

Received March 4, 2022, accepted March 23, 2022, date of publication April 11, 2022, date of current version April 27, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3166515

Coffee Maturity Classification Using Convolutional Neural Networks and Transfer Learning

MANUEL ALEJANDRO TAMAYO-MONSALVE^{1,2}, ESTEBAN MERCADO-RUIZ¹,
JUAN PABLO VILLA-PULGARIN¹, MARIO ALEJANDRO BRAVO-ORTÍZ¹,
HAROLD BRAYAN ARTEAGA-ARTEAGA¹, ALEJANDRO MORA-RUBIO¹,
JESUS ALEJANDRO ALZATE-GRISALES¹, DANIEL ARIAS-GARZON¹,
VICTOR ROMERO-CANO³, (Member, IEEE), SIMON OROZCO-ARIAS^{4,5},
GUSTAVO-OSORIO⁶, AND REINEL TABARES-SOTO¹

¹Department of Electronics and Automation, Universidad Autónoma de Manizales, Manizales, Caldas 170001, Colombia

²Faculty of Science and Engineering, Universidad de Manizales, Manizales, Caldas 170008, Colombia

³Robotics and Autonomous Systems Laboratory, Faculty of Engineering, Universidad Autónoma de Occidente, Cali, Valle del Cauca 76001, Colombia

⁴Department of Computer Science, Universidad Autónoma de Manizales, Manizales, Caldas 170001, Colombia

⁵Department of Systems and Informatics, Universidad de Caldas, Manizales, Caldas 170004, Colombia

⁶Department of Electrical, Electronic and Computation, Universidad Nacional de Colombia sede Manizales, Manizales, Caldas 170002, Colombia

Corresponding author: Reinel Tabares-Soto (rtabares@autonoma.edu.co)

This work was supported in part by the Universidad Autónoma de Manizales, Manizales, Caldas, Colombia, under Project 645-2019 TD; and in part by the Universidad Nacional de Colombia and the Universidad de Caldas through the Convocatoria Conjunta de Desarrollo Tecnológico e Innovación 2019 under the Hermes Project through Construcción del Prototipo de un Sistema de Visión Multiespectral Basado en Iluminación Light Emitting Diode (LED) under Grant 48996.

ABSTRACT This work presents a framework for coffee maturity classification from multispectral image data based on Convolutional Neural Networks (CNNs). The system leverages the use of multispectral image acquisition systems that generate large amounts of data, by taking advantage of the ability of CNNs to extract meaningful patterns from very high-dimensional data. We validated the use of five different popular CNN architectures on the classification of cherry coffee fruits according to their ripening stage. The different models were trained on a training dataset balanced in different ways, which resulted in a top accuracy higher than 98% when applied to the classification of 600 coffee fruits in 5 different stages of ripening. This work has the potential of providing the farmer with a high-quality, optimized, accurate and viable method for classifying coffee fruits. In order to foster future research in this area, the data used in this work, which was acquired with a custom-developed multispectral image acquisition system, have been released.

INDEX TERMS Coffee maturity classification, convolutional neural network, data augmentation, deep learning, multispectral images, transfer learning.

I. INTRODUCTION

Nowadays, one of the most important aspects for consumers in the agricultural industry is product quality. Traditionally, the inspection process is carried out manually, which is time-consuming, subjective, and unreliable. For this reason, a great effort is being made by the scientific community to develop automatic systems that help to improve the inspection process, from the time consumption and consistency

The associate editor coordinating the review of this manuscript and approving it for publication was Szidónia Lefkovits¹.

perspectives. Since quality control influences the viability of products, agricultural commodity producing countries are investing significant research efforts in automated quality monitoring and control [1].

One of the research approaches uses machine vision systems [2]–[5]. Thanks to significant advances in this field, including the possibility of using measurements invisible to the human eye, many applications have been developed, and sorting processes for different types of fruits and vegetables have been improved. [6], [7]. Recently, hyperspectral imaging technology has shown significant growth in the field of

produce grading [8]. These systems are considered useful for the industry from two perspectives: the first is from the artificial vision that has been widely studied and on which there is enough information available, and the second is from the capability to generate large amounts of data, which is seen as one of the challenges of Industry 4.0 [9]. The hyperspectral systems capture the spectral signature of fruits and vegetables, which provides a large amount of information on their quality [10]. This increased discriminant capacity comes at the expense of increased acquisition time, so they are not implemented in industries with massive sorting needs [11]. Therefore, an alternative that preserves space, the increased feature, and can work at higher speeds is to employ multispectral systems. These systems capture data from multiple informative wavelengths and have an appropriate speed for industrial implementations.

Multispectral and hyperspectral vision systems provide measurements about both the spatial configuration of objects and their spectral characteristics. Thus, a three-dimensional hypercube of information is generated, where two of the dimensions correspond to space in the same way as a conventional vision system, while the third dimension corresponds to the spectral response. These systems have diversified in recent years and can be divided into several groups, depending on the method for discriminating wavelengths, acquiring information from the hypercube, or the amount of data collected.

Both multispectral and hyperspectral machine vision systems have been developed to study persimmon and nectarine citrus fruits [12]–[15] with better results than traditional vision systems. Taghizade [16] compares a classical RGB system against a hyperspectral system from 400 to 1100 nm to estimate mushroom quality by calculating the L component of the Hunter LAB space. Jianwei Qin *et al.* [17] developed a multispectral citrus cancer detection system mounted on a commercial fruit sorting machine; working at a rate of 60 frames per second and joins two cameras with specific filters at 730 and 830nm to obtain an accuracy of 95.3%. Xing *et al.* [18], used a hyperspectral system to determine the optimal wavelengths for detecting mechanical damage in Golden Delicious apples and then classify the bruises using the four wavelengths that provide the most information. These wavelengths were established by a Principal Component Analysis (PCA)-based procedure. Baohua Zhang [4] evaluated the consistency of apple damage search algorithms between a 6-wavelength multispectral system and a 600 nm effective range (400 - 1000nm) hyperspectral system. Bennedsen and Peterson in 2005 [19] developed a multispectral vision system to detect surface defects in apples. The system employs two optical filters at 740 and 950 nm, respectively. Eight apple varieties with a classification rate between 78% and 92% were used.

In Colombia, coffee cultivation is an important industry, and there are currently great efforts to improve the efficiency of the harvest and post-harvest processes [20]. According to international standards, it is necessary to achieve very high

TABLE 1. Results achieved by the researchers.

Research	Number of categories	Accuracy
Bazame [30]	3	86.0%
Buitrago [31]	4	95.0%
Costa [32]	3	100%
Manrique [29]	2	99.0%
Montes [27]	4	75.7%
Ramos [26]	4	95.0%
Sandoval [25]	11	92.5%
Tamayo [28]	5	93.9%

performance in quality control to produce Colombian mild coffee. In fact, the price of coffee improves when at least 98% of ripe coffee fruits can be guaranteed [21], particularly with specialty coffees that have a higher commercial value due to their better quality, acquired by more efficient cultivation, and a rigorous selection process according to the state of ripening, to guarantee the levels of sweetness and acidity of the final product, among others [22], [23]. In practice, manual harvesting is the only possible option for coffee producers, increasing the total cost of production without having a product with guaranteed quality levels. However, technology advances in precision agriculture and agriculture 4.0 pose the challenge of involving machine learning and deep learning techniques to improve coffee processes [24]. Particularly the quality of the fruit to guarantee the production of specialty coffees.

Although most coffee classification studies are realized after the pulping process, there are some examples of cherry fruit classification in the literature, based on the color of the epidermis, according to the stage of ripening. 6 uses coffee cherries according to the number of weeks after planting, between 26 and 33 weeks. This work reports two classification methods, the first one, a Bayesian method with an accuracy of 94.5% and the second one a neural network with an accuracy of 92.5%. Ramos [26] reports a classification accuracy between 94.8%, and 99.6% for four different ripening stages (immature, semi-mature, mature and overripe). Montes [27] reports an average efficiency of 75.7% for classification among the same four different stages, with a rate of 25 fruits per second, using an FPGA as the deployment method.

