Convolutional Neural Networks for Olive Oil Classification

Belén Vega-Márquez[1(✉)], Andrea Carminati[1], Natividad Jurado-Campos[2], Andrés Martín-Gómez[2], Lourdes Arce-Jiménez[2], Cristina Rubio-Escudero[1], and Isabel A. Nepomuceno-Chamorro[1]

[1] Department of Computer Languages and Systems, University of Sevilla, Sevilla, Spain
bvega@us.es

[2] Institute of Fine Chemistry and Nanochemistry, Department of Analytical Chemistry, University of Córdoba, Córdoba, Spain
qa1arjil@uco.es

Abstract. The analysis of the quality of olive oil is a task that is hav-ing a lot of impact nowadays due to the large frauds that have been observed in the olive oil market. To solve this problem we have trained a Convolutional Neural Network (CNN) to classify 701 images obtained using GC-IMS methodology (gas chromatography coupled to ion mobil-ity spectrometry). The aim of this study is to show that Deep Learn-ing techniques can be a great alternative to traditional oil classification methods based on the subjectivity of the standardized sensory analy-sis according to the panel test method, and also to novel techniques provided by the chemical field, such as chemometric markers. This tech-nique is quite expensive since the markers are manually extracted by an expert.

The analyzed data includes instances belonging to two different crops, the first covers the years 2014–2015 and the second 2015–2016. Both har-vests have instances classified in the three categories of existing oil, extra virgin olive oil (EVOO), virgin olive oil (VOO) and lampante olive oil (LOO). The aim of this study is to demonstrate that Deep Learning techniques in combination with chemical techniques are a good alterna-tive to the panel test method, implying even better accuracy than results obtained in previous work.

Keywords: Convolutional Neural Network, Olive oil classification, GC-IMS method

1 Introduction

Olive oil is a fatty vegetable oil obtained from the fruit of the Olive tree, a traditional tree crop of the Mediterranean area. For more than six centuries [6], olive oil has been used in different daily areas: making cosmetics, medicines and soaps, used for fuel, and most famous of all, the use of olive oil in gastronomy. The increasing use in recent years of this oil has led to the search for the best quality for its use.

The current European Union Regulation [4] classifies olive oil into three classes that in descending order of quality are named as extra virgin olive oil (EVOO), virgin olive oil (VOO) and *lampante* olive oil (LOO). The difference among the three categories resides in their levels of free fatty acid, a condition that correlates with the condition of the olives when they are crushed. Damaged or old olives are high in oleic acid content, which makes a lower quality oil. It is important to clearly discriminate the types of oil that are commercialized due to two fundamental reasons: in the first place it is necessary to take into account that depending on the type of oil it can be suitable for consumption or not: extra virgin olive oil (EVOO) and virgin (VOO) are suitable, lampante (LOO) is not. In second place, the price of this product changes a lot depending on the category, including cases of fraud in which oil that was not of good quality has been sold as if it was. For this reason, the importation and exportation of this product is increasingly being controlled, with even more exhaustive quality controls.

The most common form of oil classification is currently carried out by means of a tasting by experts who identify the quality of the oil being studied via sensory attributes. This technique, although it can be considered reliable, is manual and not objective. Therefore, the development of techniques to automatically classify olive oil samples from gas chromatography coupled to ion mobility spectrometry is necessary. In this work we use deep learning as a good alternative to consider to advise these experts and the community behind the world of olive oil.

Previous studies have addressed this problem from a number of different perspectives. The first of them carried out by [2] takes the image obtained by GC-IMS analysis and applies several techniques, such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and finally the K Nearest Neighbor (k-NN) algorithm to perform the classification. Secondly, in our previous study [12] we carried out the classification by extracting interesting areas as preprocessing phase from the images, called markers to which we then applied Deep Learning techniques. Although this approach was quite accurate, the marker extraction phase is made by an expert and it was very expensive and subjective, as it was the expert who pointed at them by eye.

