Computers and Electronics in Agriculture 163 (2019) 104858

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Peach growth cycle monitoring using an electronic nose

Henike Guilherme Jordan Voss[a], Sergio Luiz Stevan Jr.[a,b,*], Ricardo Antonio Ayub[c]

[a] Graduate Program in Applied Computing (PPGCA), State University of Ponta Grossa (UEPG), Ponta Grossa (PR) 84030-900, Brazil
[b] Graduate Program in Electrical Engineering (PPGEE), Federal University of Technology of Parana (UTFPR), Ponta Grossa (PR) 84016-210, Brazil
[c] Graduate Program in Agronomy (PPG), State University of Ponta Grossa (UEPG), Ponta Grossa (PR) 84016-210, Brazil

ARTICLE INFO

Keywords:
Electronic nose (E-nose)
Volatile Organic Compounds (VOCs)
Peach
Flavor
Maturation

ABSTRACT

In regions with the predominance of agriculture, an inspection of the quality and fruit maturity index in the orchard is usually analyzed by the farmer's experience, which can be subject to errors and generate a greater cost of time and money. Thus, monitoring equipment that generates a rapid and accurate response to the growth cycle of the peaches in the crop is desirable, together with a low marketing cost. For this purpose, electronic noses prove to be the most suitable equipment, since it allows online monitoring of the VOCs (Volatile Organic Compounds) generated by the crop. In this context, a prototype was developed to perform the classification of the fruit growth cycle (pre-harvest and post-harvest). Models with the 13 gas sensors made with a metal oxide semiconductor (MOS) and the reduction to 7 sensors were studied with the aid of the Pearson's Chi-square test, for comparison. Samples with 4 growth stages were used for the training and construction of the model. The accuracy of 99.23% in the validation step and 98.08% in the sample test step using the Random Forest method with linear discriminant analysis for the reduced data set for 7 sensors shows that the device is promising for monitoring of areas with an intense emission of VOCs.

1. Introduction

Fruits produce different volatile organic compounds (VOCs) and their quality measured by aroma, taste and color changes constantly throughout the growth and maturation phase, which occurs in the pre-harvest period (Baietto and Wilson, 2015). When measuring these properties, some instrumental methods used are manual and destructive. Non-destructive measurement of internal fruit quality is becoming important for industry and for consumers (Rajkumar and Wang, 2012). The development of sensor technology allowed the electronic noses to be presented as simple devices with high detection accuracy, and these devices are increasingly being used as an alternative to traditional methods (Wu et al., 2017; Zhu et al., 2017; Lin and Zhang, 2016; Jiang and Wang, 2016; Men et al., 2018). Moreover, electronic noses can also be applied to the monitoring of air quality (Deshmukh et al., 2015; Abraham and Pandian, 2013; Bagula et al., 2012; Peterson et al., 2017; Kim and Hwangbo, 2018; Blanco-Novoa et al., 2018; Laref et al., 2018), gases emitted by the soil (Bieganowski et al., 2018; Sudarmaji and Kitagawa, 2016; Dorji et al., 2017; Pineda and Pérez, 2017), food quality (Aleixandre et al., 2015; Wojnowski et al., 2017; Chen et al., 2018; Srivastava et al., 2019; Mishra et al., 2018), beverage quality (Aleixandre et al., 2018; Wei et al., 2017; Santos and Lozano, 2015; Ab Mutalib and Jaswir, 2013; Nurul et al., 2017; Ragazzo-Sanchez et al., 2008), in the medical field (Huang et al., 2018; Voss et al., 2012; Li et al., 2017; Lorwongtragool et al., 2014), among others (see Fig. 1).

Electronic noses are made up of an array of chemical sensors based on metal oxide materials (MOS). In the same way as human noses, these devices are not able to identify the substances separately in each sample. It may make an analogy of the data obtained by these sensors with the fingerprint since it is very difficult to find two different substances with the same pattern, that is, it is possible to classify substances according to their patterns (Pineda and Pérez, 2017). These devices consist of three parts: sensor array, signal processing unit and pattern recognition. These three parts simulate, respectively, the acquisition of information by sensory neurons of the human olfactory receptor, the encoding of the olfactory nerve (bulb), brain memory and information processing by the human olfactory system (Huang et al., 2018).

The biological mechanism is a common indication of fruit ripeness. In places where agriculture predominates the inspection of the quality and maturation of the fruits is given by the experience of the farmer, which can be subject to errors and carelessness (Chen et al., 2018). In addition, if a specific period for fruit harvesting is considered, there may be situations where the fruits are of different qualities at different points in the planted area, which is part of precision agriculture. In view of this, it is desirable to have equipment that monitors the

* Corresponding author at: Graduate Program in Applied Computing (PPGCA), State University of Ponta Grossa (UEPG), Ponta Grossa (PR) 84030-900, Brazil.
E-mail address: sstevanjr@utfpr.edu.br (S.L. Stevan).

https://doi.org/10.1016/j.compag.2019.104858
Received 23 February 2019; Received in revised form 12 June 2019; Accepted 13 June 2019

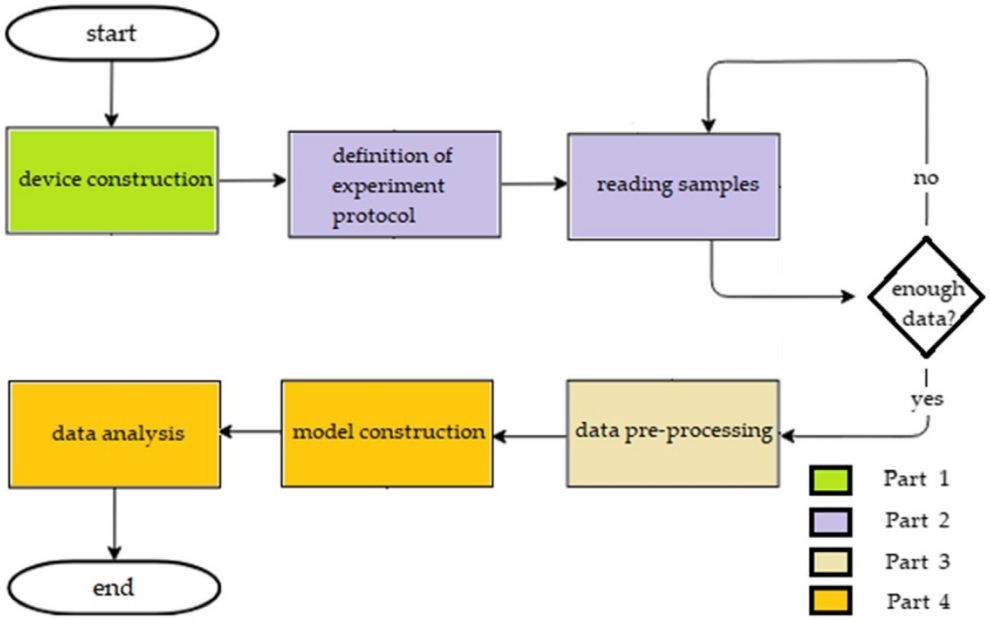

Fig. 1. An illustrative flowchart of the adopted methodology.

different degrees of maturation at the cultivation points and provided this can be identified by VOCs, the electronic noses are a robust and low-cost alternative for this purpose, besides allowing the monitoring of culture.

Based on gas chromatography and mass spectrometry (GC–MS) (Horvat et al., 1990; Baraldi et al., 1999) recognized the main VOCs exhaled by peach during and after the full bloom phase. Table 1 shows the main gases generated as by-products in this process.

Currently, the peach crop has expanded, with the planting of more than 500 thousand seedlings/year in southern Brazil (Dutra et al., 2002). Considering the physicochemical characteristics of the peach fruit growth, for this work was carried out the monitoring of the cultivar Eragil in the city of Ponta Grossa, Paraná. Thus, non-destructive equipment was used, applied near the crop and capable of detecting the gases exhaled by peach, to identify changes in the different cycles, from growth, maturation, and post-harvest.