Tamayo [28] shows a KNN sorter with an efficiency of 93.9% with a division into five different stages, including the dry fruit. Manrique [29] ensures a classification accuracy higher than 99.0% in the mature stage using multispectral information and an LDA classifier.

Table 1 shows a compilation of some results reported by different researchers, which seek to classify cherry coffee fruits at different stages of ripening, with an emphasis on color.

Artificial neural networks have been a widely used tool to identify the main features present in the large amount of data generated by spectral systems [33]. Particularly Convolutional Neural Networks (CNNs) [34] have seen great development in imaging because they contain several hidden layers [35] with hierarchies to detect the basic mor-

phologies of images in the first layers, and deeper layers that recognize complex shapes and unconventional color structures. In Lukas Cavigelli [36] they achieve a vehicle classification rate of over 99%, by incorporating spectral images and convolutional networks into the classification task. Jingxiang Yang [37] shows how to efficiently fuse information from multispectral and hyperspectral systems through feature extraction with fully connected layered neural networks. Li and Liu [38], proposed a compression method for multispectral images based on CNNs, where the compression improves computational efficiency without affecting the image quality. Reinel Tabares [39] proposed to compare multiple deep learning and machine learning algorithms for multiclass classification tasks.

In this study, the convolutional neural network architectures VGG16 [40], VGG19 [41], Inception-ResNetV2 [42], InceptionV3 [43], and DenseNet201 [44] are explored in different experiments to extract the characteristics of the spectral images of coffee fruits in different stages of ripening to determine which of them achieves the best results compared with the traditional classification carried out by experts who evaluate the color tonalities present in the skin of the fruits at the moment of harvesting. For this purpose, 4 experiments were carried out, implementing the techniques of unbalance balancing, subsampling, oversampling, and weighting on the training data. Once the database was balanced, the Deep Learning models mentioned above were trained with and without applying transfer learning (TL) on a model pre-trained on the popular ImageNet dataset [45]. Some of our TL experiments achieved an accuracy of 100% in a 10-fold cross-validation procedure.

Based on these criteria, we consider that this paper presents a significant advance for the agriculture of the region since having the certainty of the state of maturation of the harvested fruits increases the final quality of the product, which again is a trend of precision agriculture and agriculture 4.0, especially for Colombian specialty coffees. Our multispectral database and GitHub repository will be available to the public as a final contribution so that the community can use the same data to replicate the results presented in this research, as well as using it future research and algorithms.

The paper is organized as follows. Section II presents the acquisition system, the dataset, and the proposed methodology. Section III provides the experimental results, which are then discussed in Section IV. Finally, the paper concludes in Section V.

II. METHODOLOGY

A. CHERRY COFFEE

We performed a series of experiments using a total of 640 cherry coffee fruits of Arabica type Caturra variety [46], [47], which is a variety grown in the department of Caldas, Colombia. These fruits were harvested during the first harvest period of the year 2020 and were classified by expert coffee growers, into 5 different categories, as shown in Figure 1.

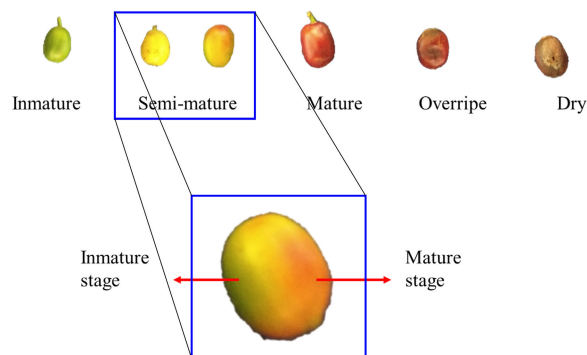


FIGURE 1. Different stages of ripening in cherry coffee fruit.

in this technique, the traditional way of classifying the fruit an expert observes the changes of coloration in the epidermis of the fruit [48]. These color changes span from green in the immature state, go through different shades of yellow and orange during semi-mature stages until reaching the reds in the mature state. Once the fruit has passed the mature stage, they show violet shades until reaching a dark brown fruit when they are dry. Although color sorting is the most widely implemented due to its ease and speed, it has several problems, among them the non-homogeneous color that is present throughout the skin, due to non-uniform maturation [31]. This phenomenon can be observed when zooming in on one of the semi-mature fruits shown in Figure 1. Color change correlated with maturation starts at the lower part of the fruit until it reaches the peduncle attached to the tree. Due to this lack of homogeneity, the color should be observed on the fruit's entire surface so the correct ripening stage can be inferred, showing the importance of the use of cameras.

B. ACQUISITION SYSTEM

Traditional RGB image capture systems capture only three wavelengths corresponding to the colors red, green, and blue. With this information, they can reproduce the color of objects in a similar way to the human eye. By contrast, a multispectral vision system can acquire a larger number of wavelengths to generate an increased feature space that can improve classification processes. These systems require a wide spectrum camera to capture information from the visible spectrum to some wavelengths outside of it [49]. Also, they require an illumination source that contains all the wavelengths to be captured and different filters to separate this information [50].

The proposed multispectral system is a proprietary design, which is calibrated and validated within a spectral range of 400 to 1000 nm and can be seen in detail in [51]. This system has three main elements: The broad electromagnetic spectrum camera, a controlled illumination environment to improve image quality, and a narrow bandwidth LED illumination corona to generate the necessary illumination at different wavelengths.

The wideband electromagnetic spectrum camera is a monochrome Flea3 FL3-GE-03S1M-C (FLIR PointGrey, Wilsonville, Oregon, USA). It has a 1/4" Sony ICX618

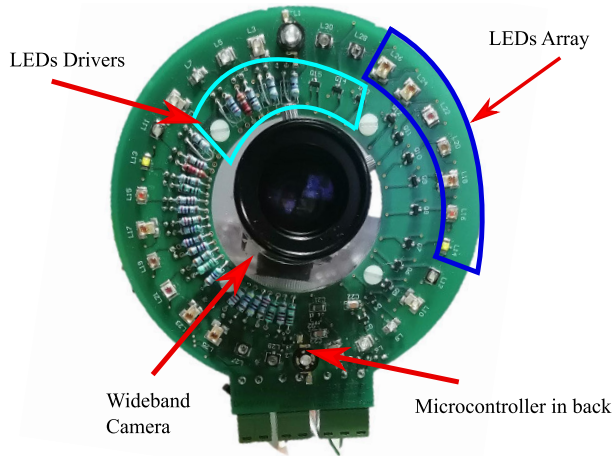


FIGURE 2. LED crown with spectral camera.

CCD sensor with 0.3 megapixels (640×480 pixels) and an acquisition rate of 120 frames per second. It captures information between 300 and 1000 nm, comprising the entire visible spectrum and part of the near-infrared NIR.

The illumination space is controlled to improve the images' quality, reducing problems generated by glare and shadows, and avoiding the influence of ambient light. This space was designed based on the application to perform the training of the machine learning systems with the best possible image quality; this structure has walls that eliminate the external light and reflect the internal light, so that the generated illumination is dispersed and, brightness and shadows are reduced, thus improving the image quality by generating more useful information.

The third element, the illumination system, is responsible for generating illumination at different wavelengths separately to identify the reflectance of each one. This system is designed as a circular crown with 15 different wavelengths between 400 and 1000nm, 30 different power LEDs of 1 watt with a bandwidth (λ) of less than 20nm. Figure 2 shows the camera attached to the LED crown, with the array of LEDs, the drivers, and in the back the microcontroller in charge of generating a pulse width modulation (PWM) control to modify the amount of light emitted; a camera trigger control to facilitate synchronization between the illumination; and a serial communication port to perform the configuration of the light trigger times, in order to adjust to different environments and generate the best possible image. The crown is designed to multiplex each of the wavelengths independently with different trigger times to be calibrated to generate a uniform response at each wavelength. For the coffee tests, the 13 wavelengths within the visible spectrum represent the different features of color, and an additional pair of wavelengths in the near-infrared is used to study additional non-visible characteristics. The table 2 shows the wavelengths selected and the manufacturers' bandwidths; both were selected so that they cover the greatest percentage of the visible spectrum.

