An evaluation of machine learning algorithms coupled to an electronic olfactory system: a study of the mint case

Ali Amkor, Kamal Maaider, Noureddine El Barbri
Laboratory of Science and Technology for the Engineer, LaSTI-ENSA, Sultan Moulay Slimane University, Khouribga, Morocco

Article Info

Article history:

Received Jul 9, 2021
Revised Jan 12, 2022
Accepted Apr 14, 2022

Keywords:

Classification methods
Data analysis
Machine learning algorithms
Metal oxide semiconductor gas sensors
Multi-sensors system
Regression methods

ABSTRACT

The aim of this investigatation is to compare the utility of machine learning algorithms in distinguishing between untreated and processed mint beside in predicting the spray day of the insecticide. Within seven days, mint treated samples with the malathion insecticide are collected, and their aromas are Studied using a laboratory-manufactured sensor array system based on commercial metallic semiconductor (MOS) gas sensors. To distinguish the mint type, some results of machine learning algorithms were compared to know the decision trees (DT), Naive Bayes, support vector machines (SVM), and ensemble classifier. Furthermore, to predict the treatment day support vector machines regression (SVMR) and partial least squares regression (PLSR) were compared. Regarding the best results, in the discrimination case, a success rate of 92.9% was achieved by the ensemble classifier while in the prediction case, a correlation coefficient of R=0.82 was reached by the SVMR. Good results are achieved if the right gas sensor array system is designed and realized coupled with a good choice of the appropriate machine learning algorithms.

Corresponding Author:

Ali Amkor
Electrical Engineering Department, Sultan Moulay Slimane University
Bp 7, Avenue Beni Amir, ENSA Khouribga, Morrocco
Email: amkorali@gmail.com

1. INTRODUCTION

In recent years, the capacity to differentiate between herbs pesticides-processed has attracted considerable attention from consumers due to the rates of poisoning and chronic illnesses caused by consuming contaminated food. This task is done using chemical materials like gas chromatography alone or coupled with mass spectrometry [1], analysis by gas chromatography coupled with olfactometry [2], or ion mobility spectrometry (IMS) [3]. These techniques are expensive and require a trip to the laboratory, hence the idea of simplifying the quality control process. Lately, the multi-sensor system is widely used to evaluate the quality of food, it is a non-destructive procedure using machine learning algorithms, this tool has been used by several researchers in the agro-processing industry, to control the quality of fruits [4], vegetable [5], meats [6], and fish [7]. It also makes it possible to identify counterfeits, for example for, meats, it makes it possible to distinguish mixtures of meats, mutton with pork [8], or adulterated beef with pork [9].

In electronics and especially multi-sensors systems, machine learning algorithms are a fundamental and crucial component. They are used to process large and complex data for the purpose of extracting useful information and to improve decision making. They are used for data classification as well as for regression. Machine-learning algorithms are supervised [10], unsupervised [11] or semi-supervised [12] but the best known and the most used in the classification are supervised algorithms such as decision tree (DT), naive Bayes, k-nearest neighbours algorithm (KNN), support vector machines (SVM), and some artificial neural networks (ANN) algorithms, like multilayer perceptron (MLP) [13]. These methods try to establish a relation between the input and the given target output in the first phase of training (learning) to form the model and use it during the test to make decisions. For the regression we find linear regression, logistic regression, poisson regression, partial least squares regression (PLSR), support vector machines regression (SVMR) and back propagation artificial neural network (BP-ANN) [14], [15]. There isn't a method that is effective for all the problems studied but each application requires carrying out several comparisons between the methods in order to determine the most suitable classifier.

In order to get the best result, many studies have compared machine learning methods in order to choose the most efficient algorithms, such as in the case of the soil salinity approximation from hyperspectral data where they compared three PLSR algorithms, the SVM and deep learning techniques [16], and they found that the PLSR is the most suitable for this case. In another study where the objective was the estimation of forest parameters, they used the classification and regression tree (CART), the SVM, the ANN, and the random forest (RF). The RF algorithm turned out to be the most efficient [17]. In another case for the estimation of gas concentration based on an electronic nose, they used partial least squares (PLS) and SVM regression. SVM regression provided better generalization and precision [18]. Another study for the comparative evaluation and analysis of three machine learning algorithms in a controllable environment. They confirmed that basic propagation (BP) works better without being limited to severely polluted air conditioning but in addition to moderately polluted air conditioning compared to radial basis function (RBF) neural network and support vector regression (SVR) [19].