2. State-of-the-art

There are many papers in the literature applying electronic noses to qualitatively discriminate peaches. Brezmes et al. (2000) observed the ripening process of peach fruits. Based on an arrangement of chemical tin oxide sensors and pattern recognition techniques based on neural networks, the designed olfactory system can classify fruit samples into three different maturation states. Measures made with peaches show a success rate above 93% (green, ripe and super mature). An additional feature of the system is the ability to accurately predict the number of days the fruit is stored since harvest. Measurements made with peaches show a maximum error of 1 day. Benedetti et al. (2008) used a commercial electronic nose to classify four peach cultivars and to evaluate the post-harvest maturation stage. Principal Component Analysis (PCA) and linear discriminant analysis (LDA) were used to investigate whether the electronic nose was able to distinguish between four different cultivars. The sensor responses were adjusted with a sigmoid transition function, allowing the definition of three different stages of maturation (immature, mature, super mature). The analysis of the classification and regression tree (CART) was applied to characterize the peach samples in the three classes. The decision tree analysis classified samples in each respective group with a cross-validation error rate of 4.87% and test error of 2.44%. Guohua et al. (2012) also applied an electronic nose to predict peach freshness. The device was produced with an arrangement of eight MOS gas sensors. Principal component analysis (PCA) and stochastic resonance (SR) were used for measurement data analysis. The results show that e-nose can distinguish peaches between fresh and overmature conditions. The validation of the results of the experiments demonstrates that the prediction accuracy of this model is 85%. Zhang et al. (2012) established a quality index model able to describe the different dates of the harvest of peaches. The partial least squares regressions (PLS) and principal component regression model (PCR) represented a good ability to describe the quality indices of three selected sets of peaches (green, ripe and mature). The results showed that the PLS model was able to represent good ability to predict quality indices, with R = 0.86 and SEP (Standard prediction error) = 8.77 for CF; R = 0.83 and SEP = 0.297 for SC; R = 0.83 and SEP = 0.2 for pH. The PCR model showed R = 0.84 and SEP = 7.33; R = 0.82 and SEP = 0.44; R = 0.78 and SEP = 0.21; for CF, SC and pH, respectively.

As previously shown, applications tend to perform fruit measurement after harvesting and do not evaluate fruit throughout the growing stage in the orchard. Therefore, the work seeks to build an electronic nose that monitors the entire period from post-flowering to post-harvest, with field monitoring. The main classification methodologies that will be addressed during the paper, such as PCA, LDA and Artificial Neural Networks (ANN).

Table 1
VOCs identified by GC/MS by (GC–MS) (Horvat et al., 1990; Baraldi et al., 1999) in the process of growing peaches.

Classe	Compound	Classe	Compound
Aldehydes	Hexanal		Terpinolene
	(E)-2-hexenal		Linalool
	Benzaldehyde	Lactones	γ-hexalactone
Alcohols	1-Hexanol		γ-heptalactone
	(E)-2-hexenol		γ-octalactone
	(Z)-3-hexenol		γ-nonalactone
Esters	Hexyl acetate		γ-decalactone
	(E)-2-hexenyl acetate		γ-dodecalactone
Ketones	Acetoin		δ-decalactone
Terpenoids	α-terpinene		δ-dodecalactone
	γ-terpinene		

Table 2
Sensors used in the experiment and their sensitive gases.

Sensor	Sensitive gases
MQ-2	H_2, LPG, CH_4, CO, ethanol, propane, butane, and methane
MQ-3	Ethanol, benzene, CH_4, hexane, LPG and CO
MQ-4	LPG, CH_4, H_2, CO and ethanol
MQ-5	H_2, LPG, CH_4, CO, ethanol, *iso*-butane, and propane
MQ-6	LPG, H_2, CH_4, CO, ethanol, *iso*-butane, and propane
MQ-7	CO, H_2, LPG, CH_4 and ethanol
MQ-8	H_2, LPG, CH_4, CO and ethanol
MQ-9	CO, CH_4, and LPG
MQ-135	NH_3, benzene, ethanol, CO_2, CO, andNH_4
TGS822	Acetone, n-hexane, benzene, ethanol, *iso*-butane, CO and methane
TGS2600	H_2, CO, methane, *iso*-butane, and ethanol
TGS2602	H_2, NH_3, ethanol, H_2S and toluene
TGS2603	H_2, H_2S, ethanol, methyl mercaptan and trimethylamine

3. Materials and methods

The methodology performed in this work is basically divided into four stages: equipment construction, execution of the experiment protocol, pre-processing and data analysis.

3.1. Construction of equipment

This part is subdivided into the following ones: the selection of the sensors to be used in the sensor array, hardware construction, signals preprocessing, database construction that will serve as input for the pattern recognition through the algorithms of machine learning.

In total, 13 MOS sensors were used in electronic nose device; 9 MQ family sensors manufactured by Hanwei Electronics Co., Ltd (MQ-2, MQ-3, MQ-4, MQ-5, MQ-6, MQ-7, MQ-8, MQ-9, MQ-135) and 4 sensors manufactured by TGS Figaro (TGS 822, TGS 2600, TGS2602, TGS 2603). The chemical reactions between the surface molecules of sensors and gases provide a change in the sensor response as the change in ambient gas concentration. Therefore, the resistivity of the sensors increases in the presence of air and decreased in the presence of sensitive gases. Table 2 shows the most sensitive sensors and gases thereof according to the manufacturers' datasheets Figaro® and Hanwei®.

Many sensors have been selected, which will create a fingerprint for each sample. Since each sensor responds in a way and the set of responses will generate a response almost unique to each type of gas, coupled with the fact that each sensor works with one set of gases different from the others. Therefore, the goal is to scan all possibilities, identify the most important sensors and if necessary remove the redundant sensors.

Arduino Mega 2560 platform, which has an ATMega 2560 microcontroller, is used in the construction of the system hardware. A 16-channel Cd74hc4067 analog multiplexer was also used. Although the microcontroller has a number of analog inputs that would be sufficient for the number of sensors used in the study, it is necessary to consider that the device is designed to be used in various types of platforms and microcontrollers. In addition, parallel and continuous reading of all sensors provides a total current consumption by the ATMega 2560 above that tolerated and therefore the multiplexer is shown as a suitable alternative as it reduces the number of current-consuming pins to only one, through channel selection. Therefore, the use for only one analog port of the microcontroller has been reduced. In addition, the pressure and temperature sensor BMP180 and the humidity sensor HIH-4030 were also used. The entire system is powered by an automotive battery with a nominal capacity of 40Ah. Fig. 2 shows the equipment developed and Fig. 3 shows the block diagram of the circuit.

The developed equipment is illustrated in Fig. 2. This one, by his turn, has a fan that has the function of ambient air circulation. The air will be drawn into the device in the sensor array and the side holes allow the air outflow of the device. This constant flow allows a smoother response by the gas sensors. It is worth mentioning that the gas sensors require an initial preheating period (about ten minutes) until they reach an ideal constant operating temperature. Therefore, the stabilization of the response of the sensors will be reached after this time and these transient data are discarded from the analysis stage of the responses.

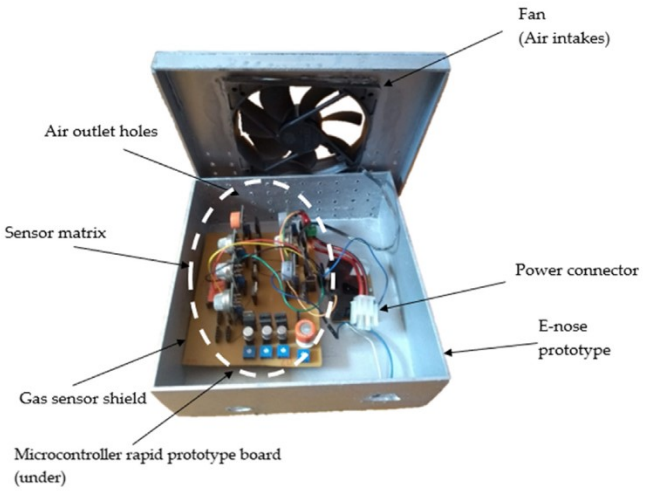

Fig. 2. Equipment developed.