The main problem of the previous studies is that the classification phase was not direct, but before there were a preprocessing phase to perform, assuming a higher cost. The aim of our study is to provide a standard technique that reduces time and complexity in the classification of olive oils, eliminating the need for an expert to extract markers from images.

The article is organized as follows, Sect. 2 describes the data used in the study and the algorithm and methodologies used to the classification task. Section 3 shows the results obtained with the techniques explained in the Methodology section. Finally, Sect. 4 provides some conclusions obtained after the study.

2 Methodology

The goal of this study is to use Deep Learning techniques in combination with chemical methods to provide, on the one hand, an alternative to the standardized sensory analysis according to the panel test method for the classification of olive oil samples and, on the other hand, to provide a simpler and quicker technique for carrying out this task, trying to improve the results obtained in previous investigations.

2.1 Olive Oil Samples

A total of 701 olive oil samples were chosen to constitute the final sample of this study. The samples were taken from olives from two separate harvests. The first harvest covered the years 2014–2015 and the second the years 2015–2016. Figure 1 shows the proportion of classes of the two harvests in the study. In the first one there are a total of 292 instances of which 35 were LOO, 98 EVOO and the remaining 159 VOO. Secondly, the harvest covering the years 2015–2016 included a total of 409 examples, 121 LOO, 92 EVOO and 196 VOO.

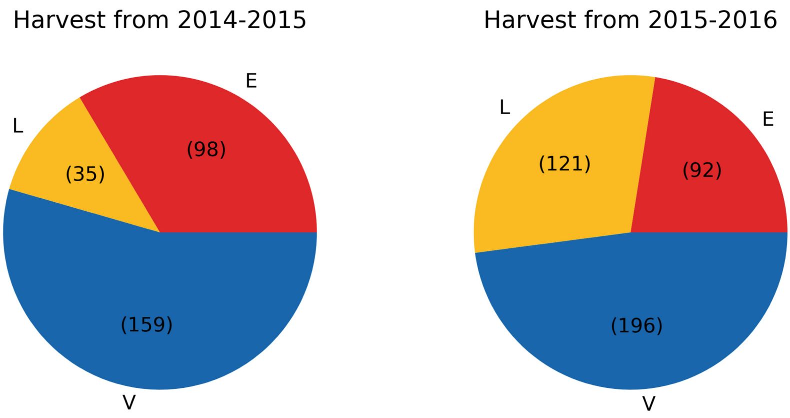

Fig. 1. Instances of oilive oil by class in the two harvests of the study

As you can see the number of instances for each of the harvests is relatively small, therefore, since the two harvests had been extracted in the same way we decided to merge them into a single set of data. Figure 2 illustrates the final proportion of classes in the dataset. We have a total of 701 samples where 156 are LOO, 190 EVOO and 355 VOO.

2.2 Chemical Methods for Data Acquisition

GC-IMS method was used for data acquisition. The GC-IMS method for olive oil analysis was obtained from a previous work by [2]. This method combines gas

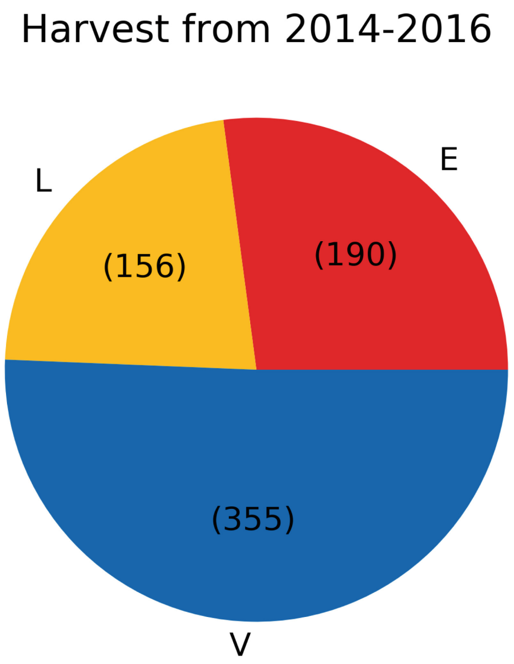

Fig. 2. Instances of olive oil by class in the final dataset

chromatography (GC) with ion mobility spectrometry (IMS) in order to detect sensitively and selectively complex compunds that can be vaporized without decomposition [8].

