Classification of hyperspectral images of the interior of fruits and vegetables using a 2D convolutional neuronal network

J S Barrera¹, A Echavarría¹, C Madrigal¹, and J Herrera-Ramirez¹
¹ Laboratorio de Visión Artificial y Fotónica, Instituto Tecnológico Metropolitano, Medellín, Colombia
E-mail: alejandroechavarria96@gmail.com

Abstract. This work evaluates the performance of a Deep Learning technique for classification of challenging hyperspectral images of the interior of fruits and vegetables when they are combined. Some of these samples have low contrast, similar colour features, and their skins or characteristic shapes are lost when cut to expose their interiors. We implemented a two-dimensional convolutional neural network for this classification task and compared their results against the technique of support vector machines. We randomly selected a group of 13 hyperspectral images from a public database containing information of the interior of 42 fruits and vegetables. Using parts of these 13 selected images, we constructed three artificial hyperspectral images merging these parts differently. We applied the two proposed techniques over the three of them. The comparison of the classification results shows that the two-dimensional convolutional neural network over-performs the support vector machine in all three composite images. The two-dimensional convolutional neural network exceeded 98% classification accuracy in all of them. These results show that the two-dimensional convolutional network benefits from the spatial and spectral data in the images obtaining proper levels of classification even in samples mixed in complex contexts, as it can occur in the food or pharmaceutical industries.

1. Introduction
Currently, the optics and photonics have advanced in the analysis and processing of hyperspectral images, allowing their use with satisfactory results in various technological and research fields ranging from agriculture [1], through mining [2], to medicine [3]. The valuable hyperspectral information contained in the image at multiple wavelengths can provide a more refined description of the components on the surface of the materials and therefore improve their detection, discrimination and characterization [4]. For example, Robert Ennis et al. [5] generated a database of 42 hyperspectral images of fruits and vegetables, captured employing an acquisition system based on an object line scanner. This hyperspectral system captures spectral bands from 376.20 nm to 821.62 nm, obtaining images with a spatial resolution of ~57 px/deg by 800 pixels at a wavelength resolution of ~1.12 nm [5].

For the classification of hyperspectral images, various methods of artificial intelligence have been proposed, specifically from the area of machine learning (ML) [6] such as the support vector machines (SVM), the classification based on K-nearest neighbors [6] or neural networks, which are alternatives based primarily on the recognition of conventional patterns [7]. Currently, algorithms within the ML field known as deep learning [8] have shown great utility in multiple applications, including the computer vision area [8]. Accordingly, implementations of deep learning architectures based on
convolutional neural networks (CNN) [9] for the classification of standard red-green-blue (RGB) images have shown high performances. Examples are: AlexNet [10], GoogLeNet [11] and ResNet [12] architectures implemented for classification over the ImageNet database. Additionally, there has been an increase in the study of the classification of hyperspectral images taking advantage of these CNN structures [8,13], and also using other types of deep learning architectures such as long short-term memory (LSTM), multi-grain network (MugNet), or two-channel deep convolutional neural network [13–15]. However, most of these studies have been focused essentially on remote sensing applications and databases such as Indian Pines, the University of Pavia, Pavia city, the Salinas and the Kennedy Space Center [13]. But other databases are also interesting in other applications like pharmaceutical or food industries. In this work, we use a database constructed with samples of value for the food industry. This database is composed of samples with the interior of fruits or vegetables, which represents different challenges like the similarity that may exist among samples, the loss of information of the texture and the typical outstanding colors of their skins.

The aim of this work is the use of the deep learning technique for the classification of 13 types of hyperspectral images of the interior of fruits and vegetables [5]. We implement a 2D-CNN architecture [8] since it can exploit both the spatial and spectral context included in the hyperspectral images [8]. Therefore, the combination of spectral information with spatial information can provide precision in the classification of foods with similar characteristics of texture or geometry.

