Classification of potatoes according to their cultivated field by SVM and KNN approaches using an electronic nose

Ali Amkor, Noureddine El Barbri

Laboratory of Science and Technology for the Engineer, LaSTI-ENSA, Sultan Moulay Slimane University, Khouribga, Morocco

Article Info	ABSTRACT
	In this article, we propose a homemade electronic nose to distinguish between two types of potatoes: the first type is traditionally treated with donkey and sheep manure, and the other type is treated with chicken manure. The proposed tool consists of a network of commercial metal oxide sensors, a data acquisition card, and a personal computer for data pre-processing and processing. Two methods were used, namely, support vector machines (SVM) and k-nearest neighbors (KNN) with 5-fold cross-validation and which achieved the same success rate of 97.5%. These results demonstrate that our concept, which is quick, simple, and inexpensive, can discriminate between potatoes based on the method of fertilization used in the field.

Corresponding Author:

Ali Amkor
Laboratory of Science and Technology for the Engineer, LaSTI-ENSA, Sultan Moulay Slimane University
Bp 7, Avenue Beni Amir, ENSA Khouribga, Morrocco
Email: amkorali@gmail.com

1. INTRODUCTION

Potato (Solanum tuberosum L.) is one of the world's most important crops along with wheat, rice, and maize because it is grown in almost all types of climates in the world [1]. This agricultural crop originated in South America, but its consumption is widespread everywhere on the planet due to its high nutrient benefit, which includes a high proportion of starch, dietary fiber, crude protein, and a range of vital vitamins and minerals [2]. With the increasing global population growth, the demand for agricultural products is also increasing, forcing agricultural experts and farmers to look for new ways to increase crops and meet the needs of consumers.

In the past and until now, farmers have resorted to using farm animal waste, such as goats, sheep, cows, and donkeys, as organic fertilizers, which gave very good results, despite the fact that the inordinate utilization of these compounds has a harmful impact on the air and groundwater [3]. With the increase in the demand for meat as well, and farmers resorting to fattening farmyard animals with medicines [4], [5], fertilization with the waste produced from these animals has become a great danger to the health of the consumer because the medicines contain heavy metals that were found [6], according to a study that was done, in agricultural products [7].

Returning to the potato, the subject of our study, the latter, due to the large demand for it, is also exposed to the use of fertilizers in order to increase its yield and meet the increasing demand for it. To avoid the negative repercussions resulting from the consumption of treated potatoes, many studies have been conducted based on the smell of potato samples in order to distinguish between treated and untreated tubers using heavy and expensive equipment such as gas chromatography (GC). This technique was employed in conjunction with a variety of detectors, including nitrogen phosphorus detector (NPD), mass spectrometry

(MS), flame photometric detector (FPD), flame ionization detector (FID), and electron capture detector (ECD) [8].

To make the matter of distinguishing between consumables easy, a multi-sensor device was used that simulates the biological nose in its structure, a simple and inexpensive system, called the electronic nose [9], [10]. It has been used in many studies related to agricultural products and has proven its ability to distinguish between treated and untreated, to mention but not limited to, such as mint [11]–[13], tea [14], apples [15], and cherry tomatoes [16].

In the same context comes our study, which aims to distinguish between potato samples derived from soil that has traditionally been treated with sheep and donkey manure that have never been fattened from those derived from soil treated with fattened chicken manure on the basis of smell and using an electronic nose. The remainder of the article is arranged as below: the initial part (Materials and Methods) discusses the electronic nose design and measurement protocol followed by the sample preparation process, and finally, the data analysis part and machine learning algorithms used. The findings of the experiments will be presented and discussed in the second part (results and discussion), beginning with the analysis of sensor reaction and progressing to the results of the employment of machine learning methods. Finally, we will conclude this article.

2. MATERIAL AND METHODS

2.1. Electronic nose design and measurement protocol

This research looks at the ability of an electronic nose to identify potato samples depending on the treatment they have undergone: traditionally treated with sheep and donkey manure that have never been fattened or treated with fattened chicken manure. Our homemade instrument is made up of three major components: the sensor array, the data collection, and the data analysis. The sensor array is mostly composed of the five-metal oxide (MOX) sensors listed in the Table 1. These sensors are installed on a PCB and have potentiometers to control the voltages provided. For the collection of data, is carried out utilising the DAQ 1901USB card and a program in LabVIEW software. The designed electronic nose's hardware comprises a ventilator with a fixed flow, a sample bottle, a sensor chamber, the DAQ 1901USB card, and a computer. Figure 1 depicts the measurement procedure schematically.