Mint is an aromatic herb widely consumed in the world and it is the subject of this study given the consumption rate of this vital substance either in cooking or as a therapeutic plant used in the preparation of medicines as well as in the beautifying products preparation, but the increasing demand for this herb drives the farmers to use insecticides to protect their harvests and to avoid the damage caused by the pests of this sensitive plant in order to meet the demand. The lifespan of insecticides product on plants can range from a week and sometimes up to a month, depending on the insecticides and the plant being treated, but nescience, voracity, and growing demand cause farmers to miss deadlines pre-registered. The scope of this research is mainly to compare the performance of the classification and regression algorithms, on the one hand, to differentiate the processed mint and the untreated one, and on the other hand to expect the treatment day with precision using different variations of machine learning algorithms. In our investigation we have employed for the classification: DT, naive bayes, SVM and ensemble classifier. For the regression: SVMR and PLSR. This paper is prepared in accordance with: the section "Materials and methods" depicts the mint used and sampling, the multi-sensor system, the data pre-processing, and the data analysis which we will present the machine learning methods used. The section "Results and discussion" presents the results of the different choice algorithms for classification and regression.

2. METHODS AND PROCEDURES

2.1. Mint used

Samples of our hand-grown mint were used within the walls of the ENSA Khouribga, Morocco, and picked freshly after ripening. The first seedlings were purchased locally. The transplanted field is divided into two, first preserved part remains unprocessed and another part is treated with the insecticide product Malyphos 50 containing a dosage of 500 g/l of the dangerous substance malathion. a dosage of 4 ml of this product diluted in one litre of potable water was used. Malathion is a carcinogenic organophosphate neurotoxic compound [20] causing many cases of poisoning forbidden in several countries such as the European Union (decision 2007/389/EC). Its chemical formula is $C_{10}H_{19}O_6PS_2$.

2.2. Electronic olfactory system and sampling

The multi-sensors system's main idea is to imitate the principle of the human olfactory system operation, the block diagram in Figure 1 shows the sensing concept of the multi-sensors system. A multi-sensors system is typically composed of three principal parts: the sensor matrix, the data preprocessor, and the data interpretation. The odour sensor array is constituted of: the odour of samples and sensor array, the data pre-processor is made up of: raw signals and signal pre-processing, and finally, the data interpretation is constituted of: data analysis and processing which we can do using software with the pattern-recognition algorithm to arrive at the classification and the prediction. As shown in Figure 1, the multi-sensor system made in our research laboratory compose of a ventilator, a sample space, a sensor space, an acquisition card, and a laptop. The main component of this system is the sensor network, it is composed of seven metal oxide gas sensors listed in Table 1.

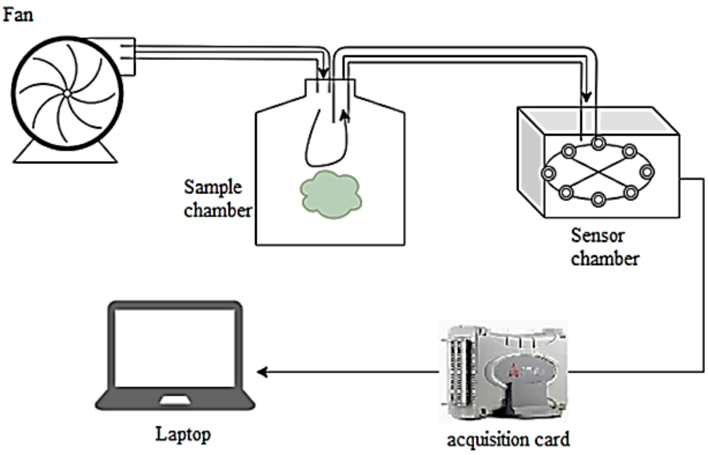


Figure 1. Schematic representation of our multi-sensors system

Table 1. Sensors network utilized in the multi-sensors system

Sensor	Target gas
TGS 822	Vapours of organic solvents
TGS 826	Ammonia, Air, Isobutane, Hydrogen and Ethanol
TGS 2620	Volatile organic vapours
TGS 4161	Carbon dioxide
MQ-7	Carbon monoxide
MQ-136	Hydrogen sulphide, ammonia, Air and carbon monoxide
MQ-214	Methane LPG, Air, Isobutane and propane

2.3. Measurement process and data collection

Mint samples weighing 5±0.1 g of each kind is placed in the sample chamber from which it undergoes an airflow coming from the ventilator for eight minutes, the airflow provided with the mint volatile organic compounds will then be channelled to the sensor chamber. The mint volatile organic compounds enter into a chemical reaction with the sensitive layer of each sensor for 8 minutes. Each sensor delivers a signal according to its sensitivity to the chemical compounds of the odour. After each measurement, the sensor array is exposed to ambient air for 10 minutes to purify it until the responses of the sensors are observed returned to their initial position.

The sensors signals will be analyzed using pattern recognition algorithms and data processing techniques that play an essential role in making a decision. The sensors' reaction is recorded utilizing an acquisition card namely ADLINK USB 1901 DAQ and a program on the labview software. Three datasets were extracted:

X_1: the dataset of untreated mint.
X_2: the dataset of mint treated.
X_T: the full dataset.