This document presents the following order: section 2 presents the 2D-CNN type model taken for the classification of hyperspectral images of fruits and vegetables. Section 3 shows comparatively the results of classification of the deep learning model and the SVM technique. Finally, section 4 presents the conclusions of the work.

2. Classification model
In general, it can be said that deep learning has gained great importance in different processes, such as classification, because of the high performance it offers, and this is due to its capacity for learning and representation of characteristics [16]. CNNs, as part of these schemes, have these same characteristics. Specifically, the CNN is a multilayer neural network composed in its basic structure by an input layer, a convolution layer (conv layer), an activation layer (relu layer), a grouping layer (pooling layer), and a fully connected layer (FC). The input layer is where the images are received (data matrix); The Convolutional layer is the layer where image characteristics are obtained using a filter; the activation layer is the layer where an activation type function is applied; the grouping layer is the layer where the amount of data is reduced, that is, this layer selects the most common parameters; and the fully connected layer is the layer that delivers the result of the network.

![2D Convolutional Operation](image)

Figure 1. Operation of the 2D-CNN model. The red box shows the space kernel to perform the 2D convolutional operation. Next, the extraction of characteristics obtained by the 2D convolutional operation to produce characteristic maps. Finally, the classification is given as output.
2.1. Two-dimensional convolutional neural network

The two-dimensional convolutional neural network (2D-CNN) architecture is generally proposed in three phases [8]: patch extraction, feature extraction and tag identification (Figure 1). In this way, the objective is to predict the labels of each pixel of the hyperspectral image. The 2D-CNN focuses on travelling the image spatially in each of the spectral bands to extract the spatial characteristics. At the end of the architecture, the prediction is given according to the extracted characteristics.

In our implementation, the 2D-CNN architecture has the input layer, three 2D convolutional layers, three activation layers, a fully connected layer, and a softmax function layer to display the general prediction results. In this proposal, the 2D-CNN architecture does not have the typical grouping layer, since it is required to maintain as much data as possible, that is, as many pixels as possible of the image (Figure 2). For the procedures of training and classification test, we used training and test sets with 80% and 20% of data, respectively.

![Figure 2](image_url)

**Figure 2.** Deep learning architecture implemented: two-dimensional convolutional neural network (2D-CNN).

3. Database and classification tests

The data used in this work was extracted from a public database of hyperspectral images of fruits and vegetables [5]. These images correspond to the inner part of the fruits and vegetables. Randomly, from a total of 42 images contained in the database, 13 images were chosen among fruits (11) and vegetables (2). Thus the 13 classes in our database are coconut, blackberry, kiwi, lemon, orange, cucumber, raspberry, tomato, pumpkin, red chili, paprika, potato and radish.

To analyze the performance of the 2D-CNN over different spatial distributions of the classes, three types of hyperspectral images were generated as databases. Additionally, these images were also classified using the SVM technique for comparison. These three images (databases) were formed including regions of the thirteen classes but distributed differently in each of them. Finally, the generated hyperspectral images had the following dimensions: 145 high x 145 wide x 200 depth. Two spatial dimensions (145x145) by one spectral dimension (200 bands). This number of spectral bands cover the wavelengths from 465 nm to 710 nm. The class tags or labels in these images are as follows: 1-Blackberry, 2-Coconut, 3-Cucumber, 4-Potato, 5-Kiwi, 6-Radish, 7-Paprika, 8-Tomato, 9-Pumpkin, 10-Raspberry, 11-Lemon, 12-Orange, and 13-Red Chili

3.1. Database (generated spectral images)

3.1.1. Hyperspectral image as a combination of areas of regular shape. For the first combined hyperspectral image, squared and rectangular areas of the sample images were trimmed and reorganized into the final image without any separation among them. We call this image, hyperspectral image of regular zones (RZ). Figure 3 shows the division of the labeled data for the 13 classes of samples contained in the RZ hyperspectral image.
3.1.2. Hyperspectral image as a combination of areas of irregular shape. For the second combined hyperspectral image, irregular areas of the sample images were trimmed. We call this image hyperspectral image of irregular zones (IZ). Figure 4 shows the image of labels for the 14 classes contained in the IZ hyperspectral image. This image has 14 classes to include a new class with the label 0 (null class), which accounts for those areas where there are no samples and it is equivalent to areas of separation among the cropped areas.

