Data processing from electrical signals acquired by an E-nose system used for quality control of cocoa

A Florez\(^1\), C Durán\(^1\), and J Carrillo\(^1\)

\(^1\) Sistemas Multisensoriales y Reconocimiento de Patrones (GISM), Universidad de Pamplona, Pamplona, Colombia

E-mail: alex82co@gmail.com, cmduran@unipamplona.edu.co

Abstract. This research consists of the implementation of different pattern recognition methods applied to discriminate and classify the physical signals (electrical) acquired by an artificial electronic nose system composed of a gas sensor array of 10 units coupled with a data acquisition board that were used to perform a sensor chamber that receives the electrical information from volatile organic compounds generated by cacao beans. A concentration chamber for samples conditioning of fermented beans was used and the samples were fermented around 72 and 144 hours while the cacao samples infected with monilia were over-fermented. For obtaining the temperature inside of the sensor chamber, a digital temperature sensor was implemented by using a Peltier Cell mechanism which was controlled through a classical algorithm. At the design stage, a data acquisition system composed of an Arduino card and a graphical interface made in LabView was developed for data storing, controlling, and signals monitoring. For the electrical signals treatment and data analysis, two pattern recognition models were applied by using Python software where two signals pre-processing methods such as Euclidean normalization and Roboust Scaling were used afterward with data processing techniques as principal component analysis and clusters analysis, obtaining a 96.51% of variance in the two first components.

1. Introduction

In recent times, the quality control of cocoa beans has become a challenge in international trade that importers have requested a review of current cocoa quality standards and testing methods. Nowadays, there are two methods for evaluating cocoa quality. The first is a set of tests performed on raw beans, and the second comprises tests performed by end-users. For instance, for dried cocoa beans, the sample randomly selects a significant percentage of the bags for inspection and a sharpening iron is used to collect several beans from the selected bags. Exporting countries and cocoa consuming countries must participate in the development of cocoa quality assessment procedures, which are sound both from a scientific and commercial point of view [1]. The fermentation process is the most important stage of cocoa as it influences the formation of aroma precursors, the chocolate flavor, and the last quality of the derived products. However, fermentation is affected by several factors such as origin cocoa genetics, intervals between harvests, amount of cocoa to ferment, amount of pulp in the seed, the fermentation method, and the conditions of the environment where the process is carried out [2]

Currently, the data processing methods from the physical signals acquired are good approach to determine the quality control of cocoa fermentation degree that comprises splitting the bean longitudinally, visualizing the color appearance of the cotyledon and electrical signals that allows determining a correct fermentation. Nevertheless, cutting tests require a special instrument with considerable price and trained personnel. Additionally, there is a technique called sensory evaluation
that allows the development of flavor profiles. The disadvantage of this method is that it requires expert personnel, training of the judges, basic conditions such as cupping cabins, sample preparation, etcetera. Other researchers have used techniques such as gas chromatography adapted to a mass spectrometer, and high-resolution liquid chromatography, which has allowed the simultaneous separation, identification, and quantification of compounds [3]. For farmers it is difficult to acquire this equipment as it is relatively expensive, it requires trained personnel, supplies, facilities, and tests cannot be performed on-site. Electronic olfaction systems are being implemented in different applications such as agro-industry, environment, security, and medicine [4]. Recent advances in microelectronics, sensors, and signal processing from electrical signals have made it possible to manufacture and optimize the electronic nose (E-nose) performance equipped with greater integration capacity in portable or mobile, robotic and intelligent platforms [5-8]. Therefore, in this study we developed and implemented different data analysis algorithms on electrical signals to enhance the E-nose performance for discriminating the cacao beans.

2. Materials and methods

Figure 1 illustrates the electronic olfaction system used for measuring and discriminating the volatile organic compounds generated during the fermentation process of cocoa beans at different times. The multisensory system was based in a concentration chamber, temperature control system made up of an LM35 digital sensor, and a Peltier cell can be seen. The Peltier cell measures the thermoelectric effect to heat the samples inside the gas concentration chamber. The measurement chamber was developed of 10 gas sensors array from the manufacturer MQ which were conditioned. The data acquisition system was based on Arduino MEGA hardware associated with signal conditioning card that allows the action of 3 airflow control valves and an electric pump.