The resulting matrix, so, is made up of three columns (three sensor responses) and seventy lines (7 numbers of days×2 mint types×5 samples).

2.4. Data pre-processing

After data collection, the raw signals have been managed to have signals referenced to zero for reading ease the responses by (1):

$$V_r = \frac{V_m - V_0}{V_0} \quad (1)$$

With V_0, V_m and V_r are successively the starting voltage, the real voltage and the output relative voltage. Then, the signals will be observed to extract the pertinent characteristics (features) from the sensor responses. Normalization is applied to the features using the column normalization by dividing each column by its maximum value:

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

Xij is the ith sample of the jth sensor, *Xi* comprises all the p rractions for the sensors of the ith sample [21]. The normalized features extracted from the recorded data will be used in the data classification and prediction thanks to the different algorithms of machine learning. Regarding the treated mint discrimination from untreated one, data classification algorithms were used. As for the mint treatment day prediction, regression algorithms were exploited.

2.5. Classification methods

2.5.1. Decision trees algorithm

The DT is a supervised machine learning algorithm that can be utilized either for classification or regression problems [22]. It is a simple method that is founded on the prediction of the response by following decisions in the tree from the root node (start) to a leaf node. Each branch is linked to decision criteria. first, the model is formed from the training data, and then it will be used to predict the class of the data in the case of classification or the value of the variable in the case of regression. In our case, we search to discriminate the mint treated from untreated one that is why we used the Statistics and machine learning toolbox™ trees which are binary. Each step in a prediction takes into check the value of one predictor (variable).

2.5.2. Naïve Bayes classifier

Naive Bayes classifier is a widespread supervised machine learning algorithm used for classification. The algorithm exploits bayes' theorem which is founded on the prior knowing of the circumstances related to an happening that can describe the probability of this event occurring. Bayes theorem is defined as (3):

$$P(X|Y) = \frac{P(X)P(Y|X)}{P(Y)} \quad (3)$$

where *X* and *Y* are independent. *P(X|Y)* is the probability of *X* when *Y* has already happened. *P(X)* and *P(Y)* are the probabilities of two independent *X* and *Y*. *P(Y|X)* is the chance of *Y* when *X* has already happened [23].

2.5.3. Support vector machines (SVM)

SVM method is a supervised learning algorithm that classifies the data using the linear or non-linear kernel function, this function attempts to distinct classes by hyper-planes [24]. The hyperplane passing through the centre of the maximum distance separating two hyperplanes is the best result. Suppose we have scores separated by two planes P1 and P2 whose maximum distance, and the optimal hyperplane P0 separating P1 and P2. P0 is defined by (4):

$$\mathrm{P_o}: W.X + b = 0 \quad (4)$$

where *X* is a point lying on the hyperplane, *W* is normal to the hyperplane, b is the bias. The minimization of *W* leads to the maximization of the margin.

2.5.4. Ensemble classifier

Ensemble classifiers are supervised learning algorithms whose purpose is to combine the predictions of several basic learners who are members of the set into a single output that often performs better on average than any other member of the set with uncorrelated error in the target data sets [25]. This leads to building a more precise classification decision. Newer algorithms contain error-correcting output coding, Bagging, and boosting but the Bayesian averaging remains the original ensemble method.

2.6. Regression methods

2.6.1. Partial least squares regression (PLSR)

PLSR is a multivariate statistical supervised method that discoveries a linear regression model by projecting predicted and observable variables into a new space. The principle of PLSR is similar to that of principal components regression (PCR), is to extract the orthogonal factors from the latent variables (LVs) and to create the regression relationship among the dataset and the corresponding reference value [26]. For example, for input data matrices X and output Y, the PLS model will try to find the multidimensional direction in the X space that describes the direction of maximum multidimensional variance in the Y space [27]. the objective of this regression is to find a linear function such that:

$$B = A\beta + C \quad (5)$$

The general original model of multivariate PLS is:

$$A = TP^T + E \tag{6}$$

$$B = UQ^T + F \tag{7}$$

where A is an n×m matrix of predictors, B is an n×p matrix of responses; T and U are n×l matrices that are, respectively, projections of A (the A score, component or factor matrix) and projections of B (the B scores); P and Q are, respectively, m×l and p×l orthogonal loading matrices; and matrices E and F are the error terms, assumed to be independent and identically distributed random normal variables. The decompositions of A and B are made so as to maximize the covariance between T and U.

2.6.2. Support vector machines regression (SVMR)

Unlike the previously mentioned support vector machines which are used for classification, SVMR is utilized to predict values [28]. The regression learner application in statistics and machine learning toolbox ™ used for this study has a specific SVM regression algorithm named linear epsilon insensitive SVM (ε-SVM). In the ε-SVM regression, the training dataset contains predictor variables and observed response values. The aim is to find a function B which diverges from yn by a value not more than ε for each learning point x, and at the same time as flat as possible.

