidity is correct, which makes the measurement taken might be subjective. To achieve this, growers need years of experience to determine the tonality of the approximate color, which can lead to considerable loss of money. Although there are instruments that measure moisture in coffee beans, they are too expensive and potential stakeholders are not aware of them. This shows how unfamiliar farmers are with emerging technologies, which causes considerable losses from not using them, even though technologies can bring big profits to this market (AgraTronix, 2020; Kett electric laboratory, 2020)

Emerging technologies with a focus on agriculture have proven to be very effective, because they allow farmers to have better results in crop productivity, rising the profits (Quesada, 2019). This is why a search is carried out for tools to manage the responsible and appropriate use of pesticides, fungicides and fertilizers, as these can lead to high costs in addition to polluting (Puerto, 2014). A sample of this is software research that seeks to determine the condition of plants for certain diseases, pests, and the land state (Tovar, 2019; Calvo, 2018; Pamplona, 2017).

Moisture spoils the coloration and quality of grains by inducing visual patterns (Pashley, 2017), which can be classified using computer vision techniques. Visual representation systems, such as HSV, TSL, LAB, and YCBCR, commonly use Gaussian filters that soften images; Then, the methods such as histogram and OTSU analysis, among others, can be used to detect a color pattern by computing a threshold (Oliveros et al., 2009). Although these methods have been effective for agricultural applications such as disease detection, water stress calculation, among others, they heavily depend on the photometric conditions of the image, hindering good results.

In the literature there are different disease pattern detection research of plants, most of them have neural network techniques that have proven to be robust algorithms allowing better image pattern classification compared to other Machine Learning techniques. However, in the state of the art there are not researches that obtain the moisture of coffee by images detection, so we ask ourselves, can coffee moisture be determined by using images through machine learning techniques to verify whether unthreshed grains can continue to the next phase of the process or not?

Machine learning provides methods for detecting patterns in databases and then using them for classification (Murphy, 2012), which are used in applications such as automatic vehicles, robots, information detection in images, analysis of behaviors and productivity, among others. Some of the most commonly used algorithms are vector support machines, decision trees, logistics and linear regressions, neuronal and grouping networks (APD,2019).

Convolutional Neural Networks (CNN) are powerful tools that allow precision farming applications such as leaf and stem counting, measuring leaf size, root localization and plant recognition. Due to the importance of the growing process, the most commonly used application is disease detection (Prasanna et al., 2016). One study (Ramcharan et al., 2017) trained an Inception-v3 network to classify infections in cassava plants, achieved 93% efficiency. The authors of (Prasanna et al., 2016) used a public domain dataset to train the AlexNet and GoogleNet architecture to identify 26 diseases that affect 14 types of bodies,

accurately identifying diseases in 99.35% of cases. Although the accuracy rates for these studies are high, the tests were conducted under laboratory conditions where there were no alterations caused by image noise. These classification models must satisfy all disturbances that would generate robust systems with high possibilities for land deployment.

In recent years, some researches have been carried out in order to improve caficulture, this has been achieved from the use of techniques such as machine learning and Deep Learning that allow the extraction of grain characteristics. For example: Through convolutional neural networks (CNN) green coffee beans were classified through their defects (Pinto et al., 2017); By regression of least squares some research get the prediction of components such as sucrose, caffeine and trigonelin in green coffee beans (Caporaso et al., 2017); in (Arboleda et al., 2018), they use image processing and artificial neural network to sort three species of coffee beans.

Because of the above, we come up with the idea of creating a database containing information that was extracted from images of coffee beans with different humidity, the classification of each sample should be within the percentages of 10% to 12% (Class 1) or higher (Class 2) and allow machine learning models to be implemented for the detections of new photographs, where the result would indicate whether the grains are ready to start with the next stage of production or not.

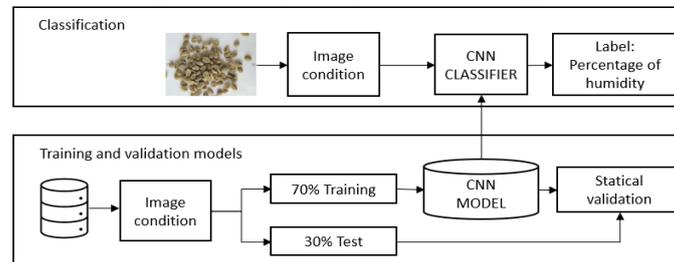
This project seeks to strengthen and support this sector of agriculture, aiming to analyze the possible coffee drying technification. By determining coffee moisture through digital image processing and machine learning techniques, the result obtained would indicate whether the grains are ready to start with the next stage of production or not.