Table 1. The sensors utilized in the electronic nose system

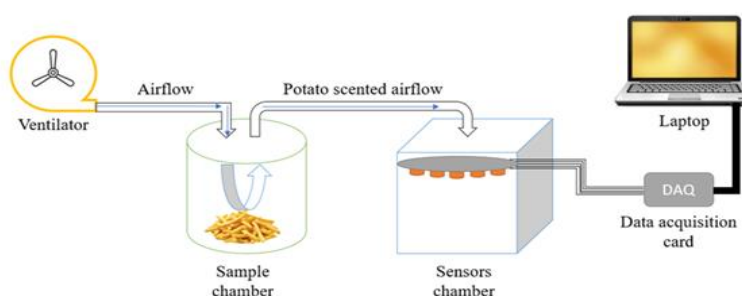| Sensor	Target gas
MQ-7	Carbon monoxide
MQ136	Hydrogen sulfide, ammonia, Air, and carbon
TGS821	Hydrogen
TGS822	Vapors of organic solvents such as ethanol
TGS2620	Volatile organic vapors

Figure 1. Measurement tool

Regarding the measurement process, the experiments begin by placing the samples in the sample chamber first and then operating the ventilator that produces the airflow, as the latter reaches the sample chamber and works to transfer the air enhanced with the volatile organic components of the potato sample subject of the experiment to the sensors chamber, which interacts by producing voltage depends on their sensitivity to smell. These reactions are gathered using the DAQ 1901 USB acquisition card and registered in an Excel file by means of a program previously created in LabVIEW for post-processing by machine learning algorithms.

2.2. Sample preparation

In this study, samples of potato tubers were used from two fields of the same place (33°53′42″N 5°33′17″W Meknes Morocco) with different fertilization methods, a field that was fertilized in the traditional way based on the manure of farm animals from sheep and donkeys, and a field that was fertilized using the manure of fattened chickens. The best and finest samples were selected, while samples that were immature or had chlorophyll spots, or showed evidence of disease were excluded, even those that were destroyed in the course of harvesting. The selected samples were washed, dried, and then cut into French fries. A weight of 100g from each sample type was placed in each 1L glass Erlenmeyer flask before being enclosed, labelled, and stored for seven days at a temperature of 25±2°C. Figure 2 depicts ones of the samples exploited in this research.

Figure 2. Picture of some samples

2.3 Data analysis

Data preprocessing is the manipulation of data in order to prepare and transform it into the most appropriate form for analysis. This is a very important and necessary step for structuring the data because proper pre-processing can create a difference between a useful and useless model as well as improve performance. After the raw sensor responses are collected, they are converted from their analog forms into digital signals that can be interpreted by a computer. Then the data is centered in (1) [17], extracting and selecting key features, and eliminating the scaling effect by normalization. In the present study, we used the column normalization method [18], which involves dividing each column by its greatest value (2).

$$V_r = \frac{V_m - V_0}{V_0} \tag{1}$$

where V_r denotes the resulting relative voltage, V_m is the measured voltage, and V_0 is the initial voltage.

$$X_{ij} = \frac{X_{ij}}{Max(X_i)} \tag{2}$$

X_{ij} represents the ith sample of the jth sensor, while X_i contains all of the p responses for the ith sample's sensors. Following that, the data is ready for multivariate analysis.

2.3.1. K-nearest neighbors algorithm

First proposed by Silverman and Jones [19] in 1951 and modified by Cover in 1967 [20], the k-nearest neighbours (KNN) method is a simple and effective technique that assigns the input sample to the class that the majority of the k samples closest to the training set belong to [21]. With a limited samples size, this supervised nonparametric method could indeed perform nonlinear classification.

The choice of the key elements, namely the number of neighbors K and the distance metric D, affects the performance of this approach. For distance measurement, although a variety of distance measures can be used to calculate the distance between two points, the Euclidean distance is the most used metric, it is calculated as (3):

$$D = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2} \tag{3}$$

2.3.2. Support vector machines algorithm

This technique initiated by Vapnik [22] in 1995 is a supervised learning algorithm that uses the linear or nonlinear kernel function to classify data; this function tries to distinguish classes using hyperplanes. classifies data using the linear or nonlinear kernel function, this function tries to distinguish classes by hyper-

planes. hyperplanes are found such that they do not include interior data and provide maximum headroom between class pairs. The data points located on these hyperplanes are called support vectors. The hyperplane that passes by the center of the maximum distance separating the two said hyperplanes is the best result, here equation is given by:

$$y(x) = w^T \varphi(x) = \sum_{j=1}^{k} w_j \varphi_j(x) + w_0 = 0 \qquad (4)$$

where $\varphi(x) = [\varphi_0(x), \varphi_1(x), \ldots, \varphi_k(x)]^T$ with $\varphi_0(x) = 1$ and w is the network's weight vector $w = [w_0, w_1, \ldots, w_k]^T$. If the data vector x meets the criterion $y(x) > 0$, it is considered a member of the first class, and if $y(x) < 0$ it is considered a member of the opposite class [23].