TABLE 2. Selected wavelangeths LEDs.

Wavelength (nm)	Color	λ
410	Violet	30
450	Royal blue	20
470	Blue	20
490	Azure	20
505	Cyan	30
530	Green	30
560	Lime	10
590	Yellow	20
600	Amber	20
620	Red-Orange	10
630	Red	20
650	Deep Red	20
720	Far Red	30
840	NIR	50
950	NIR	50

TABLE 3. Number of images per ripening stage.

Ripening stage	Number of images
Inmature	130
Semi-mature	160
Mature	160
Overripe	112
Dry	78

C. DATABASE

The database has 640 images of cherry coffee fruits in different stages of ripening. These images were captured with dimension of 480×640 , and 15 channels (each channel represents a different wavelength). They were also resized to $224 \times 224 \times 15$ to be accepted by the computational models in order to apply transfer learning and to optimize the training process. The number of fruits per ripening phase is shown in Table 3.

The database has been released for use by the scientific community and will be available at the following link: <https://doi.org/10.5281/zenodo.4914786>

D. CLASSIFICATION

Traditionally, image classification problems have been solved using CNNs. This paper compares the performance of different CNN based models present in the Keras library which includes MobilNet, Xception and EfficieNet, among others, where all models were tested and only the 5 best performing models were chosen, then these were used carefully in the four experiments. The experiments take advantage of TL to improve both training resource requirements and accuracy and their convergence. The CNN architectures explored in this work are as follows:

VGG-16 and VGG-19 network structures are very regular, there are not so many hyperparameters, and they focus on building a simple and deep network, with several convolutional layers in a row [40] [41]. On the other hand, Inception-V3 is based on inception modules. These use a series of parallel convolutions with different kernel sizes for feature extraction. The input image is projected through a sequence of convolutional and pooling layers for feature extraction. [43] Inception-ResNet-V2 uses a sophisticated architecture to retrieve the essential features from the images. The initial

layers of the network consist of standard convolutional layers followed by a maximum pooling layer. The next stage consists of simultaneously convolving one input using different filter sizes for each convolution and then concatenating them. The next parts of the network repeat 10 or 20 times the inputs and the network uses dropout layers to make the filter values equal to 0 to avoid overfitting [42].

Finally, DenseNet was proposed to solve the leakage gradient problem, which it solves using a ResNet, as it preserves the information through additive identity transformations, thus increasing the complexity of the model. DenseNet uses layer-to-layer connectivity and connects each previous layer. It uses dense blocks, and feature maps from all later layers are used as inputs to all later layers. [44]

E. TRANSFER LEARNING

The models were trained with and without using TL. The weights of the base model were obtained by training on ImageNet, a dataset with millions of images, and 1000 possible labels [45]. Since we work on 15-channel data, it is not possible to apply TL on the default configuration of the neural networks with the pre-trained weights of ImageNet. Therefore, fine tuning is performed on the first convolutional layer was configured so that the neural network could read 15-channel images. This was achieved by averaging the three pre-trained weights and then replicating it 15 times in this convolutional layer.

The neural networks were configured so that the first layers of the network were not trained, in order to take advantage of the general characteristics learned from the models with the ImageNet dataset and leaving the deeper layers as trainable, thus extracting the more complex characteristics of the coffee fruits, optimizing the training process and helping the model to converge faster and achieve better results.

Figure 3 shows the architecture of the VGG16 model used for the classification of the different stages of coffee ripening.

The red boxes are the blocks that were left untrained in order to use TL and allow the model to obtain better feature extraction properties with respect to our database. The other blocks were set as trainable, allowing the model to adapt in a better way to the database and better classify the coffee fruits.

In addition, for each network used, a variation of the trainable layers was performed in order to obtain the appropriate layers for our problem. Therefore, the trainable and non-trainable layers of the models used are mentioned below:

- VGG16 has 16 trainable layers, of which the first five were left fixed and the other layers were left trainable, modifying the fully connected layers by 256 and 128.
- VGG19 has 19 trainable layers of which the first 7 were left untrained and the others were left trainable, modifying the fully connected layers by 256 and 128.
- DenseNet201 has 706 trainable layers of which the first 150 were left untrained and the others were left trainable, also modifying the modules of the fully connected layers.

- InceptionV3 has 310 trainable layers of which the first 150 were left untrained and the other layers were left trainable, also modifying the modules of the fully connected layers.
- Inception-ResNetV2 has 728 trainable layers of which the first 150 were left un-trained and the others were left trainable, also modifying the modules of the fully connected layers.

F. HYPERPARAMETER OPTIMIZATION

Initially, we focused on fine tuning, this step is explained in detail in Section II E Transfer Learning. We then carried out an exhaustive search to determine the appropriate number of fully connected hidden layers for the model classification layers. Figure 3 shows the number of fully connected hidden layers of the VGG16 model which is 256 for its first dense block and 128 for the second one. Both blocks use ReLu as non-linear activation functions. On the other hand, we use batch normalization layers in the dense blocks of the models, applying a transformation that keeps the output mean close to 0 and the output standard deviation close to 1, this type of normalization modifies its variance between 0 and 1. That said, we use the ReLu activation function after batch normalization for the fully connected layer, as shown in Figure 3 allowing the deeper networks to converge more easily. We also searched for the best optimizer for the five chosen models, among multiple optimizers such as Adam, Adamax, Adamgrad, SGD, among others provided by Keras, finally choosing the Adam optimizer. The output layer was adjusted to the number of classes contained in the dataset, and with a Softmax activation function. Finally, the Learning Rate Scheduler function allows the model to decrease the learning rate as time passes. Thus, the model learns more slowly which allows to find with higher probability a local minimum or global minimum, as a consequence the model converges more quickly. This model was trained for an initial learning rate of 0.0001 for the initial epoch and finally, the Learning Rate Scheduler function allowed us to modify the learning rate down to 0.000001 after 100 epochs.

G. CROSS-VALIDATION

Machine learning models often do not generalize adequately when trained on a small database, or often the results vary significantly, so splitting the data into training, validation and testing is not the best approach when you have a database with little information. Therefore, cross-validation (CV) is used to assess model generalization when a small volume of data is available. CV divides the dataset into equal amounts, also known as folds, defined by the experimenter. If an experimenter decides to take 5 folds, the data set is divided into 5 equal parts training is performed with 4 folds and testing with 1. Subsequently, the test set is passed to the training set and one fold from training is passed to testing. This is repeated until all possible combinations are completed, ensuring that all data pass at least once through the training

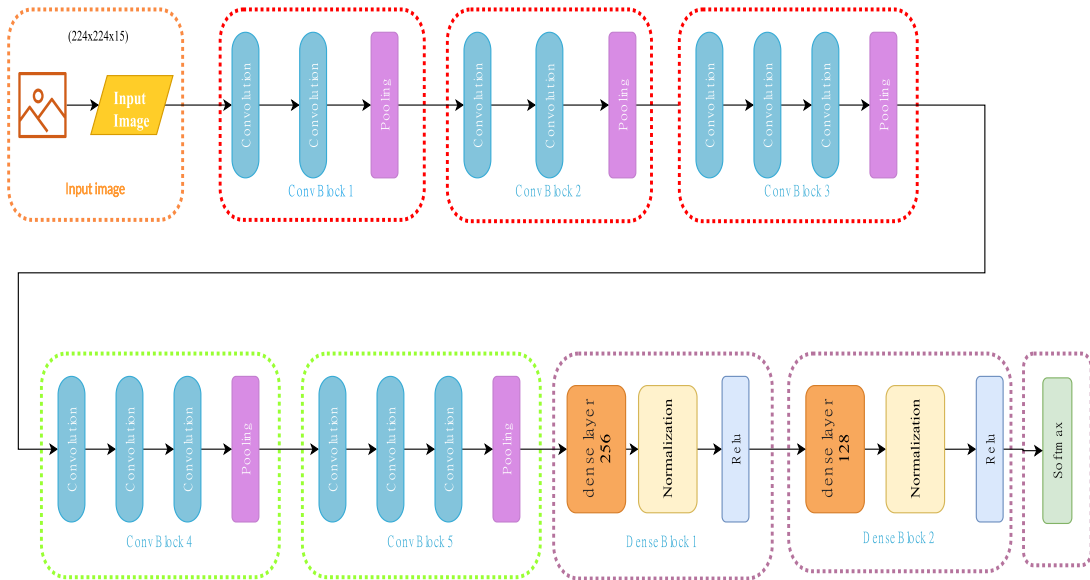


FIGURE 3. Trainable and non-trainable layers of the VGG16 model used by TL training.

and once through the test, evaluating the generalization of the model [52].

