Received: 3 October 2021 | Revised: 13 December 2021 | Accepted: 15 December 2021
DOI: 10.1002/cem.3388

SPECIAL ISSUE - RESEARCH ARTICLE

Journal of CHEMOMETRICS WILEY

Combining computer vision and deep learning to classify varieties of *Prunus dulcis* for the nursery plant industry

Sergio Borraz-Martínez[1,2] | **Joan Simó**[1,3] | **Anna Gras**[1] | **Mariàngela Mestre**[2] | **Ricard Boqué**[4] | **Francesc Tarrés**[5]

[1]Department of Agri-Food Engineering and Biotechnology, Universitat Politècnica de Catalunya, Castelldefels, Spain
[2]Center of Initial Materials, Agromillora Iberia SLU, Sant Sadurní d'Anoia, Spain
[3]Fundació Miquel Agustí, Castelldefels, Spain
[4]Department of Analytical Chemistry and Organic Chemistry, Universitat Rovira i Virgili, Tarragona, Spain
[5]Signal Theory and Communications Department, UPC Barcelona Tech, Barcelona, Spain

Correspondence
Sergio Borraz-Martínez, Department of Agri-Food Engineering and Biotechnology, Universitat Politècnica de Catalunya, Campus Baix Llobregat, Esteve Terrades 8, Castelldefels 08860, Spain.
Email: sergio.borraz@upc.edu

Funding information
Spanish Ministry of Economy and Competitiveness, Grant/Award Number: AGL2015-70106-R, AEI/FEDER, UE; Industrial Doctorates Plan of the Secretariat of Universities and Research of the Department of Economy and Knowledge of the Generalitat of Catalonia, Grant/Award Number: DI-COF 2017

Abstract

Varietal control to avoid unwanted varietal mixtures is an important objective for the nursery plant industry. In this study, we have developed and analyzed the capabilities of a computer vision system based on deep learning for the control of plant varieties in the nursery plant industry and for evaluating its capabilities. For this purpose, three datasets of nursery plant images were compared. The datasets came from two varieties of almond trees (*Prunus dulcis*) named *Soleta* and *Pentacebas*. Each dataset contained images with three different scales: whole plant, leaf, and venation. The Gradient-weighted Class Activation Mapping (Grad-CAM) technique was used to unveil the most important features to discriminate between both varieties. The three datasets provided classification accuracies above 97% in the test set, being the leaf dataset, with a 98.8% accuracy, the one providing the best results. Concerning the most important features of the plants, the Grad-CAM showed that they are located in the center of the leaf, that is, the venation. In conclusion, we have shown that computer vision is a promising technique for the control of plant varietal mixtures.

KEYWORDS

computer vision, convolutional neural network, deep learning, nursery plant, varietal mixture

1 | INTRODUCTION

Rapid discrimination between vegetal varieties is a key requirement in nursery plant production.[1] The appearance of varietal mixtures within a batch, which should be homogeneous, is an important trouble, not only because the customer receives unwanted vegetal material but also because nursery plant companies may face expensive legal suits and the deterioration of their corporative image.[2] In this context, we propose using computer vision based on deep learning techniques (CV-DL) as a possible solution to sort out this problem.

Computer vision is an old discipline that has recently undergone a qualitative improvement thanks to the use of deep learning algorithms. The goals of computer vision are to achieve a higher level of operational efficiency and productivity, as well as a higher level of automatization.[3] Moreover, it is one of the new technologies that drives the so-called Industry 4.0.

Deep learning is a specific subfield of machine learning, which in turn is included in the field of artificial intelligence. Deep learning could be defined as a modality of learning representations from data that emphasize learning successive layers of increasingly meaningful representations.[4] Each layer may be viewed as a nonlinear transform of the outputs of the precedent layer. The transformation parameters (weights) are learned using a supervised strategy. In other words, it consists of training a model from examples of the classes that we want to characterize, in our case the two varieties of almond trees. During the training process, the model calculates the classification error, which is represented as a mathematical function named *loss*, and with a small set of instructions, it autonomously modifies parameters (weights) of the model to reduce the loss, which allows improving the accuracy of the classification.[5] This process is known as model *learning*. The *deep* in *deep learning* is a reference to the large number of interconnected successive layers, stacked on top of each other and composed of computational neurons, which is called *neural network*. Nowadays, neural networks involve tens or even hundreds of layers. The structure of deep learning neural networks is divided into three sections. The first is the *input layer*, which receives the input images that feed the model. The middle section contains the *hidden layers*, a group of successive layers where the representation of the data takes place. This part constitutes the body of the neural network. Finally, the last section receives the name of *output layer*. This layer predicts the class to which a sample belongs based on the representation made by the hidden layers.

