UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

Juan Carlos Rodriguez Gamboa

Tese

# RAPID DETECTION APPROACH FOR ELECTRONIC NOSE SYSTEMS USING DEEP LEARNING MODELS

Recife - PE

30 de junho de 2020

**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO**
**PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO**
**PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA**

Juan Carlos Rodriguez Gamboa

Tese

# RAPID DETECTION APPROACH FOR ELECTRONIC NOSE SYSTEMS USING DEEP LEARNING MODELS

Trabalho apresentado ao Programa de Pós-Graduação em Biometria e Estatística Aplicada do Departamento de Estatística e Informática da Universidade Federal Rural de Pernambuco como requisito parcial para obtenção do grau de Doutorado em Biometria e Estatística Aplicada.

**Orientador: Prof. PhD. Wilson Rosa de Oliveira**

**Coorientador: Prof. Dr. Adenilton José da Silva**

Recife - PE

30 de junho de 2020

# UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO

## PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO

## PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

**Rapid Detection Approach for Electronic Nose Systems Using Deep Learning Models.**

*Juan Carlos Rodriguez Gamboa*

Tese julgada adequada para obtenção do título de Doutorado em Biometria e Estatística Aplicada, defendida e aprovada por unanimidade em 18/02/2020 pela Comissão Examinadora.

Orientador:

Prof. PhD. Wilson Rosa de Oliveira
Universidade Federal Rural de Pernambuco

Banca Examinadora:

Prof. Dr. Rafael Ferreira Leite de Mello
Universidade Federal de Pernambuco

Prof. Dr. Antonio Samuel Alves da Silva
Universidade Federal Rural de Pernambuco

Prof. Dr. Fernando Maciano de Paula Neto
Universidade Federal de Pernambuco

Profa. PhD. Teresa Bernarda Ludermir
Universidade Federal de Pernambuco

I dedicate this thesis to all my family, friends, and teachers who gave me the necessary support, motivation, and inspiration to get here and reach this big goal.

# Acknowledgements

Firstly, thank you, God, for blessing me each day of my life. You always give me the opportunities, placing the right people, and show me the path.

I want to express my gratitude to my family, especially to my mother and my sister, because they gave me the best example of constant and continuous struggle. To my wife and daughter, for being my motivation and inspiration. Also, to my in-laws, my brothers in law, my nephews for being an essential part of my life. As well, to all those people who have crossed my path and have become my family.

Also, I want to thank to my advisors and teachers Adenilton José da Silva, Wilson Rosa de Oliveira, Tiago Alessandro Espínola Ferreira, Paulo José Duarte Neto, Tatijana Stosic, Borko Stosic, Moacyr Cunha Filho, Luciana Leite de Andrade Lima, Glauco Gonçalves, Victor Medeiros, Claudio Tadeu Cristino, Anna Sikov, and José Antônio Aleixo da Silva. They made it possible to move forward and conclude my doctorate. I am sure that this would have never become truth, without their guidance.

I am grateful to some advisors that, even from a distance, supported my work and decisions: Cristhian Manuel Duran Acevedo, Edilson Delgado Trejos, and Jorge Alberto Jaramillo Garzon.

Additionally, I want to thank my colleagues, partners, and friends who welcomed me in their country. All of them contributed and supported me in several ways, thank you especially to Jonas, Elaine and her family, André and his family, Carla, Carlos, Edneide, Wesley, Kerolly, Fabio, Patricia, Ikaro, Lêda, Yarah, Eucymara, Lucas, Luisa, Adelino and his family. Besides, the foreign people who welcomed me to Brasil: Sara, David, Alvaro, Rodrigo, Robert, Ana, and Carlos.

“It is not the strongest of the species that survives, nor the most intelligent that survives.

It is the one that is most adaptable to change.”

Charles Darwin

# Resumo

O presente trabalho de tese teve como objetivo propor um método para acelerar a identificação das amostras em sistemas de nariz eletrônicos. Como a abordagem convencional de processamento de dados usada em E-Nose é baseada em um estágio inicial de pré-processamento de sinal, são aplicadas técnicas para executar a extração de recursos (características dinâmicas e estáticas) para obter a impressão digital do odor. Além disso, em alguns casos, é necessário implementar um método de seleção de recurso para escolher os melhores atributos antes das tarefas de classificação usando métodos de reconhecimento de padrões. Por exemplo, uma SVM (Support Vector Machine) é um dos métodos de processamento mais comuns para reconhecimento de odores pelos sistemas olfativos eletrônicos.

Portanto, o uso da abordagem tradicional faz uso de toda a medição para obter para obter os principais parâmetros de odorantes que envolvem técnicas de pré-processamento, o que representa um desafio para o reconhecimento de odores em tempo real. Assim, neste trabalho é apresentada uma abordagem para processamento de dados de sistemas olfativos eletrônicos focada no tratamento de dados brutos com base em um protocolo de janela ascendente para encontrar uma porção inicial dos sinais do sensor com o melhor desempenho de reconhecimento.

Comparamos a abordagem proposta com um método tradicional (usando todas as curvas de resposta, aplicando técnicas de pré-processamento para extrair os recursos e posteriormente processando-os usando um algoritmo SVM) em um problema real com as medidas adquiridas com nosso sistema desenvolvido. Além disso, para validar o uso da abordagem proposta em diferentes configurações de sistemas olfativos eletrônicos, realizamos mais testes com vários conjuntos de dados e usando técnicas de aprendizado profundo como a rede neural convolucional CNN. Os resultados mostraram uma precisão de desempenho superior à da abordagem tradicional, com a vantagem de usar uma parte inicial das respostas dos sensores, reduzindo o tempo necessário para fazer previsões.


**Palavras-chave:** electronic nose, rapid detection, real-time classification, deep learning

# Abstract

The present thesis work focused on proposing a method to accelerate the identification of the specimens in electronic nose systems. Since the conventional data processing approach used in E-Nose is based on an initial stage of signal preprocessing applying techniques to perform the feature extraction (dynamic and static characteristics) for obtaining the odor fingerprint. Besides, in some cases, it is required to implement a feature selection method to choose the best attributes before the classification tasks using pattern recognition methods. For example, a Support Vector Machine (SVM) is one of the most common processing methods for odor recognition by the electronic olfactory systems.

Therefore, the use of the traditional approach needs the whole measurement to obtain the main odorant parameters involving preprocessing techniques, which represents a challenge when aiming to perform real-time odors recognition. Thus, in this work is presented an approach for electronic olfactory systems data processing focused on the treatment of raw data based on a rising-window protocol to find an early portion of the sensor signals with the best recognition performance.

We compared the proposed approach against a traditional method (using the entire response curves, applying preprocessing techniques to extract the features and later processing them using an SVM algorithm) in a real application with measures acquired with our developed system. Further, to validate the use of the proposed approach at different settings of electronic olfactory systems, we conducted more tests with several datasets and using deep learning techniques like convolutional neural network CNN. The results showed outperformance accuracy compared with the traditional approach with the advantage of using an early portion of the responses of the sensors, reducing the necessary time to make forecasts.



**Keywords:** electronic nose, rapid detection, real-time classification, deep learning

# Contents

1. Introduction ........ 10

1.1 The sense of smell ........ 12

1.2 The electronic nose systems ........ 12

1.3 The principal elements in an electronic nose system ........ 14

1.3.1 Gas sensor array ........ 14

1.3.2 Adequacy of the gas mixture and sampling system ........ 14

1.3.3 Processing system ........ 14

1.4 Objectives ........ 15

1.4.1 General objective ........ 15

1.1.2 Specific objectives ........ 15

1.5 Research Roadmap ........ 15

2. Portable and Compact E-Nose for Agroalimentary Purposes ........ 18

2.1 Background ........ 18

2.2 Description of the E-Nose Device ........ 20

2.3 Detailed Description of the E-Nose Device ........ 23

2.4 Main contributions Related to Developed E-Nose ........ 28

3. Electronic Nose Dataset for Detection of Wine Spoilage thresholds ........ 29

3.1 Main contributions related to collected database ........ 30

3.2 Published paper in Data in Brief Journal: Electronic nose dataset for detection of wine spoilage thresholds ........ 31

Abstract ........ 31

Value of the Data ........ 31

Specifications Table ........ 32

Data ........ 33

Experimental Design, Materials, and Methods ........ 33

Acknowledgments ........................................................................................................37

References ...............................................................................................................37

4. Wine Quality Rapid Detection Using a Compact Electronic Nose System: Application Focused on Spoilage Thresholds by Acetic Acid ....................................................................38

4.1 E-nose for the Wine Industry.................................................................................38

4.2 Rapid detection Approach for E-Nose...................................................................39

4.3 Main Contributions Related to the Wine Quality Rapid Detection Approach. .............40

4.4 Published paper in LWT journal: Wine quality rapid detection using a compact electronic nose system: application focused on spoilage thresholds by acetic acid.............................41

Abstract.....................................................................................................................41

4.4.1 Introduction .........................................................................................................41

4.4.2 Materials and methods.........................................................................................43

4.4.3 Results ................................................................................................................50

4.4.4 Discussion...........................................................................................................54

4.4.5 Conclusions ........................................................................................................55

Acknowledgments ........................................................................................................56

References ...............................................................................................................56

5. Validation of the rapid detection approach for enhancing the electronic nose systems performance ...............................................................................................................59

5.1 Deep Learning Models Used in E-Nose .................................................................59

5.2 Main Contributions Related to the Validation of the Proposed Rapid Detection Approach ...............................................................................................................................60

5.3 A preprint of the research paper published at ArXiv.org and pendant to submission for a scientific journal: Validation of the rapid detection approach for enhancing the electronic nose systems performance, using different deep learning models and support vector machines ...............................................................................................................................61

5.3.1 Introduction .........................................................................................................62

5.3.2 Materials and methods.........................................................................................63

5.3.3 Results and Discussion .........................................................................................68

5.3.4 Conclusions ....................................................................................................... 72

Acknowledgments ....................................................................................................... 73

References ....................................................................................................... 73

6. Bibliographical Production....................................................................................................... 75

7. Conclusion....................................................................................................... 77

References ....................................................................................................... 79

# CHAPTER 1

## 1. Introduction

The senses are the physiological mechanisms of perception that enable humans as well as animals to perceive the surrounding environment (GUERRINI et al., 2017). Bio-inspired by the olfactory system, the development of artificial devices that combine chemical sensor array with pattern recognition techniques, commonly termed as "electronic nose" (E-Nose), have been explored for recognition and sensing of Volatile Organic Compounds (VOCs).

Real-time gas classification is an issue and challenge in settings such as food and beverage quality control, accident prevention in industrial environments, among others. One way to afront the mentioned issue is by using E-Nose. So far, the conventional approach for data processing in the Electronic Nose implies using the entire response curves of the gas sensors array, including the rising state, steady-state, recovery phases, and even others. Besides, this approach includes steps such as signal preprocessing and feature generation/extraction before performing the classification tasks, which entails the selection of a suitable method for each stage, increasing the necessary time to find the appropriate classification and forecasting models (LIU; MENG; ZHANG, 2018a; QI; MENG; ZENG, 2017a).

Nowadays, some researches have focused their energies on reducing the steps and the essential know-how for model generation. Such is the case of the work presented by (LIU; ZENG; MENG, 2019a), which the authors proposed a bio-inspired data processing method based on a neural network to imitate the mammalian olfactory system with great results, but using the whole measurement curves. In the same way, the authors by (LÄNGKVIST et al., 2013a) proposed a rapid detection system for meat spoilage using an unsupervised technique (i.e., stacked restricted Boltzmann machines and auto-encoders) that considers only the transient response. Although, in this case, the obtained models allow benefits because the features are learned from data instead of being hand-designed, it may provide low suitable and inaccurate models due to the unsupervised method. Moreover, some authors have explored an approach based on raw data treatment (PENG et al., 2018a; WEI et al., 2019a). Although this approach reduces the steps and the development time, they only tested with the whole response curves, requiring to fulfill all measurement processes. So, it can take a long time to get results.

Concerning the earlier mentioned, we proposed a novel approach on (RODRIGUEZ GAMBOA et al., 2019a), based on processing an early portion of raw signals (while the measurement process is still running.) We also tested the proposed method in a wine quality application, with outstanding results against the conventional methodology. A deep MLP classifier was trained with the raw data acquired from an E-Nose composed of an array of six MOX gas sensors. We achieved results 63times faster (Equation 1) compared with a conventional approach (using the entire response curves, applying preprocessing techniques to extract the features and later processing them using an SVM algorithm.)

$$relation\ of\ measurement\ time = \left(\frac{measurement\ time\ from\ the\ starting\ gas\ injection\ to\ the\ finish}{necessary\ time\ for\ making\ a\ forecast\ or\ window\ size}\right) \quad \text{Eq. (1)}$$

Support Vector Machines (SVM) is one of the most applied techniques for classification in E-Nose. Other used methods are K-Nearest Neighbor (KNN), Naive Bayes (NB), Linear Discriminant Analysis (LDA), and Adaptive Resonance Theory Map (ARTMAP)(JHA et al., 2019). However, the more novel methods are based on neural networks, implementing variations about the architectures, learning techniques, for instance. In recent years the deep learning DL algorithms have appeared at the electronic nose field, such as the case of (LIU; MENG; ZHANG, 2018a; PENG et al., 2018a; QI; MENG; ZENG, 2017a; WEI et al., 2019a), where the authors have investigated the Convolutional Neural Networks (CNN) in different contexts.

The main objective of this thesis was to propose and validate a rapid detection approach suitable to be applied in diverse E-Nose settings. Also, performing several experimental trials with different deep learning techniques such as Convolutional Neural Networks (CNN) against a more classical method like SVM for classification tasks in E-Nose using the proposed approach.

## 1.1 The sense of smell

The senses are the physiological mechanisms of perception that enable humans as well as animals to perceive the surrounding environment (GUERRINI et al., 2017). The olfaction is a primal sense that allows vertebrates and other organisms with olfactory receptors to identify food, mates, predators and provides pleasure as well as warnings of danger via emission and detection of volatile chemical compounds (JHA et al., 2019; LEFFINGWELL, 2002). The smell sense (**Figure 1**) is essential in the interaction and communication among animals and even humans and deserves to be studied in several ways. One example of its importance is the olfactory receptors found in ants and how such mechanisms are essential in social interaction (PASK et al., 2017).

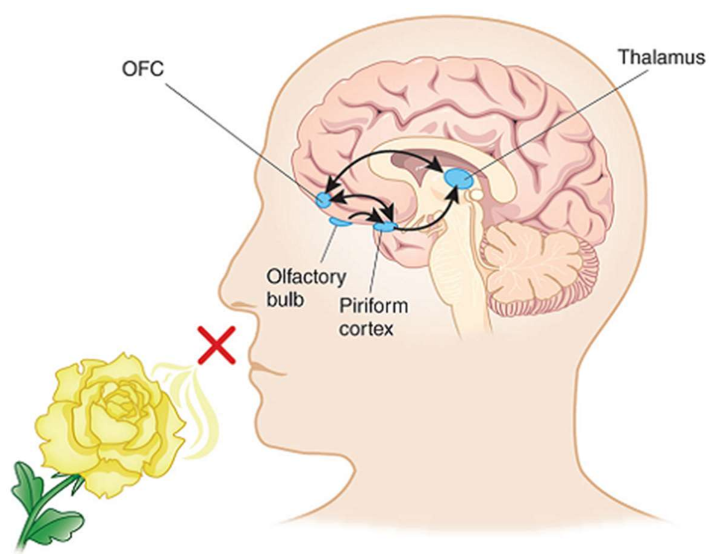


**Fig. 1.** After receiving smell information from the nose's sensory receptors, the olfactory bulb relays the information to a circuit of brain regions for processing. **Source:** BrainFacts.org, Making Sense of Scents: Smell and the Brain, 2015.

## 1.2 The electronic nose systems

Bio-inspired by the olfactory system, the development of artificial devices that combine chemical sensor array with pattern recognition techniques, commonly termed as "electronic nose" (E-Nose), have been widely investigated for recognition and sensing of Volatile Organic Compounds (VOCs).