3.   RESULTS AND DISCUSSION
3.1. Sensor responses
In order to distinguish the samples of potatoes according to their origins, we took in our experiment samples according to the treatment they underwent: samples treated with chicken manure and samples traditionally treated with donkey and sheep manure, this is the most common treatment. Data from potato samples were collected in three steps: In the first stage, the sensor room was ventilated for 10 minutes to clean the sensors and reach baseline; in the second stage, the sensor room was closed, and the stability of the sensor responses was checked for 30 seconds; in the third and final stage, the potato headspace was incorporated into the tools and subjected to an airflow emitted by the fan and it was injected for 8 minutes into the sensor chamber, then the experiment is repeated. It is noteworthy that each type of sample was subjected to 20 experiments, bringing the total to 40 experiments. An example of the recorded sensor responses is shown in Figure 3.

According to these responses, the first thing we noticed was that the intensities of the sensors' responses increased once the measurement was triggered, which proves the sensors' ability to recognize the samples' existence and the change in the composition of the smell provided in the sensor chamber. Also, by observing the responses, we noted that the responses relating to the samples of potatoes treated with chicken manure are in most cases greater than the responses of the samples of potatoes treated traditionally.
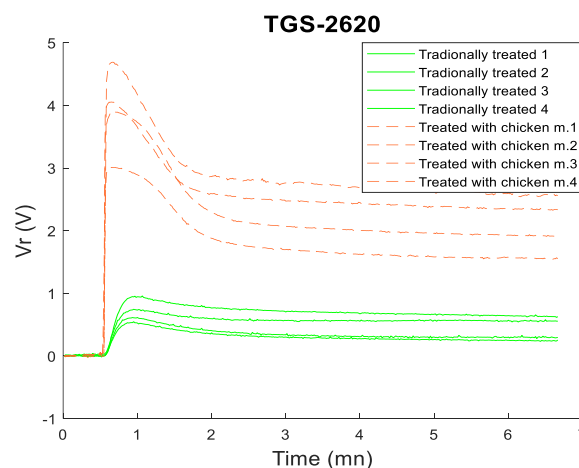

Figure 3. TGS-2620 responses

3.2. Features extraction
Raising features requires careful examination of the response figures, taking into account the purpose of the study. In this study, the aim is to distinguish between groups of samples according to the type of fertilization. In this context, we have identified some features that will inevitably assist us to reach the desired goal: i) maximum value (Vmax), ii) signal-occupied area (area), iii) fixed signal value (MS) from 450 to 500 seconds. As a consequence, our data matrix will contain 15 columns (3 features * 5 sensors) and 40 rows (20 experiments * 2 types of potatoes).

3.3. Classification of potato samples with the KNN and the SVM methods

Support vector machines (SVM) and KNN methods are the most useful classification algorithms, they are supervised classification methods in which the groups of training data are predetermined [24]. To prevent overfitting and obtain an accurate classification, the 5-fold cross-validation method was exploited to evaluate each model. This means that the data set is randomly divided into 5 folds of roughly the same size, where each time one is used for testing and the others for training until all folds have passed testing and training.

For the SVM, the linear kernel is employed; it is among the most commonly utilized kernels. As for KNN, the classification was performed with a neighbor count of 5, the distance metric was Euclidean, and the distance weight was equal. The two methods produced the same result shown in Figure 4. The confusion matrix is utilised to compute the detection models' performance parameters. The first group of potatoes traditionally treated has a success rate of 95%, while the second group of potatoes treated with chicken manure had a 100% success rate. Thus, they achieved an overall result of 97.5% success.