It was decided to use CV-DL because, in other computer vision systems, the final classification is based on the features designed by the engineer. Then, the final quality of the system depends on this previous knowledge of the best features. However, in the case of study, the most important features are very difficult to determine. Therefore, it was considered more appropriate to use CV-DL.

In the agronomic sector, the use of computer vision along with deep learning algorithms is recent but is becoming a promising technique with growing popularity.[6] This technology is fast and does not modify or destroy the samples. Moreover, it can be easily implemented or adapted to the production line, enabling the collection of useful information that can be used as a quality control tool or to monitor a process. These advantages make computer vision an attractive tool for the agricultural industry. Despite being an emerging technology, there are several studies where CV-DL has been applied, for example, for crop type classification[7,8]: crop yield estimation,[9,10] plant phenological estimation,[11,12] or even plant recognition.[13,14] In short, CV-DL techniques have great potential in the automation of the agronomic industry, and their capacities and efficiency have been improved in the last decade.[15,16]

The objective of this study was to develop a methodology to implement computer vision system for the control of plant varieties in the nursery plant industry and evaluate its capabilities. The study was developed working with two varieties of *Prunus dulcis*, *Pentacebas* and *Soleta*, which are genetically close and almost morphologically indistinguishable. Three datasets of different image scales were used, namely, whole plant, leaf, and venation. It is important to note that this study focuses on the automatic control of nursery plants in the greenhouse, which means we have avoided the manipulation of the plants. This way, we tried to simulate a possible automatic line of image acquisition without sampling. The specific objectives were (1) to compare between different image scales: whole plant, leaf, and venation, to determine which is most suitable for varietal discrimination; (2) to determine which morphological characteristics are most relevant for the classification of varieties; and (3) to evaluate the neural network complexity required for the case of study.

2 | MATERIAL AND METHODS

2.1 | Image acquisition

The RGB images used were acquired using a Nikon D5300 camera. To keep homogeneous light conditions, plants were introduced inside a professional photo studio box with a dimension of $80 \times 80 \times 80$ cm. The photo studio had integrated two LED lamps (color temperature 5,500 K, color rendering index of 93%, and 13,000 lumens), which were placed at the bottom of the box. The idea was to illuminate the leaves from the bottom up to highlight their venous architecture. A textile diffuser was positioned between the canopy plant and the light source to avoid the reflectance of the LED lights on the leaves. The camera was placed over the box and subjected by a holder; this way, apical images were collected. This method was developed in a previous study.[17] A balanced sampling was performed and 412 images per variety were used, which means 824 images for each dataset. In total, 2,472 images were employed in the study. Table 1 shows more details about the number of samples used in the study.

2.2 | Image preparation

The whole plant image dataset was cropped in 4,000 × 4,000, 3,000 × 3,000, and 2,500 × 2,500 pixels; the leaf dataset was composed of images with a range between 2,000 × 2,000 and 1,125 × 1,125 pixels; finally, the images of the venation dataset were cropped in a range of 600 × 600 and 300 × 300 pixels. Different images sizes were used to increase the robustness of the model against scale variations. The leaf and venation datasets were obtained from the same set of original images by different cropping factors, so that leaf and venation datasets were constituted by cropping the images contained in the whole plant dataset. This strategy was chosen to reduce the acquisition time of the photos and to increase the robustness of the comparison. In addition, the resolution of all images in all sets was reduced to 224 × 224 pixels before introducing them into the model. Moreover, the pixels were normalized (1.0/255) with numeric values between zero and one. The reduction of resolution was performed to reduce the computational cost and the normalization to accelerate learning in the deep learning models, thereby decreasing computational time. Figure 1 shows examples of the images of the three data sets used.

The set of images used in this study was randomly split into a training set (75% of the images), a validation set (15% of the images), and a test set (10% of the images). Table 1 shows more details about the datasets. The validation set was used to verify the correct convergence of the model and to optimize the selection of hyper-parameters. The test set was used for the final assessment of the network performance.

2.3 | Data augmentation

Overfitting tends to occur when neural networks work with small datasets (<1,000 images), and the model fits very specific features of the training images that are not representative of the whole class. Therefore, it reduces the capability of the model to correctly classify images that do not belong to the training dataset. Hence, overfitting means that the system achieves excellent accuracy in modeling the training set at the cost of reducing accuracy in the prediction of validation and test datasets. To reduce overfitting, data augmentation was employed.

