Automatic Spectrum Recognition

Simona KLUČKOVÁ, Maroš KOLLÁR, Jakub KUČEČKA, Zuzana POPOVCOVÁ, Mária RAJNÍKOVÁ, Alena VALOVÁ*

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovicova 2, 842 16 Bratislava, Slovakia
mikasa.fiit@gmail.com

1 Introduction

Olive oil is a fundamental component of the Mediterranean diet because of its nutritional values. It is also rich in antioxidants, and therefore consumption of olive oil can help prevent cellular damage caused by free radicals. However, not all olive oils are equally beneficial for the human body because they do not contain the same amount of antioxidants and nutrients [1].
The quality of olive oil depends on the technological processes of its production and natural factors. Important natural factors include the type of soil and its composition or climate.
According to the regulations from European Union standards, there are three categories of olive oil, "extra virgin" (EVOO), "virgin" (VOO), and "lampante" (LOO) [2]. The quality of olive oils is evaluated by a panel test, which assesses taste, appearance, and aroma. In the panel test, the experts assign a score to the oils. Subsequently, the statistical analysis is applied to the score, which the experts gave to the sample. Finally, they classify olive oil into one of three classes. It is a demanding, expensive and time-consuming task, so there is a need to automate this process.
We propose the system for automatic spectrum recognition, which is using the methods of machine learning. In comparison to the other approaches, the application can also visualize the olive oil samples.

2 Related works

The authors of *Quality control of olive oils using machine learning and electronic nose* [3] dealt with the recognition of olive oil authenticity. They used standard classification methods for identifying the olive oil and falsification, for instance, sunflower oil. The accuracy of the proposed method was in the range of 56-70%,

* Master study programme in field: Intelligent Software Systems
Supervisor: Ing. Marta Prnová, PhD., Mgr. Martin Sabo, PhD., Institute of Informatics, Information Systems and Software Engineering, Faculty of Informatics and Information Technologies STU in Bratislava

and the best results were achieved by classifier Naive Bayes.
In the *Laser-based classification of olive oils assisted by machine learning* [4], the authors obtained the data with Laser-Induced Breakdown Spectroscopy (LIBS). They chose to combine LIBS and machine learning algorithms, specifically SVM, LDA, and random forest, to classify olive oil and its origin. According to them, it was used for the first time, and they achieved competitive results. The results of the used classification methods were in the range of 90 and 99.2%.

3 Our approach

The main goal of our project is to create a system to analyze the quality of olive oils. To achieve that goal, we propose the architecture of the application named ASR for automatic spectrum recognition, which can process data from AMIS (Advanced Ion Mobility Spectrometer). The AMIS allows us to analyze not only gaseous substances but also liquid and solid substances. The application is developed in cooperation with research company MaSa Tech, s.r.o[1], which provides the data from the spectrometer. The proposed application has three main functions, i.e., data processing, classification, and data display. The system architecture is shown in Figure 1.

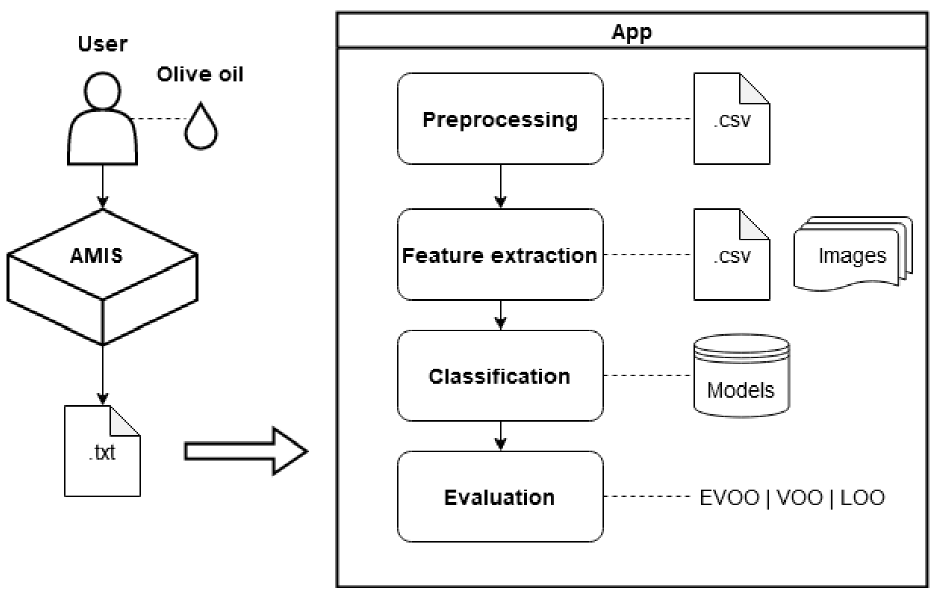

Fig. 1. System architecture.

[1] https://www.masatech.eu/

The trigger in the application is loading text data by a user. Until then, all functions are blocked. A user interface allows user to:

- load text data generated by a spectrometer,
- process data,
- show images from data,
- evaluate the class of the sample with one or multiple classification models,
- train new models or retrain the existing model,
- save models, images, and processed data.