Fig. 3 illustrates the block diagram of the electronic nose circuit.

As illustrated in Fig. 2, the equipment, with the dimensions (20 cm × 20 cm × 7 cm), has a fan for the airflow that will allow the best contact of the sample to be measured with the 13 gas sensors, the sensor temperature and pressure, and the air humidity sensor. In addition, the USB cable allows data transmission to the computer.

3.2. Experimental protocol

The experiments were carried out in the city of Ponta Grossa in the state of Paraná, which has a temperate climate with mild summer, where it presents significant rainfall in all months of the year and mild temperatures in the warmer months, located at latitude 25°05′49″S and longitude of 50°03′11″W.

The measures comprised the three periods of fruit growth and a fourth post-harvest period of the fruit. Considering the periods described previously, since the flowering occurred in late August/early September, stage I comprises until the day of October 05, 2018. Stage II occurred close to October 26, 2018. Stage III includes until the date the fruit was harvested, on November 17, 2018. After that, stage IV occurred until the date of the last experiment (12/13/2018). Table 3 summarizes each stage of growth in addition to its days based on full bloom.

The experiments were performed on 08/23/2018 (23 °C average), 09/06/2018 (21 °C average), 09/21/2018 (24 °C average) – stage I; 10/05/2018 (25 °C average), 10/11/2018 (17 °C average), 10/19/2018 (24 °C average) – stage II; 10/26/2018 (21 °C average), 11/08/2018 (19 °C average), 11/15/2018 (27 °C average) – stage III, 11/30/2018 (27 °C average), 12/07/2018 (24 °C average) and 12/13/2018 (32 °C average) – stage IV. All experiments were performed at the same time, starting at 3:00 p.m.

The equipment was positioned in the same position for all experiments, at the base of a peach tree in the middle of the orchard, as shown in Fig. 4. Therefore, the VOCs exhaled by fruits still intact in the tree were detected directly by the internal sensors to an equipment installed near the trees, which in turn collected the air (with the volatile compounds of the fruits) to into the box through a fan. This box has small air vents that allow constant air exchange and also keeps the internal pressure constant.

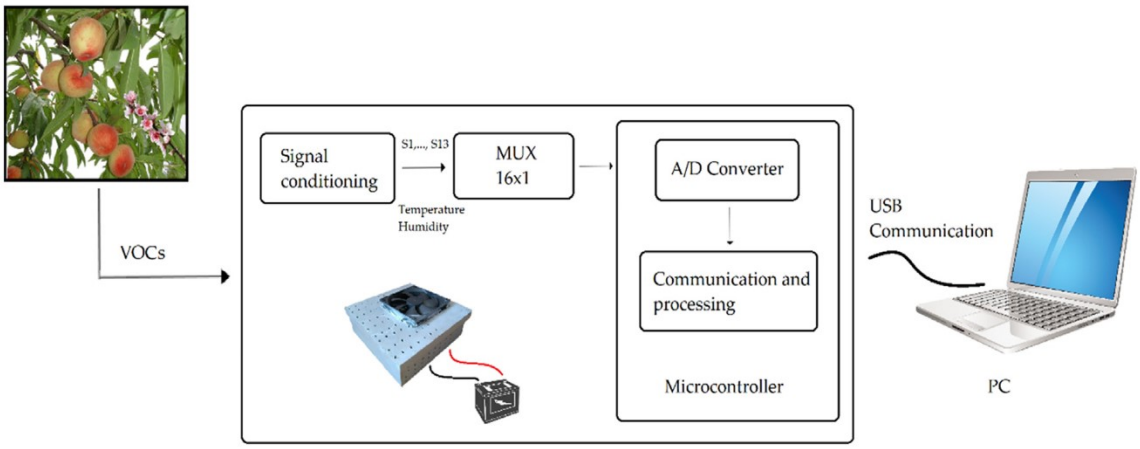

Fig. 3. Circuit block diagram.

Table 3
Duration of each stage of peach growth according to (Keske, 2009; Cunha Junior et al., 2007; Lilien-Kipnis and Lavee, 1971; Barbosa et al., 1993; Barbosa, 1990; Couto, 2006).

Maturation	Days after full bloom
Stage 1	0–35
Stage 2	36–49
Stage 3	51–70
Stage 4	71–100

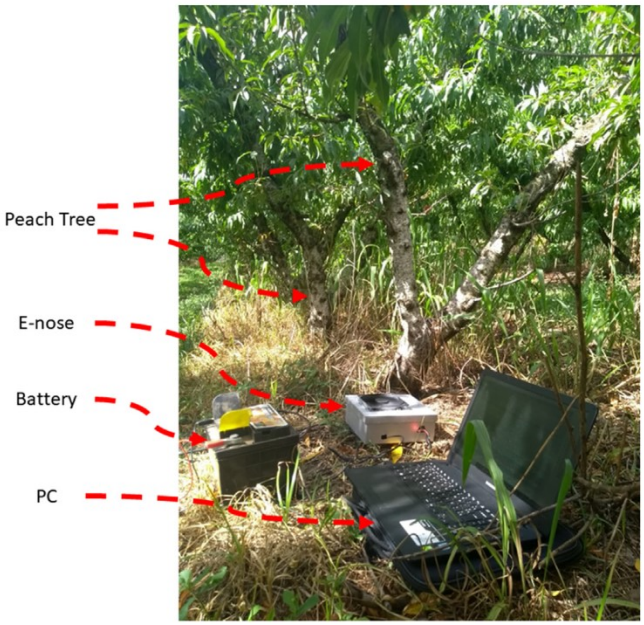

Fig. 4. Equipment mounted on cultivar.

Each day involved measuring 15 samples for 2 min. As the preheating and transient state (about 10 min) were disregarded, it took 40 min of data reading. As measurements were taken every second, a total of 1800 records were obtained for each day. For each sample, the arithmetic mean of the responses of each of the sensors was performed, and therefore these responses were used as inputs in the algorithms. Then, the database comprises a total of 180 (15 samples × 12 days) records or samples. The algorithm model, therefore, was constructed based on this database.

At the training stage, 70% of the records were used (128 samples), and for the test step, the other 52 samples were used. It is worth mentioning that the k-fold applied with 10 partitions performs the validation process only with the samples used for training, that is, for each of the 10 iterations 13 of the 128 registers were tested.

3.3. Pre-processing

A modified moving average filter was used in the preprocessing step of the sensor signals. This filter was used because it is the simplest, fastest, robust and easy to implement. A simple moving average over n elements consists of the unweighted averages of the subsets of n elements in a data set. The original value of n was 50. The change made was that only the values of the mean with a maximum difference of three times the standard deviation is entered the moving average equation, i.e.: $|p_i - \mu| \leq 3\sigma$. Therefore, the value of n in Eq. (1) will be: $1 \leq n \leq 50$, and within the normal distribution curve, 99.74% of the data will be around the mean plus 3 standard deviations. This change provides more smoothing in the data set, removing possible signal peaks due to the noise coming from the circuit.

In addition, the normalization step (p_{inorm}), which is given by Eq. (1):

$$p_{inorm} = \frac{p_i - min(p)}{max(p) - min(p)} \quad (1)$$

where min is the minimum value and max is the maximum value of the set p.

The gas sensors vary depending on the temperature and the relative humidity of the air. The manufacturers' datasheets show some characteristic curves of the sensor behaviors according to changing these two variables. These curves use the resistance of the sensors at ambient temperature and clean air (R_o) and in varying conditions (R_s) as the basis. The load resistance (R_L) changes the sensitivity of the sensors, and for the study here, all were set to 10 kΩ. Based on the curves data reported by the manufacturers, the response compensation was performed as a function where the input parameters were the measured temperature and the relative humidity of the air, and the output of the function is the compensated resistance.