Data augmentation allows increasing the training data set from the available data. This method consists of generating additional images by modifying the original ones using pre-established random transformations.[18] In this way, it is possible to increase the diversity and the number of images during the training process. Data augmentation reduces overfitting and helps the model to better predict validation/test data and to improve its classification accuracy and robustness.[19] Data augmentation was only applied to the training set, and it consisted of shear in a range of 0.2, zoom in a range of 0.4, horizontal flip, random rotation in a range of 180°, width and height shift in a range of 0.2.

2.4 | Convolutional neural network

A convolutional neural network (CNN) is a type of network architecture that is widely used in image analysis; for this reason, CNN is commonly used in computer vision for image recognition.[20] Its structural design made it possible to

TABLE 1 Description of the study

	Percentage %	Pentacebas	Soleta	Overall dataset	Overall study
Training set	75	309	309	618	1,854
Validation set	15	62	62	124	372
Test set	10	41	41	82	246
Total		412	412	824	2,472

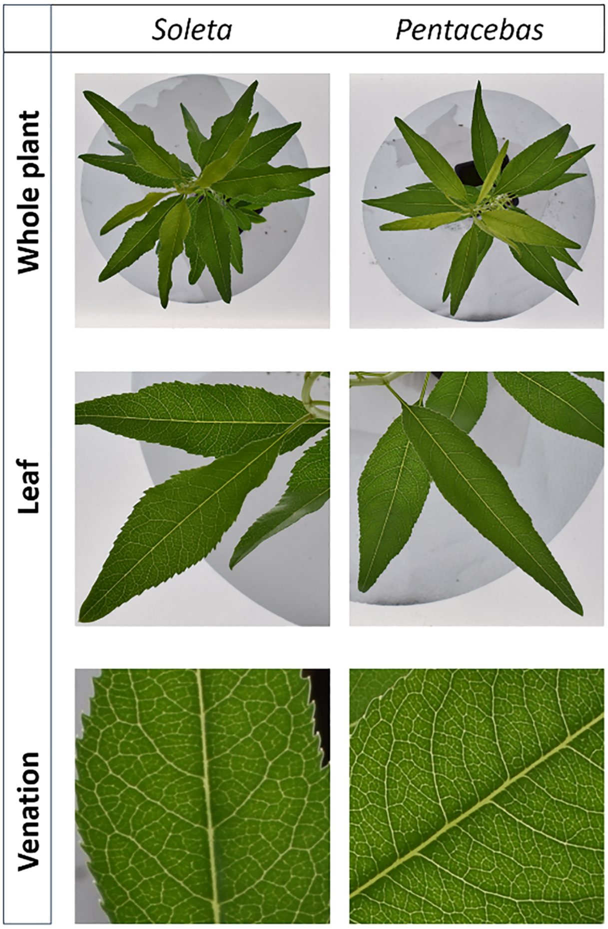

FIGURE 1 Examples of images used in the study

reduce the number of parameters of our model, which means a significant reduction in cost and computation time. The layers of a convolutional network have neurons arranged in three dimensions, so the layers have width, height, and depth. Neurons in a convolutional layer are applied using convolution to small local regions of the previous layer, so we avoid wasting completely connected neurons.[5] The convolution is translationally invariable, and kernels may detect different types of features at different scales in every layer. Furthermore, CNN also contains pooling layers that reduce the amount of information obtained from the convolutional layers, creating a condensed version of the information they contain.

In this study, two convolutional neural networks have been used: the VGG16 convolutional neural network[21] and the Shallow network. Both networks have the same types of layers, but with different structural complexity. Figure 2 shows the scheme of both neural networks.

The Shallow CNN was created specifically for this study. The goal was to build a CNN with less structural complexity than the VGG16 network, that is, with a lower number of parameters. The convolutional base of the Shallow CNN is organized in five blocks. Blocks 1 and 2 contain two convolutional layers, and blocks 3–5 contain one convolutional layer. Concerning the number of filters, blocks 1 and 2 have 32 and 64 filters, respectively. Blocks 3–5 have 128 filters. Every convolutional layer uses a filter with a receptive field of 3×3. The convolution stride is one pixel, and the spatial padding of the convolutional layer input is such that the spatial resolution is preserved after convolution. In addition, one max-pooling layer over a 2×2 pixel window is located at the end of each block. The classifier is composed of two fully connected layers. The first one has 512 neurons with ReLu activation. The last one has one neuron with sigmoid activation.