![Figure 1. Electronic olfaction system for quality control in the cocoa fermentation process.](image)

2.1. Gas concentration chamber

The main function of the concentration chamber is to house the cocoa samples, generate the headspace, preserve the tightness, and controlling the temperature of the sample. It consists of an approximate volume of 0.785 liters and an iron material with external electrostatic paint coating. The cover includes two holes that extract the volatile emissions with a wire connection (connector), 1/4-inch thread.

2.2. Temperature control system for the concentration chamber

The temperature control was developed from a novel thermoelectric heating system based on the Seebeck effect composed of a Peltier cell. This device can supply a specific temperature that lets the samples to be kept in a controlled environment for the measurement. As mentioned before, an LM35 temperature sensor was applied where the output behavior of the device was linear. Each Celsius degree is equivalent to a variation in the output signal of 10 millivolts. The measurement range was defined from -55 °C, up to 150 °C. Chamber temperature control was set as 33 °C for all measurements.
2.3. Measurement chamber
Figure 2 illustrates the measurement system which was after coupled with the concentration chamber. The electronic olfaction system comprises of 10 gas sensors which can detect a variety of volatile compounds. The sensor chamber is composed of MQ-2 sensors suitable for detecting LPG, propane, methane, alcohol, hydrogen, and smoke being more sensitive to LPG and propane. MQ-3 is a sensor sensitive to alcohol and less sensitive to benzene. It is also lowy sensitive to gases such as LPG, Hexane, CO, CH₄, which can be neglected if there is little concentration of these. MQ-4 is a sensor to detect Methane Gas (Natural Gas) in the air with concentrations from 300 ppm to 10000 ppm. MQ-5 is used in gas equipment leak detection in consumer applications and industry.

This sensor is suitable for the detection of LPG, natural gas, coal gas. MQ-6 is a gas sensor convenient for detecting the presence of LP Gas, composed mainly of Propane and Butane and Natural Gas (Methane) in the air. MQ-7 is highly sensitive to carbon monoxide (CO) and H₂ while MQ-8 is a hydrogen gas sensor appropriate for detecting hydrogen concentrations in air. It can detect gas, hydrogen concentrations anywhere from 100 ppm - 10000 ppm. MQ-9 is a Carbon Monoxide (CO) and flammable gas sensor fitting for detecting concentrations in the air. MQ-3A is highly sensitive to alcohol, and gases such as GLP, Hexane, CO and CH₄. MQ-135 is used in air quality control equipment for buildings and offices. They are suitable for the detection of NH₃, NOₓ, alcohol, benzene, smoke, CO₂.

![Figure 2. Sensor chamber.](image)

2.4. Data acquisition stage
Because of the number of gas sensors and the temperature sensor that was implemented, an acquisition card was selected including 11 analog input signals and 5 digital output signals, where 3 correspond to the activation of the valves, 1 for the activation of the pump, and 1 PWM output for the temperature control associated with the Peltier cell. For this reason, the Arduino MEGA 2560 board was selected and the communication was made through the USB port. The application was designed in high-level graphic language, under the LabVIEW 2018 software platform.

2.5. Supervision and data acquisition software
Once the data were acquired, pattern recognition techniques were implemented using the Python software tool, version 3.7.3 64-bit | Qt 5.9.6 | PyQt5 5.9.2 | Windows 10, along with the Anaconda Toolkit and Spyder 3.3.6 Console Editor. One of the methods used for data pre-processing was Euclidean normalization: It is defined as the distance between two points’ p and q. In Cartesian coordinates, the Euclidean distance is calculated using the Pythagorean Theorem. For example, in a two-dimensional space in which each point is defined by the coordinates (x, y), the Euclidean distance between p and q is given by Equation (1) [9].

\[ d_{euc}(p, q) = \sqrt{(x_p - x_q)^2 + (y_p - y_q)^2}. \] (1)
This equation can be generalized to a Euclidean n-dimensional space, where each point is defined by a vector of n coordinates: \( p = (p_1, p_2, p_3, ..., p_n) \) and \( q = (q_1, q_2, q_3, ..., q_n) \). Equation (2) relates the formula that determines the Euclidean distance between n dimensions.

\[
d_{\text{euc}}(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2}. \tag{2}
\]

Equation (2) can be generalized for a Euclidean space n-dimensional where each point is defined by a vector of n coordinates. One approach for standardizing the input variables in the presence of outliers is to ignore the outliers from the calculation of the measurement and standard deviation.