The electronic nose has a recent history, although its origins go back to the primitive system assembled in the 1960 decade. The term "electronic nose" and the first intelligent system of electronic smell did not appear until the second half of the 1980 decade (GARDNER; BARTLETT, 1994; WILKENS; HARTMAN, 1964). Its usage as chemical detectors devices is an emerging research area, playing a critical function by mimicking the biologic olfaction sense to recognize different smells. Its applicability in a wide range of fields and purposes includes environmental monitoring, disease diagnosis, public security affairs, agricultural production, food industry, among others (DURÁN-ACEVEDO et al., 2018; HU et al., 2018; JHA et al., 2019).

An E-nose system is a measuring instrument for identifying odors, based on a gas sensor array designed to detect variations in the presence of VOCs. The sensor array connected to a computer system with adequate pattern recognition software, it can perform odor classification of specimens according to their organoleptic characteristics. A comparison between a biologic smell sense and the classical electronic nose architecture is depicted in **Figure 2**. It is important to remark that the use of a sensor array instead of a single sensor helps to detect a wide variety of odors, making it possible to obtain the best depiction of the measured specimen (HINES; LLOBET; GARDNER, 1999; RODRÍGUEZ-GAMBOA; ALBARRACÍN-ESTRADA; DELGADO-TREJOS, 2011).

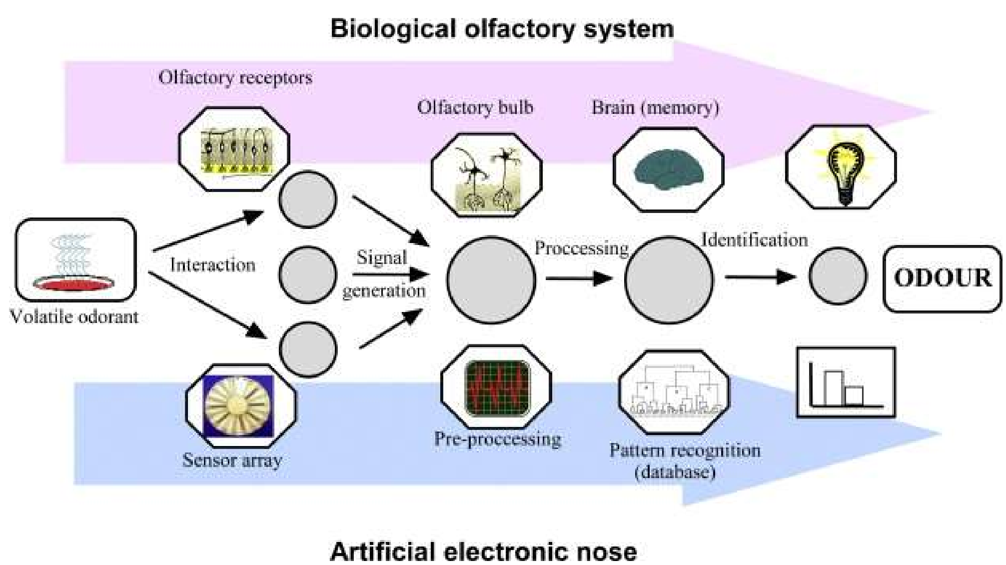


**Fig. 2** Similarity between the biological olfactory system and an e-nose. **Source:** M.Ghasemi-Varnamkhasti et al., Potential application of electronic nose technology in brewery, 2011.

## 1.3 The principal elements in an electronic nose system

An E-Nose comprises different modules that work together to recognize odors. An instrument of this type has at least three parts, each one with specific functions detailed below.

### 1.3.1 Gas sensor array

The gas sensor array or the detection module is the main component of an electronic nose. It is the reactive part of the device, and when the sensor array is in contact with VOCs, the gas sensors react by altering an electrical property such as electrical resistance (ODOTECH; OFF-SITE, 2013).

### 1.3.2 Adequacy of the gas mixture and sampling system

Sometimes the sample is manually injected into the sensors chamber. But in an automatic device, there is a module responsible for transporting volatiles emitted by the specimen towards the sensors chamber. Usually, this module involves air pumps that push the VOCs with the injection of inert gas or air to carry the gases (RODRÍGUEZ-GAMBOA; ALBARRACÍN-ESTRADA; DELGADO-TREJOS, 2011).

### 1.3.3 Processing system

The datalogger device is composed of both previous modules, and the collected data must be processed using pattern recognition techniques to identify the odor fingerprint. The variety of machine learning techniques to perform classification, mapping, or clustering tasks can be supervised, unsupervised, and reinforcement methods.

The most used supervised classification methods are Back-Propagation Neural Networks (BPNN), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Naive Bayes (NB), Linear Discriminant Analysis (LDA), and Adaptive Resonance Theory Map (ARTMAP), among others. An unsupervised method can group the chemical compounds using the sensor array response or their extracted features. The most common unsupervised techniques used in E-nose are the self-organizing maps (SOM), hierarchical cluster analysis (HCA), k-means clustering, and fuzzy clustering. The third category, based on reinforcement learning, does not require the obvious sensor response to a chemical compound and their class information. The operation is based on exploring the array response space in some tunable way, such as greedy search.

Usually, better class recognition efficiency is achieved with the SVM classifier compared to the mentioned classification methods. SVM method is extensively used for classification, feature extraction, clustering, outlier removal, and regression tasks in various disciplines. SVM is usually applied after performing preprocessing with PCA and KPCA, for instance (JHA et al., 2019).

## 1.4Objectives

### 1.4.1 General objective

Propose a novel approach to accelerate the way to make a forecast in an electronic nose system with different settings.

### 1.1.2 Specific objectives

- Define the electronic nose background, advantages and disadvantages, uses, challenges, and available databases.
- Identify an issue in the context that can be resolved using an E-Nose.
- Implement a solution for the selected topic.
- Propose a methodology seeking to accelerate the response time in an electronic nose.
- Assemble the proposed approach in a real environment (issue chosen).
- Compare the proposed rapid detection approach against the conventional method in the context.
- Validate the proposed fast detection approach against the traditional approach in different E-Nose settings.

## 1.5 Research Roadmap

The following sections explain the research work conducted during the doctoral studies. Chapter two describes the electronic nose device developed for agroalimentary applications, showing the details and characteristics of that system with enough depth. Chapter three reports the real use of O-Nose and the database collected using wines and ethanol samples. Section four describes the motivation to realize the wine quality control application and the proposed rapid detection approach for the electronic nose field. Chapter five explains the validation of the proposed method using several datasets and the comparison among diverse classifications techniques, including SVM and Deep Learning architectures. Section six shows the bibliographical productions and the relation with this thesis. The published or submitted articles were accommodated according to the related part of this thesis. Finally, the last sections display the conclusion and references for this work.

The research roadmap depicted in **Figure 3** summarized the research process split into three lines: the electronic nose device development, the wine quality control application, and the proposed rapid detection approach, displaying the products and results of each research line. The mentioned results are joined to realize a new contribution, the validation of the proposed rapid detection approach.

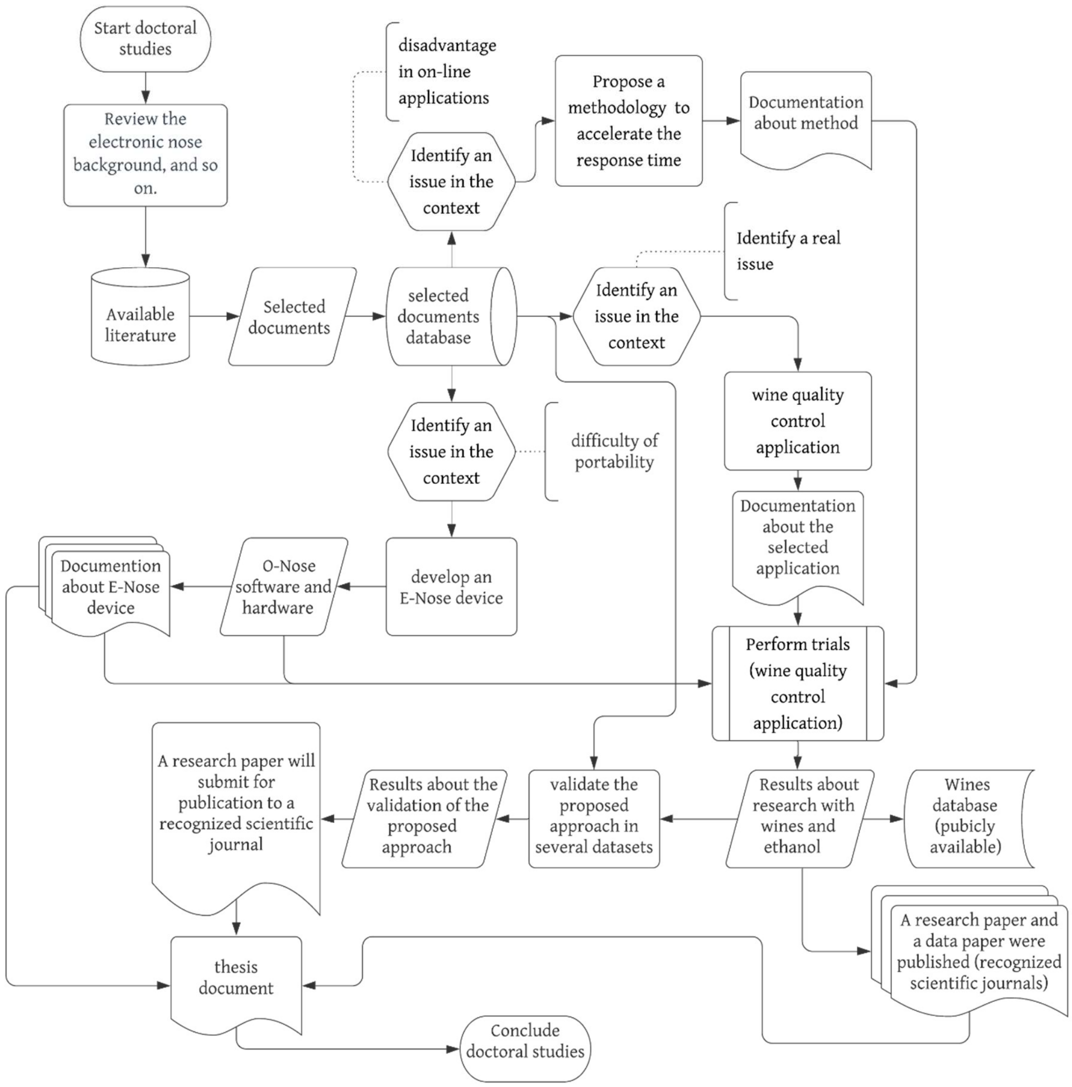


**Fig. 3** Research roadmap during the doctoral studies.

# CHAPTER 2

## 2. Portable and Compact E-Nose for Agroalimentary Purposes

The present electronic nose called O-Nose is mainly focused on a portable and compact device with low complexity for agroalimentary applications, able to work standalone, or with cloud computing functions.

### 2.1 Background

In the past few decades, many studies have been conducted to develop systems for artificial olfactory. Some of those systems are mass spectrometry (MS) based instruments that present the difficulty of portability, which is a disadvantage in on-line or in situ applications (MARTÍ et al., 2005; ZHANG et al., 2008). Secondly, "Most electronic nose systems are based only on one type of gas sensors. However, different types of sensors have different advantages and disadvantages. For example, the response of semiconductor gas sensors has a large drift, slow recovery time, but are sensitive to virtually almost any volatile substance. The amperometric sensor has much smaller drift and faster recovery time but is generally less sensitive" (JASINSKI et al., 2018). Another significant disadvantage of MOS gas sensors devices is "the logarithmic dependence of the sensor response on the gas concentration, causing problems in the presence of high concentrations of detectable species (e.g., ethanol in alcoholic beverages). Other potential problems have also been reported when MOS devices are used with food products (e.g., the baseline recovery is slow when the sensors are exposed to high molecular weight compounds" (MARTÍ et al., 2005). The researchers try to deal with the mentioned issues, generating in some cases systems with high power consumption, incrementing the instrumentation complexity, involving the use of additional equipment, as well as procedures or reagents, which increase the time and cost of analysis and require the presence of experienced personnel. In line with this scenery, in recent years, the efforts have addressed to develop electronic nose systems to solve the mentioned limitations.

Looking to overcome the mentioned limitations, in the literature exists some researches that disclosed methods based on wireless communication between the sensor system and the data process unity to improve portability. However, the complexity of the instrumentation and the necessity of additional equipment in most of them decreases the expected reach. Such is the case of the patent (刘云翔刘天伟姚文斌杜駉骏郑力刘续博李晓丹, 2016), which disclosed a data acquisition system for identifying the quality of edible spices, including a network transmission module for connecting the end of the PC and the ARM board (the acquisition and preprocessing system of the gas sensing devices.) As well, the patent (王俊姜水, 2016) disclosed an invention in the quality inspection of agricultural products field for monitoring the transport process of fruits and vegetables using a wireless electronic nose system that comprises a monitor node and a mobile terminal. The monitor node includes a control chip, a gas sensor array module, a signal processing module, a serial to Wi-Fi module, power modules, a real-time clock module, and an SD card storage module. The mobile terminal is a mobile device with Wi-Fi network functions connected through a serial to the Wi-Fi module to establish a wireless network monitoring node.

Moreover, in (刘步中陈亮李朝林徐耀庄海军杨永徐通, 2016) was disclosed a detection system for environmentally harmful gases in a mine based on wireless sensors. The authors claim that the system can include a plurality of detection units and send the data to a central computer through radio communication with the detection unit interface uploaded to the testing center computer. In another way, a possible solution to improve portability seems to use embedded devices to acquire, to preprocess and process the information in the electronic nose system. For instance, the patent (郑丽敏杨璐任发政朱虹田立军詹小琳王智凝, 2015) that disclosed an embedded electronic nose detection system and method. This approach is interesting, but the main limitation is the reduced capacity of the memory of this kind of device. A possible solution for this limitation could be to use cloud computing services. In this way, the patent (田逢春谢鑫张健吕博黄扬帆陈建军杨先一廖海林陈小娟, 2016) disclosed an electronic nose system and method for air quality monitoring based on cloud computing. The system is provided with a cloud terminal, a cloud monitoring center, and a mobile phone client.

Although this approach is fascinating, the additional equipment reduces its portability and increase the power consumption and system cost.

## 2.2 Description of the E-Nose Device

The developed device aims to provide a method for developing an artificial olfactory system with general purposes more specifically for agroalimentary applications with a compact and portable design, low complexity, reduced power consumption, lower price, and able to work standalone with internet access using cloud computing.

This compact and portable electronic nose system uses state-of-the-art low cost embedded devices, Arduino Nano, and Node-MCU ESP8266. Note that the mentioned devices are only an example, and this device is not limited to them, other replacement with similar o better characteristics could be used for this purpose. This approach is thought for the Internet of Things (IoT) and agroalimentary applications, with the advantage of decentralizing the measurement process in E-Nose and making possible the data access without the permanently connected to a computer system, such as it occurs in a classic device.

The prototype of this design was developed thinking to be used for general purpose applications and was tested in wine quality control to detect some defects in wine samples. In this case, we use 1 ml samples for each measure. We made several measures of different wines with defects, and without defects, then it was trained a classifier aiming to detect those defects in samples of wine. The compact and portable design of the proposed device has the advantage of easy transportation. This lets to be used in vineyards laboratories, warehouses of wine, customs to detect adulterated wines, restaurants, supermarkets, and other final users. The low power consumption allows being used in places with adverse conditions using batteries. The cloud computing functions allow access to the information from any place and with the possibility of using pattern recognition technics that can be updated more frequently because the data is processed in a cloud server. Useful for industries or customers that use several devices in different locations and with the benefit of more users means sharing knowledge.

The current E-Nose is based on a measurement process with three stages, as shown in **Figure 4** and detailed below:

1. Concentration: It is the first stage of the measurement process, aiming to concentrate the specimen volatiles inside the concentration chamber.

2. Data acquisition or measuring: It is the second stage of the process, aiming to acquire the signals from the sensors.

3. Purge: It is the third and last stage of the measurement process. The goal is to clean and remove volatile residues of the previous measure.