VGG16 was developed by the Geometry Group at the University of Oxford to obtain the state-of-the-art results in the Large Scale Visual Recognition Challenge (LSVRC-2014) competition.[21] This architecture has become very popular and widely used in many image recognition problems. It consists of 16 layers divided into five blocks, which are named convolutional base, together with a fully connected layer named classifier. Every block in the VGG16 is formed by a concatenation of two or three convolutional layers followed by a pooling layer with a stride of 2×2. The classifier is

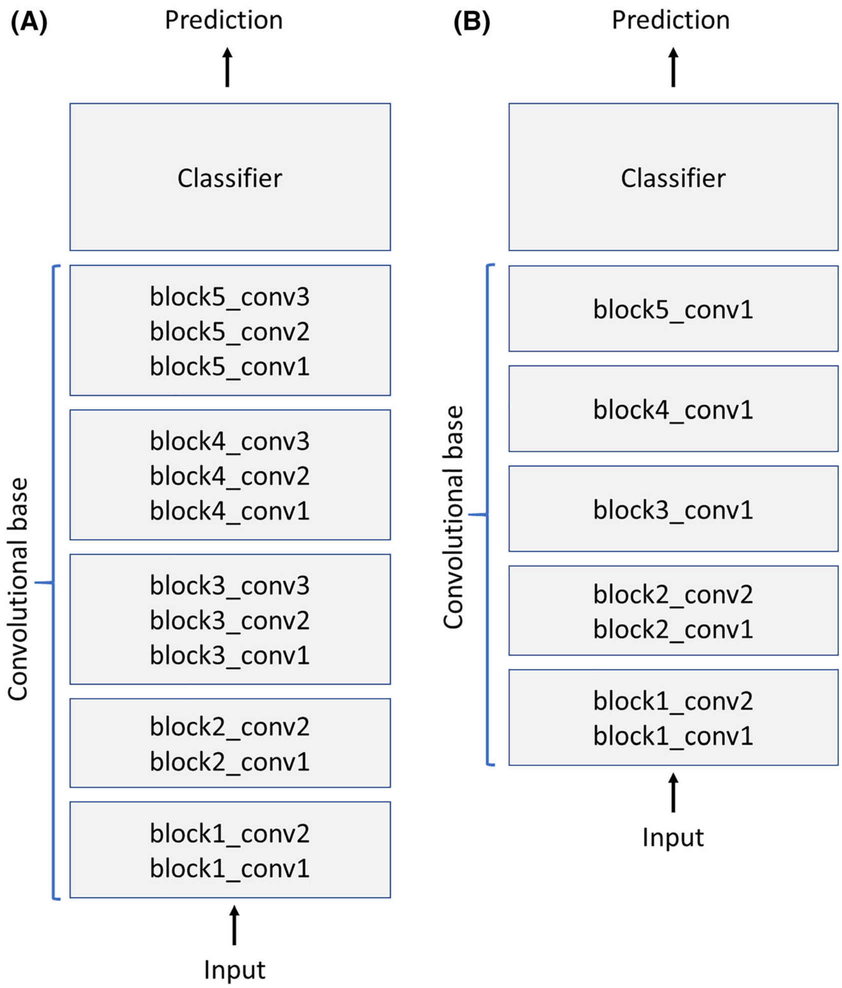

FIGURE 2 Schema of the two convolutional neural networks used. (A) VGG16; (B) Shallow

composed of three fully connected layers. The convolutional base of the VGG16 neural network was downloaded from Keras (https://keras.io/). Then, it was inserted into a new classifier, which consisted of one flattened layer followed by two dense layers. The convolutional neural networks learn local features of the input image. For this reason, the purpose of these fully connected layers is to collect the local information. The flatten layer allows transforming the 3D tensor of the CNN to a 1D tensor, which is required by the dense layer. The first dense layer contained 256 neurons with ReLU activation. This layer integrates all the local features learned by the last CNN layer, taking into account nonlinear combinations. Otherwise, without this inner layer, outputs would be a form of regression of the convolutional layers. The second dense layer had one neuron with sigmoid activation. The sigmoid activation means that the output of the neural network will be in a range between zero and one. As in this case, there are two classes; each image was encoded with a value above 0.5 if the image belonged to the *Soleta* class and below 0.5 if the image belonged to the *Pentacebas* class.

The complexity of the structure increases the time and the difficulties during the training process. Therefore, it was considered interesting to compare both CNN to assess the complexity required to develop the plant varietal control tool.

2.5 | The training steps

2.5.1 | Training of the VGG16 CNN

To facilitate the training process and try to get better results, the transfer learning technique was used. Transfer learning[22] is one of the most important techniques in deep learning as it allows us to use a pre-trained neural network.[23,24]

This is an advantage because training a neural network from scratch requires a huge dataset and a lot of computation time due to the large number of parameters that need to be adjusted. Through transfer learning, it is possible to download a pre-trained network and use its parameters as an initial good weight guess, and the fine tune the model by training it with our dataset. In this way, the representations previously learned by the network can be used to extract interesting features from new samples. In this study, a VGG16 convolutional network model, pre-trained with the dataset of ImageNet, was downloaded. This database includes 1.4 million labeled images from 1,000 different classes. ImageNet is one of the most common datasets used to pre-train CNNs in computer vision.