To get a linear function (B=Aβ+C) and confirm that it is as plane as possible, find B with the minimal norm value (β'β). This is formulated as a convex optimization problem to reduce:

$$J\,(\beta) = \frac{1}{2} * \beta'\beta \tag{8}$$

All residuals must have a value less than ε. In the absence of a function B that can verify these constraints for all the points. This problem can be resolved by entering the slack variables ξn and ξ*n for each point:

$$J(\beta) = \frac{1}{2} * \beta'\beta + C\sum_{n=1}^{n}(\xi_n + \xi_n^*) \tag{9}$$

Subject to:

$$\forall n\text{: } y_n\text{-}(x_n'\beta+b) \le \varepsilon+\xi_n,\ (xn'\beta+b)-y_n \le \varepsilon+\xi^*n, \xi^*n \ge 0,\ \xi_n \ge 0$$

The constant C is the box constraint, a positive numeric value that controls the penalty obligatory on observations that lie outside the epsilon margin (ε) and helps to avoid overfitting (regularization). This value determines the trade-off between the flatness of B and the amount up to which deviations larger than ε are tolerated. The parameters mean squared error (MSE) and correlation coefficient (R) were utilized to evaluate the model performance of each method [29]. The correlation coefficient (R) measures the relationship between the outputs and the targets, if the value R is equal to 1 it means that the correlation is close while 0 means that the correlation is random. The mean squared error (MSE) is the average squared difference between outputs and targets. Lower values are better.

$$R = \frac{\sum_{i=1}^{n} X_i Y_i}{\sqrt{\sum_{i=1}^{n} X_i^2 \sum_{i=1}^{n} Y_i^2}} \tag{10}$$

$$MSE = 1/n\sum_{i=1}^{n}(X_i - Y_i)^2 \tag{11}$$

Where X_i is the experimental value and Y_i is the forecasted value while n is the value number.

3. RESULTS AND DISCUSSION

The insecticide malathion was used to treat part of the mint field and another was reserved untreated. Afterwards, the mint samples were grown fresh every day throughout one week and but in the sample enclosure. The gas sensors response in the existence of the mint headspace was recorded. It is noteworthy that only three sensors responded well to the presence of mint, it is the sensors MQ136, TGS822, and TGS2611, and only their data that will be processed.

In the responses collected from the sensors as in Figure 2 during the seven days of experiments, the headspace circulated in the sample chamber exhibited a global behaviour of increasing voltage detected at the sensors output with a noteworthy variance between the reactions of the two types of samples (untreated mint, mint treated with Malyphos). It was noted that the sensors responses for the treated mint differ from those not treated at the level of the signal supreme value and of the surface occupied by the signal as well as of at the stabilized value. The remarkable change which happens between the sensors responses is surely due to the change in the volatile ingredients generated by each kind of mint. These selected features will be exploited subsequently using machine learning algorithms.

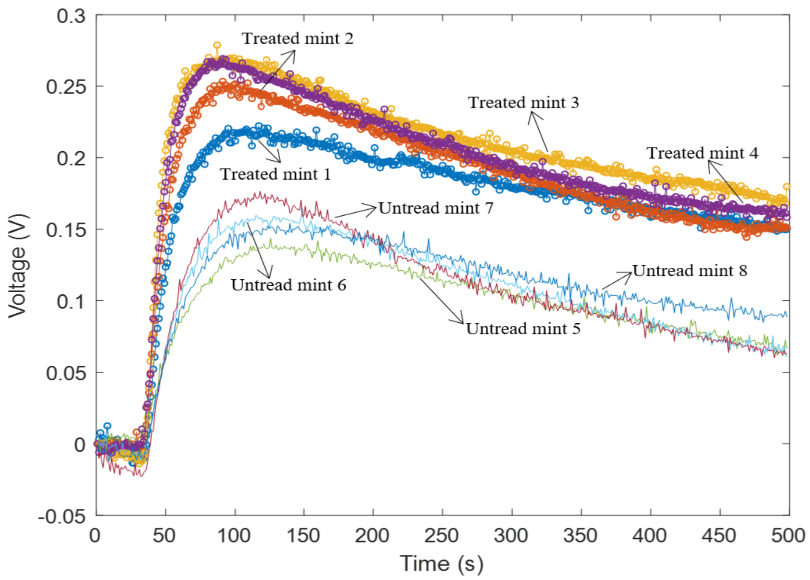


Figure 2. Example of responses collected from a sensor for one day

3.1. Discrimination of treated mint from untreated one based on classification algorithms

In this study, the classification learner app from statistics and machine learning toolbox in MATLAB 2019a was used. It is a toolbox that provides functions and applications for describing, analysing, and modelling data. To evaluate the model, a 5-folds cross-validation method was applied to avoid overfitting and achieve classification accuracy. Figure 3 shows the result obtained for four methods.