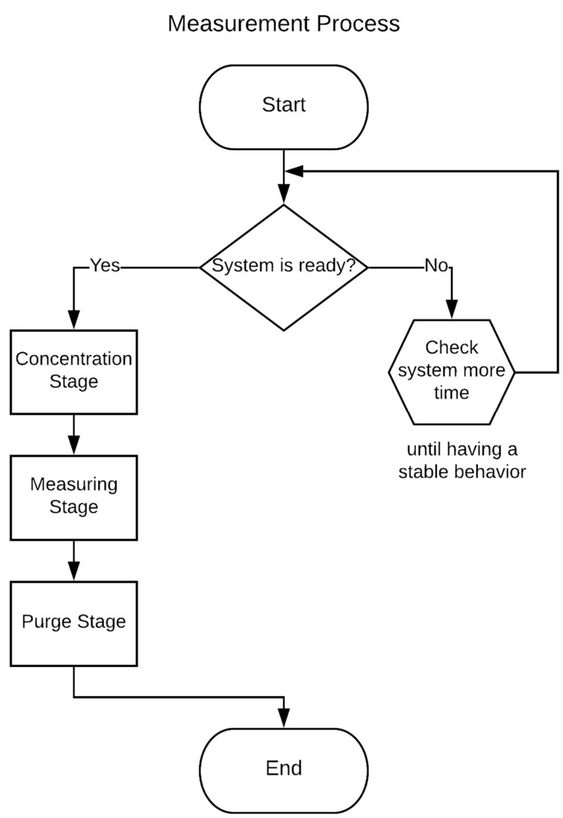


**Fig. 4** Measurement Process showing the three stages (concentration, measuring and purge.)

The present device has three modes of operation shown in **Figure 5** and detailed below:

Mode 1. The system can work traditionally, linking the system with a computer that has the proper software to process the information and display the outcome of the measurement. In this operation mode, the user has access to a graphics interface to start and monitoring the entire process.

Mode 2. The system can work standalone, to make the entire measurement process: to acquire, to preprocess and to process the information and then return the outcome of the measurement, using only the embedded devices to process the information and a display device to show the output. In this operation mode, the user must press the start button to begin the measurement process.

Mode 3. The system can work as a data logger to acquire and preprocess the information in the electronic nose system, then to publish the information in a cloud server via state-of-the-art publish/subscribe protocols, for instance, MQTT protocol, wherein the information will be processed. The outcome is returned to the user. In this operation mode, the user must press the start button to begin the measurement process. However, whether the system has no internet access, the device switches to work in operation mode 2.

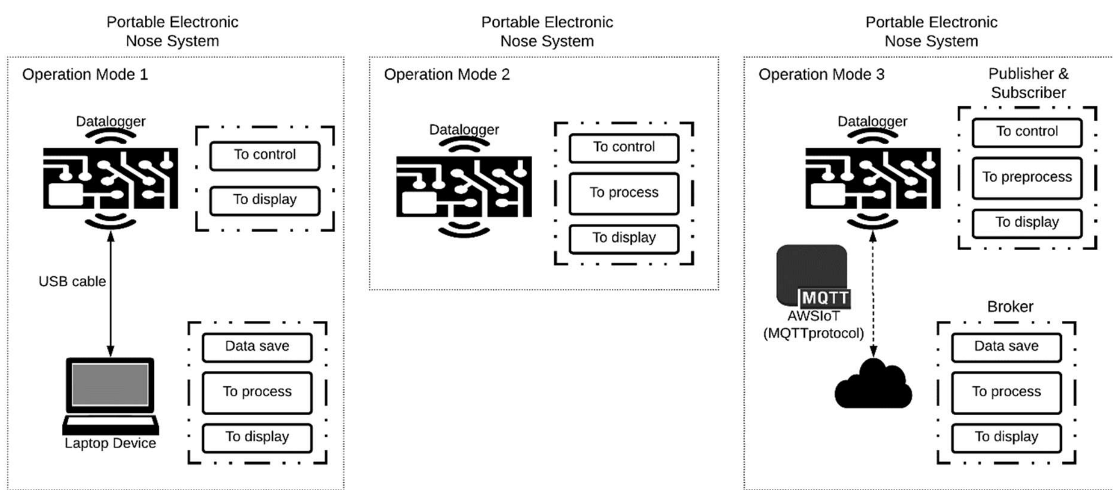


**Fig. 5** Operation modes of the system.

Whatever be the operation mode, the equipment has the same components and is adapted to work in any of the three modes.

This device has five parts, the concentration chamber, the sensors chamber, the actuators, the data logger circuit, and the power supply or battery.

## 2.3 Detailed Description of the E-Nose Device

The stages of the measurement process are concentration, data acquisition, and purge. The first stage is performed aiming to accumulate the analyte volatiles inside the concentration chamber, to achieve this, it must be activated the valve one and deactivated the valve two for isolating the concentration chamber interior of the external environment, activating the air pump. In **Figure 6**, the dashed line indicates the airflow at this stage; a time less than 5 minutes usually is enough for this stage, in the mentioned wine application above, we use a time of 30 seconds.

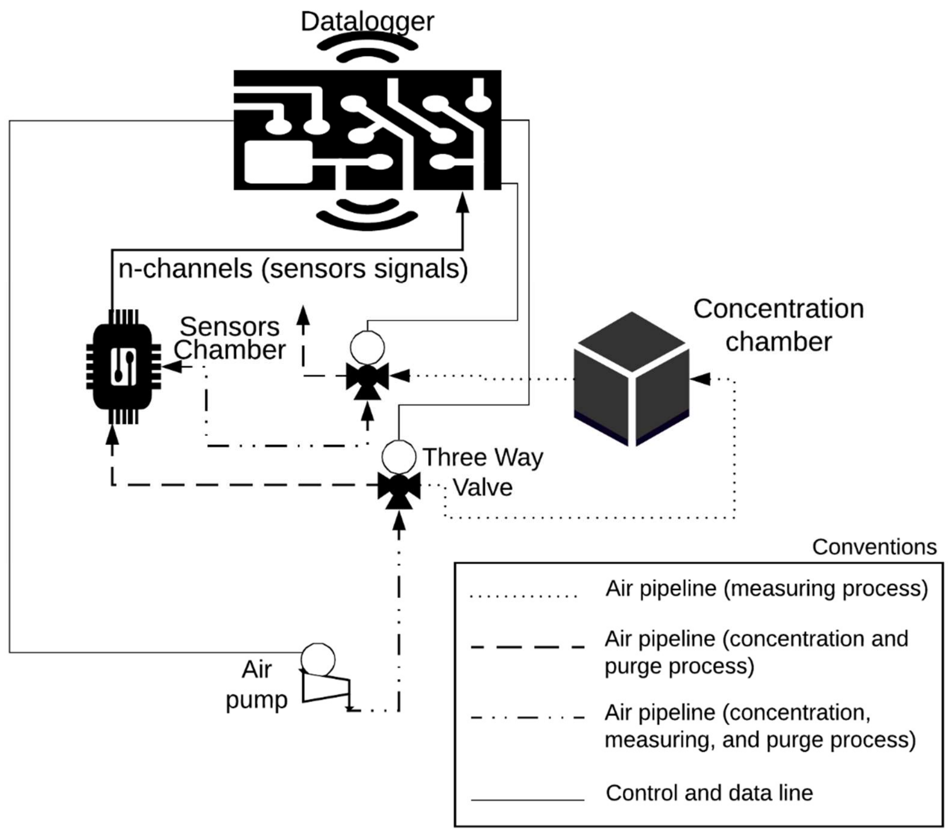


**Fig. 6** General schematic diagram for system stages.

The measuring stage is performed aiming to collect the signals from the gas sensors, to achieve this, it must be deactivated the valve one and activated the valve 2, activating the air pump to direct the airflow from the concentration chamber dragging the volatiles towards the sensors chamber. In **Figure 6**, the dotted line indicates the airflow at this stage; a time about 3 minutes usually is enough for this stage. Precisely 3 minutes was the time used in the wine application for this stage.

The goal of the purge stage is to clean and remove volatile residues from the previous measure, and to achieve this, it must be activated the valve one and deactivated the valve two, the same way that for the concentration stage activating the air pump. In **Figure 6**, the dashed line indicates the airflow at this stage; for this stage, a time about 10 minutes usually is enough. Precisely 10 minutes was the time used in the wine application for this stage.

The variables containing the times of each stage are saved in the embedded device nonvolatile memory EEPROM and can be updated using the graphical user interface in the computer software or via state-of-the-art publish/subscribe protocols, changing the values of each variable. When the system is running in the operation mode 3, the system takes the time values from the corresponding variables in the server and update the variables in the embedded device.

The concentration chamber (**Figure 7**) is a container with two holes to connect the pneumatic hoses, for instance, Urethane SHM 4, note that the mentioned references are only an example, and this device is not limited to them. The container could have a volume of 50 ml or higher. In the experimental process were tested 50 ml, 100 ml and 200 ml containers, the volume of the selected container depends on the specific application and its size, for the mentioned wine application we use the container of 100 ml. The hoses are connected each one to a three ways solenoid valve (**Figure 6**), to isolate the chamber in the concentration stage and to transport the volatiles in the measuring stage toward the sensors chamber.

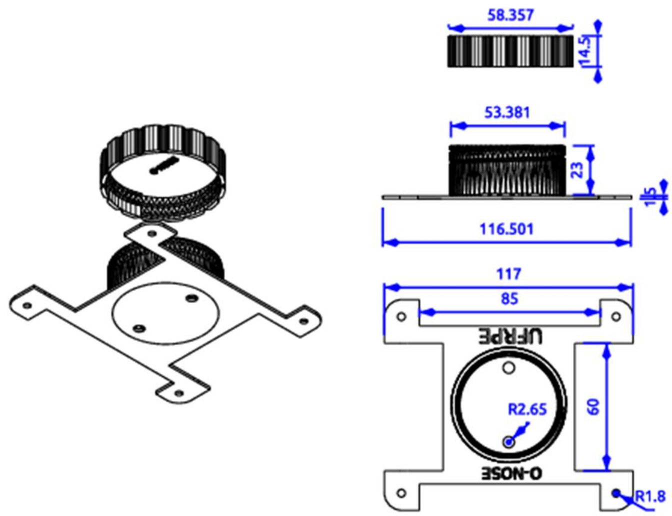


**Fig. 7** Concentration chamber.

The sensors chamber is a container with two holes to connect the pneumatic hoses. The container could have a volume of 150 ml or higher. In the experimental process were tested 150 ml, 200, and 300 ml containers, **Figure 8** shown the 200 ml container used in the mentioned wine application. The hoses are connected each one to a three ways solenoid valve to clean the chamber in the concentration and purge stages and to transport the volatiles in the measuring stage from the concentration chamber (**Figure 7**). In this chamber, the circuit board acts as a cover, which is fixed with the top of the sensors chamber, this to guarantee the airtightness. Some screws are required to fix the bottom and top parts of the sensors chamber.

The actuators proposed to use in this E-Nose are two 3-ways solenoid valves and one air pump (could be two for increment the airflow power). For instance, the ZHV 0519 reference for the solenoid valves, and the PM201U reference for the mini air pump, these actuators work with +5 VDC, the same way of all elements in the system.

**Figure 9** shows the use of MOSFET transistors to control the actuators and gas sensors; this reduces the power consumption and size of the circuit board. The PWM signals applied to the gas sensors allow controlling the voltage in the heater element, this for experimental purposes and to reduce the power consumption (this option can be suppressed by reducing the price and size even more).

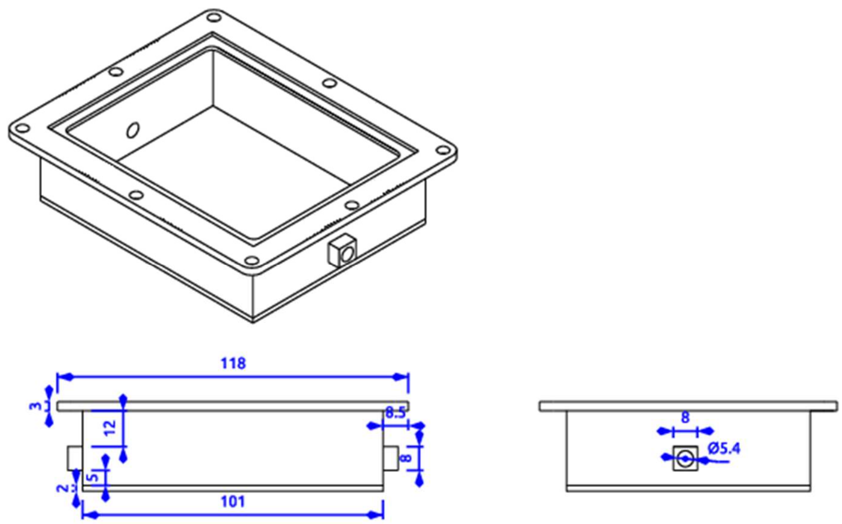


**Fig. 8** Sensors chamber.

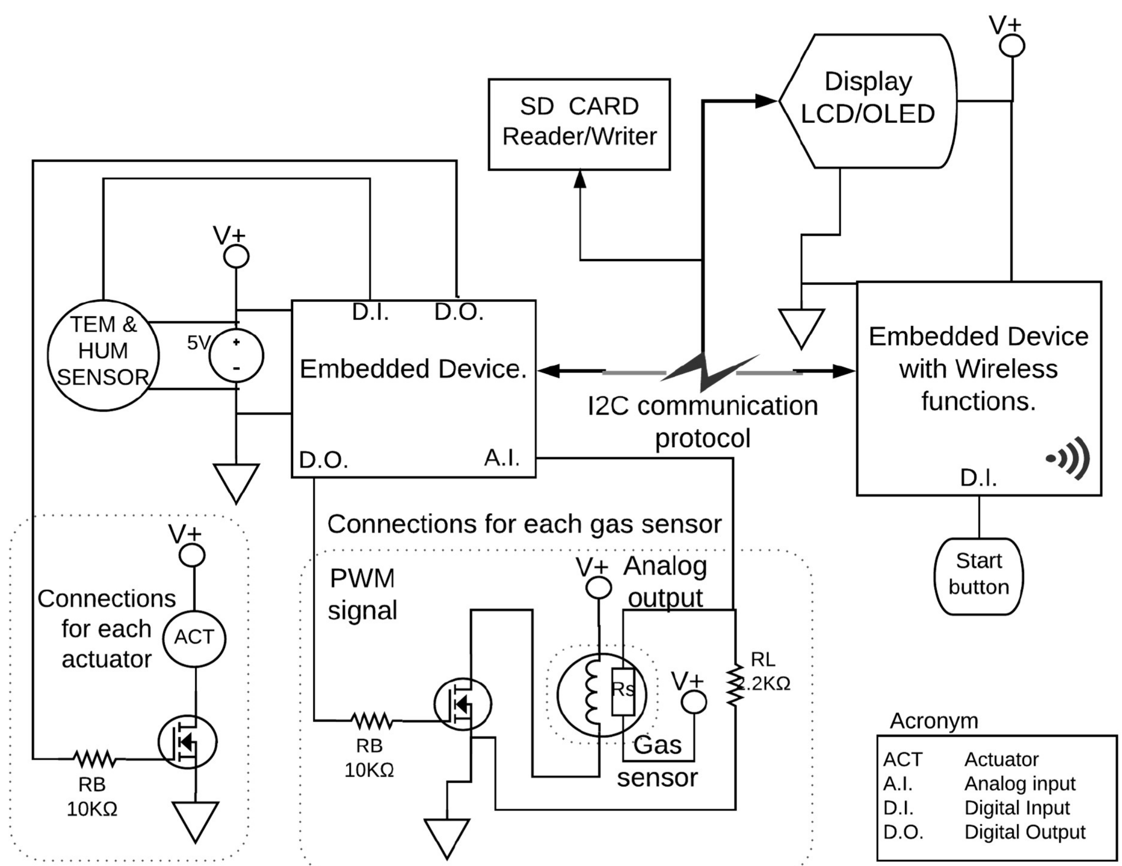


**Fig. 9** Inside the Data logger circuit.

The data logger circuit board has an embedded device (**Figure 9**), as the main component, this device is responsible for acquiring the gas sensors electrical signals and the temperature and humidity sensor signal. Also, it must control the voltage of the heater element in the gas sensors using PWM, the timing control for the solenoid valves and the air pump, as well to establish the communication with the computer (to process the information and display the outcome in the operation mode 1) and with another embedded device with Wireless functions.