To reduce the number of sensors, Pearson's Chi-square test was applied to verify the association between each of the input variables (gas sensors) and the output variable (maturation stage). This test outputs the Cramer V coefficient, which is a measure of association between two nominal variables, giving a value between 0 (no association between variables) and 1 (complete association). For this study, it was adopted that if the variable has the coefficient greater than 0.75 it will be maintained, and if it is smaller 0.50 its will be removed. Variables with values ranging from 0.50 to 0.75 went through the exhaustive search method before dimensionality reduction and

classification steps to confirm if the presence of each of them would influence the variance in the reduction components of PCA and LDA, and therefore whether they were kept in the data analysis. Thus, the search space, through the exhaustive search, will be considerably reduced, since only a part of the sensors will be considered. Therefore, in this study, we studied 2 behaviors in the responses of the sensors: the data with the 13 sensors and the data with the reduced sensors.

3.4. Data analysis

In the implementation of the data mining algorithms, the data of the transient regime, which is the period necessary to stabilize the data of the sensors after the sample is placed for the reading, were disregarded. This transient regime period takes about 10 min.

Before data classification, the PCA and LDA were used to identify possible clusters and to data reduction, which are two of the main methods present in the literature. After that, we used four supervised machine classification algorithms: K-nearest neighbors KNN which is the simplest and most intuitive classification algorithm (Mucherino et al., 2009), Support Vector Machine (SVM) that has the advantages of generating non-linear boundaries and the use of state space representation (Ruiz et al., 2018), Random Forest (RF) that has received increasing interest because of its precise prediction and noise robustness compared to single classifiers, in addition to being computationally faster than other set methods of trees (Li et al., 2017) and the artificial neural network with a simple hidden layer called Extreme Machine Learning (ELM). These algorithms were selected because as in (Chen et al., 2018; Brezmes et al., 2000) good results were reached in the fruit maturation classification.

SVM was used with the linear kernel. Since the work involves a multiclass problem, the “one-against-one” approach was used. For this approach, SVM classifiers have all combinations of class pairs. Therefore, for k classes, $k(k-1)/2$ binary classifiers are trained. The output of each classifier in the form of a label is obtained by a voting scheme in which the most frequent label is assigned. In the case of a tie, a tiebreaker strategy is adopted. The tie-breaking strategy adopted was to select a class randomly.

RF is not necessary to perform the pruning method of some nodes to avoid overfitting. Some recent papers show that the use of this technique in e-noses can improve the accuracy of regression/classification (Li et al., 2017; Men et al., 2018). The operation scheme of a Random Forest algorithm is illustrated in Fig. 5. This algorithm is a set of several decision trees, in which each of these trees performs the classification from a subset of randomly selected attributes. Then, the final classification is obtained by means of the majority voting method, in which the class that obtained the highest number of classifications by the trees will be selected.

The neural network ELM presents better training performance compared to traditional algorithms, besides facilitating the construction of the online model (Huang et al., 2012). Unlike traditional learning algorithms, the ELM not only tends to achieve the smallest training error but also lower weights. The only parameters to be adjusted are the hidden layer size and the learning function. Fig. 6 shows the structure of a feedforward neural network of the ELM type, in which it is composed of three layers (input, concealment, and output). After the adjusted weights of the input layer, the signals of the hidden layer are passed by the activation functions, which in turn are combined linearly to the output layer.

The step of constructing the model involves doing the classification algorithm and then train the algorithm with the data collected from the database. Then, an ideal model (classifier) will be constructed that will be responsible for predicting the new data inserted in the database.

To validate the results obtained with the classifiers, the k-fold cross-validation method was used. Cross-validation k-fold is a commonly used technique that takes a set of m examples and partitions them into K sets of equal sizes (folds) of size m/K. For each set, a classifier is trained in the other sets (Langford, 2005).

All algorithms were implemented using R and Java languages, which have a high range of libraries and functions that help in the process of constructing the models. For the data storage was used the relational database PostgreSQL

4. Results and discussion

Fig. 7(a) illustrates typical sensor responses for an experiment measured on the peach-tree in stage 1. Remembering that the first 600 s comprise the preheating time and the transient state period and were removed from the data analysis. The figure shows the response obtained by the first set of data.

Fig. 7(b)–(d) illustrates the typical responses of the sensors to an experiment measured on the peach in stage 2, stage 3, and stage 4, respectively. In the same way, the initial 600 s comprise the preheating time and the transient state period.

It is can see that from the obtained data that the initial stages allow a previous stabilization of the sensors. The higher concentrations of gases in the later stages of maturation cause the sensors to take a longer

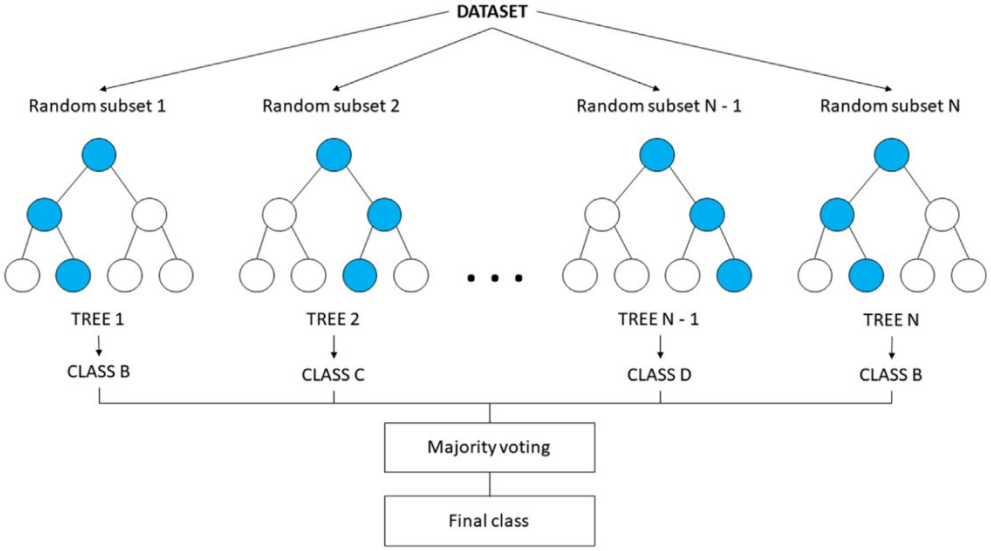

Fig. 5. Random forest ensemble method operation.

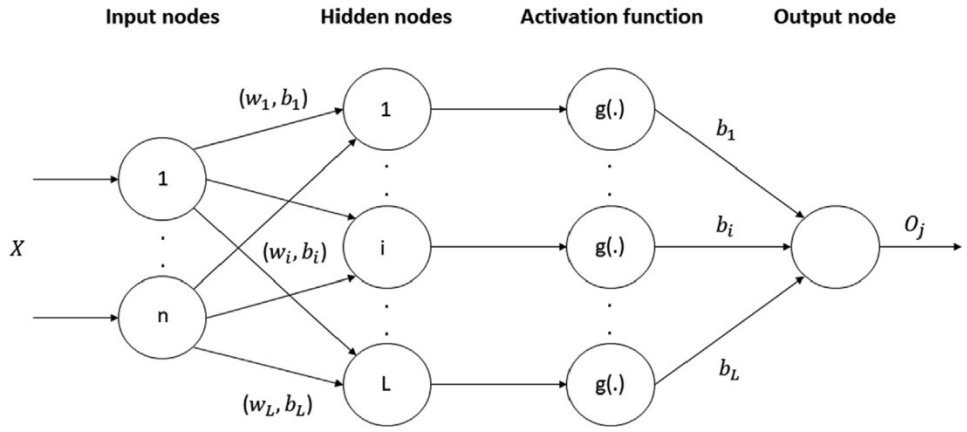

Fig. 6. Structure of an ELM neural network.

time for stabilization. Fig. 7 shows the difference in response of the sensors during these four phases. In addition, it is noted that the MQ-5 sensor obtained a higher response in the three initial stages in relation to the final stage. This is probably because this sensor has a higher sensitivity for the classes of VOCs detected in the initial stages, such as alcohols, aldehydes, and terpenoids.