The resulting variable has a mean and median of zero and a standard deviation of 1; although it is not biased by outliers and outliers are still present with the same relative relationships with other values; Equation (3), defines the formula for the calculation of robust standardization [10].

\[
\text{Valor} = \frac{\text{valor} - \text{mediana}}{p_{75} - p_{25}}. \tag{3}
\]

In the data processing phase, the statistical method of pattern recognition PCA was used, which is a technique frequently used in applications and multivariate analysis. The PCA method determines the components that best represent the data according to the minimum square error [10].

2.6. Conditioning of sample

This research began from the extraction of the cocoa beans from the cob of the ICS-95 clone obtained from La Vereda Restauración, located in the municipality of Puerto Santander, Norte de Santander Department, Colombia. In the fermentation process, the flavor and aroma of the product develops and contributes to forming a brown grain swollen color with a good appearance.

A cocoa mass of approximately 2.2 kilos was placed into a wooden barrel. For the selection of the study samples and data processing, samples with 72 hours of fermentation, 144 hours of fermentation with the absence and presence of Monilia. It should be noted that Cacao moniliasis is a disease caused by the fungus Moniliophthora Roreri.

The over fermented samples were exposed to a time of 288 hours. As complementary information for the analysis, each sample submitted to the concentration chamber weighted 20 grams of each class.

3. Results and discussion

Figure 3 illustrates the sensor responses to the detection of the compounds emitted for a sample of fermented cocoa at 72 hours after starting the process. It could be analyzed that the sensors MQ135, MQ9, MQ7, MQ2, and MQ3 responded in greater amplitude.

The remaining sensors recorded a lower amplitude but responded to the different volatile compounds. As can see in Figure 3(a), the required time for the acquisition stage guarantees the transient responses, getting stability in the reading of the signals, which allows validating the tightness of the system.

Afterward, the cleaning process begins and tries to carry the system to the baseline or initial amplitude; however, the sensors generate a memory effect leaving a small amplitude of the voltage. In Figure 3(b), the response for a cocoa fermented sample during 144 hours and a decrease in the amplitude of the signals of the MQ135, MQ9, MQ7, MQ2, MQ3 sensors was observed.

Figure 3(c) shows the response for a sample infected with monilia, it was found that the response of the MQ135 sensor illustrates an amplitude high and the MQ6 and MQ3 sensors try to respond with a similar amplitude.

Finally, Figure 3(d) shows a sample of over-fermented cocoa, where it was observed that the amplitude of the signals MQ135, MQ9, MQ7, MQ2, MQ3 is much lower in contrast to the cocoa samples fermented at 72 and 144 hours, and infected.
Figure 3. Sensor responses to the detection of the compounds emitted; (a) trace fermented cocoa mass 72 hours, (b) trace fermented cocoa mass 144 hours, (c) trace fermented cocoa mass 144 hours with monilia and (d) Trace over-fermented cocoa mass.

The Figure 4(a) shows a comparison of the data before scaling and after scaling, using robust scaler data pre-processing; Figure 4(b) illustrates the PCA plot made with Python where the clusters related with cocoa samples fermented with 144 hours (blue colour) and 72 hours (red colour) infected with monilia (orange colour) and over fermented (green colour) can be observed. In the plot, it can be deduced that by implementing a robust data pre-processing, is evidence of overlap between samples related with the fermented classes at 144 hours and over fermented. The percentage of variance in the PC1 = 97.7% and the PC2 = 0.14%.

Figure 4. PCA plot, (a) data pre-processing and scaling implementing robust scale, (b) PCA plot by using robust scaler normalization.
Figure 5(a) depicts a comparison of the data before scaling and after scaling, using data pre-processing Euclidean normalization. Subsequently, the dimensionality reduction of variables and data discrimination were carried out implementing PCA analysis, where the sklearn and decomposition library was also used.

Figure 5(b) illustrates the clusters of cocoa samples fermented during 144 hours, (blue color) 72 hours (red color), infected with monilia (orange color), and over-fermented (green color). In this plot, it can be deduced that by implementing a pre-processing of Euclidean data, the discrimination between associated samples is demonstrated.

The percentage of variance in the principal component PC1 = 89.6% and the PC2 = 6.91%. A cluster analysis was also implemented in this study. Cluster Analysis, known as Cluster, is a multivariate statistical technique that seeks to group elements (or variables) trying to achieve maximum homogeneity in each group and the greatest difference between groups [11,12].