This second embedded device must preprocess the information and then publish in the cloud server in the operation mode 3, or even to process the information and display the outcome in the operation mode 2. The two mentioned embedded devices could be replaced by a unique embedded device with the minimum requirements: analog inputs (one for each gas sensor), digital outputs (at least 3 for the actuators, one more for each sensor if the PWM wants to be implemented), digital inputs (at least 2 for the temperature sensor and another for the start button), available ports for communicates with the display and the SD card reader/writer using a state-of-the-art protocol, for instance, I2C communication protocol, and wireless communication via Wi-Fi and publish/subscribe protocols.

The user can start the measurement process by pressing the start button shown in **Figure 9** or launching the process from the software on the computer. If the user presses the start button, the system tries to work in the operation mode 3, but whether it is not possible the internet connection, the system will be running in operation mode two. The user can launch the process from the computer, choosing the operation mode via the graphical user interface.

In operation mode three, the information is preprocessed in the embedded devices before sending it to the server, reducing the quantity of information. For instance, instead of sending all sampling values of each sensor, it is better to send the extracted characteristics from each sensor. This reduces the transmission time and optimizes the storage capacity. Then, these characteristics are used as the inputs of the pattern recognition software, e.g., an SVM classifier or a neural network running on a cloud server.

The power supply must provide the voltage and current necessary for the operation of the system. In this case, the three actuators used in total consume approximately 1000mA, of which, 800mA correspond to the current consumed by the two solenoid valves and 200mA required for the operation of the air pump. Additionally, the current consumption of each of the sensors is estimated at approximately 150mA, having an array of six gas sensors, these elements expect a current consumption of about 1A, and finally, an additional 100mA is estimated in the other components of the system. In consequence, the power supply must provide the circuit with a minimum of 2.1A holding 5 VDC in this setup for its correct operation.

In accordance with what was explained in the previous paragraph and to guarantee the correct work of the present E-Nose, it is necessary that the power supply has the minimum features mentioned below.

Universal AC input range from 85Vac to 264VAC without power

Nominal Output Voltage 5 VDC

Output current > 3A

Output Power > 15W

For example, the power supply PMC-05V015W1AA of Delta Company has these minimum requirements.

The power supply could be replaced by one battery or a portable charger, increasing the portability. This battery must supply 5V and minimum 3A, for example, the portable charger RAVPower 26800. Note that the mentioned devices are only some examples, and this E-Nose is not limited to them, other devices with similar o better characteristics could be used for these purposes.

## 2.4 Main contributions Related to Developed E-Nose

- - The developed E-Nose has the followings advantages: no reagents, no occupational risk, minimal training, and smooth operation when compared with other alternatives of the artificial olfactory systems.
- - The device could be used for general purposes but is focused on agroalimentary applications, and initially, it was tested in a wine quality control problem.
- - The E-Nose has greater portability due to the size and arrangements.
- - The device has three useful modes of operations.

# CHAPTER 3

## 3. Electronic Nose Dataset for Detection of Wine Spoilage thresholds

The recorded database corresponds to time series obtained for an application of wine quality detection focused on spoilage thresholds, containing 235 recorded measurements of wines divided into three groups and labeled as high quality (HQ), average quality (AQ) and low quality (LQ), in addition to 65 ethanol measurements. Which was collected using an electronic nose based on Metal Oxide Semiconductor (MOS) gas sensors, self-developed at the *Universidade Federal Rural de Pernambuco* (Brazil). A data paper with the details about was published in Data in Brief journal (RODRIGUEZ GAMBOA et al., 2019b), and also can be accessed publicly at the repository: (RODRIGUEZ GAMBOA et al., 2019c).

We used 22 bottles of commercial wines of different varieties and vintages, elaborated in four wineries of the *São Francisco* valley (Pernambuco-Brazil). The spoiled samples obtained from 13 of the 22 bottles were randomly selected and left open for six months before starting the measurements (low- quality LQ wines). Besides, four bottles were opened two weeks before beginning the data collection (average-quality AQ wines), and the remaining five bottles opened at the starting time of each measurement (high-quality HQ wines). Also, we measured isolated ethanol in concentrations (v/v): 2, 5, 10, 20, 30, and 40ml of ethanol diluted in distilled water to make solutions of 200 ml. These concentrations allow guaranteeing a range that covers the different possible values in wines with and without spoilage. To ensure the repeatability of the experiments using O-NOSE, we collected between 10 and 11 samples of 1mL of each wine bottles, and around 10 and 12 of the ethanol samples at their different concentrations. Therefore, the database contains 235 wines measurements divided into three groups: high quality (HQ), average quality (AQ), and low quality (LQ), with 51, 43, and 141 measurements, respectively, and 65 ethanol measurements (RODRIGUEZ GAMBOA et al., 2019b).

## 3.1 Main contributions related to collected database

- The database is available as a benchmark of E-Nose applications, focused on wine spoilage thresholds studies.
- This database is useful for testing classifiers and pattern recognition methods with comparison purposes in studies related to E-Nose applications, including but not limited to food quality monitoring.
- To the best of our knowledge, this database is the first one publicly available regarding commercial wines measurements acquired with E-Nose.
- These data are suitable to support E-Nose applications, helping in the decision-making of winemakers and consumers in routine tasks of wine quality control.

## 3.2 Published paper in Data in Brief Journal: Electronic nose dataset for detection of wine spoilage thresholds

**Authors:** Juan C. Rodriguez Gamboa[1], Eva Susana Albarracin E.[1], Adenilton J. da Silva[2], Tiago A. E. Ferreira[1]

**Affiliations**

[1] Departamento de Estatística e Informática, Universidade Federal Rural de Pernambuco - UFRPE, Recife, PE, Brazil
[2] Centro de Informática, Universidade Federal de Pernambuco - UFPE, Recife, PE, Brazil

**Corresponding author(s):** Juan C. Rodriguez Gamboa (juan.gamboa@ufrpe.br)

**Abstract**

In this data article, we provide a time series dataset obtained for an application of wine quality detection focused on spoilage thresholds. The database contains 235 recorded measurements of wines divided into three groups and labeled as high quality (HQ), average quality (AQ) and low quality (LQ), in addition to 65 ethanol measurements. This dataset was collected using an electronic nose system (E-Nose) based on Metal Oxide Semiconductor (MOS) gas sensors, self-developed at the Universidade Federal Rural de Pernambuco (Brazil). The dataset is related to the research article entitled "Wine quality rapid detection using a compact electronic nose system: application focused on spoilage thresholds by acetic acid" by Rodriguez Gamboa et al. [1]. The dataset can be accessed publicly at the repository: https://data.mendeley.com/datasets/vpc887d53s/

**Keywords**

electronic nose; chemical sensing; machine learning; beverage quality control; wine spoilage

**Value of the Data**

- The dataset is available as a benchmark of E-Nose applications, focused on wine spoilage thresholds studies.
- This dataset is useful for testing classifiers and pattern recognition methods with comparison purposes in studies related to E-Nose applications.
- To the best of our knowledge, this dataset is the first one publicly available regarding commercial wines measurements acquired with E-Nose.

- These data are suitable to support E-Nose applications, helping in the decision-making of winemakers and consumers in routine tasks of wine quality control [2].

## Specifications Table

| | |
|---|---|
| **Subject** | Food Science; Computer Science Applications; Signal Processing |
| **Specific subject area** | Wine quality assessment using electronic nose technology |
| **Type of data** | Text files |
| **How data were acquired** | By using an electronic nose system (E-Nose) based on six Metal Oxide Semiconductor (MOS) gas sensors (MQ-3, MQ-4, MQ-6; two of each type). |
| **Data format** | Raw data, time series data |
| **Parameters for data collection** | In each experiment was used a 1 ml sample to amass the volatiles during 30 seconds inside the concentration chamber. The recorded data for each measurement corresponds to 180 seconds with 18.5 Hz sample rate. Then, the sensors were exposed to clean air for 600 seconds after the sample presentation. |
| **Description of data collection** | We collected wine samples categorized into three spoilage thresholds: low-quality (LQ), average-quality (AQ), and high-quality (HQ). In addition, we collected ethanol measurements in concentrations of 1%, 2.5%, 5%, 10%, 15%, and 20% (v/v). |
| **Data source location** | Institution: Universidade Federal Rural de Pernambuco<br><br>City/Town/Region: Recife, PE<br><br>Country: Brazil<br><br>Latitude and longitude (and GPS coordinates) for collected samples/data: Latitude: 8° 1' 2.68" Longitude 34° 56' 52.211" (Latitude: -8.017852 \| Longitude: -34.94785) |
| **Data accessibility** | Repository name: Mendeley Data<br>Data identification number: DOI: 10.17632/vpc887d53s.3<br>Direct URL to data: https://data.mendeley.com/datasets/vpc887d53s/ |
| **Related research article** | J.C. Rodriguez Gamboa, E.S. Albarracin E., A.J. da Silva, L. L. de Andrade Lima, T.A. E. Ferreira, Wine quality rapid detection using a compact electronic nose system: application focused on spoilage thresholds by acetic acid, LWT - Food Science and Technology. 108 (2019) 377–384. doi:10.1016/j.lwt.2019.03.074. |

**Data**

The recorded time series was acquired at the sampling frequency of 18.5 Hz during 180 seconds, resulting in 3330 data points per sensor. Each file in the dataset has eight columns: relative humidity (%), temperature (°C), and the resistance readings in kΩ of the six gas sensors: MQ-3, MQ-4, MQ-6, MQ-3, MQ-4, MQ-6.

We organized the database in three folders for the wines: AQ_Wines, HQ_Wines, LQ_Wines, and one folder for the ethanol. Each folder contains text files that correspond to different measurements.

In the wines folders, each filename identifies a wine measurement as follows: the first 2 characters of the filename are an identifier of the spoilage wine threshold (AQ: average-quality, HQ: high-quality, LQ: low-quality); characters 4-9 indicate the wine brand; characters 11-13 indicate the bottle, and the last 3 characters indicate the repetition (another sample of the same bottle). For example, file LQ_Wine01-B01_R01 contains the time series recorded when low-quality wine of the brand 01, bottle 01, sample 01 was measured.

In the Ethanol folder, each filename identifies an ethanol measurement as follows: the first 2 characters of the filename are an identifier of Ethanol (Ea); characters 4-5 indicate the concentration in v/v (C1: 1%, C2: 2.5%, C3: 5%, C4: 10%, C5: 15%, C6: 20%); and the last 3 characters indicate the repetition. For example, file Ea-C1_R01 contains time series acquired when Ethanol at 1% v/v of concentration, sample 01 was measured.

In **Fig. 1**, we depicted the time series for several measurements collected in this work. The measurements displayed at the top of the figure are in resistance units (Ω), and at the bottom side are the same measurements in conductance units (S).

**Experimental Design, Materials, and Methods**

*Experimental setup*

The dataset was collected with an E-Nose self-developed, that was named O-NOSE. We designed the datalogger for operating linked to a computer that has the proper software for data recording and processing, as shown in **Fig. 2**.

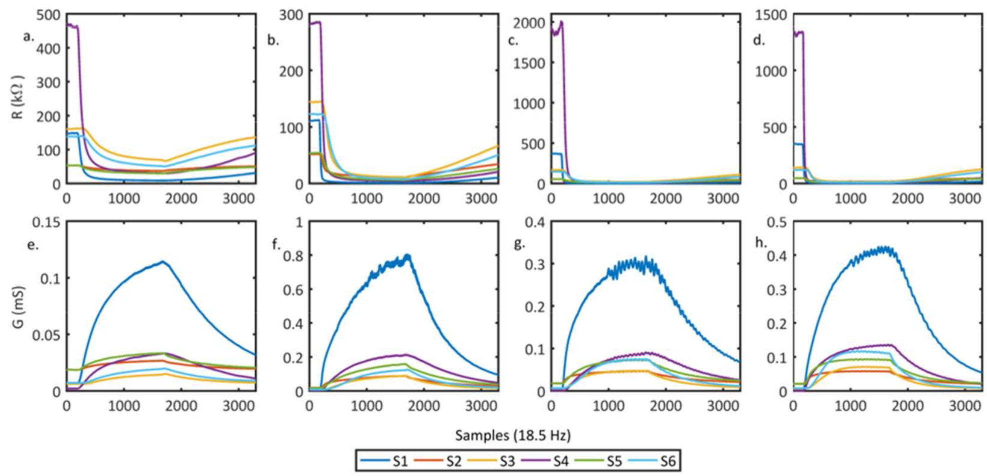


**Fig. 1.** Measurements acquired with our E-Nose, where S1, S2,…, S6 represent the gas sensors outputs; a. and e. correspond to the dataset file EaC1R10 (ethanol measurement); b. and f. correspond to LQWine02B01R09 dataset file; c. and g. correspond to AQWine01B01R07 dataset file; d. and h correspond to HQWine05B01R01 dataset file.

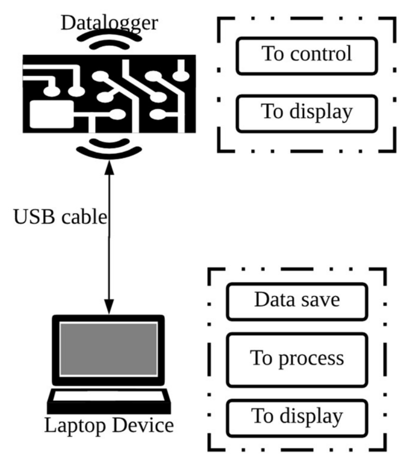


**Fig. 2**. Operating general diagram of O-NOSE system.

The operating mode of O-NOSE is depicted with more details in Fig. 3. The device contains two mini three-way solenoid valves ZHV 0519, and two mini air pumps PM201U (these actuators work with +5 VDC in the same way of all elements in the system) controlled by an embedded device: microcontroller Arduino Nano. The microcontroller takes charge of the data acquisition from the gas sensors and the temperature and humidity sensor DHT11 located into the sensors chamber. As well, of the timing control of the solenoid valves and the air pump, and the communication with the computer.

It used a 100 ml concentration chamber, where is placed the specimen to be analyzed. The sensor array of six MOS gas sensors manufactured by Hanwei Sensors (MQ-3, MQ-4, and MQ-6; two of each) is located into a 200 ml chamber connected to pneumatic hoses that carry the volatiles. The gas is sensed by its effect on the sensitive layer of tin dioxide ($SnO_2$), resulting from changes in conductivity brought about by chemical reactions on the surface of the tin dioxide particles [3,4].

The stages of the measurement process are concentration, data acquisition, and purge. The first stage aims to accumulate the analyte volatiles inside the concentration chamber for 30 seconds, to achieve this, the microcontroller activates the valve 1 and the air pump; simultaneously, deactivates the valve 2 for isolating the concentration chamber interior of the external environment. In Fig 3, the dashed line indicates the airflow at this stage.

The data acquisition stage that lasts three minutes aims to collect the signals from the gas sensors, to achieve this, the microcontroller deactivates the valve 1 and activates the valve 2 and the air pump to direct the airflow from the concentration chamber dragging the volatiles towards the sensors chamber. In **Fig 3**, the dotted line indicates the airflow at this stage.

The goal of the purge stage is to clean and remove volatile residues from the previous measurement during 10 minutes. Hence, the microcontroller activates the valve 1 and the air pump; simultaneously, deactivates the valve 2, the same way that for concentration stage. In **Fig 3**, the dashed line indicates the airflow at this stage.