H. METRICS

The metrics used to evaluate the generalization performance of the models in this work are the same as those utilized in the literature [53]:

- Accuracy: this measures the overall percentage of samples that the model has correctly classified, allowing us to measure the quality of the model and the number of hits.
- Average recall: the recall metric calculates what portion of the true positives the model correctly classified as positive. This metric is used to select the best model when there is a high cost associated with false negatives.
- Precision: this is the fraction of all relevant instances divided by the instances obtained. It is used to measure the quality of the model, identifying the positive predictions that were actually correct.
- F1-score: this is used to combine the precision and recall measures into a single value. This is practical because it makes it easier to compare the combined performance of accuracy and completeness(recall) between various solutions, regardless of whether the test set is balanced or not. It should also be noted that it is the most widely used metric to measure the classification capability of a model in unbalanced databases.
- Standard deviation [σ]: this number describes how spread out the values are. A low standard deviation (SD) means that most of the numbers are close to the mean (average) value. A high SD means that the values are spread out over a wider range. [54]

III. EXPERIMENTS AND RESULTS

Our experimental protocol included one training and one inference run with the dataset as it was collected. The original

dataset is slightly unbalanced, so in addition to the experiment with the unbalanced dataset, other experiments were carried out utilizing as input the dataset balanced using the following three techniques:

- 1) Balancing to the smallest class (78 images per category).
- 2) Balancing to the largest class, generating an artificial increase of data (160 images per category).
- 3) Adjusting the model parameters by obtaining weights for the training data, penalizing the class with more data, and giving more importance to the class with fewer data.

The Keras class weight library can set the class weight for each class when the data set is unbalanced. For example, if we have 5000 samples of a class “x” and 45000 samples of a class “y”, then the weighting obtained by class weight would be 0:5, 1:0.5. That gives class “x” 10 times the weight of class “y”, which means that in its loss function it assigns a higher value to these instances.

The loss becomes a weighted average when the weight of each sample is specified by class weight and its corresponding class.

Using this technique, the following weights were obtained for each class in Table 4:

For the best performing experiment, a distribution of 70% for training data and 30% for test data was guaranteed. And finally, 20% of the training data was used for validation.

A. EXPERIMENT 1: UNBALANCED

All CNNs were trained with an unbalanced data distribution as shown in Table 3. Table 5 shows a comparison between the results of the 10-fold cross-validation without applying TL and applying TL.

TABLE 4. Weights calculated by Keras class weighth library for balancing data.

Ripening stage	Class weight
Inmature	0.9955
Semi-mature	0.8000
Mature	0.7593
Overripe	1.2108
Dry	1.6592

B. EXPERIMENT 2: DOWNSAMPLING

For this experiment, the models were trained to use a dataset balanced by subsampling according to the minority class, which in this case was the dry class. The balanced data set contains 78 images per class. The results achieved by this experiment are shown in Table 6.

C. EXPERIMENT 3: UPSAMPLING

For this experiment, the data were artificially up-sampled, so that 160 images per class were available at training time. This data balancing was achieved by randomly rotating, translating, and reflecting the images corresponding to all classes except the majority one. This sort of data preprocessing allowed our model to achieve 97.18% accuracy when transfer learning was applied and 97.81% accuracy without applying TL, as shown in Table 7.

D. EXPERIMENT 4: WEIGHTING

In this experiment, instead of balancing the dataset before training, the training procedure was modified by utilizing a loss function weighted according to the class frequencies in the dataset and thus allowing the training samples to have an influence that is proportional to these frequencies: the under represented class ends up being more influential. The assigned weights are shown in Table 4. The accuracies obtained are as follows: up to 98.75% when using 10-fold cross-validation without TL and 99.84% by applying TL as shown in Table 8.

E. BEST CONFIGURATION

The best results were obtained during **experiment 1**, in which the model **DenseNet201** was trained by **applying TL** with the unbalanced data, as shown in Table 3. Fine tuning the model, previously trained on ImageNet, made training converge faster and with better results, compared to the training without TL, as shown in Table 5.

Figure 4 shows that the prediction improves as a function of the epochs in both training and validation thus, showing that the model has learned to classify coffee beans, with a high percentage of accuracy, according to their ripening stages.

As can be seen in Figure 5, both the learning and validation loss curves decrease in a very similar way as the epochs increase, from which it can be deduced that the model does not overfit.

In order to evaluate the proposed model for each class, precision, recall and F1-score metrics were calculated as shown in Figure 6.

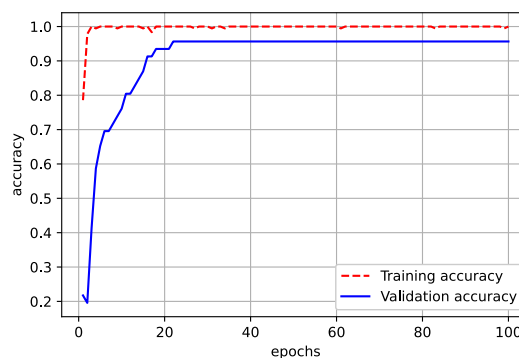


FIGURE 4. Evolution of training and validation for DenseNet201 accuracy with unbalanced database and applying TL.

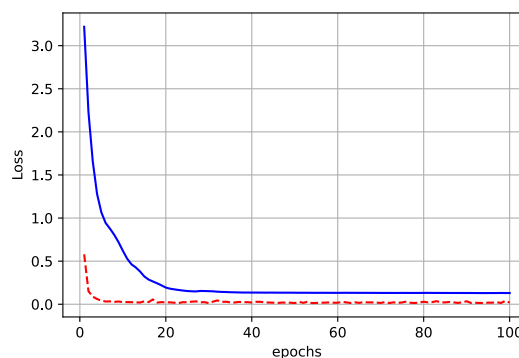


FIGURE 5. Loss function curve during DenseNet201 training and validation with an unbalanced database and applying TL.

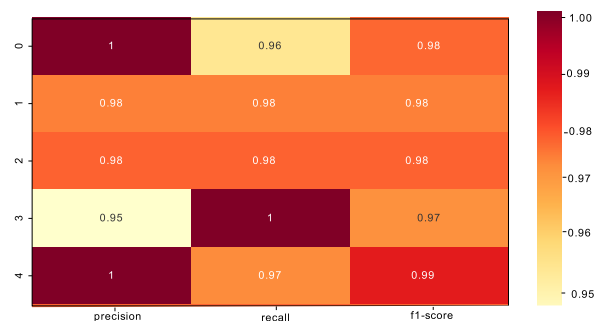


FIGURE 6. Precision, recall and f1-score calculated for each class applying transfer of learning in the DenseNet201 model.

The confusion matrix shown in Figure 7 helps us to observe the errors that the model has and for which classes it is making mistakes more frequently.