To reduce the number of sensors, the Chi-square test showed that the variables TGS2603 and MQ-5 obtained a Cramer V coefficient greater than 0.75, and therefore considered important for the model. The variables MQ-6 and MQ-9 were removed from the model construction since they obtained a coefficient lower than the threshold of 0.50. For the remaining sensors, through the exhaustive search before step of dimensionality reduction and classification, it was found that the MQ-3, MQ-4, TGS822, TGS2600 and TGS2602 attributes resulted in a higher variance in the first 3 principal components compared to MQ-2, MQ-7, MQ-8, and MQ-135, and therefore were maintained in the reduced analysis model. Thus, in the reduced analysis model, 7 sensors (MQ-3, MQ-4, MQ-5, TGS822, TGS2600, TGS2602 and TGS2603) were used and the complete analysis model used the 13 initial sensors. Table 4 shows the values of Cramer V coefficient obtained by each sensor.

From the data obtained, the first step was to apply the data reduction methods. Since we have responses for 13 sensors, a three-dimensional reduction was performed using the PCA. As shown in Fig. 8(a), the three major components have a total variance of 84.1% of the original set. It can be verified that the post-harvest stage (orange) obtained greater discrimination compared to the other phases. The same procedure for three dimensions was performed with the LDA method.

Clearly, the second method was able to discriminate better, compared to PCA. The total variance obtained with LD1, LD2, and LD3 was 100%, that is, with only three variables it is possible to represent the set of data obtained by all the gas sensors.

Table 4
Results obtained through the Chi-square test.

Sensor	Cramer V coefficient
MQ-2	0.6905
MQ-3	0.6171
MQ-4	0.6369
MQ-5	0.7794
MQ-6	0.0000
MQ-7	0.6982
MQ-8	0.7324
MQ-9	0.4584
MQ-135	0.5297
TGS822	0.5340
TGS2600	0.7001
TGS2602	0.6163
TGS2603	0.8653

The same procedure was performed for the reduced input set. As shown in Fig. 9(a), the three principal components have a total variance of 93.31% of the original set. Compared to Fig. 8(a), there is a greater variance and discrimination of maturation stages.

It can also be noted that the LDA compared to PCA has been able to cluster more clearly with the classes, and this is proven with the variance in LD1, LD2, and LD3 of 99.99%. Compared to the original dataset there was also a slight improvement in the performance of the LDA algorithm since the LD1 and LD2 are responsible for 97.4% of the data variability against 95.07% of the total set. For the 7 sensors, 2 components of LDA would be sufficient, but to compare and standardize it was adopted to use 3 components for all cases.

Then, the supervised classification was performed with the four algorithms (KNN, SVM, RF, and ELM). As input from these methods, the

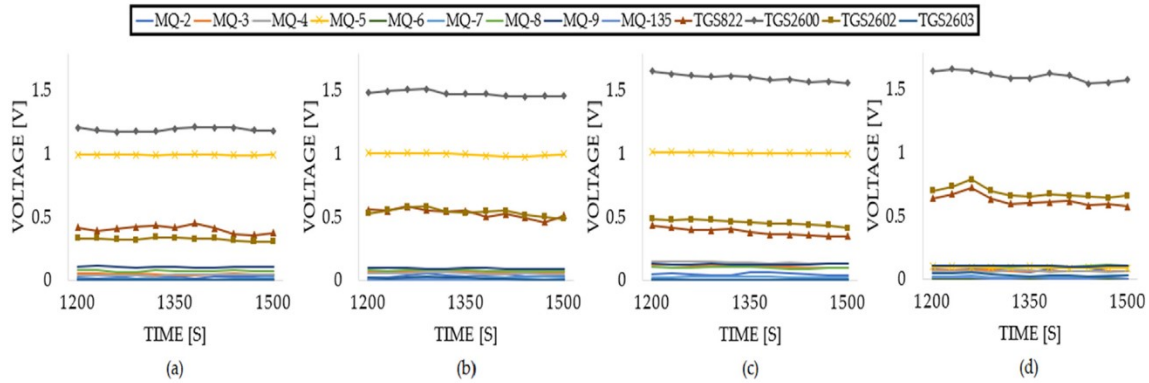

Fig. 7. The typical response was given by the sensors: (a) samples of stage 1, (b) samples of stage 2, (c) samples of stage 3 and (d) samples of stage 4.

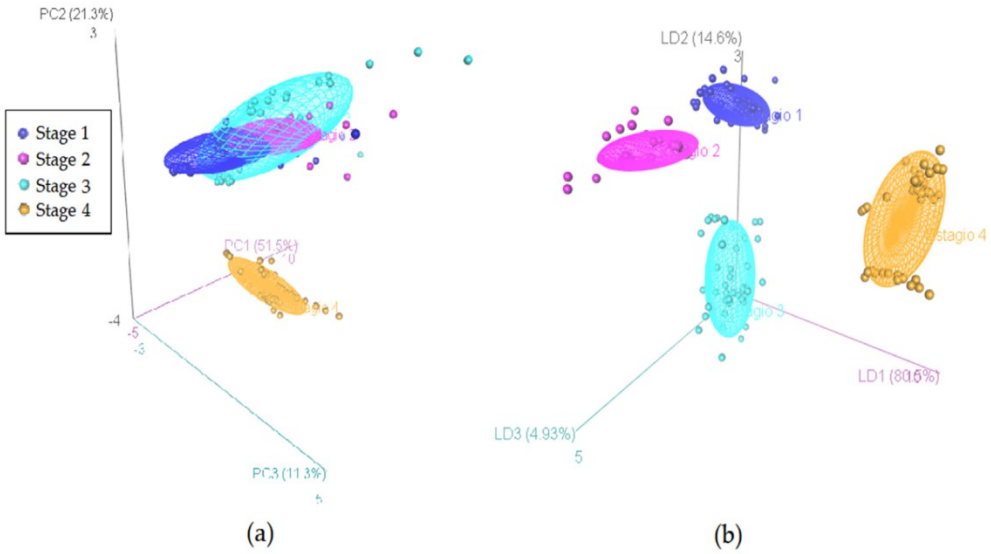

Fig. 8. Three-dimensional data obtained with 13 sensors: (a) by PCA and (b) LDA.

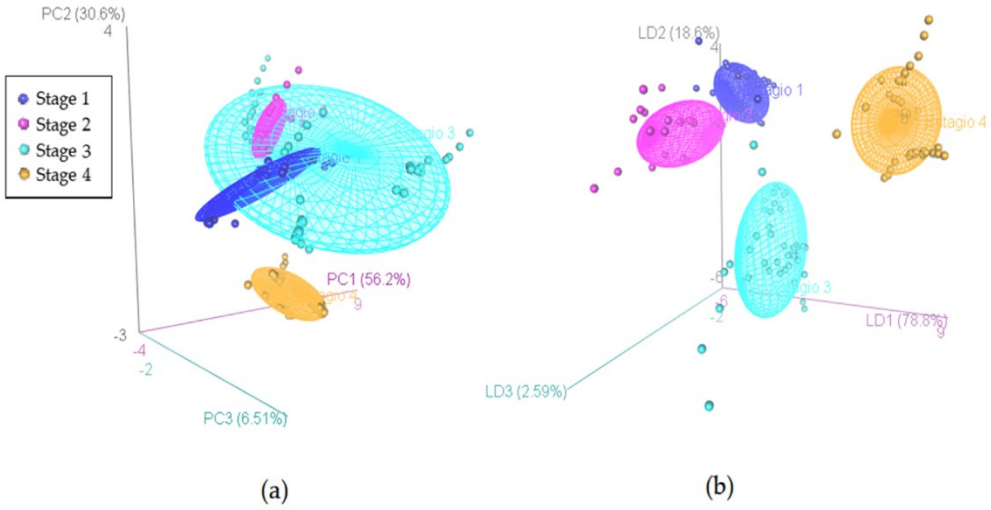

Fig. 9. Three-dimensional data obtained with 7 sensors: (a) by PCA and (b) LDA.

Table 5
Results obtained with the classifiers for the dataset with 13 sensors.