There are two modalities of transfer learning, *feature extraction* and *fine-tuning*. Based on the results obtained in previous tests,[17] fine-tuning was chosen. *Fine-tuning*[4] consists of freezing the parameters from the initial layers and training the last layers of the convolutional base, specifically the layers from block 5, together with the classifier. In this way, it is possible to slightly adjust the more abstract representations of the model being reused, in order to make them more relevant for the problem at hand.[4] Through this strategy, the trainable parameters were reduced by 45.4%, from more than 16 million to 9 million.

The three datasets were trained using the same strategy, except for the number of epochs, which was chosen based on the convergence between training and validation, shown by the accuracy and loss. The model of the whole plant dataset was trained during 90 epochs because it tended to suffer from overfitting above 100 epochs. In contrast, the leaf and venation models were trained during 110 epochs. The batch size was 30 for the three datasets, and a learning rate of 10^{-5} was used. Regarding the optimizer, an Adam with a beta 1 = 0.9, beta 2 = 0.999, and epsilon = 1e−0.7 was used. Besides, a dropout of 0.5 was applied to avoid the overfitting and to increase the robustness of the models.

2.5.2 | Training of the shallow CNN

Since the Shallow network has fewer parameters than VGG16, it was preferred to train it from scratch, allowing us to simplify the process. The strategy followed consisted of training the model using an RMSprop optimizer with a learning rate of 10^{-5} during 40 epochs. Then, the CNN was trained again with 30 epochs but changing the RMSprop optimizer by the Adam optimizer (beta 1 = 0.9, beta 2 = 0.999, and epsilon = 1e−0.7). Throughout the training process, a batch size of 30 epochs was used.

2.6 | Gradient-weighted class activation mapping (grad-CAM)

It is not just important to know what type of dataset works best, but it is also interesting to assess why this happens. In the same way, visualization of the activation maps may help to understand missclassifications. The interpretation of how CNN works to assign a label is usually difficult.[25] Grad-CAM[26] is a technique that consists of producing heat maps of class activation over input images. A class activation heat map is a 2D grid of scores associated to a specific output class, computed for every location in any input image, indicating how important each location is for the class under consideration.[4] This way, false color images are used to highlight the most significant regions of the image used in the final classification. Therefore, we relied on the information provided by the Grad-CAM technique to know what features the model used to classify the two classes of each dataset.

3 | RESULTS AND DISCUSSION

3.1 | Comparison between datasets

The images were collected simulating a typical production line that could be found in any nursery plant company. The idea was to mimic an automatic image acquisition process. Table 2 shows the results obtained in the deep learning models.

The model for the leaf dataset provided the best result in the test. The plant and venation datasets got the same accuracy, slightly below the leaf dataset.

Before starting the discussion, it is important to remark that there is a resolution reduction step just before the images are introduced in the CNN. So that, in the images of the whole plant dataset, the features related to the

TABLE 2 Results obtained in the comparison of datasets

Dataset	Test accuracy (%)
Whole plant	97.6
Leaf	98.8
Venation	97.6

edge of the leaves and the venation are negligible. This means that the whole plant only provides information about the color, the shape, the relation between length/width of the leaf, and the disposition of the leaves. Concerning the images of the leaf dataset, they contain information about the shape, edge features, and principal veins of the leaves. These characteristics seem to be the most critical to discriminate between the two classes, *Soleta* and *Pentacebas*. Using the leaf dataset, a 98.8% of accuracy was achieved in the test, which means the model only misclassified one image. This misclassification corresponds to a *Pentacebas* image that was erroneously assigned to the *Soleta* class with a value of 0.61. It is important to remark that the rest of the images, which were correctly classified in both *Soleta* and *Pentacebas* classes, had confidence values very close to one and zero, respectively. This result is surprisingly good considering that they are two genetically very close varieties of *P. dulcis*. This agrees with the botanical science that uses leaf feature taxonomic keys for plant identification.[27,28] Regarding venation, some studies highlight the usefulness of the venation pattern for the identification of vegetal species.[29] In our study, the venation model provided good results but worse than the ones of the leaf dataset. We believe that the methodology chosen to collect the venation images might not be the most suitable. Although the source of light was positioned under the plant to emphasize the venation characters, the illumination was not homogeneous in all the leaves. This lack of homogeneity may have increased the variability within a class, thus increasing the difficulty of varietal identification. Perhaps this approach is underestimating the real potential of venation for varietal discrimination, but this study focused on the automatic acquisition of plant images. Thus, we did not consider sampling individual leaves to obtain better images of the venation.