To get a good idea of the data, we chose to view them in a two-dimensional plane, taking as the horizontal axis the stabilized value obtained from the MQ-7 sensor and as the vertical axis the maximum value obtained from the TGS-822 sensor. The result is shown in Figure 5. According to Figure 5, individuals from each group of samples were correctly allotted to their data classes, with the exception of one sample from the group of potatoes traditionally treated, which was misclassified. Nonetheless, the success rate was excellent. Considering this very good result, and in comparison, with those obtained with other materials which are heavy, difficult to handle, and onerous such as gas chromatography-mass spectrometry (GC-MS) [25], [26], Headspace solid-phase microextraction/gas chromatography–mass spectrometry (HS-SPME/GC–MS) and isotope ratio mass spectrometry (IRMS) [27], [28] and near-infrared reflectance spectroscopy (NIR) [29], [30], used in different applications related to the study of potatoes, we can say that our electronic nose which is made up mainly of commercial gas sensors could be used as an efficient device for the classification of potatoes according to their crop fields.
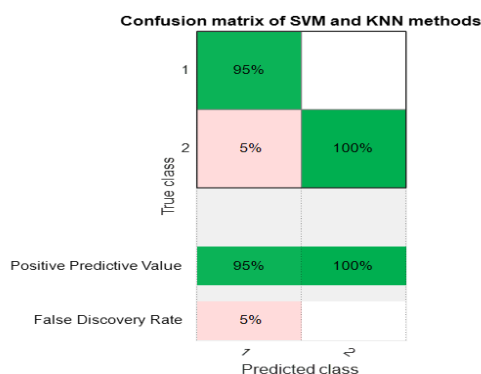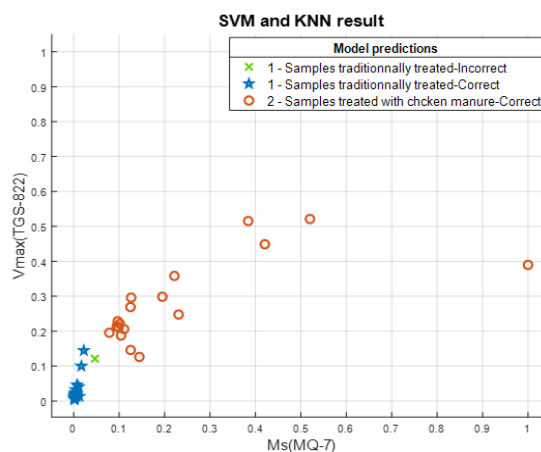

Figure 4. confusion matrix of the algorithms used

Figure 5. The KNN and SVM result for the classification of potatoes according to their cultivated field

4. CONCLUSION

The consumption of food treated with manure from medication-fattened poultry represents a serious danger to the health of consumers, it is with this in mind that we decided to manufacture an electronic nose made up mainly of commercial gas sensors for monitoring the quality of potatoes by distinguishing the samples treated with this chicken manure. Using two machine learning algorithms in this case KNN and SVM with 5-fold cross-validation we arrived at a success rate of 97.5% in the classification of the samples, which means our homemade tool is able to distinguish between potatoes traditionally treated with the sheep and donkeys' manure and potatoes treated with the chicken manure. Without forgetting that the strong points of our tool lie in its modest cost and its simple design.

REFERENCES

[1] H. Zhang, F. Xu, Y. Wu, H. Hu, and X. Dai, "Progress of potato staple food research and industry development in China," *Journal of Integrative Agriculture*, vol. 16, no. 12, pp. 2924–2932, 2017, doi: 10.1016/s2095-3119(17)61736-2.

[2] K. A. Beals, "Potatoes, Nutrition and Health," *American Journal of Potato Research*, vol. 96, no. 2, pp. 102–110, 2018, doi: 10.1007/s12230-018-09705-4.

[3] S. Savci, "Investigation of Effect of Chemical Fertilizers on Environment," *{APCBEE} Procedia*, vol. 1, pp. 287–292, 2012, doi: 10.1016/j.apcbee.2012.03.047.

[4] T. Bunmee, S. Jaturasitha, M. Kreuzer, and M. Wicke, "Can calcium chloride injection facilitate the ageing-derived improvement in the quality of meat from culled dairy cows?," *Meat Science*, vol. 96, no. 4, pp. 1440–1445, Apr. 2014, doi: 10.1016/j.meatsci.2013.11.002.

[5] V. H. M. Visschers, D. M. Iten, A. Riklin, S. Hartmann, X. Sidler, and M. Siegrist, "Swiss pig farmers' perception and usage of antibiotics during the fattening period," *Livestock Science*, vol. 162, pp. 223–232, Apr. 2014, doi: 10.1016/j.livsci.2014.02.002.