In the decision tree results in Figure 3(a), a low success rate of 87.1% was obtained, nine misclassified scores, five from the untreated mint and four from the treated one. For the naive Bayes results in Figure 3(b), the success rate achieved is 90% with six scores of the untreated mint and one of the treated one are misclassified. For support vector machines as shown in Figure 3(c), it arrived at a success rate of 91.4% with four scores of untreated mint, and two of treated are misclassified. Finally, for the ensemble classifier as shown in Figure 3(d), it arrived at the best results by achieving a success rate of 92.9% and just five misclassified scores two untreated and three treated. Table 2 summarizes the results obtained.

These results illustrated that the ensemble classifier provides better performance against decision tree, naive Bayes, and support vector machines in the discrimination of mint kind. It should be noted that in our first study on mint with the same insecticide but just in the first four days, the said study was crowned by the first articles [30], [31], a hybrid method with the use of scores from the statistical method of principal component analysis in the SVM method was used and the result was a success rate of 87.50% and the results of this current study are very better than the first one.

3.2. Treatment day prediction based on regression algorithms

In the prediction case, only the dataset of the mint treated according to the sampling day was used, it is composed of 9 columns (3 characteristics*3 sensors) and 35 rows (5 samples taken*7 days). The prediction requires the use of regression algorithms, PLSR and SVMR are very well known and widely used methods. For the PLSR, the hold-out validation was used, the dataset has been subdivided into: 80% for the training, 20% for the test, the three-fist components were selected. In the case of the SVMR, the choice was to use the 5-folds cross-validation.