3.2 | Important features

The class activation heat map for each class is shown in Figure 3. For the whole plant, high activation regions of the CNN can be appreciated. These areas were located mainly in the center of the leaves. Leaves with greater activation than others can also be seen, which is because the grad-CAM image represents the features analyzed by one neuron. Other CNN neurons generate complementary activation, so this heat map is extrapolated to the other leaves.

Concerning the leaf dataset, a high activation was observed in the center and edges of the leaves. This indicates that the shapes of the edge of the leaves, along with the venation, are interesting features for identification and discrimination between the two classes. By comparing the heat map activation of leaves and whole plants, it is easy to find similarities. This implies the model searches and uses the same features in both cases. The main difference between leaf and whole plant datasets is the scale. In the whole plant dataset, more leaves can be used, each with its own features, but the clarity of the edge shape and the venation is less than in the leaf dataset due to the reduced resolution. In contrast, in the leaf dataset, there is less information due to the limited vision field that is focused on fewer leaves, but the edges and venation are observed with a better resolution than when the image is of the entire plant. Plant-scale images provide a global image compared to leaf-scale images. Despite this, the leaf scale ones offer better information to classify the images correctly because they show better the characteristics that differentiate both classes.

The venation dataset did not show any high activation zone, in contrast to what was observed in the other two cases. This indicates that there are no specific features useful for classification other than the entire image. Thus, all the information contained in the image is important for discrimination. By looking at Figure 3 from top to bottom, it can be observed that the model is focused on the center of the leaves in the three datasets, which means the most important characteristics are located in the center of the leaf, that is, the venation. Despite the result obtained with the venation dataset being good, it is not the best due to the variability caused by the position of the leaves in respect to the source of light, which affected the visualization of the veins of the leaves. The way the beam of light pass through the leaves makes the veins are clearly observed in some images but are difficult to appreciate in others. This issue could have made classification difficult as there was no clear pattern. The leaf dataset was the best in the comparison, which can

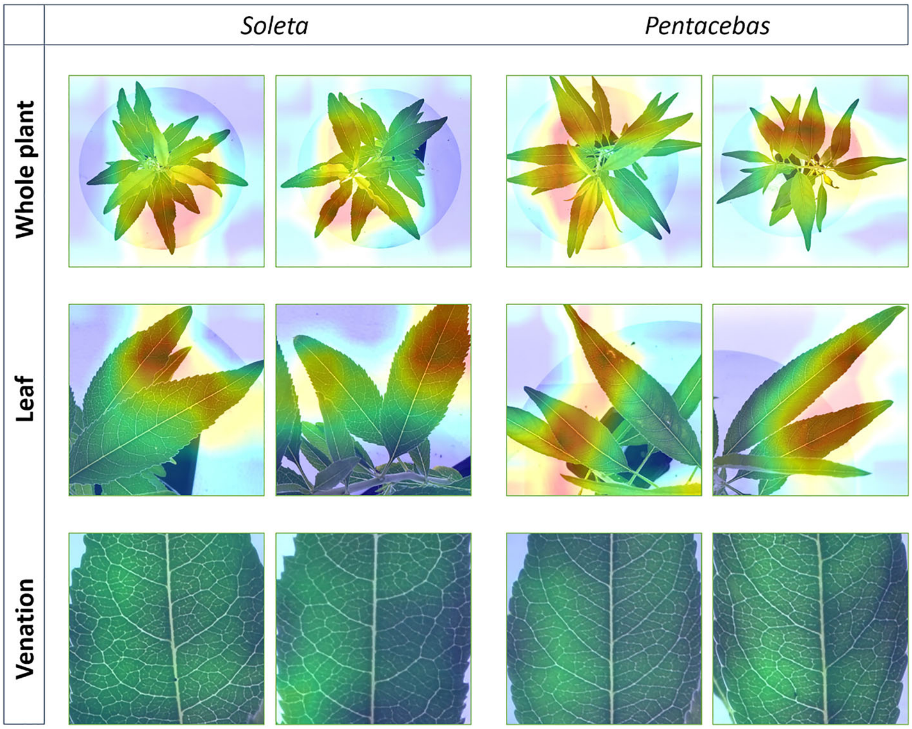

FIGURE 3 Heat map activation from the three datasets

TABLE 3 Comparison between Shallow and VGG16 networks

	Accuracy	Number of images misclassified	Number of parameters (x10^6)
VGG16	98.8	1	13.5
Shallow	93.9	5	1.5

be explained because leaves contain information extracted from the venation, together with other important features, such as the shape of the edge. Due to the method chosen to collect the images, edge shape was a more robust characteristic than venation, and this can help to improve the discrimination between *Soleta* and *Pentacebas* varieties.