[6] F. Zhang, Y. Li, M. Yang, and W. Li, "Content of Heavy Metals in Animal Feeds and Manures from Farms of Different Scales in Northeast China," *International Journal of Environmental Research and Public Health*, vol. 9, no. 8, pp. 2658–2668, Jul. 2012, doi: 10.3390/ijerph9082658.

[7] C. Manyi-Loh, S. Mamphweli, E. Meyer, and A. Okoh, "Antibiotic Use in Agriculture and Its Consequential Resistance in Environmental Sources: Potential Public Health Implications," *Molecules*, vol. 23, no. 4, p. 795, Mar. 2018, doi: 10.3390/molecules23040795.

[8] S. T. Narenderan and S. N. Meyyanathan, "Sample Treatment and Determination of Pesticide Residues in Potato Matrices: a Review," *Potato Research*, vol. 62, no. 1, pp. 47–67, Jul. 2018, doi: 10.1007/s11540-018-9396-x.

[9] A. Loutfi, S. Coradeschi, G. K. Mani, P. Shankar, and J. B. B. Rayappan, "Electronic noses for food quality: A review," *Journal of Food Engineering*, vol. 144, pp. 103–111, Jan. 2015, doi: 10.1016/j.jfoodeng.2014.07.019.

[10] A. Sanaeifar, H. ZakiDizaji, A. Jafari, and M. de la Guardia, "Early detection of contamination and defect in foodstuffs by electronic nose: A review," *{TrAC} Trends in Analytical Chemistry*, vol. 97, pp. 257–271, 2017, doi: 10.1016/j.trac.2017.09.014.

[11] A. Amkor and N. El Barbri, "Identification of potatoes harvested from the treated field with NPK fertilizers with an electronic system," pp. 31–41, doi: 10.1007/978-981-19-0039-6_3.

[12] A. Amkor and N. El Barbri, "A measurement prototype based on gas sensors for detection of pesticide residues in edible mint," *Journal of Food Measurement and Characterization*, vol. 15, no. 1, pp. 170–180, 2021, doi: 10.1007/s11694-020-00617-8.

[13] A. Amkor, N. El Barbri, and K. Maaider, "A comparison between PLSR, SVMR and NARX network for the mint treatment day prediction based on multisensor system," *2021 International Conference on Optimization and Applications, ICOA 2021*, 2021, doi: 10.1109/ICOA51614.2021.9442652.

[14] A. Sanaeifar, X. Li, Y. He, Z. Huang, and Z. Zhan, "A data fusion approach on confocal Raman microspectroscopy and electronic nose for quantitative evaluation of pesticide residue in tea," *Biosystems Engineering*, vol. 210, pp. 206–222, 2021, doi: 10.1016/j.biosystemseng.2021.08.016.

[15] Y. Tang *et al.*, "A novel electronic nose for the detection and classification of pesticide residue on apples," *{RSC} Advances*, vol. 11, no. 34, pp. 20874–20883, 2021, doi: 10.1039/d1ra03069h.

[16] L. Feng, M. Zhang, B. Bhandari, and Z. Guo, "A novel method using MOS electronic nose and ELM for predicting postharvest quality of cherry tomato fruit treated with high pressure argon," *Computers and Electronics in Agriculture*, vol. 154, pp. 411–419, 2018, doi: 10.1016/j.compag.2018.09.032.

[17] K. Mahmodi, M. Mostafaei, and E. Mirzaee-Ghaleh, "Detection and classification of diesel-biodiesel blends by LDA, QDA and SVM approaches using an electronic nose," *Fuel*, vol. 258, no. August, p. 116114, 2019, doi: 10.1016/j.fuel.2019.116114.

[18] J.-M. Jo, "Effectiveness of Normalization Pre-Processing of Big Data to the Machine Learning Performance," *The Journal of the Korea Institute of Electronic Communication Sciences*, vol. 14, no. 3, pp. 547–552, 2019.

[19] B. W. Silverman and M. C. Jones, "E. Fix and J.L. Hodges (1951): An Important Contribution to Nonparametric Discriminant Analysis and Density Estimation: Commentary on Fix and Hodges (1951)," *International Statistical Review / Revue Internationale de Statistique*, vol. 57, no. 3, p. 233, 1989, doi: 10.2307/1403796.

[20] T. Cover and P. Hart, "Nearest Neighbor Pattern Classfication," *IEEE TRANSACTIONS ON INFORMATION THEORY*, vol. I, pp. 21–27, 1967, doi:10.1109/tit.1967.1053964.