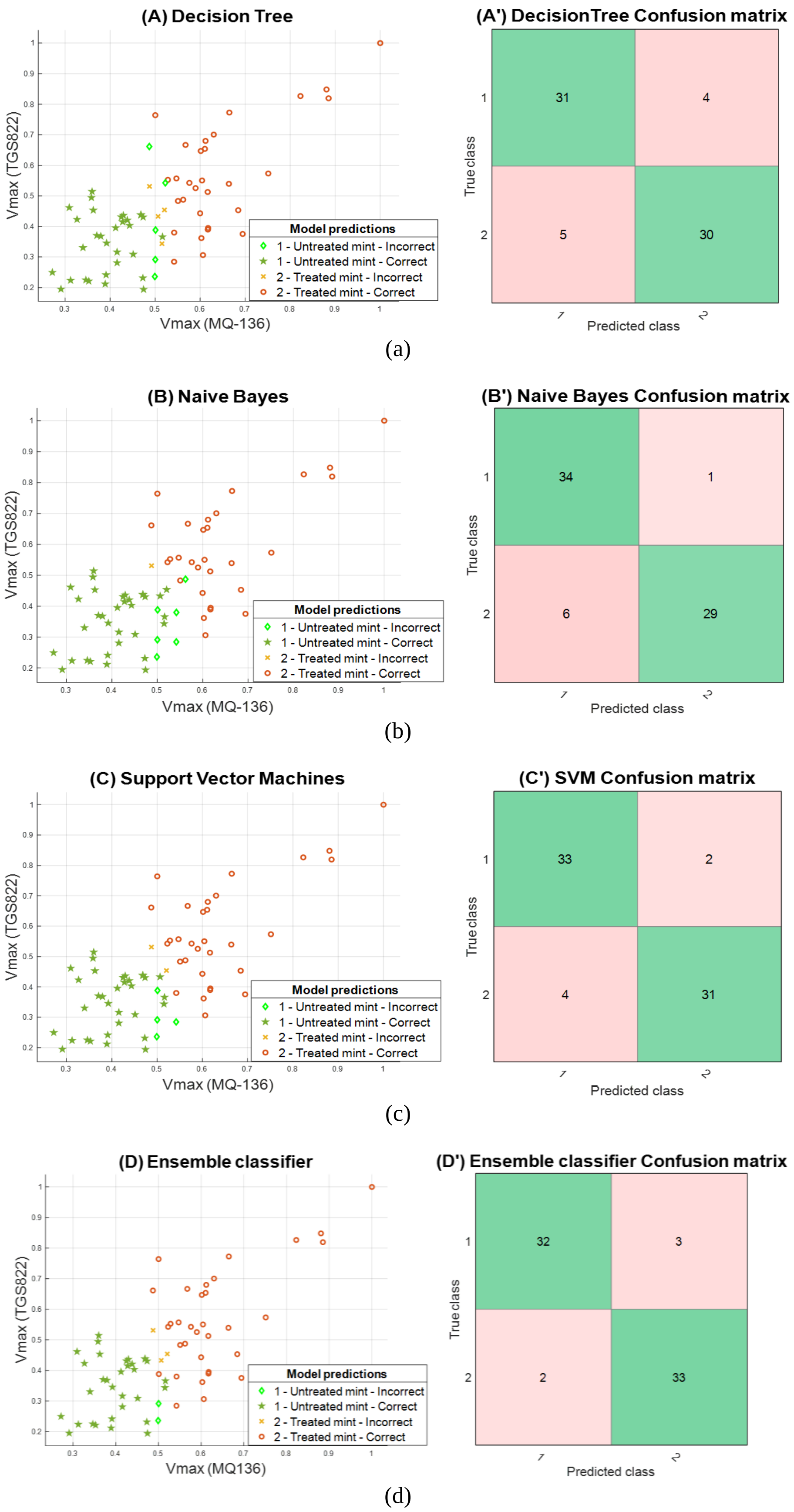


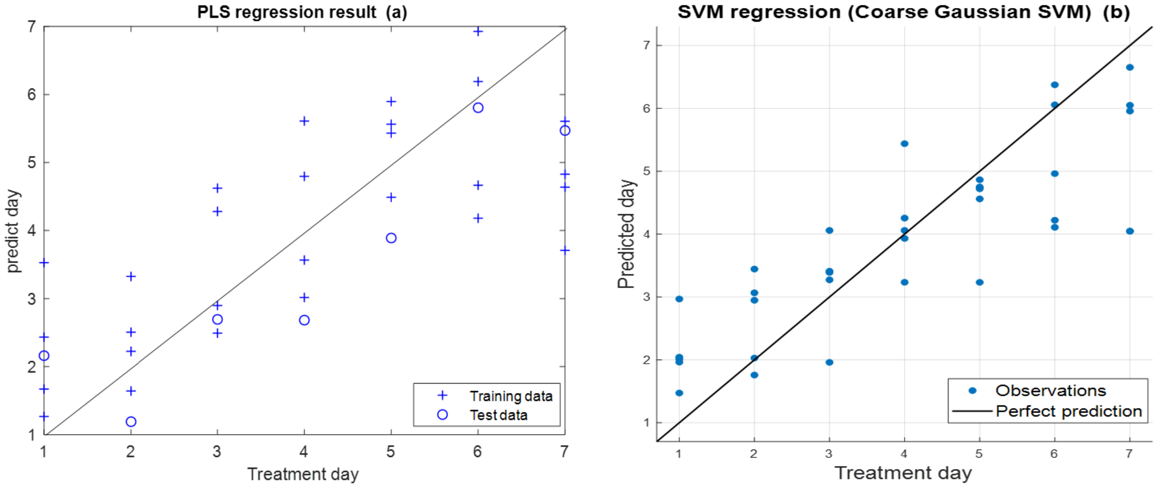
Figure 3. Different classification results: (a) A and A' DT results, (b) B and B' naive Bayes results, (c) C and C' SVM results, and (d) D and D' ensemble classifier results

Table 2. Methods and their success rate

Method	Success rate
Decision tree	87.1%
Naïve Bayes	90%
Support vector machines	91.4%
Ensemble classifier	92.9%

Figures 4(a) and 4(b) and Table 3 illustrates the results obtained by said algorithms. According to the figure and the table, PLSR did not succeed in predicting the days with great precision, the correlation coefficient is small, it is 0.67 with a big MSE greater than 2.0958. Whereas the SVMR attained good precision, the correlation coefficient is 0.82 higher than the correlation coefficient of PLSR with a mean square error (MSE) of 1.3615 much lower than that of PLSR.

These results show the superiority of SVMR in predicting the day of mint treatment with malathion. To summarize this rich study, a multisensor system carried out with commercial gas sensors achieved good results in discriminating mint treated with malathion from untreated one using the ensemble classifier algorithms, and for the treatment day prediction, the SVMR is best suited given its results.



(a) (b)

Figure 4. The results of the regression (a) using PLSR algorithm and (b) using SVMR algorithm

Table 3. Methods and their correlation coefficient with the MSE

Method	Correlation coefficient (R)	Mean square error (MSE)
Patrial least squares regression (PLSR)	0.67	2.0958
Support vector machines regression (SVMR)	0.82	1.3615

4. CONCLUSION

Chronic diseases and poisoning cases associated with the consumption of food contaminated with insecticides are constantly increasing, which requires vigilant monitoring. In the present paper, to analyse the mint headspace treated with the malathion insecticide, a multi-sensor system containing three commercial gas sensors has been manufactured and utilized. Firstly, four artificial intelligence (AI)-based classification methods namely DT, naive bayes, SVM, and ensemble classifier were inspected to distinguish the mint treated with malathion from the untreated one. The results demonstrated that the ensemble classifier reached the greatest result with a high success rate of 92.9% compared to the others. secondly, two regression methods were compared namely SVM regression and PLS regression for the treatment day prediction. Numerical results show that SVM regression provides a better correlation with a coefficient of about 0.82. The results of this study show that is possible to discriminate between the mint kinds (treated or not) and predicted the mint treatment day with good precision if the right machine learning algorithm was well adopted using just a simple, portable, and inexpensive multi-sensors system prototype designed in the laboratory.