3.1.3. Hyperspectral image as a combination of irregular areas within other irregular areas. For the last combined hyperspectral image, irregular areas of the sample images were trimmed. In this case, smaller irregular zones of another sample class were allowed to be included within zones of another class. This image will be called hyperspectral image of irregular zones within another zone (IZZ). With this specific spatial distribution of the samples, we wanted to verify whether the two-dimensional convolutional neural network changed in its classification performance due to the increase in the complexity of the distribution and the size of the samples. Figure 5 shows the image of labels for the 14 kinds of samples contained in the IZZ hyperspectral image. As in the IZ image, in this IZZ image there is a class with label 0, since there are some spaces among trimmed areas.
4. Results of classification tests
As mentioned earlier, the results obtained by the 2D-CNN architecture were compared with the results of the SVM technique. We used the accuracy percentage as a quantitative descriptor for the evaluation of the methods. This accuracy percentage is a typical metric in classification tasks, and it is calculated as the number of correct predictions made by the model over all kinds of predictions made. The results for the classification of hyperspectral images, in the respective order presented in the database, are shown below. The classification result for the RZ hyperspectral image is illustrated in Figure 6. Figure 6(a) shows the result for 2D-CNN classification with 99.92% of accuracy, and Figure 6(b) presents the result for SVM technique with 85.52% accuracy.

![Figure 6](image1.png)

**Figure 6.** Results of classification test over the RZ hyperspectral image. (a) 2D-CNN classification with 99.92% of accuracy; (b) SVM classification with 85.52% of accuracy.

Figure 7(a) and Figure 7(b) show the results of the 2D-CNN and SVM technique over the IZ hyperspectral image with accuracy percentages of 98.56% and 99.57%, respectively. The results for the IZZ hyperspectral image are given in Figure 8(a) and Figure 8(b) with classification accuracy of 99.24% and 28.62% for 2D-CNN and SVM, respectively.

![Figure 7](image2.png)

**Figure 7.** Results of classification test over the IZ hyperspectral image. (a) 2D-CNN classification with 98.56% of accuracy; (b) SVM classification with 99.57% of accuracy.

![Figure 8](image3.png)

**Figure 8.** Results of classification test over the IZZ hyperspectral image. (a) 2DCNN with 99.24% of success; (b) SVM with 28.62%.

As it is comparatively shown in Table 1, the ability of the 2D-CNN architecture to classify hyperspectral images with samples of different shapes and distributions or mixtures (such as the RZ, IZ and IZZ images generated in this work) is superior to the performance of the classical machine learning
technique of SVM. The SVM technique only obtains good results in cases where the classes are clearly spatially separated, as in the IZ image. It can be said that in an image of greater complexity such as the IZZ image, SVMs are unable to create nonlinear classification boundaries that account for this complexity. This shows that the 2D-CNN architecture is more robust since it takes into account not only spatial information, but also spectral information and it is able to combine both properly to classify samples in much more spatially complex contexts. This work is important because it shows the potential of the hyperspectral images in combination with the data analysis provided by the CNNs in challenging samples related with the food industries.

| Classification accuracy | SVM     | 2D-CNN  |
|-------------------------|---------|---------|
| RZ hyperspectral image  | 85.52%  | 99.92%  |
| IZ hyperspectral image  | 99.57%  | 98.56%  |
| IZZ hyperspectral image | 28.62%  | 99.24%  |

5. Conclusions
The implementation of this convolutional 2D neural network showed that deep learning for the classification of fruits and vegetables through the 2D-CNN architecture is highly efficient. The results give a clear sample of the power that these architectures possess and their robustness for the classification in hyperspectral images of the interior of fruits and vegetables. The three trained models manage to classify in a general percentage higher than 98%.