[21] H. R. Estakhroueiyeh and E. Rashedi, "Detecting moldy Bread using an E-nose and the KNN classifier," *2015 5th International Conference on Computer and Knowledge Engineering, ICCKE 2015*, pp. 251–255, 2015, doi: 10.1109/ICCKE.2015.7365836.

[22] V. Vapnik, "Support-Vector Networks," *machine learning*, vol. 20, no. 3, pp. 273–297, 1995, doi: 10.1007/BF00994018.

[23] K. Brudzewski, S. Osowski, T. Markiewicz, and J. Ulaczyk, "Classification of gasoline with supplement of bio-products by means of an electronic nose and SVM neural network," *Sensors and Actuators B: Chemical*, vol. 113, no. 1, pp. 135–141, Jan.

2006, doi: 10.1016/j.snb.2005.02.039.

[24] G. Kesavaraj and S. Sukumaran, "A study on classification techniques in data mining," in *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, Jul. 2013. doi: 10.1109/icccnt.2013.6726842.

[25] Z. Juhász, G. Dancs, F. Marincs, M. Vossen, S. Allefs, and Z. Bánfalvi, "Vitamin C, B5, and B6 contents of segregating potato populations detected by GC-MS: a method facilitating breeding potatoes with improved vitamin content," *Plant Breeding*, vol. 133, no. 4, pp. 515–520, Feb. 2014, doi: 10.1111/pbr.12169.

[26] T. Fukuda, K. Takamatsu, T. Bamba, and E. Fukusaki, "Gas chromatography-mass spectrometry metabolomics-based prediction of potato tuber sprouting during long-term storage," *Journal of Bioscience and Bioengineering*, vol. 128, no. 2, pp. 249–254, 2019, doi: 10.1016/j.jbiosc.2019.01.016.

[27] F. Longobardi, G. Casiello, D. Sacco, L. Tedone, and A. Sacco, "Characterisation of the geographical origin of Italian potatoes, based on stable isotope and volatile compound analyses," *Food Chemistry*, vol. 124, no. 4, pp. 1708–1713, Feb. 2011, doi: 10.1016/j.foodchem.2010.07.092.

[28] H. Ayvaz, M. Plans, K. M. Riedl, S. J. Schwartz, and L. E. Rodriguez-Saona, "Application of infrared microspectroscopy and chemometric analysis for screening the acrylamide content in potato chips," *Analytical Methods*, vol. 5, no. 8, p. 2020, 2013, doi: 10.1039/c3ay00020f.

[29] X. Ding, Y. Ni, and S. Kokot, "{NIR} spectroscopy and chemometrics for the discrimination of pure, powdered, purple sweet potatoes and their samples adulterated with the white sweet potato flour," *Chemometrics and Intelligent Laboratory Systems*, vol. 144, pp. 17–23, 2015, doi: 10.1016/j.chemolab.2015.03.004.

[30] R. Tierno *et al.*, "Phytochemicals determination and classification in purple and red fleshed potato tubers by analytical methods and near infrared spectroscopy," *Journal of the Science of Food and Agriculture*, vol. 96, no. 6, pp. 1888–1899, Jul. 2015, doi: 10.1002/jsfa.7294.

BIOGRAPHIES OF AUTHORS

Ali Amkor ⓘ 🔾 SC ↻ receive his Ph.D. in physics specialty in electrical engineering, at the Faculty of Sciences and Techniques of Béni-Mellal, Morocco in 2022 and a Master Science and Technology degree in engineering and industrial management from the Faculty of Sciences and Techniques of Settat Morocco in 2017. His research interest includes electronic, multi-sensor systems, electronic noses, data analysis, and artificial intelligence. He can be contacted at email: amkorali@gmail.com.

Noureddine El Barbri ⓘ 🔾 SC ↻ is an Associate Professor in the Electrical Engineering Department at the National School of Applied Sciences of Khouribga, Morocco since 2010. He received his Ph.D. in physics, Electronics, and Artificial Intelligence operation, at the Faculty of Science Moulay Ismail University, Morocco in 2008. In 2010, he became an Assistant Professor at the National School for Applied Sciences, Khouribga, Morocco, where he became an Associate Professor in 2015. His research interests include design and realization of multi-sensor systems, electronic nose, data analysis, image processing, artificial vision, artificial intelligence, food processing, electronics of emblematic systems. He can be contacted at email: elbarbri.noureddine@yahoo.fr.