REFERENCES

[1] P. Vazquez-Roig and Y. Pico, "Gas chromatography and mass spectroscopy techniques for the detection of chemical contaminants and residues in foods," in *Chemical Contaminants and Residues in Food*, Elsevier, 2012, pp. 17–61.

[2] S. M. van Ruth and J. P. Roozen, "Gas chromatography-olfactometry analysis and its importance in food quality control," in *Advances in Experimental Medicine and Biology*, Springer US, 2004, pp. 155–165.

[3] W. Vautz, D. Zimmermann, M. Hartmann, J. I. Baumbach, J. Nolte, and J. Jung, "Ion mobility spectrometry for food quality and safety," *Food Additives and Contaminants*, vol. 23, no. 11, pp. 1064–1073, Nov. 2006, doi: 10.1080/02652030600889590.

[4] J. Brezmes *et al.*, "Evaluation of an electronic nose to assess fruit ripeness," *IEEE Sensors Journal*, vol. 5, no. 1, pp. 97–108, Feb. 2005, doi: 10.1109/JSEN.2004.837495.

[5] W. Lu, W. Yu, C. Gan, Q. Liu, and J. Li, "Application of electronic nose technology in the detection of wheat quality," in *2015 International Conference on Intelligent Transportation, Big Data and Smart City*, Dec. 2015, pp. 133–136, doi: 10.1109/ICITBS.2015.39.

[6] Z. Haddi *et al.*, "Instrumental assessment of red meat origins and their storage time using electronic sensing systems," *Analytical Methods*, vol. 7, no. 12, pp. 5193–5203, 2015, doi: 10.1039/C5AY00572H.

[7] N. El Barbri *et al.*, "Selectivity enhancement in multisensor systems using flow modulation techniques," *Sensors*, vol. 8, no. 11, pp. 7369–7379, Nov. 2008, doi: 10.3390/s8117369.

[8] X. Tian, J. Wang, Z. Ma, M. Li, and Z. Wei, "Combination of an E-nose and an E-tongue for adulteration detection of minced mutton mixed with pork," *Journal of Food Quality*, vol. 2019, pp. 1–10, Apr. 2019, doi: 10.1155/2019/4342509.

[9] F. Han, X. Huang, J. H. Aheto, D. Zhang, and F. Feng, "Detection of beef adulterated with pork using a low-cost electronic nose based on colorimetric sensors," *Foods*, vol. 9, no. 2, Feb. 2020, doi: 10.3390/foods9020193.

[10] A. Singh, N. Thakur, and A. Sharma, "A review of supervised machine learning algorithms," in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, 2016, pp. 1310–1315.

[11] M. Khanum, T. Mahboob, W. Imtiaz, H. A. Ghafoor, and R. Sehar, "A survey on unsupervised machine learning algorithms for automation, classification and maintenance," *International Journal of Computer Applications*, vol. 119, no. 13, pp. 34–39, Jun. 2015, doi: 10.5120/21131-4058.

[12] G. Huang, S. Song, J. N. D. Gupta, and C. Wu, "Semi-supervised and unsupervised extreme learning machines," *IEEE Transactions on Cybernetics*, vol. 44, no. 12, pp. 2405–2417, Dec. 2014, doi: 10.1109/TCYB.2014.2307349.

[13] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: a review of classification and combining techniques," *Artificial Intelligence Review*, vol. 26, no. 3, pp. 159–190, Nov. 2006, doi: 10.1007/s10462-007-9052-3.

[14] A. Dasgupta, Y. V Sun, I. R. König, J. E. Bailey-Wilson, and J. D. Malley, "Brief review of regression-based and machine learning methods in genetic epidemiology: the genetic analysis workshop 17 experience," *Genetic Epidemiology*, vol. 35, no. S1, pp. S5–S11, 2011, doi: 10.1002/gepi.20642.

[15] L. Xu, L. Wencong, J. Shengli, L. Yawei, and C. Nianyi, "Support vector regression applied to materials optimization of sialon ceramics," *Chemometrics and Intelligent Laboratory Systems*, vol. 82, no. 1–2, pp. 8–14, May 2006, doi: 10.1016/j.chemolab.2005.08.011.

[16] W. Zeng, D. Zhang, Y. Fang, J. Wu, and J. Huang, "Comparison of partial least square regression, support vector machine, and deep-learning techniques for estimating soil salinity from hyperspectral data," *Journal of Applied Remote Sensing*, vol. 12, no. 2, Jan. 2018, doi: 10.1117/1.JRS.12.022204.