This method works as follows: first, compounds that have been previously separated into a GC column are ionized and passed through the Ion Mobility Spectrometer (IMS). Ionized compounds are conducted by a magnetic field in the opposite direction to a drift flow. Depending on the charge, mass and shape of the ions, they reach the collector at different speeds. This speed is collected in an intensity form and allows a two-dimensional matrix to be drawn. Figure 3 shows an example of olive oil image obtained with GC-IMS method.

2.3 Deep Learning Approach

The use of Deep Learning techniques is not something new in the field of gastronomy or agriculture, for example, it has already been used to detect the quality of wine using taste sensors and neural networks [10], in the detection of plant diseases with leaf images and computer vision [5] and in food classification [3].

The case of study for a classification task has been performed with a convolutional neural network (CNN). Convolutional neural networks are computational systems inspired by the human nervous system [13]. The fundamental structure of any neural network is formed by interconnected nodes that simulate the neurons of the brain. These nodes can be combined in layers that will be trained

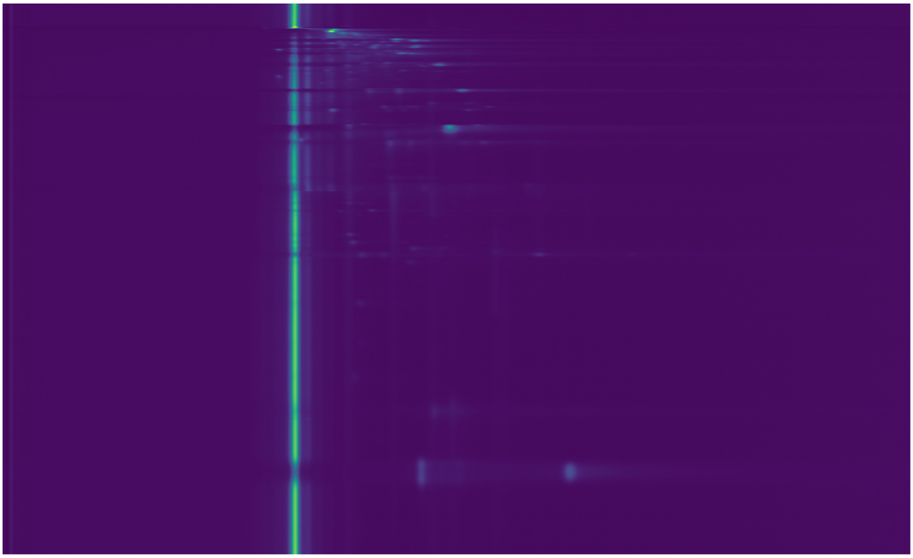

Fig. 3. Example of image obtained from olive oil instance with GC-IMS method

after an iterative process to learn from the input data and be able to give the desired output.

2.4 Software and Experimental Setting

The CNN used in this study has been implemented with the Keras library [1] running on top of Tensorflow and written in Python code. The data visualization of the images has been made with Python. The train test split method has been carried out with scikit-learn library [9]. Due to the amount of data available, the runnings of the code were performed on an Intel machine, specifically Intel(R) Core(TM) i7-8700 CPU @ 3.20 Ghz, with 64 GB of RAM and 12 cores.

3 Results

3.1 Architecture of the Model

The architecture of a neural network is the determinant of the precision and accuracy of the model, therefore, defining a good architecture beforehand is the fundamental part in the learning process.

The network used for oil classification is a convolutional neural network. These networks have proved to be the pioneers in image classification tasks, with the best in world image classification competitions [7]. CNNs are comprised of four types of layers. Firstly, the input layer that holds the pixel values of the input image. Secondly, the convolutional layer which determines the output of neurons which are connected to local regions of the input through the calculation of the scalar product between their weights and the region connected to the input. In the third case, pooling layers reduce the dimensionality of the representation to reduce the number of parameters to assist the learning process by significantly reducing the number of parameters and the complexity of the model. Finally, the fully-connected layers that perform the classification task with class scores, these scores indicate the probability that the input data belongs to one class or another.