This article describes a method for detecting different levels of moisture in coffee beans using a methodological variation of the LeNet CNN architecture. Training network parameters was performed by using a database created under controlled conditions. This study consists of two parts: in Section II, we describe the contents of each stage including the methodological development of the models of machine learning models, the construction of the annotated database and statistical validation processes. In Section III, we describe our experiments and discuss the results. In Section IV we present what we find from this study and all of its benefits.

2. Materials Y Methods

The proposed methodology for performing moisture detection on coffee beans was the development of a CNN machine learning algorithm, which allow such measurement to be carried out by capturing and processing images. Below this paragraph, we show the classification process in the figure 1.

Figure 1: Diagram of the methodology used.



2.1 Classification model

we use four activities to classify images of coffee beans from their humidity, these are presented with greater detail in the information below.

2.1.1 Database for Severity Level Detection

In the state of the art, it can be verified that there are different methods to perform the measurement of coffee moisture, where the most commonly used consists of an empirical activity where the farmer threshes the grains and through his experience determines whether they are ready to continue with the next stage of production or not. However, there are not image databases founded under capture protocols to classify coffee at its different stages of moisture.

Because of the above, a database was built to accomplish a certain image capture protocol which is explained below:

- To ensure that the moisture measurement of coffee beans is correct, it is necessary to use a photography studio. Use a studio is recommended because in this process it is feasible to take a small sample of grains and handle it in a controlled lighting environment, which ensures the repeatability of the experiment. In this case, we recommend the example shown in Figure 2.

Figure 2: Photography studio used to capture images of coffee beans.



- It was verified that the light conditions were adequate, i.e., in this case it is recommended that the place have the best lighting, capturing the images in the time slot from 9:00 a.m. to 4:00 p.m.

- It was verified that the surface of the photography studio, where the coffee beans are put, was completely clean, as this ensures that there are not major alterations and measurement failures.
- The image of the coffee beans should be captured at a distance of approximately 20cm as shown in Figure 3.

Figure 3: How to capture images of coffee beans.



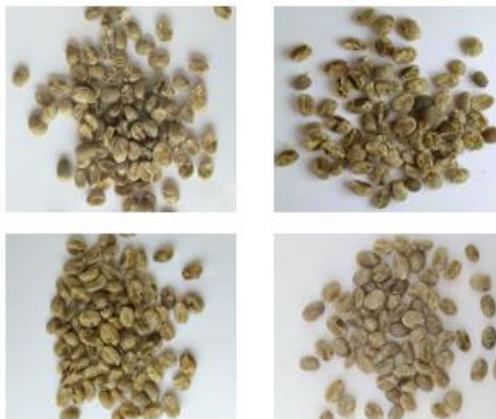
It should be noted that each of the images captured were labeled according to the moisture presented by the coffee beans, this measurement was obtained from the Grain Moisture Tester PM-410 expert meter shown in the figure 4. This procedure ensures that the database does not present any subjectivity errors, on the other hand, the meter was previously calibrated to guarantee the measurements performed.

Figure 4: Coffee moisture meter PM-410.



Sampling was performed with low mid-range cell phones whose cameras have a resolution of 12MB (4000x3000) and 13Mp (4160x3120). Figure 5 shows some examples of the captured database.

Figure 5: Some of the images in the database.



The database described can be found in the following link, which consists of a total of 6264 images where the moisture range of coffee beans is between 9% and 20%, however, it should be noted that the images are at their highest resolution. https://drive.google.com/drive/folders/1BO_P1MgVe_H86nM1lD_qZmsLLXXHNgYT?usp=sharing

The name of the folders in the database, is due to the percentage of moisture of the coffee, i.e., labels 9 and 20 will correspond to 9% and 20% humidity, it extends for each of the folders, so there are 12 classes that describe the advance of coffee drying.

2.1.2 Image conditioning

At this stage the image was resized because classification with neural networks requires lower resolution compared to photos originally captured where a size of 96 x96 (Kumar et al., 2018) is recommended.

In this process, all images are initially resized to a size of 500 x 500 pixels. Subsequently, the region of interest was determined to obtain the target information, i.e. coffee beans, where it was decided to trim them from the center towards the ends, resulting in the dimensions of 96 x 96 pixels. we show the process in the figure 6.