Finally, the ROC curve shown in Figure 8 represents the true positives and false positives, which demonstrates the success rate of our model for the five classes corresponding to the stages of coffee ripening.

IV. DISCUSSION

In this paper, four experiments were carried out to demonstrate the potential of CNNs in the classification of five different coffee ripening stages; these different experiments aimed to demonstrate the level of reliability of some of the most popular convolutional neural network architectures and how each can find different features, important to perform a classification process from multispectral images. In addition, this

TABLE 5. Metrics calculated in the 10 folds cross-validation for the CNNs with the unbalanced database, without applying and applying TL.

Experiment 1: Unbalanced Dataset												
CNN	Without Applying TL						Applying TL					
	Accuracy		Sensitivity		F1-score		Accuracy		Sensitivity		F1-score	
	[%]	SD	[%]	SD	[%]	SD	[%]	SD	[%]	SD	[%]	SD
VGG16	68.43	21.21	68.12	20.87	64.05	25.90	84.79	18.65	83.53	22.00	83.87	20.72
VGG19	62.34	13.39	65.64	13.30	58.41	15.60	92.39	2280	90.60	28.20	90.96	27.09
Inception-ResNetV2	99.06	1.25	99.05	1.23	98.99	1.34	100.00	0.00	100.00	0.00	100.00	0.00
InceptionV3	99.06	1.87	99.08	1.80	99.12	1.78	97.19	5.94	98.57	2.71	97.90	4.14
DenseNet201	94.21	15.27	93.70	16.81	92.54	20.38	100.00	0.00	100.00	0.00	100.00	0.00

TABLE 6. Metrics calculated in the 10 folds cross-validation for the CNNs with the database balanced to the minor class (downsampling) without applying transfer learning and applying transfer learning.

Experiment 2: Downsampling												
CNN	Without Applying TL						Applying TL					
	Accuracy		Sensitivity		F1-score		Accuracy		Sensitivity		F1-score	
	[%]	SD	[%]	SD	[%]	SD	[%]	SD	[%]	SD	[%]	SD
VGG16	71.01	13.44	71.04	13.16	67.88	14.73	91.09	18.77	91.76	19.88	90.55	21.40
VGG19	62.99	23.86	62.16	24.16	57.89	28.87	93.90	9.79	96.24	5.27	93.92	8.71
Inception-ResNetV2	97.36	3.74	97.67	3.65	97.36	3.77	99.37	1.87	99.49	1.52	99.44	1.67
InceptionV3	98.22	2.77	98.33	2.54	98.29	2.64	96.56	3.26	96.33	3.95	96.19	3.59
DenseNet201	94.66	7.83	94.78	7.64	93.91	9.42	100.00	0.00	100.00	0.00	100.00	0.00

TABLE 7. Metrics calculated in the 10 folds cross-validation for the CNNs with the database balanced to the class with the highest amount of data (oversampling) without applying transfer learning and applying transfer learning.

Experiment 3: Artificially Increasing Data												
CNN	Without Applying TL						Applying TL					
	Accuracy		Sensitivity		F1-score		Accuracy		Sensitivity		F1-score	
	[%]	SD	[%]	SD	[%]	SD	[%]	SD	[%]	SD	[%]	SD
VGG16	91.87	7.32	92.35	9.36	89.66	10.92	80.00	19.15	81.75	20.10	78.49	22.71
VGG19	70.93	13.42	72.49	19.30	67.82	18.36	85.15	17.03	86.44	20.49	84.41	19.31
Inception-ResNetV2	94.68	7.50	96.86	3.26	95.26	6.57	97.18	5.62	97.88	4.18	97.40	5.28
InceptionV3	71.87	20.39	76.73	19.49	67.99	22.64	91.71	4.74	93.52	3.38	92.32	4.02
DenseNet201	97.81	3.64	98.04	3.39	97.81	3.72	97.18	3.54	97.40	3.65	97.19	3.75

TABLE 8. Metrics calculated in the 10 folds cross-validation for the CNNs implementing the weighting method without applying transfer learning and applying transfer learning.

Experiment 4: Weighting												
CNN	Without Applying TL						Applying TL					
	Accuracy		Sensitivity		F1-score		Accuracy		Sensitivity		F1-score	
	[%]	SD	[%]	SD	[%]	SD	[%]	SD	[%]	SD	[%]	SD
VGG16	83.28	20.24	87.23	17.89	81.54	22.59	85.00	19.70	87.31	17.93	82.51	22.63
VGG19	90.15	21.85	90.01	22.22	89.55	23.92	85.93	22.64	88.39	23.62	85.17	25.31
Inception-ResNetV2	98.75	1.53	98.52	1.70	98.55	1.61	99.84	0.46	99.87	0.37	99.87	0.36
InceptionV3	97.03	6.34	97.59	5.41	96.94	6.40	94.06	5.21	95.24	4.15	93.72	5.28
DenseNet201	98.75	1.95	98.91	1.59	98.84	1.60	99.06	2.33	99.17	2.07	99.07	2.32

study allows us to identify, from efficiency measures such as precision and the F1-score, the impact of data augmentations and different tradeoffs on the original database, in order to present a possible benchmark for future research in the field of agriculture-oriented machine learning. As a regulatory test in this type of work, a 10 folds cross-validation is presented, so that a generalization of the result can be identified and no bias is incurred.

This work demonstrates the use of validated CNN architectures, TL techniques, and data scaling, on the classification of coffee fruit ripening from multispectral images. To the best of our knowledge, these technologies had not been applied to the coffee fruit ripening classification problem as reported in the literature. In this study, 10 folds cross-validation was implemented in order to validate the trained models and calculate performance metrics. Given the low number of captured images in the database, analyzing these results shows

a significant improvement in the mentioned metrics when implementing TL with the ImageNet pre-trained weights.

The best results were obtained with an unbalanced data set after applying TL, as shown in Table 5. 100% accuracy and 0% standard deviation were achieved using 10-fold cross-validation for the Inception-ResNet v2 and the DenseNet 201 architectures. These results are very encouraging because the experiment does not need synthetic data augmentation, which on many occasions can prove to be negative because of the potential biases introduced by data augmentation. In Table 6, the DenseNet 201 model also had 100% accuracy after applying TL, but in this case, a data balancing procedure was performed with respect to the lowest class, which means that information was removed from the dataset, a process that is not normally recommended in Deep Learning tasks due to the biases and overfitting that a small dataset may cause.

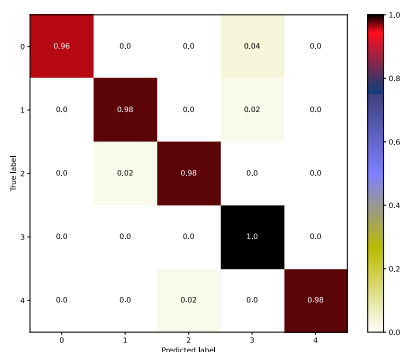


FIGURE 7. Confusion matrices of the proposed TL models, where classes 0 to 4 represent ‘dry’, ‘mature’, ‘semi-mature’, ‘overripe’ and ‘immature’, respectively.

In Table 7 the accuracies are approximately 97% but with standard deviations between 3.64% and 3.54% (with or without TL) for the DenseNet201, which shows that an increase of synthetic data in this type of problem does not guarantee the best accuracies or the best generalization of the model. We claim that this is due to the fact that generating a data increase by adding synthetic samples, which may not align with the real data generation process can cause certain biases, worsening the results. Table 8 shows that when a class penalty balancing is performed, the results are very encouraging with TL, with an accuracy close to 98.75% and a standard deviation close to 1.95%, leaving as evidence that this type of balancing can be a good option for future research, in cases where the imbalances are much larger than those tackled in this publication. In Tables 5 to 8, it can be seen that the best architectures, without applying TL are Inception V3 and DenseNet 201. Applying TL, the best models are Inception ResNetV2 and DenseNet 201. By observing the trend in the results of the DenseNet 201 network and taking as reference Table 5 with TL, a deep analysis of the experiment showed that, as illustrated in Figures 5 and 6, during the training process, the percentages in the validation set are close to 100% without generating overfitting. Additionally, it is observed that the training process has quickly converged, after just 30 epochs. Finally, Figures 7, 8 and 9, show that despite the class imbalance in the experiment, the behavior of the different metrics, the confusion matrix, and the ROC curves are very stable and accurate for each class independently, evidencing the good generalization of the models and discarding the possibility of overfitting.