The comparison with the SVM technique shows how this classical technique is highly dependent on the spatial distribution of the samples, lowering its performance to percentages of classification accuracy of 28.62% for a hyperspectral image with high spatial complexity such as the IZZ image. In short, the use of combined spatial and spectral information made by the 2D-CNN architecture gives it a superior ability to deal with this type of tasks.

This work shows the possibilities of combining deep learning and hyperspectral imaging in these kind of classification tasks, where the type of samples can be challenging, or in industries where the recognition or quantification of components in products based on mixtures are common, like in the pharmaceutical or food industries.

Acknowledgments
A Echavarri thanks the Instituto Tecnologico Metropolitano de Medellin for the support given under the program for undergraduate young researchers.

References
[1] Bock C H, Poole G H, Parker P E, and Gottwald T R 2010 Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging CRC. Crit. Rev. Plant Sci. 29 59
[2] Kuenzer C, Bluemel A, Gebhardt S, Quoc T V, and Dech S 2011 Remote sensing of mangrove ecosystems: A review Remote Sens. 3 878
[3] Kulcke A, Holmer A, Wahl P, Siemers F, Wild T, and Daeschlein G 2018 A compact hyperspectral camera for measurement of perfusion parameters in medicine Biomed. Tech. 63 S19
[4] Burger J E, and Gowen A A 2011 The interplay of chemometrics and hyperspectral chemical imaging 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS) (Lisbon: IEEE)
[5] Ennis R, Schiller F, Toscani M, and Gegenfurtner K R 2018 Hyperspectral database of fruits and vegetables J. Opt. Soc. Am. A 35 2 B256
[6] Pereira F C, and Borysov S S 2019 Machine Learning Fundamentals Mobility Patterns, Big Data and Transport Analytics (Chichester: Elsevier)
[7] Melgani F, and Bruzzone L 2004 Classification of hyperspectral remote sensing images with support vector machines IEEE Trans. Geosci. Remote Sens. 42 1778
[8] Yang X, Ye Y, Li X, Lau R Y K, Zhang X, and Huang X 2018 Hyperspectral image classification with deep learning models IEEE Trans. Geosci. Remote Sens. 56 5408

[9] Deng L 2014 A tutorial survey of architectures, algorithms, and applications for deep learning APSIPA Trans. Signal Inf. Process. 3 e2

[10] Krizhevsky A, Sutskever I, and Hinton G E 2012 ImageNet classification with deep convolutional neural networks 26th Conference on Neural Information Processing Systems (CNIPS) (Nevada: Curran Associates Inc.)

[11] Szegedy C, Wei L, Yangqing J, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, and Rabinovich A 2015 Going deeper with convolutions IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Boston: IEEE)

[12] He K, Zhang X, Ren S, and Sun J 2016 Deep residual learning for image recognition IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Las Vegas: IEEE)

[13] Kang X, Li S, and Benediktsson J A 2017 Hyperspectral image classification: A benchmark IEEE International Geoscience and Remote Sensing Symposium (IGARSS) (Fort Worth: IEEE)

[14] Ma A, Filippi A M, Wang Z, and Yin Z 2019 Hyperspectral image classification using similarity measurements-based deep recurrent neural networks Remote Sens. 11 (2) 194

[15] Zhou F, Hang R, Liu Q, and Yuan X 2019 Hyperspectral image classification using spectral-spatial LSTMs Neurocomputing 328 39

[16] Yong-Qiang Z, and Jingxiang Y 2015 Hyperspectral image denoising via sparse representation and low-rank constraint IEEE Trans. Geosci. Remote Sens. 53 296