Method	Validation accuracy (10-fold)	Validation Cohen's kappa coefficient (10-fold)	Test accuracy	Test Cohen's kappa coefficient
KNN (k = 5)	93.32%	0.9112	92.31%	0.8974
SVM	98.54%	0.9805	98.08%	0.9744
RF (nvar = 11)	93.87%	0.9181	96.15%	0.9487
ELM (nhid = 5, actfun = tansig)	78.80%	0.7171	61.54%	0.4872
PCA – KNN (k = 5)	91.02%	0.8801	88.46%	0.8462
PCA – SVM	78.61%	0.7139	76.92%	0.6923
PCA – RF (nvar = 2)	91.24%	0.8824	90.38%	0.8718
PCA – ELM (nhid = 5, actfun = tansig)	78.08%	0.7073	73.08%	0.6410
LDA – KNN (k = 5)	97.92%	0.9721	98.08%	0.9744
LDA – SVM	98.42%	0.9790	98.08%	0.9744
LDA – RF (nvar = 2)	98.47%	0.9796	98.08%	0.9744
LDA – ELM (nhid = 5, actfun = purelin)	89.80%	0.8633	86.54%	0.8205

data obtained by the 13 sensors and the 7 sensors selected were placed. For each algorithm, the data reduced by PCA and LDA were entered as input. Tables 5 and 6 summarize the accuracy percentages in the validation (test with the trained data) and test (black box), for the total and reduced data, respectively, where nvar is the number of variables sampled randomly as candidates in each division in the RF method; nhid is the number of neurons in the hidden layer of an ELM; actfun is the activation function used in neurons; and k is the maximum number of nearest neighbors.

These results show that using 7 sensors and no dimensionality reduction method as input, all classifiers improved their performances in relation to the 13, except for the LDA-RF that obtained a slight deterioration for the validation and maintained the hit rate for the set of tests. Using the PCA as the input of the classifiers, the set with 7 sensors

Table 6
Results obtained with the classifiers for the dataset with 7 sensors.

Method	Validation accuracy (10-fold)	Validation Cohen's kappa coefficient (10-fold)	Test accuracy	Test Cohen's kappa coefficient
KNN (k = 5)	94.81%	0.9307	94.23%	0.9231
SVM	99.18%	0.9891	98.08%	0.9744
RF (nvar = 2)	94.10%	0.9213	94.23%	0.9231
ELM (nhid = 5, actfun = purelin)	85.29%	0.8034	90.38%	0.8718
PCA – KNN (k = 5)	93.15%	0.9085	92.31%	0.8974
PCA – SVM	81.60%	0.7544	76.92%	0.6923
PCA – RF (nvar = 3)	93.73%	0.9162	92.31%	0.8974
PCA – ELM (nhid = 5, actfun = tansig)	76.03%	0.6793	86.54%	0.8205
LDA – KNN (k = 5)	98.69%	0.9825	96.15%	0.9487
LDA – SVM	99.23%	0.9897	98.08%	0.9744
LDA – RF (nvar = 2)	97.20%	0.9626	98.08%	0.9744
LDA – ELM (nhid = 5, actfun = purelin)	92.96%	0.9059	92.31%	0.8974

obtained a better hit rate for all classifiers, except for the PCA-ELM that obtained a deterioration for the validation set. With the LDA as input, there was an improvement in all classifiers for the set of 7 sensors compared to the 13 (except for the LDA-KNN in the test step), and this is the LDA for the 7 managed to cluster better than for 13, as shown in Fig. 8(a) and (b).

Comparing the classifiers for the sets of sensors (7 and 13), the LDA imputation was better than the PCA and no imputation, because as shown in Figs. 8 and 9, this method besides reducing the dimensionality, facilitates the work for the classifiers. Therefore, the LDA provided an improvement in accuracy for almost all classifiers, both in the validation and in the test. Therefore, the LDA, besides improving classification allows a considerable reduction of the sensor data set, which is interesting when working with embedded devices and for online classification.

As previously described, 30% of the sample set was used for testing. With a total of 52 samples, each of the four stages was left with 13 samples for testing. Tables 7 and 8 show the percentages of test samples correctly classified for each stage and classifier algorithm in each dataset.

In summary, the RF classifier with LDA imputation for the reduced set of 7 sensors was the one that obtained the best hit rate with 99.23% in the validation step and 98.08% of the hit in the test step. This demonstrates that the initial reduction with the Chi-square test for 7 sensors, the subsequent reduction of dimensionality with the LDA for 3 dimensions and the use of the correct algorithm improves the classification rate of the model and contributes to the construction of an ideal device for the studied application.

As shown in (Benedetti et al., 2008; Huang et al., 2012), and based on Table 1, during stage 1 (the first days after full bloom), the alcohols (1-Hexanol, (E)-2-hexenol e (Z)-3-hexenol) are the main components exhaled by the peach. The aldehydes (Hexanal, (E)-2-hexenal e Benzaldehyde), esters (Hexyl acetate e (E)-2-hexenyl acetate) and acetone have a stage 2 emission peak and decreases after the onset of Stage 3

Table 7
Percentage of correctly classified samples to the set of 13 sensors.

Method	Stage 1	Stage 2	Stage 3	Stage 4
KNN	100%	76.92%	92.30%	100%
SVM	100%	92.30%	100%	100%
RF	100%	84.61%	100%	100%
ELM	61.53%	61.53%	38.46%	100%
PCA – KNN	92.30%	100%	61.53%	100%
PCA – SVM	76.92%	69.23%	61.53%	100%
PCA – RF	76.92%	100%	84.61%	100%
PCA – ELM	76.92%	84.61%	30.76%	100%
LDA – KNN	100%	100%	92.30%	100%
LDA – SVM	100%	100%	92.30%	100%
LDA – RF	100%	100%	92.30%	100%
LDA – ELM	76.92%	92.30%	84.61%	92.30%

Table 8
Percentage of correctly classified samples to the set of 7 sensors.

Method	Stage 1	Stage 2	Stage 3	Stage 4
KNN	92.30%	100%	84.61%	100%
SVM	100%	100%	92.30%	100%
RF	100%	76.92%	100%	100%
ELM	61.53%	100%	100%	100%
PCA – KNN	100%	92.30%	76.92%	100%
PCA – SVM	84.61%	100%	53.84%	100%
PCA – RF	92.30%	100%	76.92%	100%
PCA – ELM	84.61%	100%	61.53%	100%
LDA – KNN	92.30%	100%	92.30%	100%
LDA – SVM	92.30%	100%	100%	100%
LDA – RF	100%	92.30%	100%	100%
LDA – ELM	92.30%	76.92%	100%	100%

(day 51 after full bloom). Terpenoids (α-terpinene, γ-terpinene e Linalool), as the study in (Visai and Vanoli, 1997) are exhaled at a more abundant rate at day 61 after full bloom, which already comprises the third stage of fruit growth. Finally, lactones (γ-hexalactone, γ-heptalactone, γ-octalactone, γ-nonalactone, γ-dodecalactone, γ-decalactone, δ- decalactone, δ-dodecalactone) will be identified only in stage 4, after 80 days of full flowering. Therefore, through the stage of peach growth, it is possible to identify which VOCs are being exhaled with greater abundance in each period.

Table 9 compares the results of the related work with that developed in this paper. It can be noticed that the technique used here (LDA-RF) allows the reduction of the data set generated by the sensors, besides the highest classification rate of the studies with the objective of measuring the maturation of the peach fruit.

To compare, the device built here, besides allowing measurements in the open environment, has a reduced error rate, in comparison to the other works with electronic noses to measure the maturation of the peaches found in the literature, since the prototype constructed in this work obtained a test rate of 98.08% in the best case, against 93% of Brezmes et al. (2000), 97.56% of Benedetti et al. (2008), and 85% of Guohua et al. (2012). The methodology adopted by Zhang et al. (2012) shows a considerable SEP of 8.77.

5. Conclusions

To develop equipment capable of classifying peach growth phases, the equipment developed with the 13 MOS sensors and the ventilation system with ambient air was successful in discriminating these phases. Samples with the four growth stages that were used for the training and construction of the model prove to be an adequate methodology in the application of a device for this purpose. Thus, farmers have the possibility of obtaining a robust and low-cost device capable of determining the harvesting point in the peach trees.