Along with this research article, the database and code used for experimentation have been published at: <https://doi.org/10.5281/zenodo.4914786> and <https://github.com/BioAITeam/Coffee-Maturity-Classification-using-Convolutional-Neural-Networks-and-Transfer-Learning>. Thanks to the experience gained in recent years by the different members of the workgroup, we have been able to record and recognize different combinations of techniques that can be applied to classification processes from CNNs, and that can improve classification systems such as the one presented in this article. In this way, multispectral systems that generate large amounts of information can benefit, both from the

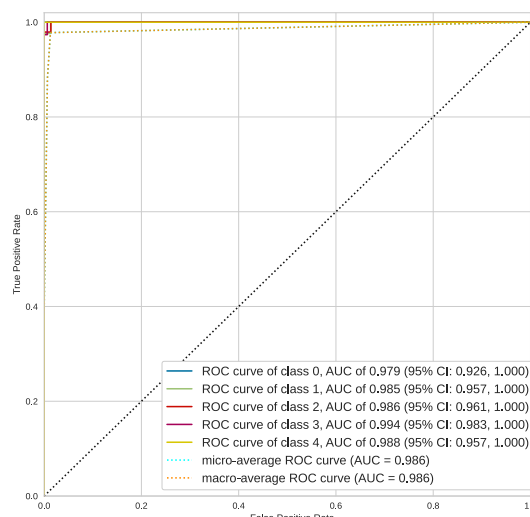


FIGURE 8. ROC curves of the proposed TL models with 95% confidence interval. Note that classes 0 to 4 represent ‘dry’, ‘mature’, ‘semi-mature’, ‘overripe’ and ‘immature’, respectively.

results found and from the possibility of comparison with a reference.

V. CONCLUSION

In this work, we developed a total of four experiments for the classification of cherry coffee fruit ripening stages, each experiment comparing five different convolutional neural network architectures and the application of TL. In these experiments, the TL allows us to train the models more efficiently, in addition to identifying the different techniques for hyperparameter, optimization, class leveling and TL techniques to experiment in training the models when classifying the different stages of coffee ripening.

Five models well recognized for their efficiency were used, namely VGG16, VGG19, Inception-ResNet-V2, Inception-V3, and DenseNet201, where the best results were achieved by DenseNet201, in which it was not necessary to use data augmentation, which can generate biases or overtraining, but it is important to use TL. DenseNet201 under different accuracy metrics popularly used in this type of experiments achieved up to 98% accuracy on the dataset and 100% accuracy on cross-validation.

With the publication of the multispectral images of the coffee database and the modifications made in the CNN architectures, we define a benchmark for future research in the area. We show precise results to the scientific community about the potential of joining the CNN with multispectral systems. It also provides a tool for precision agriculture to aim for better quality when classifying the coffee fruit automatically, guarantees a better quality than traditional coffees, and provides more characteristic data of the fruit when seeking a special coffee production.

As future work, we propose a deeper study of the networks, including the testing of more complex and newly designed CNN architectures, oriented to precision agricultural classification problems based on spectral images. We also propose to implement a color reproduction from the multispectral

database. We then seek to apply these algorithms to different data extracted from multispectral images, such as the relationship of the true color of fruits with traditional classification processes, or the possibility of extracting information outside the visible spectrum to allow early identification of physicochemical conditions and even early detection of diseases all this through the improvement of the multispectral system and the increase of the database. Also we propose to increase the database through using artificial data generated by generative antagonistic networks (GANs) and including information from different coffee varieties to improve the results presented in this research and further apply the benchmark.