Figure 6: Some of the images in the database



2.1.3 Development of the convolutional network

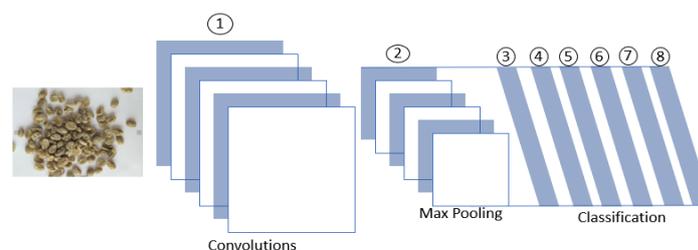
The proposed classifier corresponds to the LeNet CNN topology (Rasdi et al., 2016), which consists of stages of extracting the characteristics of the database and its corresponding classification.

The extraction of features was done by a convolution layer which applies filters in the form of a sliding window whose purpose is to extract color patterns, target information from images. The kernel used in this case is 5x5 and activation stage was the hyperbolic tangent, where, through the latter, nonlinearity is introduced into the network to generate a significant separability of the data (Xu et al., 2015).

The calculation of the convolution is presented in equation (1), where Y_j is the output of the neuron, y_j is the matrix of the scalar product between the Y_i image and the convolution of the K_{ij} nucleus, b_i corresponds to the bias parameter of each neuron and g is the activation function.

In the second layer, a subsampling of the image was implemented according to the maximum grouping process, at this point the size of the image is reduced without generating variance in the rotations of the images.

Figure 7: Topology LeNet.



Finally, the image is flattened to obtain the feature vector that is classified from connected layers as shown in Figure 7 listed from 3 to 8. The first 5 contain hyperbolic tangent activation functions to increase the model weights of neuron model and introduce nonlinearity. The last layer uses a softmax activation function (Liu et al., 2016) which estimates the probability of belonging to a certain class, i.e., the percentage of moisture in coffee grams.

2.1.4 Training and validation of the model

As seen in the methodology item, an algorithm was developed in order to detect moisture in coffee beans by using CNN. In order to compare the performance of the proposed method with other techniques in the literature, we proposed to use Vector Support Machine-SVM (Camargo & Smith, 2009) and a CNN- Inception-v3 (Gomez, 2019). These techniques are chosen because they are currently one of the most widely used techniques to solve this type of problems.

The training of the proposed method was carried out using the resized database mentioned above. The proposed network consisted in a CNN-LeNet model, which learned weight values through the use of the Adadelta algorithm, the cross-entropy loss function was used to improve the performance of the algorithm.

The models obtained were evaluated from the efficiency of each of them, for this, the database was divided into two parts: 70% of the data for training and 30% for validation, a process that was repeated randomly using a Monte Carlo experiment. To run the experiment,

100 iterations were performed, and the confusion matrix was calculated for each iteration in order to determine the efficiency to calculate the percentage of moisture at each of its levels. Subsequently, all the confusion matrices were used to compute an average hit value for each class, as well as their respective standard deviation. All experiments were computed in the Python programming language, on a machine with 32 GB of RAM, 16 cores at 3.5 GHz and Linux operating system.

3. Results

In order to verify the performance of the proposed method, we performed a comparative analysis between the proposed method and other methodologies such as CNN Inception-v3 [34] and SVM [33], which have been successfully implemented in pattern detection applications, in addition to being widely used in classification problems.

The SVM setup for training is a polynomial kernel and wave transform feature descriptor in RGB and HSV color spaces. Furthermore, the CNN Inception-V3 network performs the classification by internal activation layers, which transform the images and by Softmax layer that allows the classification as equal as the CNN LeNet network.

The CNN LeNet method has greater precision in determining the moisture in coffee beans, with an accuracy rate of $98.7\% \pm 0.002\%$. On the other hand, the CNN Inception-v3 method showed a competitive result with $98.7\% \pm 0.003\%$. Finally, SVM presented a considerably low performance compared to neural network models whose accuracy was approximately $66\% \pm 0.013\%$.

Table 1: Accuracy of the proposed classifiers

Model	CNN LeNet	CNN Inception-v3	SVM
Accuracy	$98.7\% \pm 0.002\%$	$98.7\% \pm 0.003\%$	$66.3\% \pm 0.013\%$
Computational time (Seconds)	521.6	717.8	9.91