The equipment developed was very promising, showing a good

Table 9
Summary of the results of the related works of peach maturation.

Objective	Techniques used	Parameters	Reference
Maturity stage at post-harvest	ANN	Error = 7% (Test)	Brezmes et al. (2000)
Maturity stage at post-harvest	PCA and CART	Error = 2.44% (Test)	Benedetti et al. (2008)
Maturity stage at post-harvest	PCA and SR	Error = 15% (Test)	Guohua et al. (2012)
Maturity stage at harvest	PCR and PLS	SEP = 8.77 (Test)	Zhang et al. (2012)
Maturity stage/growth in the orchard	LDA-RF	Error = 1.92% (Test)	This work

sensitivity for the detection of VOCs emitted in the peach environment. Therefore, rural producers will be able to identify the maturation stage in a clear and precise way, reducing the risk of fruit losses at undesirable maturation stages for commercialization. Therefore, the 180 samples of the 4 stages of growth (from full bloom to harvest), with 99.24% accuracy in model validation and 98.08% in the test using the LDA-RF classifier and data reducer, for sensors reduced to 7, proves the suitability of the electronic nose for monitoring areas with an intense emission of VOCs. In addition, with the idea of creating a generic device, the step of removing the sensors through Pearson's Qui-Square test helped in the process of removing attributes that would impair the performance of the classifiers, since for most of the methods applied there were improvements in the validation and in the test. Therefore, depending on the application, different sensors can be removed.

In future works, it is desired to reconstruct the model, training it with samples of new harvests, new species of peaches, new types of fruits, besides creating a network topology with several connected electronic noses and collecting data continuously in several points of the monitored culture. In addition, to classify the samples in larger spectra (besides the four phases), applying new tools and algorithms to have similar accuracy to those obtained in this work.

Acknowledgments

The authors thank the Brazilian financial agency (CNPq) for the scholarships, grant number [312104/2015-4].

References

Ab Mutalib, N.A., Jaswir, I., Akmeliawati, R., 2013. IIUM-fabricated portable electronic nose for halal authentication in beverages. In: Information and Communication Technology for the Muslim World (ICT4M), 2013 5th International Conference On, IEEE, Rabat, Morocco, pp. 1–4.

Abraham, K., Pandian, S., 2013. A low-cost mobile urban environmental monitoring system. In: 2013 4th International Conference on Intelligent Systems, Modelling and Simulation, IEEE, Bangkok, Thailand, pp. 659–664. doi: 10.1109/ISMS.2013.76.

Aleixandre, M., Santos, J.P., Sayago, I., Cabellos, J.M., Arroyo, T., Horrillo, M.C., 2015. A wireless and portable electronic nose to differentiate musts of different ripeness degree and grape varieties. Sensors 15, 8429–8443. https://doi.org/10.3390/s150408429.

Aleixandre, M., Cabellos, J.M., Arroyo, T., Horrillo, M.C., 2018. Quantification of wine mixtures with an electronic nose and a human panel. Front. Bioeng. Biotechnol. 6, 14. https://doi.org/10.3389/fbioe.2018.00014.

Bagula, A., Zennaro, M., Inggs, G., Scott, S., Gascon, D., 2012. Ubiquitous sensor networking for development (usn4d): an application to pollution monitoring. Sensors 12, 391–414. https://doi.org/10.3390/s120100391.

Baietto, M., Wilson, A., 2015. Electronic-nose applications for fruit identification, ripeness and quality grading. Sensors 15, 899–931. https://doi.org/10.3390/s150100899.

Baraldi, R., Rapparini, F., Rossi, F., Latella, A., Ciccioli, P., 1999. Volatile organic compound emissions from flowers of the most occuring and economically important species of fruit trees. Phys. Chem. Earth Part B 24, 729–732. https://doi.org/10.1016/S1464-1909(99)00073-8.

Barbosa, W., Ojima, M., Dall'Orto, F.A.C., Martins, F.P., Lovate, A.A., 1993. Desenvolvimento dos frutos e das sementes de pêssegos subtropicais de diferentes ciclos de maturação. Pesquisa Agropecuária Brasileira 28, 701–707.

Barbosa, W., Campo-Dall'orto, F.A., Ojima, M., Sampaio, V.R., Bandel, G., 1990. Ecofisiologia do desenvolvimento vegetativo e reprodutivo do pessegueiro em região subtropical. Instituto Agronômico, Campinas.

Benedetti, S., Buratti, S., Spinardi, A., Mannino, S., Mignani, I., 2008. Electronic nose as a non-destructive tool to characterise peach cultivars and to monitor their ripening stage during shelf-life. Postharvest Biol. Technol. 47, 181–188. https://doi.org/10.1016/j.postharvbio.2007.06.012.

Bieganowski, A., Józefaciuk, G., Bandura, L., Guz, Lukasz, Lagód, G., Franus, W., 2018. Evaluation of hydrocarbon soil pollution using E-nose. Sensors 18, 2463. https://doi.org/10.3390/s18082463.

Blanco-Novoa, O., Fernández-Caramés, T., Fraga-Lamas, P., Castedo, L., 2018. A cost-effective IoT system for monitoring Indoor radon gas concentration. Sensors 18, 2198. https://doi.org/10.3390/s18072198.

Brezmes, J., Llobet, E., Vilanova, X., Saiz, G., Correig, X., 2000. Fruit ripeness monitoring using an electronic nose. Sens. Actuators, B 69, 223–229. https://doi.org/10.1016/S0925-4005(00)00494-9.

Chen, L.-Y., Wu, C.-C., Chou, T.-I., Chiu, S.-W., Tang, K.-T., 2018. Development of a dual MOS Electronic nose/camera system for improving fruit ripeness classification. Sensors 18, 3256. https://doi.org/10.3390/s18103256.

Couto, M., 2006. Efeito da temperatura durante a diferenciação de gemas, floração, crescimento e desenvolvimento de frutos em pessegueiro na região de Pelotas, RS. < http://guaiaca.ufpel.edu.br:8080/handle/123456789/2122 > .

Cunha Junior, L.C., Durigan, M.F.B., Mattiuz, B.-H., Martins, R.N., Durigan, J.F., 2007. Caracterização da curva de maturação de pêssegos' Aurora-1', na região de Jaboticabal-SP. Revista Brasileira de Fruticultura 661–665. https://doi.org/10.1590/S0100-29452007000300045.

Deshmukh, S., Bandyopadhyay, R., Bhattacharyya, N., Pandey, R.A., Jana, A., 2015. Application of electronic nose for industrial odors and gaseous emissions measurement and monitoring – an overview. Talanta 144, 329–340. https://doi.org/10.1016/j.talanta.2015.06.050.

Dorji, U., Pobkrut, T., Kerdcharoen, T., 2017. Electronic nose based wireless sensor network for soil monitoring in precision farming system. In: Knowledge and Smart Technology (KST), 2017 9th International Conference On, IEEE, Pattaya, Thailand, pp. 182–186. doi: 10.1109/KST.2017.7886087.

Dutra, L.F., Kersten, E., Fachinello, J.C., 2002. Cutting time, indolebutyric acid and tryptophan in rooting of peach tree cuttings. Sci. Agricola 59, 327–333. https://doi.org/10.1590/S0103-90162002000200019.

Guohua, H., Yuling, W., Dandan, Y., Wenwen, D., Linshan, Z., Lvye, W., 2012. Study of peach freshness predictive method based on electronic nose. Food Control 28, 25–32. https://doi.org/10.1016/j.foodcont.2012.04.025.

Horvat, R.J., Chapman Jr, G.W., Robertson, J.A., Meredith, F.I., Scorza, R., Callahan, A.M., Morgens, P., 1990. Comparison of the volatile compounds from several commercial peach cultivars. J. Agric. Food. Chem. 38, 234–237. https://doi.org/10.1021/jf00091a051.

Huang, C.-H., Zeng, C., Wang, Y.-C., Peng, H.-Y., Lin, C.-S., Chang, C.-J., Yang, H.-Y., 2018. A study of diagnostic accuracy using a chemical sensor array and a machine learning technique to detect lung cancer. Sensors 18, 2845. https://doi.org/10.3390/s18092845.