[17] Q. Zhao, S. Yu, F. Zhao, L. Tian, and Z. Zhao, "Comparison of machine learning algorithms for forest parameter estimations and application for forest quality assessments," *Forest Ecology and Management*, vol. 434, pp. 224–234, Feb. 2019, doi: 10.1016/j.foreco.2018.12.019.

[18] R. Laref, E. Losson, A. Sava, K. Adjallah, and M. Siadat, "A comparison between SVM and PLS for E-nose based gas concentration monitoring," in *2018 IEEE International Conference on Industrial Technology (ICIT)*, Feb. 2018, pp. 1335–1339, doi: 10.1109/ICIT.2018.8352372.

[19] L. Xu, J. He, S. Duan, X. Wu, and Q. Wang, "Comparison of machine learning algorithms for concentration detection and prediction of formaldehyde based on electronic nose," *Sensor Review*, vol. 36, no. 2, pp. 207–216, Mar. 2016, doi: 10.1108/SR-07-2015-0104.

[20] W. N. Aldridge, J. W. Miles, D. L. Mount, and R. D. Verschoyle, "The toxicological properties of impurities in malathion," *Archives of Toxicology*, vol. 42, no. 2, pp. 95–106, 1978, doi: 10.1007/BF00316489.

[21] P. C. Jurs, G. A. Bakken, and H. E. McClelland, "Computational methods for the analysis of chemical sensor array data from volatile analytes," *Chemical Reviews*, vol. 100, no. 7, pp. 2649–2678, Jul. 2000, doi: 10.1021/cr9800964.

[22] J. H. Cho and P. U. Kurup, "Decision tree approach for classification and dimensionality reduction of electronic nose data," *Sensors and Actuators B: Chemical*, vol. 160, no. 1, pp. 542–548, Dec. 2011, doi: 10.1016/j.snb.2011.08.027.

[23] D. R. Wijaya, R. Sarno, and A. F. Daiva, "Electronic nose for classifying beef and pork using Naïve Bayes," in *2017 International Seminar on Sensors, Instrumentation, Measurement and Metrology (ISSIMM)*, Aug. 2017, pp. 104–108, doi: 10.1109/ISSIMM.2017.8124272.

[24] A. Mathur and G. M. Foody, "Multiclass and binary SVM classification: implications for training and classification users," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 2, pp. 241–245, Apr. 2008, doi: 10.1109/LGRS.2008.915597.

[25] Y. Yang, "Ensemble learning," in *Temporal Data Mining Via Unsupervised Ensemble Learning*, Elsevier, 2017, pp. 35–56.

[26] H. Abdi, "Partial least squares regression and projection on latent structure regression (PLS Regression)," *WIREs Computational Statistics*, vol. 2, no. 1, pp. 97–106, Jan. 2010, doi: 10.1002/wics.51.

[27] A. Sharifi, "Partial least squares-regression (PLS-regression) in chemometrics," in *1th National Conference on Achievements in Chemistry and Chemical Engineering*, 2016, pp. 305–308.

[28] R. Rodríguez-Pérez, M. Vogt, and J. Bajorath, "Support vector machine classification and regression prioritize different structural features for binary compound activity and potency value prediction," *ACS Omega*, vol. 2, no. 10, pp. 6371–6379, Oct. 2017, doi: 10.1021/acsomega.7b01079.

[29] A. Eghbalzadeh, M. Javan, M. Hayati, and A. Amini, "Discharge prediction of circular and rectangular side orifices using artificial neural networks," *KSCE Journal of Civil Engineering*, vol. 20, no. 2, pp. 990–996, Mar. 2016, doi: 10.1007/s12205-015-0440-y.

[30] 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, Feb. 2021, doi: 10.1007/s11694-020-00617-8.

[31] A. Amkor, K. Maaider, and N. El Barbri, "Mint treatment day prediction using a multi-sensors system and machine learning algorithms," *Sensors and Actuators A: Physical*, vol. 328, Sep. 2021, doi: 10.1016/j.sna.2021.112787.

BIOGRAPHIES OF AUTHORS

Ali Amkor received the Master Science and Technology degree in engineering and industrial management from the faculty of Sciences and Techniques of Settat Morocco in 2017, currently pursuing a Ph.D. degree in National School of Applied Sciences Khouribga, a subsidiary of Sultan Moulay Slimane Beni Mellal University, Morocco. His research interest includes electronic, multi-sensors systems, electronic noses, data analysis, and artificial intelligence. He can be contacted at email: amkorali@mail.com.

Kamal Maaider his Ph.D. from Faculty of Science and Technology of Settat, Morocco in April 2016 in Engineering Sciences specializing in physics of matter and electrical modeling. He is an Assistant Professor in the Electrical Engineering Department at the National School of Applied Sciences of Khouribga, Morocco since 2018. His research interests are electronic engineering electrical engineering, electrical components, machine learning, artificial intelligence. He can be contacted at email: maaider2@yahoo.fr.

Noureddine El Barbri 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.