*Measurement protocol*

O-NOSE performs the measurement process in three stages: concentration, data acquisition (the recorded data corresponds to 180 seconds with 18.5 Hz sample rate) and purge [1]. Each measurement corresponds to the time-dependent output voltages of each gas sensor converted to resistance values according to the voltage-divider scheme [5] and the corresponding load resistor ($R_L$). The sensor resistance ($R_S$) value changes when the gas sensor is exposed to a certain specimen and was calculated as follows:

$$R_S = \frac{V_C - V_{R_L}}{V_{R_L}} \times R_L \qquad (1)$$

$$V_{R_L} = \frac{ADC \times V_C}{1023} \qquad (2)$$

where $V_C$ , $V_{R_L}$, $R_L$, $ADC$ are the standard voltage of microcontroller (5V), the output voltage, sensor load resistor, and the Analog to Digital Converter (ADC) reading, respectively[5].

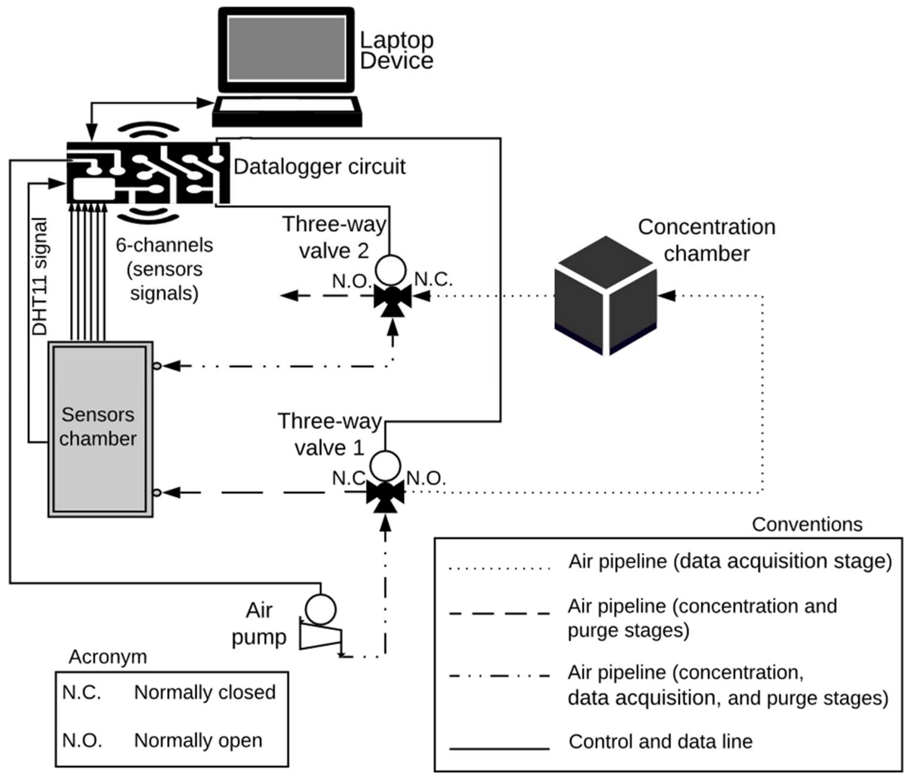


**Fig. 3**. Schematic diagram of O-NOSE displaying the operation stages.

*Samples*

We used 22 bottles of commercial wines of different varieties and vintages, elaborated in four wineries of the São Francisco valley (Pernambuco-Brazil). To obtain spoiled samples, 13 of the 22 bottles were randomly selected and left opened for six months before starting the measurements (low-quality LQ wines).Besides, four bottles were opened two weeks before beginning the data collection (average-quality AQ wines), and the remaining five bottles were opened at the starting time of each measurement (high-quality HQ wines) [1].

In addition to wines, we measured isolated ethanol in concentrations (v/v): 2, 5, 10, 20, 30, and 40 ml of ethanol diluted in distilled water to make solutions of 200 ml. These concentrations allow guaranteeing a range that covers the different possible values in wines with and without spoilage. To ensure the repeatability of the experiments using O-NOSE, we collected between 10 and 11 samples of 1mL of each wine bottles; and between 10 and 12 of the ethanol samples at their different concentrations. In this way, the database contains 235 measurements of wines divided into three groups: high quality (HQ), average quality (AQ) and low quality (LQ), with 51, 43, and 141 measurements, respectively, and 65 ethanol measurements [1].

## Acknowledgments

This work was supported by the Serrapilheira Institute (grant number Serra-1709-22626), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (CAPES) – Finance Code 001, and Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (grant number 420319/2016-6).

We want to thank Centro de Tecnologias Estratégicas do Nordeste (CETENE), Laboratório de Pesquisa em Infraestrutura Computacional (Juá Labs UFRPE), and Laboratório Interdisciplinar de Formação de Educadores (LIFE UFRPE) for contributing with the development of this project. Additionally, we thank Miolo Wine Group, Global wines, and Vale Do São Francisco winery for supplying the wines used in this work.

## References

[1] J.C. Rodriguez Gamboa, E.S. Albarracin E., A.J. da Silva, L. L. de Andrade Lima, T.A. E. Ferreira, Wine quality rapid detection using a compact electronic nose system: application focused on spoilage thresholds by acetic acid, LWT - Food Science and Technology. 108 (2019) 377–384. doi:10.1016/j.lwt.2019.03.074.

[2] M.L. Rodríguez-Méndez, J.A. De Saja, R. González-Antón, C. García-Hernández, C. Medina-Plaza, C. García-Cabezón, F. Martín-Pedrosa, Electronic Noses and Tongues in Wine Industry, Frontiers in Bioengineering and Biotechnology. 4 (2016) 1–12. doi:10.3389/fbioe.2016.00081.

[3] J. Rodríguez, C. Durán, A. Reyes, Electronic nose for quality control of Colombian coffee through the detection of defects in "Cup Tests," Sensors. 10 (2010) 36–46. doi:10.3390/s100100036.

[4] J.C. Rodríguez-Gamboa, E.S. Albarracín-Estrada, E. Delgado-Trejos, Quality Control Through Electronic Nose System, in: Modern Approaches to Quality Control, InTech, 2011: pp. 505–522. doi:10.5772/22217.

[5] D. Rahman, R. Sarno, E. Zulaika, Data in Brief Electronic nose dataset for beef quality monitoring in uncontrolled ambient conditions, Data in Brief. 21 (2018) 2414–2420. doi:10.1016/j.dib.2018.11.091.

# CHAPTER 4

## 4. Wine Quality Rapid Detection Using a Compact Electronic Nose System: Application Focused on Spoilage Thresholds by Acetic Acid

The food industry is one of the most common fields to use E-nose. There are published papers that focused on measuring specimens like meat, milk and dairy products, eggs, different grains, fruits, oils, alcoholic and non-alcoholic beverages (BERNA, 2010; EL BARBRI et al., 2008; RODRÍGUEZ; DURÁN; REYES, 2010). Food quality control is one of the many applications that can benefit from the use of an electronic nose. E-nose has potential as a useful instrument to recognize a type of product, and it can be used for classification by region, quality, time of ripening or storage, the food life span can be determined or predicted, as well as the level of deterioration or decomposition (RODRÍGUEZ-GAMBOA; ALBARRACÍN-ESTRADA; DELGADO-TREJOS, 2011).

### 4.1 E-nose for the Wine Industry

In the wine industry is crucial to have methods for real-time monitoring thresholds of acetic acid in wines, preventing its spoilage, or determining its quality (BURLACHENKO et al., 2016; LOUTFI et al., 2015). The most important technique used to assess wine quality is directly related to the evaluation of the organoleptic characteristics by trained experts (ALEIXANDRE et al., 2018; CRETIN; DUBOURDIEU; MARCHAL, 2018; SÁENZ-NAVAJAS et al., 2015). Since the analytical panels are expensive, time-consuming, and they are not always available, the wine is also characterized using gas and liquid chromatography or spectrophotometry, which require reagents and experienced personnel (MARTINS et al., 2018; PERESTRELO; RODRIGUEZ; CÂMARA, 2017; STUPAK et al., 2017; VAZALLO-VALLEUMBROCIO et al., 2017).

In this way, E-nose is an alternative to traditional methods for wines discrimination regarding the organoleptic characteristics. Their purpose is to analyze aroma profiles by registering signals produced by the mixture of gases and then comparing the responses patterns generated by different samples (LOZANO; SANTOS; HORRILLO, 2016; PERIS; ESCUDER-GILABERT, 2016; RODRÍGUEZ-MÉNDEZ et al., 2016; ZHAO et al., 2017). However, most E-nose are designed for general purposes, and sometimes its design does not allow to use on-site.

Accordingly, the issue mentioned above motivated the research about wine quality thresholds using a compact electronic nose system. The purpose of this application was the wine spoilage detection produced by acetic acid. We focused on developing a novel application and acquiring our dataset to work with original data, as was explained in the previous chapter of this document.

## 4.2 Rapid detection Approach for E-Nose

Although some E-nose appliances claim to perform real-time monitoring, the real approach to process the sensor array outcomes are achieved in an off-line way, because the measurements need to be completed before the system makes a forecast. The previous issue is significant, taking in count that the measurement process generally takes some minutes. This issue is a limitation when the idea is the massification of this technology and obtaining responses in a short time.

The traditional approach for data processing in E-Nose implies to use the entire response curves of the gas sensors array, including the rising state, steady-state, recovery phases, and even others. Besides, it includes steps such as signal preprocessing and feature generation/extraction, which involves the selection of a suitable method for each stage, increasing the necessary time to find a suitable classifier and forecast models (LIU; MENG; ZHANG, 2018a; QI; MENG; ZENG, 2017a).

In recent Literature, some researchers have focused their efforts on reducing the steps and the necessary know-how for model generation. For instance, in (LIU; ZENG; MENG, 2019a), the authors proposed a bio-inspired data processing method based on a neural network to mimic the mammalian olfactory system with excellent results but using the entire measurement curves. In another work (LÄNGKVIST et al., 2013a), the authors proposed a rapid detection system for meat spoilage using an unsupervised technique that considers only the transient response (stacked restricted Boltzmann machines and auto-encoders.) Although the obtained models offer advantages because the features are learned from data instead of being hand-designed, it may produce low suitable and inaccurate models due to the unsupervised method.

Further, in (PENG et al., 2018a; WEI et al., 2019a), the authors explored an approach based on raw data treatment. Although this approach reduces the steps and the development time, they only tested with the entire response curves, then must wait until the measurement procedure finalization.

Consequently, the mentioned issue motivated the research about a rapid detection approach for the electronic nose systems. We focused on processing an early portion of the signals to reduce the time for making forecasts, testing the proposed method in the collected database with wines samples.

## 4.3 Main Contributions Related to the Wine Quality Rapid Detection Approach.

- We provide an online solution for wine quality applications using E-NOSE.
- We accomplished reducing the time for forecasting of spoilage thresholds in a real wine appliance (as quick than 3 seconds after gas injection).

## 4.4 Published paper in LWT journal: Wine quality rapid detection using a compact electronic nose system: application focused on spoilage thresholds by acetic acid

**Juan C. Rodriguez Gamboa[a,b], Eva Susana Albarracin E.[a], Adenilton J. da Silva[b], Luciana Leite[c], Tiago A. E. Ferreira[a]**

[a]*Departamento de Estatística e Informática,* [b]*Departamento de Computação,* [c]*Departamento de Tecnologia Rural. Universidade Federal Rural de Pernambuco, Rua Dom Manoel de Medeiros s/n, Dois Irmãos, 52171-900 Recife, PE, Brazil*

**Abstract**

It is crucial for the wine industry to have methods like electronic nose systems (E-Noses) for real-time monitoring thresholds of acetic acid in wines, preventing its spoilage or determining its quality. In this paper, we prove that the portable and compact self-developed E-Nose, based on thin film semiconductor ($SnO_2$) sensors and trained with an approach that uses deep Multilayer Perceptron (MLP) neural network, can perform early detection of wine spoilage thresholds in routine tasks of wine quality control. To obtain rapid and online detection, we propose a method of rising-window focused on raw data processing to find an early portion of the sensor signals with the best recognition performance. Our approach was compared with the conventional approach employed in E-Noses for gas recognition that involves feature extraction and selection techniques for preprocessing data, succeeded by a Support Vector Machine (SVM) classifier. The results evidence that is possible to classify three wine spoilage levels in 2.7 seconds after the gas injection point, implying in a methodology 63 times faster than the results obtained with the conventional approach in our experimental setup.

***Keywords**:* beverage quality control; wine spoilage; online early detection.

### 4.4.1 Introduction

Wine flavor depends on 20 or more compounds, besides water and ethanol, that with subtle alterations in concentration determine its quality (JACKSON, 2008). The most important technique used to determine wine quality is directly related to the organoleptic characteristics evaluation by trained experts (ALEIXANDRE et al., 2018a; CRETIN; DUBOURDIEU; MARCHAL, 2018a; SÁENZ-NAVAJAS et al., 2015a). Since the analytical panels are expensive, time-consuming, and they are not always available, the wine is also characterized using gas and liquid chromatography or spectrophotometry, that require on reagents and experienced personnel (MARTINS et al., 2018a; PERESTRELO; RODRIGUEZ; CÂMARA, 2017a; STUPAK et al., 2017a; VAZALLO-VALLEUMBROCIO et al., 2017a). Besides, E-Noses are used as an alternative to traditional methods for wines discrimination regarding the organoleptic characteristics. Their purpose is to analyze aroma profiles by registering signals produced by the mixture of gases (as the human nose does) and then comparing the pattern of responses generated by different samples (LOZANO; SANTOS; HORRILLO, 2016a; PERIS; ESCUDER-GILABERT, 2016a; RODRÍGUEZ-MÉNDEZ et al., 2016b; ZHAO et al., 2017a). However, most E-Noses are designed for general purpose, and sometimes they are not portable to use on-site.

Volatile acidity (VA) measurements, generally interpreted as acetic acid content ($g \cdot l^{-1}$), are used routinely as an indicator of wine spoilage (ZOECKLEIN, B. W., FUGELSANG, K. C., GUMP, B. H., & NURY, 1995). Thereby, it is crucial for the wine industry and consumers to have methods for real-time monitoring of VA thresholds. There are previous works which the wine spoilage was characterized using E-Noses developed with special sensors or combined with other technologies and methods. Some common characteristics of those systems are the instrumentation complexity, most of them involve the use of additional equipment that requires experienced personnel, and they do not realize online detection. For instance, a metalloporphyrin based optoelectronic nose was developed in (AMAMCHARLA; PANIGRAHI, 2010) for the simultaneous prediction of Volatile Organic Compounds (VOCs) concentrations in binary mixtures (acetic acid and ethanol) using partial least square regression (PLSR) and multilayer perceptron neural network (MLP-NN). Besides, in (GIL-SÁNCHEZ et al., 2011), it is reported the wine spoilage analysis when in contact with air using a combined system of a potentiometric electronic tongue and a humid E-Nose.

The acetic acid detection was studied by (MACÍAS et al., 2012) using a commercial E-Nose for general purpose, in combination with a neural network classifier (MLP). They detected only the excessive concentrations of acetic acid, equal to or greater than $2g \cdot l^{-1}$ in synthetic wine samples (aqueous ethanol solution at 10% v/v). However, levels higher than $1.2g \cdot l^{-1}$ of VA cause that the wine takes on vinegar aromas (unpleasant), reducing its quality; hence the governments forbid their commercialization («Normative instruction N° 14», 2018; ZOECKLEIN, B. W., FUGELSANG, K. C., GUMP, B. H., & NURY, 1995). Thus, our work was aimed to detect lower levels and with a quick identification in real wine samples with several spoilage thresholds using the self-developed E-Nose, without using any reagent to reduce the environmental impact, as well with a smooth and safe operation interface (no occupational risk for the operator and with minimal training).

This study presents the self-developed E-Nose based on commercially available gas sensors for early detection of spoilage thresholds by VA in routine tasks of wine quality control. We recorded electrical signals corresponding to odorant profiles of wines samples with different spoilage levels. Afterward, we compared the conventional data processing approach used in E-Noses against our online data processing approach to accelerate the responses. In the conventional approach was applied the preprocessing and feature extraction before an SVM classifier to obtain the main odorant parameters (which requires that the measurement process had finished before data processing stage). By contrast, we focused on an online solution, that let to achieve faster results, using an early portion of the signals while the measurement process is still running. Our approach is based on the training of a deep MLP classifier using the raw data.