The combination of layers gives rise to a neural network. The one used consists of three convolutional layers which in turn consist of pooling layers, a flatten layer and four dense layers which will determine the predicted class as can be seen in Fig. 4.

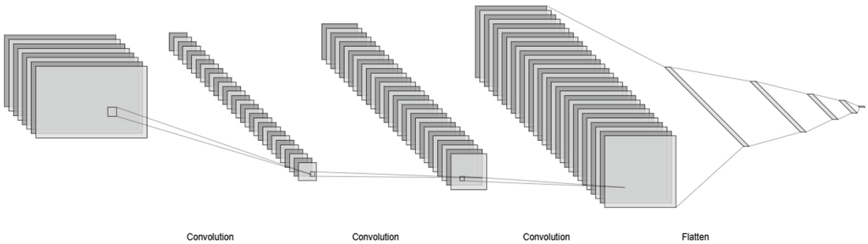

Fig. 4. Convolutional Neural Network used in this study

3.2 Training the Model

The training step in a neural network needs some parameters that will indicate the way the model is going to be trained, such as the optimizer to be used, the loss function to be taken into account, the number of epochs or the learning rate.

Table 1 shows the parameters chosen in the training step after hyperparameter tunning.

Table 1. Parameters used in the training step

Parameter	Value
Optimizer	SGD
Loss	Sparse categorical crossentropy
Metrics	Accuracy
Learning rate	0.01
Momentum	0.4
Epochs	125

We can see in Fig. 5 that with a momentum of 0.4 and 125 epochs the minimum loss is reached with a learning rate of 0.01. We have also noticed that with this methodology the training time is very fast, less than ten minutes for the full data set.

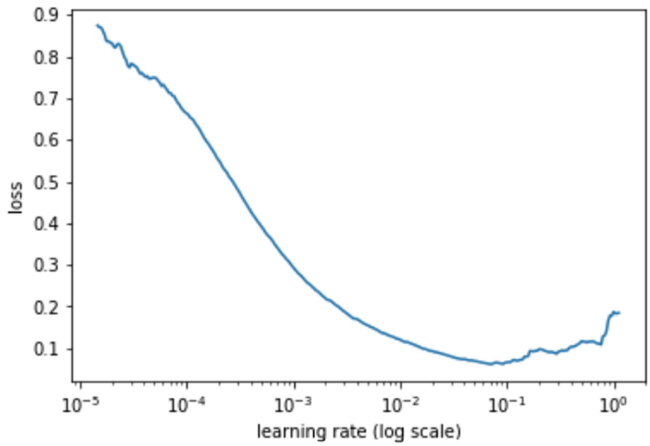

Fig. 5. Learning rate tunning

3.3 Classification Performance

A train-test split method was used for the validation of the CNN model. In order to be able to compare our work with the previous results obtained by [2], we have chosen for the set of train and test the same instances that they chose in their study, implying 80% of examples of each of the harvests for train and the remaining 20% for test. The performance of the CNN was measured by the accuracy score. A total of three experiments were made, one of them classified between the three existing classes (EVOO/LOO/VOO), and the other two classified distinguishing one class against the others (EVOO/non-EVOO, LOO/non-LOO), i.e. a ternary model and two binary models were trained. Table 2 compares the results obtained with our model to those obtained by [2] with the spectral fingerprint, i.e. the full image and those obtained with the Deep Learning techniques applied to the specific markers [12].

Table 2. Comparison of our results with previous results

	Spectral fingerprint	Specific markers	Convolutional neural net
EVOO/VOO/LOO	79.40	74.29	82.86
EVOO/non-EVOO	85.10	85.72	87.85
LOO/non-LOO	92.90	90.71	94.57

4 Discussion

In this paper we present a Convolutional Neural Network (CNN) that classifies olive oil chemical data obtained from GC-IMS method in their respective classes according to the quality that it possesses.