3.3 | Comparative between CNNs

The leaf dataset was used to compare the VGG16 and Shallow neural networks, because it was the dataset producing the best results in the previous section. Table 3 shows the results of the comparison between both CNNs.

Concerning the classification rate, the Shallow network reached a 93.9% of accuracy, which is a good result given the difficulties to discriminate both *P. dulcis* varieties. Despite the successful classification, the Shallow network provided a classification rate of almost 5% lower than the VGG16 network. This means five images misclassified by the Shallow network against only one image erroneously classified by the VGG16 network. To evaluate these results, it is important to remark that during the training process, 13.5 million parameters were adjusted in the VGG16 network, against only 1.5 million parameters in the Shallow network. Training the VGG16 network requires more time and has a higher computational cost, and definitely, we always have a trade-off between training time and accuracy of the model.

The Shallow network proved to be effective, but in the industry, accuracy is very important, even more than the time required to develop an innovation. In addition, once the VGG16 network has been trained, there is not a significant difference in the time required to do a prediction between both networks. For this reason, the VGG16 network seems to be more appropriate in this case.

4 | CONCLUSIONS

In this study, computer vision and deep learning models have been applied for plant varietal discrimination. Three datasets (whole plant, leaf, and venation) have been compared, each of them containing images with a different scale from two varieties of *P. dulcis*, *Soleta* and *Pentacebas*, in nursery plant state. The best results were obtained for the leaf data set, with a classification accuracy of 98.8%, while whole plant and venation datasets obtained 97.6% of accuracy in the test. It was shown that the edge shapes of the leaves, and especially the venation, are the most important characteristic for the identification and discrimination of the classes. Although the venation features had great importance, the image acquisition method used in the study may have limited their potential. Finally, the performances of two CNNs were compared. The VGG16 network performed better than the Shallow network, indicating that the enormous difficulty of discriminating both varieties of *P. dulcis* requires a complex neural network architecture.

These results show that computer vision together with deep learning is a promising technique for automatic control of varietal mixtures. Moreover, it will be interesting to develop a method that allows us to obtain good images of the venation of the leaves without having to take individual leaf samples.

Finally, although these models were built using images of nursery almond trees, the idea is to extend the models to other plant species.

FUNDING INFORMATION

The authors thank the financial support of the Industrial Doctorates Plan of the Secretariat of Universities and Research of the Department of Economy and Knowledge of the Generalitat of Catalonia (project DI-COF 2017) and by the Spanish Ministry of Economy and Competitiveness (project AGL2015-70106-R, AEI/FEDER, UE).

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1002/cem.3388.

DATA AVAILABILITY STATEMENT

Data available on request from the authors.

ORCID

Sergio Borraz-Martínez https://orcid.org/0000-0002-5607-9462
Joan Simó https://orcid.org/0000-0002-2853-3381
Anna Gras https://orcid.org/0000-0003-0111-7246
Ricard Boqué https://orcid.org/0000-0001-7311-4824
Francesc Tarrés https://orcid.org/0000-0003-0920-4782

REFERENCES

1. Borraz-Martínez S, Boqué R, Simó J, Mestre M, Gras A. Development of a methodology to analyze leaves from prunus dulcis varieties using near infrared spectroscopy. *Talanta*. 2019;204:320-328. doi:10.1016/j.talanta.2019.05.105
2. Borraz-Martínez S, Simó J, Gras A, Mestre M, Boqué R. Multivariate classification of prunus dulcis varieties using leaves of nursery plants and near infrared spectroscopy. *Sci Rep*. 2019;9(1):1-9. doi:10.1038/s41598-019-56274-5
3. Lu Y. Industry 4.0: a survey on technologies, applications and open research issues. *J Ind Inf Integr*. 2017;6:1-10. doi:10.1016/j.jii.2017.04.005
4. Chollet F. *Deep Learning With Python*. Manning Publications; 2018.
5. Buduma N, Locascio N. *Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms*. Vol. 29. O'Reilly Media, Inc; 2017.
6. Kamilaris A, Prenafeta-Boldú FX. Deep learning in agriculture: a survey. *Comput Electron Agric*. 2018;147(July 2017):70-90. doi:10.1016/j.compag.2018.02.016
7. Kussul N, Lavreniuk M, Skakun S, Shelestov A. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geosci Remote Sens Lett*. 2017;14(5):778-782. doi:10.1109/LGRS.2017.2681128
8. Debats SR, Luo D, Estes LD, Fuchs TJ, Caylor KK. A generalized computer vision approach to mapping crop fields in heterogeneous agricultural landscapes. *Remote Sens Environ*. 2016;179:210-221. doi:10.1016/j.rse.2016.03.010
9. Wang Q, Nuske S, Bergerman M, Singh S. *Automated Crop Yield Estimation for Apple Orchards*. Springer, Heidelberg; 2013:745-758. doi:10.1007/978-3-319-00065-7_50.
10. Dorj UO, Lee M, Yun S. An yield estimation in citrus orchards via fruit detection and counting using image processing. *Comput Electron Agric*. 2017;140:103-112. doi:10.1016/j.compag.2017.05.019