Huang, G.-B., Zhou, H., Ding, X., Zhang, R., 2012. Extreme learning machine for regression and multiclass classification. IEEE Trans. Syst., Man, Cyber. Part B (Cybernetics) 42, 513–529. https://doi.org/10.1109/TSMCB.2011.2168604.

Jiang, S., Wang, J., 2016. Internal quality detection of Chinese pecans (Carya cathayensis) during storage using electronic nose responses combined with physicochemical methods. Postharvest Biol. Technol. 118, 17–25. https://doi.org/10.1016/j.postharvbio.2016.03.016.

Keske, C., 2009. Epidemiologia da podridão parda em pessegueiros conduzidos em sistema de produção orgânico no alto vale do. Itajaí-SC.

Kim, J., Hwangbo, H., 2018. Sensor-based optimization model for air quality improvement in home IoT. Sensors 18, 959. https://doi.org/10.3390/s18040959.

Langford, J., 2005. The cross validation problem. In: International Conference on Computational Learning Theory, Springer, Bertinoro, Italy, pp. 687–688. doi: 10.1007/11503415_47.

Laref, R., Losson, E., Sava, A., Siadat, M., 2018. Support vector machine regression for calibration transfer between electronic noses dedicated to air pollution monitoring. Sensors 18, 3716. https://doi.org/10.3390/s18113716.

Li, Q., Gu, Y., Wang, N.-F., 2017. Application of random forest classifier by means of a QCM-based e-nose in the identification of Chinese liquor flavors. IEEE Sens. J. 17, 1788–1794. https://doi.org/10.1109/JSEN.2017.2657653.

Li, W., Liu, H., Xie, D., He, Z., Pi, X., 2017. Lung cancer screening based on type-different sensor arrays. Sci. Rep. 7, 1969. https://doi.org/10.1038/s41598-017-02154-9.

Lilien-Kipnis, H., Lavee, S., 1971. Anatomical changes during the development of "Ventura" peach fruitst. J. Horticult. Sci. 46, 103–110. https://doi.org/10.1080/00221589.1971.11514388.

Lin, S., Zhang, X., 2016. A rapid and novel method for predicting nicotine alkaloids in tobacco through electronic nose and partial least-squares regression analysis. Anal. Methods 8, 1609–1617. https://doi.org/10.1039/C5AY02257F.

Lorwongtragool, P., Sowade, E., Watthanawisuth, N., Baumann, R.R., Kerdcharoen, T., 2014. A novel wearable electronic nose for healthcare based on flexible printed chemical sensor array. Sensors 14, 19700–19712. https://doi.org/10.3390/s141019700.

Men, H., Jiao, Y., Shi, Y., Gong, F., Chen, Y., Fang, H., Liu, J., 2018. Odor fingerprint analysis using feature mining method based on olfactory sensory evaluation. Sensors

18, 3387. https://doi.org/10.3390/s18103387.
Men, H., Shi, Y., Jiao, Y., Gong, F., Liu, J., 2018. Electronic nose sensors data feature mining: a synergetic strategy for the classification of beer. Anal. Methods 10, 2016–2025. https://doi.org/10.1039/C8AY00280K.
Mishra, G., Srivastava, S., Panda, B.K., Mishra, H.N., 2018. Prediction of Sitophilus granarius infestation in stored wheat grain using multivariate chemometrics & fuzzy logic-based electronic nose analysis. Comput. Electron. Agric. 152, 324–332. https://doi.org/10.1016/j.compag.2018.07.022.
Mucherino, A., Papajorgji, P., Pardalos, P.M., 2009. Data Mining in Agriculture. Springer Science & Business Media.
Nurul, A., Jaswir, I., Akmeliawati, R., Ibrahim, A.M., Aslam, M., Octavianti, F., 2017. Rapid detection of ethanol in beverages using IIUM-fabricated Electronic Nose. Int. Food Res. J. 24.
Peterson, P.J., Aujla, A., Grant, K.H., Brundle, A.G., Thompson, M.R., Vande Hey, J., Leigh, R.J., 2017. Practical use of metal oxide semiconductor gas sensors for measuring nitrogen dioxide and ozone in urban environments. Sensors 17, 1653. https://doi.org/10.3390/s17071653.
Pineda, D.M., Pérez, J.C., 2017. SENose: an under U $50 electronic nose for the monitoring of soil gas emissions. Comput. Electron. Agric. 133, 15–21. https://doi.org/10.1016/j.compag.2016.12.004.
Ragazzo-Sanchez, J.A., Chalier, P., Chevalier, D., Calderon-Santoyo, M., Ghommidh, C., 2008. Identification of different alcoholic beverages by electronic nose coupled to GC. Sens. Actuators, B 134, 43–48. https://doi.org/10.1016/j.snb.2008.04.006.
Rajkumar, P., Wang, N., ElImasry, G., Raghavan, G.S.V., Gariepy, Y., 2012. Studies on banana fruit quality and maturity stages using hyperspectral imaging. J. Food Eng. 108, 194–200. https://doi.org/10.1016/j.jfoodeng.2011.05.002.
Ruiz, L.I., Beccaro, W., Evaristo, B.G., Fernandez, F.J.R., 2018. Tactile sensing glove-based system for objects classification using support vector machine. IEEE Lat. Am. Trans. 16, 1658–1663. https://doi.org/10.1109/TLA.2018.8444383.
Santos, J.P., Lozano, J., 2015. Real time detection of beer defects with a hand held electronic nose. In: Electron Devices (CDE), 2015 10th Spanish Conference On, IEEE, Aranjuez – Madrid, Spain, pp. 1–4. doi: 10.1109/CDE.2015.7087492.
Srivastava, S., Mishra, G., Mishra, H.N., 2019. Fuzzy controller based E-nose classification of Sitophilus oryzae infestation in stored rice grain. Food Chem. 283, 604–610. https://doi.org/10.1016/j.foodchem.2019.01.076.
Sudarmaji, A., Kitagawa, A., 2016. Application of temperature modulation-SDP on MOS gas sensors: capturing soil gaseous profile for discrimination of soil under different nutrient addition. J. Sens. 2016. https://doi.org/10.1155/2016/1035902.
Visai, C., Vanoli, M., 1997. Volatile compound production during growth and ripening of peaches and nectarines. Sci. Horticult. 70, 15–24. https://doi.org/10.1016/S0304-4238(97)00032-0.
Voss, A., Witt, K., Fischer, C., Reulecke, S., Poitz, W., Kechagias, V., Surber, R., Figulla, H.R., 2012. Smelling heart failure from human skin odor with an electronic nose. IEEE, San Diego, United States, pp. 4034–4037. https://doi.org/10.1109/EMBC.2012.6346852.
Wei, Z., Xiao, X., Wang, J., Wang, H., 2017. Identification of the Rice wines with different marked ages by electronic nose coupled with smartphone and cloud storage platform. Sensors 17, 2500. https://doi.org/10.3390/s17112500.
Wojnowski, W., Majchrzak, T., Dymerski, T., Gębicki, J., Namieśnik, J., 2017. Portable electronic nose based on electrochemical sensors for food quality assessment. Sensors 17, 2715. https://doi.org/10.3390/s17122715.
Wu, H., Yue, T., Xu, Z., Zhang, C., 2017. Sensor array optimization and discrimination of apple juices according to variety by an electronic nose. Anal. Methods 9, 921–928. https://doi.org/10.1039/C6AY02610A.
Zhang, H., Wang, J., Ye, S., Chang, M., 2012. Application of electronic nose and statistical analysis to predict quality indices of peach. Food Bioprocess Technol. 5, 65–72. https://doi.org/10.1007/s11947-009-0295-7.
Zhu, J., Chen, F., Wang, L., Niu, Y., Xiao, Z., 2017. Evaluation of the synergism among volatile compounds in Oolong tea infusion by odour threshold with sensory analysis and E-nose. Food Chem. 221, 1484–1490. https://doi.org/10.1016/j.foodchem.2016.11.002.