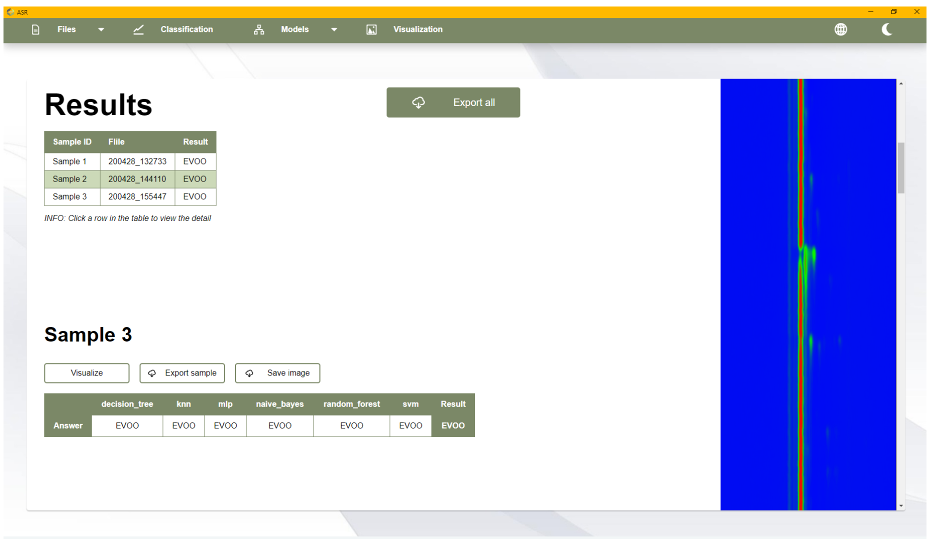

Fig. 2. Application screen preview - evaluation view.

3.1 Data

We obtain the input data with a spectrometer, which we are using for chemical analysis of olive oils. We acquire the information about spectra and measured values in them. Spectrometer records drift field intensity, pressure, temperature, drift tube length, analysis time, the time when the intensity was measured, and intensities values.

The data were classified in two panel tests, which have an accuracy of classification around 80%. Eventually, only data that had the same predicted class in both panel tests were selected. Therefore, the data has an accuracy of 96%, which we computed using Formula 1.

$$data\_accuracy = 1 - (panel1\_error * panel2\_error) \tag{1}$$

The panel_error was computed as a complement of panel accuracy (Formula 2).

$$panel\_error = 1 - panel\_accuracy \tag{2}$$

3.2 Extracting the features

To achieve more accurate results, we decided to train our models on computed features from input data. We compute some relevant information from the original data:

- the average for each oil,
- the average value for each column in a given oil sample,
- the maximum for each column in a given oil sample,
- the minimum for each column in a given oil sample,
- the median for each column in a given oil sample,
- the standard deviation for each column in a given oil sample,
- delta for each column in a given oil sample (maximum-minimum),
- the sum of the values for each column in a given oil sample.

In addition to these standard features, we came up with an idea to extract differential maps and point features. The differential maps represent the difference between the two compared oil samples. Point feature is computed as the mean and max value from significant areas, which we find during the analysis of differential maps.
We transform original data into images, which we use in an experiment with a convolutional neural network. Furthermore, the images can be displayed by users in the application for better visualization of data.
We apply feature selection on created features because of the amount of new data. Thus, the proposed classification method is trained only with data, which indicates the difference between olive oil classes.

3.3 Classification

In the classification phase, we decided to use these algorithms:

- k-nearest neighbors (KNN),
- support vector machine (SVM),
- decision tree (DT),
- random forest (RF),
- naive bayes (NB),
- artificial neural networks.

The classifiers were selected based on similar works, where they achieved excellent results in the field. We tried a convolutional neural network, specifically VGG-16, which was not pre-trained on the ImageNet dataset, but it did not achieve competitive results.

4 Results

We evaluate every model individually on test data. We acquired training and test datasets by splitting original data in a ratio of 4:1 for every class to provide even distribution. From all classification models, the random forest achieved the best results with an accuracy of 82.93%. We obtain these results on data with 96% accuracy, as we mentioned in Section 3.1. In comparison to the panel tests with an accuracy of 80%, we achieved better results.
We also predict only extra virgin olive oil from the data, which means we combined the virgin and lampante olive oils into one class. We achieve better results by using only two categories (extra virgin and combined class) of olive oil. The random forest achieved the best results again. It predicted the correct class with an accuracy of 90.24%.

5 Conclusion

Our solution provides an accurate and relatively fast classification of olive oil samples. Users can test the olive oil and find out whether it is a high-quality or lower-quality oil. They can also visualize the oil samples as images using the proposed system to imagine how the measured data looks. The application offers the possibility of training a new model or retraining the existing model with new data. Our future work could focus on scaling up an application that could be used in several areas of spectrum research.

References

1. Gaforio JJ, Visioli F, Alarcón-de-la-Lastra C, et al. Virgin Olive Oil and Health: Summary of the III International Conference on Virgin Olive Oil and Health Consensus Report, JAEN (Spain) 2018. Nutrients. 2019;11(9):2039. Published 2019 Sep 1. doi:10.3390/nu11092039
2. Beatriz Quintanilla-Casas, Julen Bustamante, Francesc Guardiola, Diego LuísGarcía-González, Sara Barbieri, Alessandra Bendini, Tullia Gallina Toschi,Stefania Vichi, and Alba Tres. Virgin olive oil volatile fingerprint and chemo-metrics: Towards an instrumental screening tool to grade the sensory quality.LWT, 121:108936, 2020
3. Emre Ordukaya and Bekir Karlik. Quality control of olive oils using machine learning and electronic nose.Journal of Food Quality, 2017:1–7, 10 2017.
4. Odhisea Gazeli, Elli Bellou, Dimitrios Stefas, and Stelios Couris. Laser-based classification of olive oils assisted by machine learning.Food Chemistry,302:125329, 20220.