### 4.4.2 Materials and methods

#### *4.4.2.1 Electronic Nose*

We used an E-Nose, that we named O-NOSE, comprising principally of an array of six metal-oxide gas sensors (**Table 1**), used to detect the volatile compounds. **Fig. 1** shows O-NOSE on the left side, and the sensors board with two layers for a compact design on the right side.

*Experimental setup*

In **Fig. 2a**, we depict the O-NOSE measurement process divided into three stages. (i) Concentration stage: we used 1ml wine samples to accumulate the volatiles for 30 seconds inside the concentration chamber. (ii) Data acquisition: ten seconds after the initialization of this stage, the VOCs push toward the sensors chamber for 80 seconds generating change in the sensor resistance (gas absorption). Subsequently, the gas injection stops, and it begins the desorption for 90 seconds. Therefore, the acquired data corresponds to 180 seconds with 18.5Hz sample rate. (iii) Purge: the goal is to clean and remove volatile residues for 600 seconds. **Fig. 2b** shows the standard block diagram for the whole experiments, the electrical signals acquired are processed using the pattern recognition techniques after finished the data acquisition stage in the conventional approach or online applying our approach.

**Table 1.** Gas sensors array setup. The sensors manufactured by Hanwei Sensors[1] are commercially available. They have been chosen because of their high sensitivity to organic, natural, ethanol, methanol, and combustible gases, as well as its simplicity of use and low financial cost.

| Number | Sensor | Description | Load resistance |
|---|---|---|---|
| 1, 4 | MQ-3 | High sensitivity to alcohol and small sensitivity to Benzine | 22kΩ |
| 2, 5 | MQ-4 | High sensitivity to CH4 and natural gas | 18 kΩ |
| 3, 6 | MQ-6 | High sensitivity to LPG, iso-butane, propane | 22 kΩ |

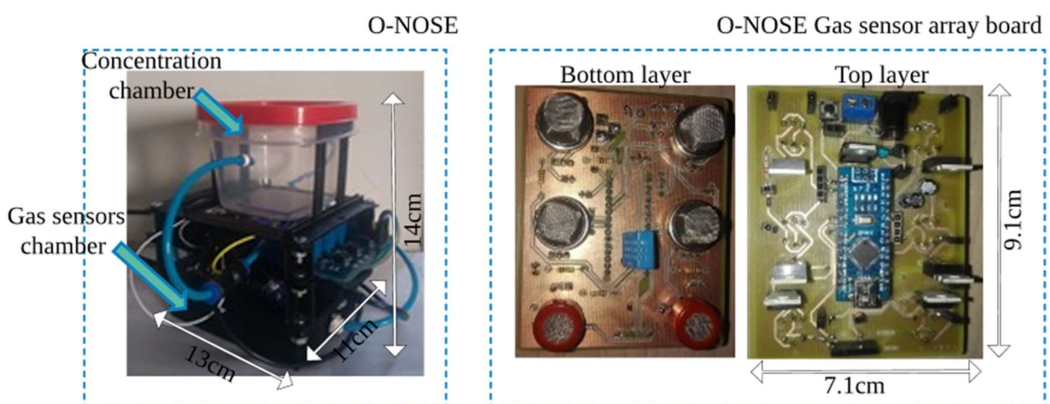


**Fig. 1.** O-NOSE system. On the left side: system appearance and dimensions. Into 100 ml concentration chamber is placed the wine sample. The sensors array is into the 200 ml chamber. On the right side: the main board with the gas sensors and the microcontroller. The gas is sensed by its effect on the sensitive layer of tin dioxide ($SnO_2$), resulting from changes in conductivity brought about by chemical reactions on the surface of the tin dioxide particles.

[1] www.hwsensor.com

*4.4.2.2 Data*

*Wine samples*

We used 22 bottles of commercial wines, and to obtain spoiled samples, 13 of the 22 bottles were randomly selected, opened and left in an uncontrolled environment six months before starting the measurements. These bottles were labeled as low-quality (LQ) wines. Besides, another four bottles were opened two weeks before beginning the data collection. These four bottles were labeled as average-quality (AQ) wines, and the remaining five bottles were labeled as high-quality (HQ) wines.

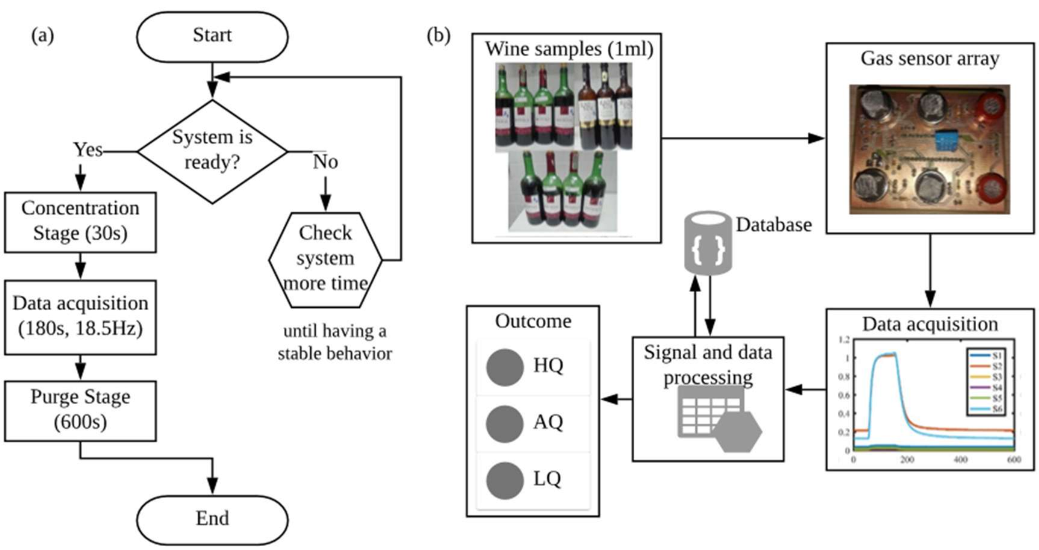


**Fig. 2.** (a) Flowchart of the measurement setup. (b) Block diagram for wine spoilage detection using 1ml samples, the outcome is according to the wine quality.

The 22 wine bottles were characterized as follows: (i) the VA quantification was performed in triplicate according to official methods for wine analysis of the International Organization of Vine and Wine (OIV. INTERNATIONAL ORGANIZATION OF VINE AND WINE, 2014). (ii) Acetic acid was identified by High Performance Liquid Chromatography (HPLC) with UV/Vis absorption detector, following the procedure detailed in (DE ANDRADE LIMA et al., 2010), and the ranges obtained are shown in **Table 2**. It is known that at normal levels in wines ($<0.3 g \cdot l^{-1}$) the VA can be a desirable flavor, adding to the complexity of taste and odor, as well, a content of less than 0.70 $g \cdot l^{-1}$ seldom imparts spoilage character. However, a progressive increment in VA gives to the wines a sour taste and taints its fragrance (Jackson, 2008; Zoecklein, B. W., Fugelsang, K. C., Gump, B. H., & Nury, 1995). Brazilian Ministry of Agriculture, Livestock and Supply (Instrução Normativa N° 14, 2018) establishes that the maximum level of VA in wine is 1.2 $g \cdot l^{-1}$.

**Table 2.** Ranges detected of volatile acidity and acetic acid according to the wine spoilage thresholds. The ranges presented correspond to the minimum and maximum values of the analysis.

| Wine quality level | Volatile acidity in $g \cdot l^{-1}$ | Acetic acid in $g \cdot l^{-1}$ |
|---|---|---|
| HQ | [0.15, 0.3] | [ND, 0.23] |
| AQ | [0.31, 0.41] | [0.24, 0.34] |
| LQ | [0.8, 3] | [0.74, 2.75] |

ND: not detected.

The database collected using O-NOSE has 235 wines measurements as follow: 51, 43, and 141 measurements of HQ, AQ, and LQ respectively. Besides, we collected 65 ethanol measurements in concentrations (v/v) of 2, 5, 10, 20, 30, and 40 ml of ethanol diluted in distilled water to make solutions of 200ml.

*4.4.2.3 Feature extraction and selection*

The most common groups of characteristics extracted from the gas sensors signals are the steady and transient state features (YAN et al., 2015). We used 23 features to capture the dynamic and static behavior of each gas sensor. So, we obtained a 138 columns characteristics matrix, where each row represents the fingerprint of one measurement. One example of the raw data (**Fig. 3a**) evidences the sensor sensitivity regarding VOCs analyzed. In **Fig. 3b-c,** we show the steady and transient features for the response of one sensor during the three intervals of the acquisition procedure explained at the end of Section 2.1.

Afterward, we applied the SVM Recursive Feature Elimination Cross Validation (RFECV) method to reduce the dimensionality, looking to generate parsimonious and robustness models (LIN et al., 2012; YAN; ZHANG, 2015). Thus, it was chosen the followings steady-state characteristics: $\Delta G = max_k g[k] - min_k g[k]$, defined as the maximal conductance change concerning the baseline, and its normalized version ($\|\Delta G\| =$

$(max_k g[k] - min_k g[k])/min_k g[k]$), as well, the area under the curve in the absorption and desorption portions of the gas, blue and gray areas in **Fig. 3b**, respectively. Additionally, we had an aggregate of features reflecting the dynamics of the rising/falling transient portion of the sensor response using an exponential moving average filter ($ema_\alpha$) that converts the transient portion into a real scalar by estimating the maximum/minimum value $y[k] = (1 - \alpha)y[k-1] + \alpha(x[k] - x[k-1])$, where $[k = 1,2, \ldots, T]$, $y[0]$ its initial condition, set to zero $(y[0] = 0,$ and the scalar $\alpha(\alpha \in \{0,1\})$ being a smoothing parameter of the operator such as was defined in (MUEZZINOGLU et al., 2009; VERGARA et al., 2012). We tested three different values for $\alpha = 0.1,\ \alpha = 0.01,$ and $\alpha = 0.001$ as shown in **Fig. 3c**; and by RFECV feature selection, it was chosen the **max** $ema_\alpha$ with $\alpha = 0.01$ as an informative transient feature.

#### *4.4.2.4 Classification methods*

We use two approaches for the classification tasks in this application. The first one consists in applying feature extraction and selection before the classifier. And the second one consists in processing an early portion of the raw data.

*Conventional approach to classification using SVM*

In this approach, it is necessary to have the whole measurement to obtain the main odorant parameters. We tested various kernels on an SVM classifier and selected a gaussian kernel; then it was trained the model. The block diagram of this approach (depicted in **Fig. 4**) exhibits the steps performed that includes a feature extraction block generating the $C_{i,j}$ vector, where $i = 1,2 \ldots,23$ is the number of characteristics and $j = 1,2 \ldots,6$ is the number of sensors. Afterward, the characteristics vector feed the feature selection block, and finally, the chosen variables are carried to the inputs of the SVM classifier.

*Rapid and online detection approach using deep MLP*

This approach is based on a neural network classifier that is feed with the raw data to perform the discrimination tasks (PENG et al., 2018a). Inspired by the mentioned approach and looking to accelerate the response, we propose a rapid detection method in wine quality control, focused on an online solution that lets to achieve faster results using only an early portion of the signals, similar to the presented in (LÄNGKVIST et al., 2013a) for a meat spoilage application, but using a supervised method: deep MLP neural network.

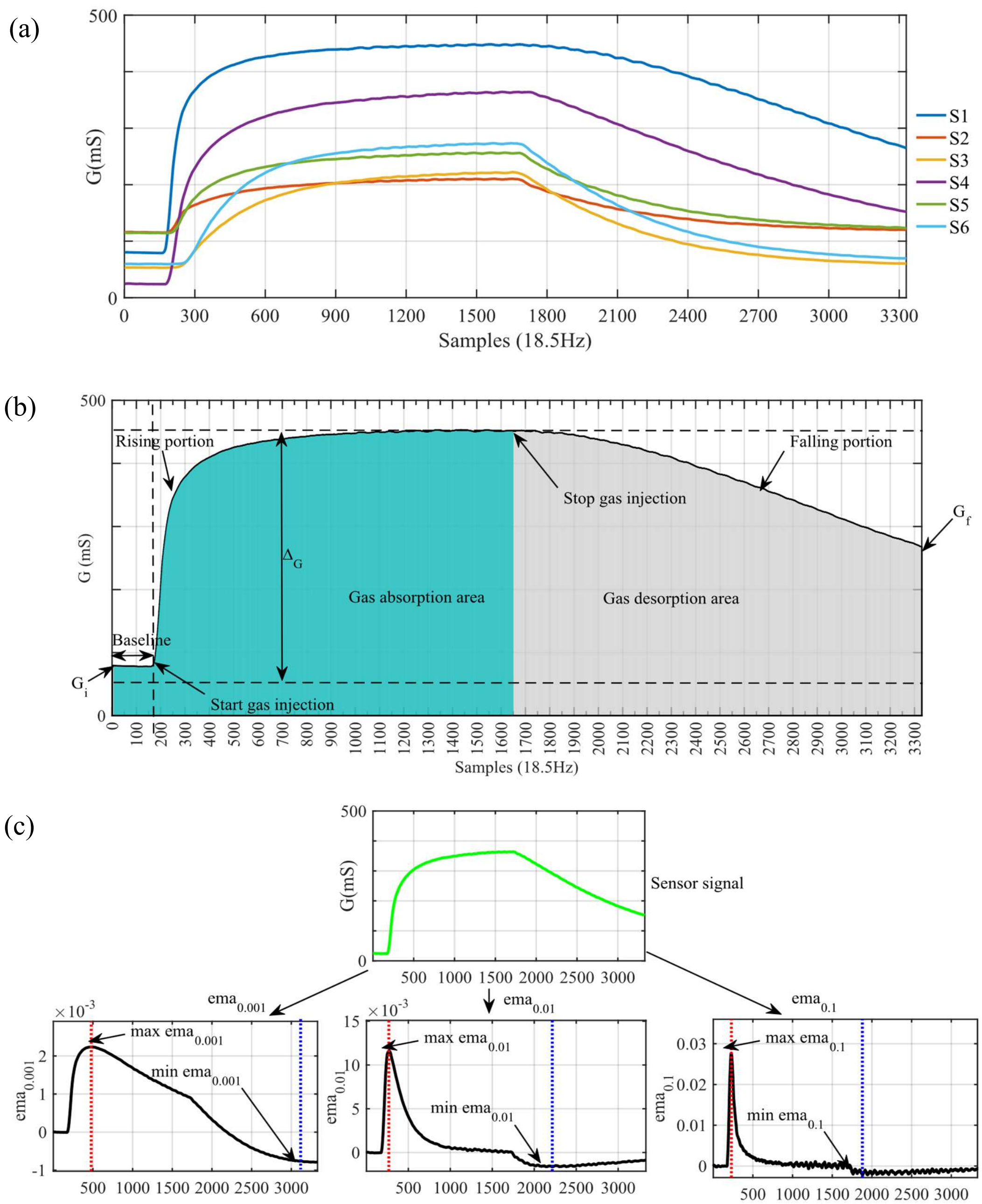


**Fig. 3.** (a) Wine measurement acquired with O-NOSE; S1, S2,…, S6: gas sensor outputs in conductance units G. (b) Output of a gas sensor; Gi: initial conductance value, Gf : final conductance value, $\Delta G$: maximal conductance change concerning the baseline. (c) Dynamics of the rising/falling transient portion using an exponential moving average filter (emaα) for α=0.1, α=0.01, and α=0.001.

The goal with this approach is to offer the possibility to make estimations a few seconds after beginning the measurement process while it is still running. Note that we did not consider the baseline of the sensor since generally in this slice there is no change. Consequently, the data processing starts instantly before the gas injection. A rising window method was applied to find the minor portion of information with the best performance of the classifier. This reduces the effort to obtain discrimination models since as complicated preprocessing techniques no need to be applied and it is feasible by the computational acceleration in the last years. **Fig. 5** depicts the approach employed.