Table 2: Accuracy of the proposed classifiers

Model Humidity label	CNN LeNet	CNN Inception-v3	SVM
9	$99.2 \pm 0.4\%$	$98.8 \pm 0.7\%$	$72.4 \pm 4.3\%$
10	$100 \pm 0\%$	$100 \pm 0\%$	$34.3 \pm 6.9\%$
11	$100 \pm 0\%$	$100 \pm 0\%$	$58.9 \pm 5.4\%$
12	$98.3 \pm 0.8\%$	$98.7 \pm 0.4\%$	$69.3 \pm 5.3\%$
13	$98.6 \pm 0.7\%$	$98.1 \pm 1.1\%$	$57.5 \pm 8.3\%$
14	$98.8 \pm 0.5\%$	$98.2 \pm 1\%$	$76.4 \pm 6.5\%$
15	$98.4 \pm 0.9\%$	$98.1 \pm 0.6\%$	$72.8 \pm 2.9\%$
16	$96.3 \pm 1.6\%$	$96.0 \pm 1.7\%$	$59.0 \pm 5\%$
18	$100 \pm 0\%$	$100 \pm 0\%$	$74.4 \pm 2.5\%$
19	$98.5 \pm 0.5\%$	$98.8 \pm 0.5\%$	$74.1 \pm 4.5\%$
20	$100 \pm 0\%$	$100 \pm 0\%$	$58.1 \pm 3.7\%$
18	$100 \pm 0\%$	$100 \pm 0\%$	$74.4 \pm 2.5\%$

As can be seen in the tables 1 and 2, the models obtained by neural networks allow to classify the humidity presented on coffee beans with high precision.

The CNNLeNet and CNN Inception-v3 algorithms were more than 95% accurate, so both are excellent for moisture classification. However, CNN Inception-v3 requires more layers to achieve the same precision, which is reflected in its computational cost and establishes that the model proposed in this document is the best for this task. On the other hand, the vector support machine model presented a considerably lower precision compared to the CNNLeNet methodology, this is because the descriptor of this method is not robust enough to generate an adequate representation space and can require additional calculations, making it an unattractive method due to the computational cost and the method could not guarantee adequate precision.

4. Conclusion

We developed an automatic method that allows to measure the moisture in coffee beans from the capture of images in a controlled environment. This methodology would allow coffee growers to have a diagnostic aid to avoid errors in the drying process.

The models were trained and validated with a database created specifically for this work. This is because, in the literature, there is no database of coffee beans images that is properly labeled based on humidity, validated by an expert or a properly calibrated meter.

The database was made available to the public, in order to facilitate comparisons and contribute to possible future research in this area. In the future, the database should be expanded to validate the proposed approach, on a broader set of coffee types.

Acknowledgment

A sincere acknowledgments to the Universidad Tecnológica de Pereira and to the members of the electronic engineering research group (GIIE), for their support.

Conflict of interests

The authors declare no conflict of interest.

Author contributions

Yurley Tovar carried out the design and implementation of a test bed to verify the proposed methodology. She conducted a literature review, implemented the model, collected and processed data sets. She conducted experiments and wrote the first version of the document under the supervision of Andrés Calvo and Arley Bejarano. Andrés Calvo and Arley Bejarano proposed the methodology, edited and expanded the manuscript.

References

- Portafolio. (2019). El café vuelve a ser el motor de la economía. [Online]. Available: <https://www.portafolio.co/economia/el-cafe-vuelve-a-ser-el-motor-de-la-539712>
- Federación Nacional de Cafeteros de Colombia. (2019). Ensayo sobre economía cafetera. [Online]. Available: <https://federaciondefcafeteros.org/app/uploads/2019/12/Econom%C3%ADa-Cafetera-No.-30>
- Pashley T. (2017). ¿Por Qué es Importante El Contenido de Humedad del Grano Verde?. [Online]. Available: <https://perfectdailygrind.com/es/2017/09/07/por-que-es-importante-el-contenido-de-humedad-del-grano-verde/>
- Oliveros C, Peñuela A. & Julieth Jurado. (2009). Fondo nacional del café. Control de humedad del café en el secado solar utilizando el método gravimet. [Online]. Available: <https://biblioteca.cenicafe.org/bitstream/10778/385/1/avt0387.pdf>
- Rodríguez V., Rivas D, Masiques J. & Rovira J. (2019). LA HUMEDAD EN EL CAFÉ. Fórum café. [Online]. Available: <https://www.revistaforumcafe.com/humedad>
- Centro de comercio internacional. (2018). Calidad del café-Grado de humedad y secado. [Online]. Available: <https://www.intracen.org/guia-del-cafe/calidad-del-cafe/Grado-de-humedad-y-secado/#:~:text=No%20hay%20una%20norma%20ideal,para%20la%20mayor%C3%ADa%20de%20caf%C3%A9s.&text=El%20caf%C3%A9%20cuyo%20grado%20de,12.5%20%25%20no%20debe%20embarcarse%20nunca.>
- AgraTronix. (2020). Operators Manual MT-PRO Grain Moisture Tester. [Online]. Available: <https://agratronix.com/wp-content/uploads/2015/11/08125-manual.pdf>
- Kett electric laboratory. (2020) PM-410 Grain and coffee moisture tester. [Online]. Available: <https://kett.com/files/brpm410.pdf> (2020)
- Quesada J. (2019). ¿Cómo mejorar la productividad agrícola a través de la transferencia tecnológica?. [Online]. Available: <https://www.tec.ac.cr/hoyeneltec/2019/06/27/mejorar-productividad-agricola-traves-transferencia-tecnologica>
- Puerto A., Tamayo S. & Palacio D. (2014). Efectos de los plaguicidas sobre el ambiente y la salud. [Online]. Available: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1561-30032014000300010
- Calvo A., Pamplona L., Bejarano A, Escudero C., Vélez V., Escalante D. & Castillo A. (2018). Diseño e implementación de un sistema automático para la detección de enfermedades en cultivos de fruta. [Online]. Available: <https://www.utp.edu.co/vicerrectoria/investigaciones/investigaciones/DetallesProyecto/1858>
- Pamplona L. (2017). Universidad tecnológica de Pereira. Software para determinar el nivel de severidad de la enfermedad antracnosis en cultivos de mora (swseacm). [Online].