We have shown that Deep Learning is a good tool that can replace the traditional method of quality evaluation, providing a more reliable and objective assessment of quality and avoiding fraud committed in the sales process.

We have verified that this technique is a great alternative to those carried out in previous studies with comparable results. The proposed technique has only one training stage, whereas in previous studies there were several stages prior to the classification of the data, such as obtaining markers or the Principal Components Analysis (PCA).

In order to take into account all possible casuistry and to be comparable with previous studies, we have trained three different models, one ternary and two binary. As expected, binary models have a higher success rate.

Deep Learning techniques stand out for their good success rate when working with a lot of data, as this means you have more reference examples to learn from. Perhaps one of the main problems of this study has been that the data was scarce, even so, the results obtained have been promising.

As future work, we study two alternative ways, on the one hand to train the developed model with more data. On the other hand, we carry out a pre-processing of the images, such as the Super-Pixel segmentation technique [11], in order to obtain the best results.

References

1. Chollet, F., et al.: Keras (2015). https://keras.io
2. Contreras, M.D.M., Jurado-Campos, N., Arce, L., Arroyo-Manzanares, N.: A robustness study of calibration models for olive oil classification: target and untargeted fingerprint approaches based on GC-IMS (2019, in press)
3. Debska, B., Guzowska-Świder, B.: Application of artificial neural network infood classification. Analytica Chimica Acta **705**(1), 283–291 (2011). A selection of papers presented at the 12th International Conference on Chemometrics in Analytical Chemistry. https://doi.org/10.1016/j.aca.2011.06.033. http://www.sciencedirect.com/science/article/pii/S0003267011008622
4. EEC: European Commission Regulation (EEC). European Commission Regulation EEC/2568/91 of 11 July on the characteristics of olive and pomace oils and on their analytical methods. Off. J. Eur. Communities **L248**(640), 1–82 (1991)
5. Ferentinos, K.P.: Deep learning models for plant disease detection and diagnosis. Comput. Electron. Agric. **145**, 311–318 (2018). https://doi.org/10.1016/j.compag.2018.01.009. http://www.sciencedirect.com/science/article/pii/S0168169917311742
6. Galili, E., Stanley, D.J., Sharvit, J., Weinstein-Evron, M.: Evidence for earliest olive-oil production in submerged settlements off the Carmel coast, Israel. J. Archaeol. Sci. **24**(12), 1141–1150 (1997). https://doi.org/10.1006/jasc.1997.0193. http://www.sciencedirect.com/science/article/pii/S030544039790193X

7. Hussain, M., Bird, J.J., Faria, D.R.: A study on CNN transfer learning for image classification. In: Lotfi, A., Bouchachia, H., Gegov, A., Langensiepen, C., McGinnity, M. (eds.) UKCI 2018. AISC, vol. 840, pp. 191–202. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-97982-3_16
8. Kanu, A.B., Hill, H.H.: Ion mobility spectrometry detection for gas chromatography. J. Chromatogr. A **1177**(1), 12–27 (2008). https://doi.org/10.1016/j.chroma.2007.10.110
9. Pedregosa, F., et al.: Scikit-learn: machine learning in python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
10. Riul, A., et al.: Wine classification by taste sensors made from ultra-thin films and using neural networks. Sens. Actuators B: Chem. **98**(1), 77–82 (2004). https://doi.org/10.1016/j.snb.2003.09.025. http://www.sciencedirect.com/science/article/pii/S0925400503007512
11. Soomro, N., Wang, M.: Superpixel segmentation: a benchmark. Sig. Process. Image Commun. **56** (2017). https://doi.org/10.1016/j.image.2017.04.007
12. Vega-Márquez, B., Nepomuceno-Chamorro, I., Jurado-Campos, N., Martín-Gómez, A., Arce, L., Rubio-Escudero, C.: Deep Learning Techniques to Improve the Performance of Olive Oil Classification (2019, in press)
13. Wu, J.: Introduction to convolutional neural networks. Technical report (2017). https://doi.org/10.1007/978-3-642-28661-2-5