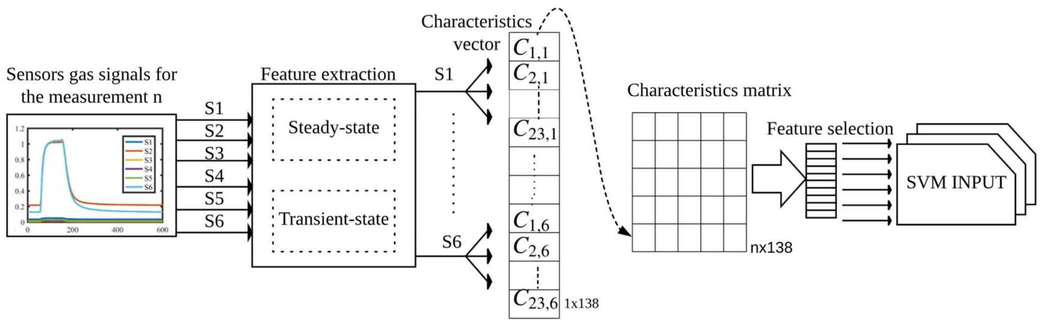


**Fig. 4.** Block diagram of the conventional approach to classification using SVM. This diagram comprises a Feature Extraction block (FE), a Feature Selection block (FS), and subsequently, the characteristics matrix feeds an SVM classifier.

The method to find the minor portion of information is as follows. Given the data series $X_j = x_1, x_2, \dots, x_N$, that represents the gas sensor response $j = 1,2 \dots,6$, their corresponding rising windows are defined as: $X_{j,t} = x_{j,1}, \dots, x_{j,t\Delta}$, where $t = 1, \dots, \left[\frac{N}{\Delta}\right]$, the step is $\Delta \leq N \wedge \Delta \epsilon \mathbb{N}$, the window size is $t\Delta$, and the operator [.] denotes taking the integer part of the argument. The time series in each window $X_{j,t}$ are used to train the deep MLP classifier. **Fig. 5** exhibits the application of the rising windows protocol in our dataset with $\Delta = 50$, hence each $X_{j,1}$ window has 50 points, each $X_{j,2}$ window has 100 points, and so on. The example architecture of the deep MLP, shown in the same figure, corresponds to the neural network used to process the data for the $X_{j,1}$ window. In this case, the input layer size corresponds to the first window ($t = 1$), six sensors, step $\Delta = 50$; then, it has 6 $(1\text{x}50) = 300$ points. The only data preprocessing applied before the deep MLP neural network was a simple data scaling in each window.

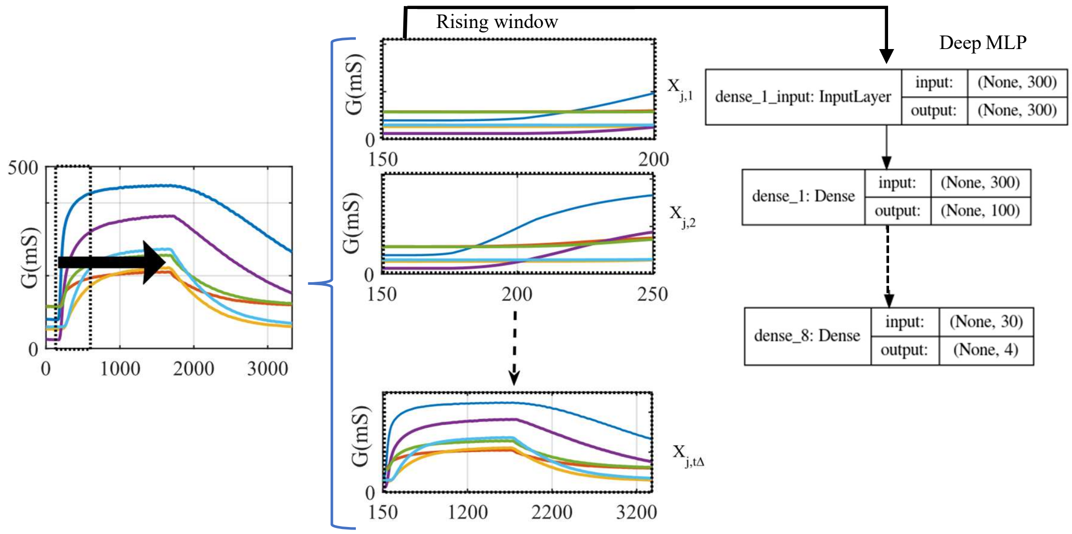


**Fig. 5.** Rapid and online detection approach. Rising window protocol applied to the raw data searching for the minor portion of data to train the deep MLP classifier with the best performance. On the right side is depicted the neural network architecture for the $X_{j,1}$ window with an input size of 300 points and four outputs (three wine spoilage levels and ethanol). The meaning of "None" is unspecified input because we reshaped the data in a flatted array.

### 4.4.3 Results

*4.4.3.1 Data exploratory analysis*

We performed the database exploratory analysis using the Principal Components Analysis (PCA). The scores for the first components (2D and 3D plots) for the wines are shown in **Fig. 6.** We also graph the PCA scores of ethanol jointly wines, as shown in **Fig. 7**.

Based on this exploratory analysis, we performed two experiments with the aim of comparing the performance when the classes are only wines with three spoilage thresholds, and when the ethanol is present as an additional class, which is evidenced as a more complex problem because the ethanol is an essential wine component. These two experiments were performed so much for the conventional approach using SVM, as for the rapid and online detection approach using deep MLP neural network, and the results were compared at the end of Section 3.

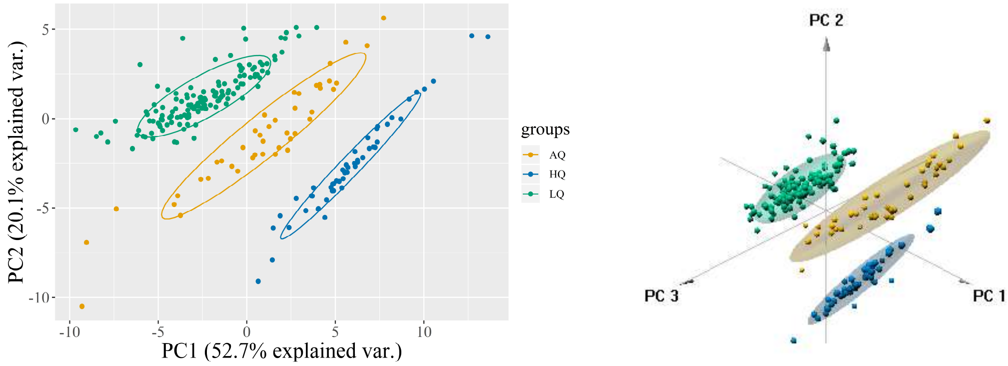


**Fig. 6.** PCA for the three wine groups HQ, AQ, and LQ. On the left side in 2D and the right side in 3D. It is revealed that O-NOSE detects differences between the three groups according to its quality and spoilage threshold. In this case, the three principal components capture a cumulative variance of 81%.

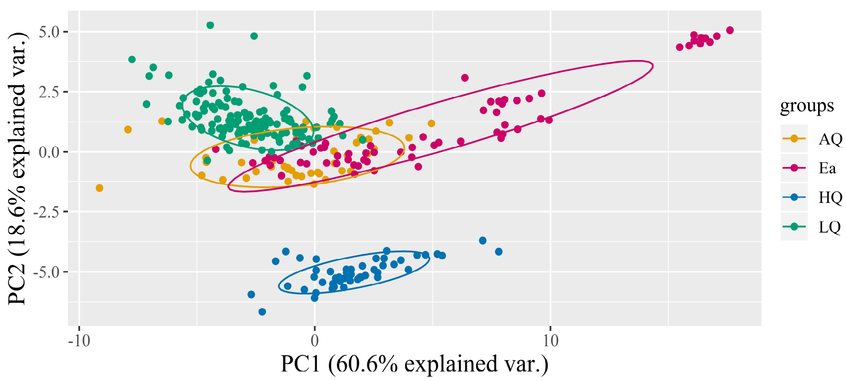


**Fig. 7.** PCA for the three wine groups (HQ, AQ, LQ) and ethanol (Ea). The close relationship between ethanol and wine is evidenced more strongly for the wines labeled as AQ. The groups labeled as HQ and LQ have greater separation regarding the ethanol. In the case of HQ, the organoleptic characteristics are rich in other elements that characterize the excellent taste. For the LQ, the taste is commonly described as vinegar or metallic taste and low level of ethanol.

#### *4.4.3.2 Conventional approach to classification using SVM*

We used an SVM classifier applying the technique known as Leave One Out (LOO), selecting the measurements of one bottle for the validation group and the remaining for the training group. Since as the dataset contains twenty two bottles, we performed this procedure that quantity of times, and we applied five folds cross-validation technique to prevent the overfitting in the training set. We implemented the scripts for this approach using Matlab R2016a and the Statistics and Machine Learning Toolbox - version 10.2; and, to ensure the integrity of the results, we repeated the procedure 100 times with data shuffling before each training. Then, we averaged the accuracy of each experiment.

In **Table 3** are shown the parameters set on the SVM classifier for the two experiments performed: experiment 1 to discriminate among the three wine thresholds (LQ, AQ, and HQ); and experiment 2 to classify among the three wine thresholds and ethanol (LQ, AQ, HQ, and Ea). The recognition accuracy for training and validation, in the first experiment, was 99.78% and 97.34%, and, for the second experiment, 98.31% and 96.23%, respectively.

We did several simulations to find an early portion of the raw data with the best recognition performance in the two experiments. To achieve this, we applied the rising window protocol searching for the minor portion of data to train the deep MLP classifier and averaging the accuracy of each experiment. In this way, we applied the LOO-protocol like the before experiments (Section 3.2), but now training the deep MLP models of eight layers with full neurons connections as detailed in **Table 4** (architecture examples of experiment 2).

**Table 3.** Parameters of the SVM classifiers used for each experiment.

| Parameter | Experiment 1 | Experiment 2 |
|---|---|---|
| Kernel function | Gaussian | Gaussian |
| Kernel parameter scale (gamma) | 8.3 | 19 |
| Box constraint level (C penalty parameter) | 10 | 10 |
| Multiclass method | One-vs-One | One-vs-One |
| Standardize data | True | True |
| Feature selection: variables used in the model | 69 | 56 |
| PCA | disabled | disabled |

*4.4.3.3 Rapid and online detection approach using deep MLP*

The original raw data have 3330 points, but as was explained in Section 2, the baseline is not considered. Thus, we defined the interval to analyze from the point 150 to the point 3300 (to ensure an integer $\left[\frac{N}{\Delta}\right]$). Since as the step was $\Delta = 50$, we trained 63 models that correspond to each $X_{j,t}$ window using python 3.5.3, repeating the procedure 100 times with data shuffling. In the first experiment with the rapid and online detection approach, the accuracy for the windows with the best performance in the training data was 100%, that occurred 97% of the times in windows with a size less or equal than $X_{j,24}$. This corresponds to the first 64.86 seconds of the raw data interval. In validation data, the accuracy for the windows with the best performance was 97.68%, that occurred 88% of the times in the first window ($X_{j,1}$). This represents only an early portion of the raw data, that is equivalent to the first 2.7s, indicating a significant reduction in the time for the recognition when compared to the conventional approach using the feature extraction/selection method.

**Table 4.** Network architecture of three models for the classification using deep MLP, where $\boldsymbol{X_{j,t}}$ is the time series in each window $\boldsymbol{t}$. The trainable parameters are computed as the multiplication between the inputs and the number of neurons in each layer plus the bias number (see the examples for the layers one and eight in the $\boldsymbol{X_{j,1}}$ model).

| Layer | Neurons | Trainable parameters | | |
|---|---|---|---|---|
| | | $X_{j,1}$ model | $X_{j,12}$ model | $X_{j,63}$ model |
| 1 | 100 | (300x100)+100=30100 | 360100 | 1.8901E+6 |
| 2 | 30 | 3030 | 3030 | 3030 |
| 3 | 30 | 930 | 930 | 930 |
| 4 | 30 | 930 | 930 | 930 |
| 5 | 30 | 930 | 930 | 930 |
| 6 | 30 | 930 | 930 | 930 |
| 7 | 30 | 930 | 930 | 930 |
| 8 | 4 | (30x4)+4=124 | 124 | 124 |

The results for the second experiment with this approach indicated that the best performance occurred in windows with a size less or equal than $X_{j,13}$, corresponding to the first 35.13 seconds of the raw data interval. The accuracy was 99.99%, and 96.34%; occurring 54% and 61% of the times in training and validation, respectively. Note that, the separability of the data in this experiment is more complex than the experiment 1 that includes only the three wine spoilage levels, causing that the early portion time necessary for the recognition task being greater. However, it is still less than using the conventional approach which consumes the whole measurement time, suggesting outperformance for the online detection approach using deep MLP.

### 4.4.4 Discussion

The comparison based on the test results between the two discussed approaches is presented in **Table 5**. We highlight the gain in timing for recognition wine quality with our approach, and the possibility of using this approach for online detection without preprocessing techniques.

The rapid and online detection approach has the highest computational time in the training. However, the training is performed offline and in most cases is performed just once. Besides, the computational time using the trained model is about a few milliseconds (<<1s) for the two approaches and experiments. Finally, to support the results obtained and assuming independence between both approach with 5% of significance level, we performed the statistical comparison tests. The results revealed that there is enough evidence to say that in the two experiments the accuracy values for the forecasting with the conventional approach is less than the accuracy values for rapid and online detection approach.

In **Table 6**, we compared the results of (PENG et al., 2018a) and (LÄNGKVIST et al., 2013a) with our results. We chose these approaches because, unlike the classical feature selection method used in artificial olfactory systems, they also used the raw data to process the gas signals. In that way, in (PENG et al., 2018a) was presented an approach based on a Deep Convolutional Neural Network (DCNN) tailored for gas classification but using the entire signal measurement of the gas sensors, resulting in a disadvantage regarding to our approach that lets to achieve faster results using only an early portion of the signals. In (LÄNGKVIST et al., 2013a), similar to the approach proposed in our work, they considered only the transient response centered on an online solution but using unsupervised learning techniques (stacked restricted Boltzmann machines and auto-encoders), although they also focused on obtaining a rapid response, the accuracy of the system is not high. Therefore, our results are better in terms of the time needed to perform the detection. The comparison suggests that it is possible to obtain better results in accuracy and time, using our method. Therefore, our approach is promising for online analyses in E-Nose with low complexity in hardware using standard gas sensors.

**Table 5.** Comparison between the conventional and the rapid detection approach.

| Summary of test results | Conventional approach | | Rapid and online detection approach | |
|---|---|---|---|---|
| | Experiment1 | Experiment2 | Experiment1 | Experiment2 |
| Average accuracy (%) | 97.34±0 | 96.23±0 | $97.68 \pm 4.6 \times 10^{-3}$ | $96.34 \pm 4.6 \times 10^{-3}$ |
| Time for recognition (s) | 171.89 | 171.89 | **2.7** | **35.13** |
| Data preprocessing | FE + FS | FE + FS | Scaling | Scaling |
| Online | NA | NA | Yes | Yes |
| Input size | 69 | 56 | 300 | 3900 |
| Time for training (s) | 16 | 27 | 99 | 130 |
| Time for validation (s) | <<1 | <<1 | <<1 | <<1 |

Average accuracy is presented as the mean ± standard deviation obtained from 100 repetitions. The Mann-Whitney-Wilcoxon test was conducted with ($P>0.05$). FE: Feature extraction; FS: Feature selection; NA: Not available.

### 4.4.5 Conclusions

In this paper, we prove that it is possible to detect wine quality thresholds in a rapid and online way using a deep MLP classifier processing an early portion of the raw data. We obtained an estimation in 2.7 seconds after the gas injection started when we classified three wine spoilage thresholds, and 35.13 seconds when we included ethanol measurements as a class. Therefore, the rapid detection method lets to make predictions 63 times faster for experiment 1, and at least five times faster for experiment 2, when compared with the conventional approach that needs the whole measurement to obtain the main odorant parameters and involves preprocessing techniques.

In this application, we employed Brazilian commercial wines. For future works, it is expected that more researches been conducted including other varieties of wines and more spoilage thresholds. Besides, the rapid detection approach could be extended to other E-Nose applications.