Figure 6 shows the distribution of structures obtained with each of the classes linked to the fermentation process. It can be seen that class (A) fermented 144 hours and over-fermented (C) obtained just one overlapped sample related to fermentation during 72 hours (D), and bad fermentation monilia samples (B), where they are discriminated by classes.

Figure 5. Comparison of the data, (a) pre-processing and scaling of data implementing Euclidean normalization, (b) PCA plot by using Euclidean normalization.

Figure 6. Cluster analysis in the cocoa fermentation process.
4. Conclusions
The robust scaler and euclidean pre-processing methods were implemented to the electrical signals acquired by E-nose and the results by using PCA analysis obtained few overlapping samples which were observed and corresponded to the fermented 144h and over fermented classes, applying robust scaler pre-processed method. Besides, comparing results with the Euclidean pre-processing method through the principal component analysis from the electrical signals, is observed that the samples associated with the 4 classes did not show overlappings, being the most suitable method for data analysis. Through the implementation the data processing methods to the data set by using Cluster Analysis in the cocoa fermentation process, it can be observed that the results showed an overlapped of samples between the 144h fermented and over fermented classes, similar to the results obtained by the PCA analysis using robust scaled data pre-processing. Besides, during the development of the research, 4 main categories were analyzed in the process which were fermented for 144 hours (desired), over fermented, fermented with monilia, and fermented 72 hours.

Through the different data processing methods used from the physical signals previously acquired of the E-nose was possible to determine the quality control of cocoa fermentation degree by using the electrical values that allowed achieved a good discrimination of the cacao beans fermentation. Therefore, this methodology might be useful to be applied in different fields of physics such as fluid mechanics, biophysics as there are various biological processes that need to explain the operation and importance of the physical variables.

Acknowledgments
Thanks to the graduate and cocoa expert Jorge A Duque, and Miller Otero, representative of Fedecacao, San José de Cúcuta, Colombia. To Emiro Villalba and engineers Jorge Villalba, José Villalba, businessmen of the cocoa sector for their support. Also to Ing Lucas Quintana, leader of the Giaunad research group, and Dr. Alejandro Cote of the Unad University from San José de Cúcuta, Colombia, for the support in spaces and socialization of the research. Finally, to the multisensor system and pattern recognition research group for their advice and support for the development of this study.

References
[1] Olunloyo V O, Ibidapo T A, Dinrifo R 2011 Neural network-based electronic nose for cocoa beans quality assessment Agricultural Engineering International: CIGR Journal 13 (4) 1
[2] Warren A et al. 2019 Forcing fermentation: Profiling proteins, peptides and polyphenols in lab-scale cocoa bean fermentation Food Chemistry 278(25) 786
[3] Phuong D T et al. 2015 Assessing cocoa aroma quality by multiple analytical approaches Food Research International 77 (3) 657
[4] Durán C, Baldovino M 2009 Monitoring system to detect the maturity of agro-industrial products through of an electronic nose Colombian Journal of Advanced Technologies 1(13) 1
[5] Fernandez M, Herrero A, Lozano R, Santos B J P, Suarez M 2017 An online odor classification system for electronic noses DYNA 92(3) 354
[6] Durán C, Velásquez C, Gualdron O 2012 Implementation of an electronic nose to detect COPD patients from the exhaled breath Engineering and Development Magazine 30(2) 143
[7] Gómez P, Durán C, 2017 Wireless electronic nose (NEI) for control of gas emissions in coal mines Colombian Journal of Advanced Technologies 1(29) 118
[8] Gustavo C, Marko M, Stan C 2018 Intelligent Digital Oil and Gas Fields Concepts, Collaboration, and Right-Time Decisions (United States of America: Elsevier Inc.)
[9] Sachin P P, Upadhyay S H 2020 Euclidean distance based feature ranking and subset selection for bearing fault diagnosis Expert Systems with Applications 154 113400:1
[10] Alexandre L 2020 Parametric PCA for unsupervised metric learning Pattern Recognition Letters 135 425
[11] Falasconi M, Pardo M, Vezzoli M, Sberveglieri G 2007 Cluster validation for electronic nose data Sensors and Actuators B: Chemical 125(2) 596
[12] Jeniffer C, Cristhian D, Ramon O 2019 Bacterial discrimination of drinking water through of an electronic nose and a volatiles extraction equipment Revista Colombiana de Tecnologias de Avanzada 1(33) 155