Available: <https://www.utp.edu.co/vicerrectoria/investigaciones/investigaciones/DetailsProyecto/2061>

- Tovar Y. (2019). Diseño e implementación de un aplicativo móvil para realizar la detección temprana de la enfermedad de la Sigatoka Negra en los cultivos de plátanos. [Online]. Available: <http://repositorio.utp.edu.co/dspace/bitstream/handle/11059/10708/T005.445%20T736.pdf?sequence=1&isAllowed=y>
- Murphy, K. (2012). Aprendizaje automático: una perspectiva probabilística. Serie de computación adaptativa y aprendizaje automático. [Online]. Available: <https://www.semanticscholar.org/paper/Machine-learning-a-probabilistic-perspective-Murphy/25badc676197a70aaf9911865eb03469e402ba57>
- APD. (2019). ¿Cuáles son los tipos de algoritmos del machine learning?. [Online]. Available: <https://www.apd.es/algoritmos-del-machine-learning/>
- Prasanna S., Peter D. & Salathe M. (2016). Using deep learning for image-based plant disease detection. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2016arXiv160403169P/abstract>
- Ramcharan A., Baranowski K., McCloskey P. et al., (2017). Deep learning for image-based cassava disease detection. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fpls.2017.01852/full?report=reader>
- Pinto C., Furukawa J., Fukai H. & Tamura S. (2017). Classification of Green coffee bean images based on defect types using convolutional neural network (CNN), [Online]. Available: <https://doi.org/10.1109/ICAICTA.2017.8090980> p.1 – 5,
- Caporaso N., Whitworth M., Grebby S. & Fisk I. (2017). Non-destructive analysis of sucrose, caffeine and trigonelline on single green coffee beans by hyperspectral imaging, [Online]. Available: <https://doi.org/10.1016/j.foodres.2017.12.031>, p.193 – 203.
- Arboleda E., Fajardo A. & Medina R. (2018). Classification of coffee bean species using image processing, artificial neural network and K nearest neighbors, [Online]. Available: <https://doi.org/10.1109/ICIRD.2018.8376326>, p.1– 5
- Kumar A., Ganapathysubramanian B., Sarkar S. & Singh A. (2018). Deep learning for plant stress phenotyping: Trends and future perspectives. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1360138518301572>
- Rasdi L., Fanany M. & Murni A. (2016). Metaheuristic algorithms for convolution neural network Computational Intelligence and Neuroscience. [Online]. Available: <https://www.hindawi.com/journals/cin/2016/1537325/>
- Xu B., Wang N., Chen T. & Li M. (2015). Empirical evaluation of rectified activations in convolution network. [Online]. Available: <https://arxiv.org/abs/1505.00853>

- Liu W., Wen Y., Yu Z. & Yang M. (2016). Large-margin SoftMax loss for convolutional neural networks. [Online]. Available: <https://www.jmlr.org/proceedings/papers/v48/liud16.pdf>
- Camargo A. & Smith J. (2009). Image pattern classification for the identification of disease-causing agents in plant. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0168169909000027>
- Gomez M., Vergara A., Ruiz H. et al. (2019). AI-powered banana diseases and pest detection. [Online]. Available: https://plantmethods.biomedcentral.com/articles/10.1186/s13007-019-0475-z?utm_source=dlvr.it&utm_medium=twitter