11. Yalcin, H. Plant phenology recognition using deep learning: deep-pheno. In 2017 6th International Conference on Agro-Geoinformatics. IEEE; 2017:1-5. doi:10.1109/Agro-Geoinformatics.2017.8046996.
12. Mochida K, Koda S, Inoue K, et al. Computer vision-based phenotyping for improvement of plant productivity: a machine learning perspective. *GigaScience*. 2018;8(1):1-12. doi:10.1093/gigascience/giy153
13. Zhu X, Zhu M, Ren H. Method of plant leaf recognition based on improved deep convolutional neural network. *Cogn Syst Res*. 2018;52: 223-233. doi:10.1016/j.cogsys.2018.06.008
14. Barré P, Stöver BC, Müller KF, Steinhage V. LeafNet: a computer vision system for automatic plant species identification. *Ecol Inform*. 2017;40:50-56. doi:10.1016/j.ecoinf.2017.05.005
15. Dutta S. An overview on the evolution and adoption of deep learning applications used in the industry. *Wiley Interdiscip Rev Data min Knowl Discov*. 2018;8(4):1-12. doi:10.1002/widm.1257
16. Cheng Y-P, Li C-W, Chen Y-C. Apply computer vision in GUI automation for industrial applications. *Math Biosci Eng*. 2019:7526-7545. doi:10.3934/mbe.2019378
17. Borraz-Martínez S, Tarrés F, Boqué R, Mestre M, Simó J, Gras A. Varietal quality control in the nursery plant industry using computer vision and deep learning techniques. *J Chemometr*. 2020:1-11. doi:10.1002/cem.3320
18. Mikołajczyk A, Grochowski M. Data augmentation for improving deep learning in image classification problem. In *2018 International Interdisciplinary PhD Workshop, IIPhDW 2018*. Institute of Electrical and Electronics Engineers Inc.; 2018:117-122. doi: 10.1109/IIPHDW.2018.8388338.
19. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. *J Big Data*. 2019;6(1):1-48. doi:10.1186/s40537-019-0197-0
20. Qin Z, Yu F, Liu C, Chen X. How convolutional neural networks see the world—a survey of convolutional neural network visualization methods. *Math Found Comput*. 2018;1(2):149-180. doi:10.3934/mfc.2018008
21. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc*. 2014.
22. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl*. 2010;22(10):1345-1359. doi:10.1109/TKDE.2009.191
23. Mohanty SP, Hughes DP, Salathé M. Using deep learning for image-based plant disease detection. *Front Plant Sci*. 2016;7(September): 1-10. doi:10.3389/fpls.2016.01419
24. Mehdipour Ghazi M, Yanikoglu B, Aptoula E. Plant identification using deep neural networks via optimization of transfer learning parameters. *Neurocomputing*. 2017;235(January):228-235. doi:10.1016/j.neucom.2017.01.018
25. Lipton ZC. The mythos of model interpretability. *Commun ACM*. 2018;61(10):36-43. doi:10.1145/3233231
26. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. *Int J Comput vis*. 2020;128(2):336-359. doi:10.1007/s11263-019-01228-7
27. Cope JS, Corney D, Clark JY, Remagnino P, Wilkin P. Plant species identification using digital morphometrics: a review. *Expert Systems with Applications Pergamon*. 2012;39(8):7562-7573. doi:10.1016/j.eswa.2012.01.073
28. Hearn DJ. Shape analysis for the automated identification of plants from images of leaves. *Taxon*. 2009;58(3):934-954. doi: 10.1002/tax.583021
29. Lee SH, Chan CS, Mayo SJ, Remagnino P. How deep learning extracts and learns leaf features for plant classification. *Pattern Recognit*. 2017;71:1-13. doi:10.1016/j.patcog.2017.05.015

How to cite this article: Borraz-Martínez S, Simó J, Gras A, Mestre M, Boqué R, Tarrés F. Combining computer vision and deep learning to classify varieties of *Prunus dulcis* for the nursery plant industry. *Journal of Chemometrics*. 2022;e3388. doi:10.1002/cem.3388