REFERENCES

- G. Gyarmati and T. Mizik, "The present and future of the precision agriculture," in *Proc. IEEE 15th Int. Conf. Syst. Syst. Eng. (SoSE)*, Jun. 2020, pp. 593–596.
- S. Cubero, N. Aleixos, E. Moltó, J. Gómez-Sanchis, and J. Blasco, "Advances in machine vision applications for automatic inspection and quality evaluation of fruits and vegetables," *Food Bioprocess Technol.*, vol. 4, no. 4, pp. 487–504, May 2011.
- K. K. Patel, A. Kar, S. N. Jha, and M. A. Khan, "Machine vision system: A tool for quality inspection of food and agricultural products," *J. Food Sci. Technol.*, vol. 49, no. 2, pp. 123–141, Apr. 2012.
- B. Zhang, W. Huang, J. Li, C. Zhao, S. Fan, J. Wu, and C. Liu, "Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review," *Food Res. Int.*, vol. 62, pp. 326–343, Aug. 2014.
- Y. A. Ohali, "Computer vision based date fruit grading system: Design and implementation," *J. King Saud Univ., Comput. Inf. Sci.*, vol. 23, no. 1, pp. 29–36, Jan. 2011.
- M. S. M. Alfatni, A. R. M. Shariff, M. Z. Abdullah, M. H. B. Marhaban, and O. M. B. Saad, "The application of internal grading system technologies for agricultural products—Review," *J. Food Eng.*, vol. 116, no. 3, pp. 703–725, Jun. 2013.
- D. Wu and D.-W. Sun, "Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review—Part I: Fundamentals," *Innov. Food Sci. Emerg. Technol.*, vol. 19, pp. 1–14, Jul. 2013.
- A. Bhargava and A. Bansal, "Fruits and vegetables quality evaluation using computer vision: A review," *J. King Saud Univ., Comput. Inf. Sci.*, vol. 33, no. 3, pp. 243–257, Mar. 2021.
- L. B. Furstenau, M. K. Sott, L. M. Kipper, E. L. Machado, J. R. Lopez-Robles, M. S. Dohan, M. J. Cobo, A. Zahid, Q. H. Abbasi, and M. A. Imran, "Link between sustainability and industry 4.0: Trends, challenges and new perspectives," *IEEE Access*, vol. 8, pp. 140079–140096, 2020.
- D. Lorente, N. Aleixos, J. Gómez-Sanchis, S. Cubero, O. L. García-Navarrete, and J. Blasco, "Recent advances and applications of hyperspectral imaging for fruit and vegetable quality assessment," *Food Bioprocess Technol.*, vol. 5, no. 4, pp. 1121–1142, 2012.
- Y.-Y. Pu, Y.-Z. Feng, and D.-W. Sun, "Recent progress of hyperspectral imaging on quality and safety inspection of fruits and vegetables: A review," *Food Sci. Technol.*, vol. 14, no. 2, pp. 176–188, 2015.
- J. Blasco, S. Cubero, R. Arias, J. Gómez, F. Juste, and E. Moltó, "Development of a computer vision system for the automatic quality grading of Mandarin segments," in *Proc. Iberian Conf. Pattern Recognit. Image Anal.* Berlin, Germany: Springer, 2007, pp. 460–466.
- J. Blasco, M. P. Diago, N. Aleixos, S. Cubero, and B. Millan, "Aplicaciones de la visión artificial en viticultura (yII)," *ResearchGate*, vol. 1, no. 2, pp. 2–6, 2014.
- S. Munera, C. Besada, J. Blasco, S. Cubero, A. Salvador, P. Talens, and N. Aleixos, "Astringency assessment of persimmon by hyperspectral imaging," *Postharvest Biol. Technol.*, vol. 125, pp. 35–41, Mar. 2017.
- S. Munera, J. M. Amigo, J. Blasco, S. Cubero, P. Talens, and N. Aleixos, "Ripeness monitoring of two cultivars of nectarine using VIS-NIR hyperspectral reflectance imaging," *J. Food Eng.*, vol. 214, pp. 29–39, Dec. 2017.
- M. Taghizadeh, A. A. Gowen, and C. P. O'Donnell, "Comparison of hyperspectral imaging with conventional RGB imaging for quality evaluation of *Agaricus bisporus* mushrooms," *Biosyst. Eng.*, vol. 108, no. 2, pp. 191–194, Feb. 2011.
- J. Qin, T. F. Burks, X. Zhao, N. Niphadkar, and M. A. Ritenour, "Development of a two-band spectral imaging system for real-time citrus canker detection," *J. Food Eng.*, vol. 108, no. 1, pp. 87–93, Jan. 2012.
- J. Xing, W. Saeys, and J. De Baerdemaeker, "Combination of chemometric tools and image processing for bruise detection on apples," *Comput. Electron. Agricult.*, vol. 56, no. 1, pp. 1–13, Mar. 2007.
- B. S. Bennedsen and D. L. Peterson, "Performance of a system for apple surface defect identification in near-infrared images," *Biosyst. Eng.*, vol. 90, no. 4, pp. 419–431, Apr. 2005.
- C. E. Oliveros-Tascón and J. R. Sanz-Urbe, "Ingeniería y café en Colombia," *Revista de Ingeniería*, no. 33, pp. 99–114, Jan. 2011.
- C. Oliveros, J. Pabón, E. Montoya, C. Ramírez, and J. Sanz, "Separación de frutos de café verdes por medios mecánicos," *Cenicafé*, vol. 61, pp. 260–269, 2010.
- S. Ponte, "Estándares, comercio y equidad: Lecciones de la industria de los cafés especiales," *Economía Mundial del café*, Centro de Investigaciones Para el Desarrollo de Copenhague, Anaheim, CA, Tech. Rep. 5 de mayo de, 2002, pp. 131–163.
- F. M. Borém, M. A. Cirillo, A. P. de Carvalho Alves, C. M. D. Santos, G. R. Liska, M. F. Ramos, and R. R. de Lima, "Coffee sensory quality study based on spatial distribution in the Mantiqueira mountain region of Brazil," *J. Sensory Stud.*, vol. 35, no. 2, 2020, Art. no. e12552.
- M. Sott, L. Furstenau, L. Kipper, F. Giraldo, J. Lpez-Robles, M. Cobo, A. Zahid, Q. Abbasi, and M. Imran, "Precision techniques and agriculture 4.0 technologies to promote sustainability in the coffee sector: State of the art, challenges and future trends," *IEEE Access*, vol. 8, pp. 149854–149867, 2020.
- Z. L. S. Ni no and F. A. P. Ortiz, "Caracterización de café cereza empleando técnicas de visión artificial," *Revista Facultad Nacional de Agronomía-Medellín*, vol. 60, no. 2, pp. 4105–4127, 2007.
- P. Ramos, J. Sanz, and C. Oliveros, "Identificación y clasificación de frutos de café en tiempo real a través de la medición de color," *Cenicafé*, vol. 61, pp. 315–326, 2014.
- N. L. M. Castrillón, "Real-time classification of coffee fruits using FPGA," Ph.D. dissertation, Dept. Cundinamarca, Universidad Nacional de Colombia, Bogotá, Colombia, 2015.
- M. A. Tamayo Monsalve, "Diseño de un sistema de adquisición de imágenes multiespectrales basado en iluminación led de potencia de ancho de banda estrecho," Ph.D. dissertation, Dept. Cundinamarca, Universidad Nacional de Colombia, Bogotá, Colombia, 2020.
- L. Manrique Naranjo, "Design of an on-line multispectral coffee fruit classification and sorting system using narrowband LEDs and MCU," M.S. thesis, Dept. Cundinamarca, Ingeniería y Arquitectura-Maestría en Ingeniería-Automatización, Universidad Nacional de Colombia, 2020.
- H. C. Bazame, J. P. Molin, D. Althoff, and M. Martello, "Detection, classification, and mapping of coffee fruits during harvest with computer vision," *Comput. Electron. Agricult.*, vol. 183, Apr. 2021, Art. no. 106066.
- J. Buitrago-Osorio, H. A. Tinoco, L. Perdomo-Hurtado, A. Rincon-Jimenez, O. Ocampo, L. V. Berrio, M. F. Pineda, and J. Lopez-Guzman, "Physical-mechanical characterization of coffee fruits *Coffea arabica* L. var. Castillo classified by a colorimetry approach," *Materialia*, vol. 21, Mar. 2022, Art. no. 101330.
- A. G. Costa, D. A. G. D. Sousa, J. L. Paes, J. P. B. Cunha, and M. V. M. D. Oliveira, "Classification of robusta coffee fruits at different maturation stages using colorimetric characteristics," *Engenharia Agrícola*, vol. 40, no. 4, pp. 518–525, Aug. 2020.
- A. Podorozhniak, N. Lubchenko, O. Balenko, and D. Zhuikov, "Neural network approach for multispectral image processing," in *Proc. 14th Int. Conf. Adv. Trends Radioelectronics, Telecommun. Comput. Eng. (TCSET)*, Feb. 2018, pp. 978–981.
- M. A. Bravo-Ortiz, H. B. Arteaga-Arteaga, R. Tabares-Soto, J. I. Padilla-Buriticá, and S. Orozco-Arias, "Cervical cancer classification using convolutional neural networks, transfer learning and data augmentation," *Revista EIA*, vol. 18, no. 35, pp. 1–12, 2021.
- T. S. Reinell, A. A. H. Brayan, B. O. M. Alejandro, M. R. Alejandro, A. G. Daniel, A. G. J. Alejandro, B. J. A. Buenaventura, O. A. Simon, I. Gustavo, and R. P. Raúl, "GBRAS-Net: A convolutional neural network architecture for spatial image steganalysis," *IEEE Access*, vol. 9, pp. 14340–14350, 2021.
- L. Cavigelli, D. Bernath, M. Magno, and L. Benini, "Computationally efficient target classification in multispectral image data with deep neural networks," *CoRR*, vol. 10, 2016.

[37] J. Yang, Y.-Q. Zhao, and J. Chan, "Hyperspectral and multispectral image fusion via deep two-branches convolutional neural network," *Remote Sens.*, vol. 10, no. 5, p. 800, May 2018.

[38] J. Li and Z. Liu, "Multispectral transforms using convolution neural networks for remote sensing multispectral image compression," *Remote Sens.*, vol. 11, no. 7, p. 759, Mar. 2019.

[39] R. Tabares-Soto, S. Orozco-Arias, V. Romero-Cano, V. S. Bucheli, J. L. Rodríguez-Sotelo, and C. Jiménez-Varón, "A comparative study of machine learning and deep learning algorithms to classify cancer types based on microarray gene expression data," *PeerJ Comput. Sci.*, vol. 6, Apr. 2020, Art. no. e270.

[40] X. Zhang, J. Zou, K. He, and J. Sun, "Accelerating very deep convolutional networks for classification and detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 10, pp. 1943–1955, Oct. 2016.

[41] J. Xiao, J. Wang, S. Cao, and B. Li, "Application of a novel and improved VGG-19 network in the detection of workers wearing masks," *J. Phys., Conf. Ser.*, vol. 1518, no. 1, Apr. 2020, Art. no. 012041.

[42] Y. Bhatia, A. Bajpayee, D. Raghuvanshi, and H. Mittal, "Image captioning using Google's inception-ResNet-v2 and recurrent neural network," in *Proc. 12th Int. Conf. Contemp. Comput. (IC3)*, Aug. 2019, pp. 1–6.

[43] A. H. Shahin, A. Kamal, and M. A. Elattar, "Deep ensemble learning for skin lesion classification from dermoscopic images," in *Proc. 9th Cairo Int. Biomed. Eng. Conf. (CIBEC)*, Dec. 2018, pp. 150–153.

[44] A. Adegun and S. Viriri, "Deep learning techniques for skin lesion analysis and melanoma cancer detection: A survey of state-of-the-art," *Artif. Intell. Rev.*, vol. 54, pp. 811–841, Jun. 2020.

[45] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.

[46] F. Y. Orozco, "Descripción de especies y variedades de café," *Cenicafé*, vol. 11, p. 29, 1978.

[47] A. Julca-Otiniano, W. Solano-Arrue, and R. Crespo-Costa, "Crecimiento de *Coffea arabica* variedad caturra amarillo en almácigos con sustratos orgánicos en Chanchamayo, selva central del Perú," *Investigación Agraria, Product Protein Vegetable*, vol. 17, no. 3, pp. 353–365, 2002.

[48] J. P. Rodríguez, D. C. Corrales, J.-N. Aubertot, and J. C. Corrales, "A computer vision system for automatic cherry beans detection on coffee trees," *Pattern Recognit. Lett.*, vol. 136, pp. 142–153, Aug. 2020.

[49] X. Ran and B. Xiao, "Multi-spectral image analysis for rescue target detection," in *Proc. Int. Assoc. Institutes Navigat. World Congr. (IAIN)*, Oct. 2015, pp. 1–4.

[50] P. Colantoni, R. Pillay, C. Lahanier, and D. Pitzalis, "Analysis of multispectral images of paintings," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2006, pp. 2–6.

[51] M. A. T. Monsalve, G. Osorio, N. L. Montes, S. Lopez, S. Cubero, and J. Blasco, "Characterization of a multispectral imaging system based on narrow bandwidth power LEDs," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.

[52] S. Manna. (Mar. 20, 2020). *K-fold cross validation for deep learning models using keras*. [Online]. Available: <http://medium.com/the-owl/k-fold-cross-validation-in-keras-3ec4a3a00538>

[53] K. P. Shung. (Mar. 15, 2018). *Accuracy, precision, recall or F1?* [Online]. Available: <https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>

[54] Z. Huo, G. Du, F. Luo, Y. Qiao, and J. Luo, "D-MSCD: Mean-standard deviation curve descriptor based on deep learning," *IEEE Access*, vol. 8, pp. 204509–204517, 2020.



ESTEBAN MERCADO-RUIZ is currently a Biomedical Engineer graduated from the Universidad Autónoma de Manizales. He has worked on projects of human pose estimation using deep learning. He has supported research to classify skin lesions and skin cancer states, and also researches on the detection of respiratory system diseases from chest X-ray images.



JUAN PABLO VILLA-PULGARIN is currently pursuing the bachelor's degree in biomedical and electrical engineering with the Universidad Autónoma de Manizales. He has been a member of the Research Group on Bioinformatics and Artificial Intelligence, since 2020. He has worked on the projects of optimized convolutional neural network models for skin lesion classification and classification of coffee using CNNs and transfer learning.



MARIO ALEJANDRO BRAVO-ORTÍZ received the degree in biomedical and electrical engineering from the Universidad Autónoma de Manizales. He has been a Member of the Research Group on Bioinformatics and Artificial Intelligence, since 2018. He has been a Professor with the Department of Electronics and Automation, since 2021. His current research interests include the application of CNNs to steganalysis, detection of cancer through deep learning, and detecting respiratory system diseases from chest X-ray imaging.



HAROLD BRAYAN ARTEAGA-ARTEAGA is currently pursuing the bachelor's degree in electronic engineering with the Universidad Autónoma de Manizales. He has been a member of the Research Group on Bioinformatics and Artificial Intelligence, since 2018. His research interests include applying deep learning to steganalysis and detecting respiratory system diseases from chest X-ray imaging, machine learning applications to predict two-phase flow patterns, and text classification based on natural language processing techniques.



MANUEL ALEJANDRO TAMAYO-MONSALVE received the B.Sc. degree in electronic engineering, the M.Sc. degree in industrial automation, and the Ph.D. degree in automatic engineering from the Universidad Nacional de Colombia sede Manizales, in 2012, 2015, and 2020, respectively. His research interests include image processing, computer vision, and electronic design.



ALEJANDRO MORA-RUBIO is currently pursuing the bachelor's degree in biomedical and electrical engineering with the Universidad Autónoma de Manizales. He has been a part of the University Research Group on Bioinformatics and Artificial Intelligence, since 2018. He has worked on projects involving electrophysiological signals classification using machine learning. He has supported projects as applying deep learning techniques in digital media steganalysis and detection of respiratory system diseases from chest X-Ray imaging.



JESUS ALEJANDRO ALZATE-GRISALES is currently pursuing the bachelor's degree in biomedical and electrical engineering with the Universidad Autónoma de Manizales. He has been a member of the Research Group on Bioinformatics and Artificial Intelligence, since 2018. He has worked on projects involving electrophysiological signals classification using machine learning. He has supported projects, such as applying deep learning techniques in digital media steganalysis, detecting Alzheimer's from 3D magnetic resonance, and detecting respiratory system diseases from chest X-ray imaging.



DANIEL ARIAS-GARZON received the degree in biomedical engineering from the Universidad Autónoma de Manizales. He has been a member of the Automatic Research Hotbed in the line of Artificial Intelligence, since 2018. He has worked on projects involving human pose estimation and artificial vision using machine learning, and also on time series and disease detection on images.



VICTOR ROMERO-CANO (Member, IEEE) received the bachelor's degree in mechatronic engineering from the Universidad Autónoma de Occidente, Colombia, in 2007, the M.Sc. degree in electrical engineering from the Universidade de Sao Paulo, Brazil, in 2010, and the Ph.D. degree from the Australian Centre for Field Robotics, The University of Sydney, Australia, in 2015. He joined Inria as a Postdoctoral Research Associate within the CHROMA Team, in 2016. Since 2017, he has been a Senior Lecturer and a Researcher with the Universidad Autónoma de Occidente. His research interests include multi-sensor robotic perception and probabilistic graphical models.



SIMON OROZCO-ARIAS received the B.S. degree in system and computing engineering and the Ph.D. degree in engineering (bioinformatics) from the Universidad de Caldas, Colombia, in 2016 and 2022, respectively. He is currently working as a Professor and a Researcher with the Universidad Autónoma de Manizales, developing projects on bioinformatics, supercomputing, and data analysis. He has contributed to the solution of biological problems, especially, related to plant DNA. His current research interest includes the application of machine learning techniques to automatically extract knowledge from DNA data.



GUSTAVO-OSORIO received the B.Sc. degree in electronic engineering and the M.Sc. degree in industrial automation from the Universidad Nacional de Colombia sede Manizales, in 1997 and 2003, respectively, and the Ph.D. degree in computer and control engineering from the Università degli Studi di Napoli Federico II, Italy, in 2008. He has been an Assistant Instructor with the Department of Electrical, Electronic and Computer Engineering, Universidad Nacional de Colombia sede Manizales, since 1998. In 2013, he became an Associate Professor. His research interests include dynamical systems and control, computer engineering, and electronic design.



REINEL TABARES-SOTO received the B.S. degree in electronic engineer from the Universidad Nacional de Colombia, in 2009, the B.S. degree in systems and computer engineer from the Universidad de Caldas, Colombia, in 2016, the M.S. degree in engineering from the Universidad Nacional de Colombia, in 2017, and the Ph.D. degree in computer science from the Universidad Autónoma de Manizales, Colombia, in 2021. He has been a Professor and a Researcher with the Department of Electronics and Automation, Universidad Autónoma de Manizales, since 2014. He has been also a data science consultant for different public and private companies in Colombia since 2017. He is currently holding a Postdoctoral position at the Adolfo Ibáñez University in the framework of the ethical and responsible algorithms project. His main research interests include steganalysis, machine learning, deep learning, bioinformatics, and high-performance computing.

...