**Table 6.** Comparison of the rapid detection approach with other similar works.

| | (Peng, Zhao, Pan, & Ye, 2018) | (Längkvist, Coradeschi, Lotfi, & Balaguru Rayappan, 2013) | | Proposed work | |
|---|---|---|---|---|---|
| | | Result1 | Result2 | Result1 | Result2 |
| Model | DCNN | DBN | | Deep MLP | |
| Method | Supervised | Unsupervised | | Supervised | |
| Application or gases | CO, $CH_4$, H, and $C_2H_4$ | Ethanol and TMA | | Wine samples and ethanol | |
| Gas sensor type | MOS | Nanostructured ZnO | | MOS | |
| Online | Not | Yes | | Yes | |
| Average accuracy (%) | 95.2 | 60±4.5 | 83.7±4.1 | 97.68 | 96.34 |
| Time for recognition (s) | 100 | 5 | 25 | 2.7 | 35.13 |
| Time for training (s) | 154 | NA | NA | 99 | 130 |

CO: carbon monoxide; $CH_4$: methane; H: hydrogen; $C_2H_4$: ethylene; TMA: thrimethylamine; DBN: Deep Belief Network; DCNN: Deep Convolutional Neural Networks; MOS: Metal oxide semiconductor; NA: Not available.

## Acknowledgments

This work was supported by the Serrapilheira Institute (grant number Serra-1709-22626), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (CAPES) – Finance Code 001, and Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (grant number 420319/2016-6).

## References

Aleixandre, M., Cabellos, J. M., Arroyo, T., & Horrillo, M. C. (2018). Quantification of Wine Mixtures with an Electronic Nose and a Human Panel. Frontiers in Bioengineering and Biotechnology, 6(February), 1–7. https://doi.org/10.3389/fbioe.2018.00014

Amamcharla, J. K., & Panigrahi, S. (2010). Simultaneous prediction of acetic acidethanol concentrations in their binary mixtures using metalloporphyrin based opto-electronic nose for meat safety applications. Sensing and Instrumentation for Food Quality and Safety, 4(2), 51–60. https://doi.org/10.1007/s11694-010-9092-2

Cretin, B. N., Dubourdieu, D., & Marchal, A. (2018). Influence of ethanol content on sweetness and bitterness perception in dry wines. LWT - Food Science and Technology, 87, 61–66. https://doi.org/10.1016/j.lwt.2017.08.075

De Andrade Lima, L. L., Alexandre, S., Guerra, N. B., Pereira, G. E., De Andrade Lima, T. L., & Rocha, H. (2010). Otimização e validação de método para determinação de ácidos orgânicos em vinhos por cromatografia líquida de alta eficiência. Quimica Nova, 33(5), 1186–1189. https://doi.org/10.1590/S0100-40422010000500032

Gil-Sánchez, L., Soto, J., Martínez-Máñez, R., Garcia-Breijo, E., Ibáñez, J., & Llobet, E. (2011). A novel humid electronic nose combined with an electronic tongue for assessing deterioration of wine. Sensors and Actuators A: Physical, 171(2), 152–158.

Jackson, R. S. (2008). Wine Science: Principles and Applications. Igarss 2014. Elsevier. https://doi.org/10.1007/s13398-014-0173-7.2

Längkvist, M., Coradeschi, S., Loutfi, A., & Balaguru Rayappan, J. B. (2013). Fast classification of meat spoilage markers using nanostructured ZnO thin films and unsupervised feature learning. Sensors (Switzerland), 13(2), 1578–1592. https://doi.org/10.3390/s130201578

Lin, X., Yang, F., Zhou, L., Yin, P., Kong, H., Xing, W., … Xu, G. (2012). A support vector machine-recursive feature elimination feature selection method based on artificial contrast variables and mutual information. Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences, 910, 149–155. https://doi.org/10.1016/j.jchromb.2012.05.020

Lozano, J., Santos, J. P., & Horrillo, M. C. (2016). Wine Applications With Electronic Noses. Electronic Noses and Tongues in Food Science. Elsevier Inc. https://doi.org/10.1016/B978-0-12-800243-8.00014-7

Macías, M., Manso, A., Orellana, C., Velasco, H., Caballero, R., & Chamizo, J. (2012). Acetic Acid Detection Threshold in Synthetic Wine Samples of a Portable Electronic Nose. Sensors, 13(12), 208–220. https://doi.org/10.3390/s130100208

Martins, N., Garcia, R., Mendes, D., Costa Freitas, A. M., da Silva, M. G., & Cabrita, M. J. (2018). An ancient winemaking technology: Exploring the volatile composition of amphora wines. Lwt, 96(November 2017), 288–295. https://doi.org/10.1016/j.lwt.2018.05.048

Muezzinoglu, M. K., Vergara, A., Huerta, R., Rulkov, N., Rabinovich, M. I., Selverston, A., & Abarbanel, H. D. I. (2009). Acceleration of chemo-sensory information processing using transient features. Sensors and Actuators, B: Chemical, 137(2), 507–512. https://doi.org/10.1016/j.snb.2008.10.065

Normative instruction N° 14 (2018). Brazil. Retrieved from http://www.agricultura.gov.br/noticias/mapa-atualiza-padroes-de-vinho-uva-e-derivados/INMAPA142018PIQVinhoseDerivados.pdf

OIV. International Organization of vine and wine. Compendium of Internacional Methods of Analysis of wine and Musts, 1 § (2014).

Peng, P., Zhao, X., Pan, X., & Ye, W. (2018). Gas Classification Using Deep Convolutional Neural Networks. Sensors, 18(1), 157. https://doi.org/10.3390/s18010157

Perestrelo, R., Rodriguez, E., & Câmara, J. S. (2017). Impact of storage time and temperature on furanic derivatives formation in wines using microextraction by packed sorbent tandem with ultrahigh pressure liquid chromatography. LWT - Food Science and Technology, 76, 40–47. https://doi.org/10.1016/j.lwt.2016.10.041

Peris, M., & Escuder-Gilabert, L. (2016). Electronic noses and tongues to assess food authenticity and adulteration. Trends in Food Science and Technology, 58, 40–54. https://doi.org/10.1016/j.tifs.2016.10.014

Rodríguez-Méndez, M. L., De Saja, J. A., González-Antón, R., García-Hernández, C., Medina-Plaza, C., Garcíia-Cabezón, C., & Martíin-Pedrosa, F. (2016). Electronic noses and tongues in wine industry. Frontiers in Bioengineering and Biotechnology, 4(OCT), 81. https://doi.org/10.3389/fbioe.2016.00081

Sáenz-Navajas, M. P., Avizcuri, J. M., Ballester, J., Fernández-Zurbano, P., Ferreira, V., Peyron, D., & Valentin, D. (2015). Sensory-active compounds influencing wine experts' and consumers' perception of red wine intrinsic quality. LWT - Food Science and Technology, 60(1), 400–411. https://doi.org/10.1016/j.lwt.2014.09.026

Stupak, M., Kocourek, V., Kolouchova, I., & Hajslova, J. (2017). Rapid approach for the determination of alcoholic strength and overall quality check of various spirit drinks and wines using GC–MS. Food Control, 80, 307–313. https://doi.org/10.1016/j.foodcont.2017.05.008

Vazallo-Valleumbrocio, G., Medel-Marabolí, M., Peña-Neira, Á., López-Solís, R., & Obreque-Slier, E. (2017). Commercial enological tannins: Characterization and their relative impact on the phenolic and sensory composition of Carménère wine during bottle aging. LWT - Food Science and Technology, 83, 172–183. https://doi.org/10.1016/j.lwt.2017.05.022

Vergara, A., Vembu, S., Ayhan, T., Ryan, M. A., Homer, M. L., & Huerta, R. (2012). Chemical gas sensor drift compensation using classifier ensembles. Sensors and Actuators B: Chemical, 166, 320–329. https://doi.org/10.1016/j.snb.2012.01.074

Yan, J., Guo, X., Duan, S., Jia, P., Wang, L., Peng, C., & Zhang, S. (2015). Electronic Nose Feature Extraction Methods: A Review. Sensors, 15(11), 27804–27831. https://doi.org/10.3390/s151127804

Yan, K., & Zhang, D. (2015). Feature selection and analysis on correlated gas sensor data with recursive feature elimination. Sensors and Actuators, B: Chemical, 212, 353–363. https://doi.org/10.1016/j.snb.2015.02.025

Zhao, Z., Yang, X., Zhao, X., Bai, B., Yao, C., Liu, N., … Zhou, C. (2017). Vortex-assisted dispersive liquid-liquid microextraction for the analysis of major Aspergillus and Penicillium mycotoxins in rice wine by liquid chromatography-tandem mass spectrometry. Food Control, 73, 862–868. https://doi.org/10.1016/j.foodcont.2016.09.035

Zoecklein, B. W., Fugelsang, K. C., Gump, B. H., & Nury, F. S. (1995). Volatile acidity. In Wine Analysis and Production (pp. 192–198). Springer, Boston, MA. https://doi.org/10.1007/978-1-4757-6978-4_11

# CHAPTER 5

# 5. Validation of the rapid detection approach for enhancing the electronic nose systems performance

Seeking to validate the rapid detection approach for electronic nose, we used five E-Nose databases that include 15 datasets. The tested databases correspond to different systems used in several settings and with distinct experimental setups, guaranteeing varied conditions.

## 5.1 Deep Learning Models Used in E-Nose

Regarding the classification methods, Support Vector Machines (SVM) is commonly used in the traditional approach. In contrast, the more novel approaches utilize neural networks, implementing variations regarding the architectures, learning techniques, for instance. Newly, the Deep Learning (DL) techniques have been explored in electronic nose applications. Such as the case of (LIU; MENG; ZHANG, 2018a; PENG et al., 2018a; QI; MENG; ZENG, 2017a; WEI et al., 2019a), these authors used the Convolutional Neural Networks CNN approach in some applications.

Searching for optimal neural network architectures is a task that can be both difficult and time consuming for researchers. Seeking a way to automate this task, the authors in (LIU; SIMONYAN; YANG, 2018) have proposed a method to perform this search using differentiation. The key reason to use differentiation as the search method is that other approaches based on reinforcement learning, evolutionary computing, or even Bayesian optimization tend to be a lot of hardware and time demanding. Besides, the main advantage of using this approach instead of performing a black-box search over a set of candidate architectures, it has to do with the search space is continuous, and the algorithm performs the architecture optimization over the validation set performance.

The Deep Learning models have shown great potential to classify and forecast data. That motivated us to compare several Deep Learning implementations against SVM (a traditional classification method) to test whether the proposed rapid detection approach for E-nose outperforms the results obtained with the conventional approach.

## 5.2 Main Contributions Related to the Validation of the Proposed Rapid Detection Approach

- We tested the proposed rapid detection approach in several datasets to validate the performance in varied conditions over different measurement settings. It is crucial to highlight that in most researches carried out in the E-Nose field the authors draw upon too few databases to validate the methods.

- We tested deep learning methods in several electronic nose datasets to perform classification tasks. We intend to make this work as a reference for further researches, encouraging the research community to conduct more studies or analysis regarding E-Nose using several datasets from different databases, as well as making public the collected databases (ARAUJO; GAMBOA; SILVA, 2020).

- Our investigation allowed finding that, in most cases, it is possible to obtain a reliable forecast using only the first 30% of the measure after the gas injection started and without applying preprocessing techniques.

## 5.3 A preprint of the research paper published at ArXiv.org and pendant to submission for a scientific journal: Validation of the rapid detection approach for enhancing the electronic nose systems performance, using different deep learning models and support vector machines

Juan C. Rodriguez Gamboa[a,*], Adenilton J. da Silva[b], Ismael C. S. Araujo[c], Eva Susana Albarracin E.[a], *Cristhian M. Duran A.*[d]

[a] *Departamento de Estatística e Informática, Universidade Federal Rural de Pernambuco - UFRPE, Recife, PE, Brazil*

[b] *Centro de Informática, Universidade Federal de Pernambuco - UFPE, Recife, PE, Brazil*

[c] *Departamento de Computação, Universidade Federal Rural de Pernambuco, Recife, PE, Brazil*

[d] *Facultad de ingeniería y arquitectura, Universidad de Pamplona, Pamplona, NdeS, Colombia*

[*] *Corresponding author. Departamento de Estatística e Informática, Universidade Federal Rural de Pernambuco - UFRPE, Rua Dom Manoel de Medeiros s/n, Dois Irmãos, CEP: 52171-900, Recife, PE, Brazil. E-mail: juan.gamboa@ufrpe.br*

**Abstract**

Real-time gas classification is an essential issue and challenge in applications such as food and beverage quality control, accident prevention in industrial environments, for instance. In recent years, the Deep Learning (DL) models have shown great potential to classify and forecast data in diverse problems, even in the electronic nose (E-Nose) field. In this work, we used a Support Vector Machines (SVM) algorithm and three different DL models to validate the rapid detection approach (based on processing an early portion of raw signals and a rising window protocol) over different measurement conditions. We performed a set of trials with five different E-Nose databases that include fifteen datasets. Based on the results, we concluded that the proposed approach has a high potential, and it can be suitable to be used for E-nose technologies, reducing the necessary time for making forecasts and accelerating the response

time. Because in most cases, it achieved reliable estimates using only the first 30% or fewer of measurement data (counted after the gas injection starts.) The findings suggest that the rapid detection approach generates reliable forecasting models using different classification methods. Still, SVM seems to obtain the best accuracy, right window size, and better training time.

***Keywords***: *Electronic nose, E-Nose, rapid detection, datasets, deep learning, real-time classification*

### 5.3.1 Introduction

The conventional approach for data processing in the Electronic Nose implies using the entire response curves (including rising, steady-state, recovery phases, and other) of the gas sensors array. Besides, this approach includes steps such as signal preprocessing and feature generation/extraction before performing the classification tasks, which requires the selection of a suitable method for each stage, increasing the necessary time to find the appropriate classification and forecasting models (LIU; MENG; ZHANG, 2018b; QI; MENG; ZENG, 2017b). Nowadays, some researches have focused their efforts on reducing the steps and the essential know-how for model generation, such as the works presented by (LIU; ZENG; MENG, 2019b) and (LÄNGKVIST et al., 2013b). In the first case, the authors proposed a bio-inspired data processing method based on a neural network to mimic the mammalian olfactory system with excellent results but using the entire measurement curves. In the second case, the authors proposed a rapid detection system for meat spoilage using an unsupervised technique (i.e., stacked restricted Boltzmann machines and auto-encoders) that considers only the transient response. Although the obtained models offer advantages because the features are learned from data instead of being hand-designed, it may produce low suitable and inaccurate models due to the unsupervised method. Furthermore, some authors have explored an approach based on raw data treatment (PENG et al., 2018b; WEI et al., 2019b). Although this approach reduces the steps and the development time, they only tested with the entire response curves, requiring to complete all measurement processes. So, it can take critical time to get results.

Concerning the above mentioned, we proposed a novel approach on (RODRIGUEZ GAMBOA et al., 2019d), based on processing an early portion of signals (while the measurement process is still running.) We also tested the proposed method in a wine quality application, with excellent results against the traditional methodology. A deep MLP classifier was trained with the raw data acquired from an E-Nose composed of an array of six MOX gas sensors. We achieved results around 63 times faster (Eq. 1) compared with a traditional method (using the entire response curves, applying preprocessing techniques to extract the features and later processing them using an SVM algorithm.)

$$relation\ of\ measurement\ time = \left(\frac{measurement\ time\ from\ the\ starting\ gas\ injection\ to\ the\ finis}{necessary\ time\ for\ making\ a\ forecast\ or\ window\ size}\right) \quad \text{Eq. (1)}$$

Support Vector Machines (SVM) is one of the most applied methods for classification in E-Nose. Other used methods are K-Nearest Neighbors (KNN), Naive Bayes (NB), Linear Discriminant Analysis (LDA), and Adaptive Resonance Theory Map (ARTMAP) (JHA et al., 2019). However, the more novel approaches are based on neural networks, implementing variations about the architectures, learning techniques, for instance. In recent years the deep learning algorithms have arrived at the electronic nose field, such as the case of (LIU; MENG; ZHANG, 2018b; PENG et al., 2018b; QI; MENG; ZENG, 2017b; WEI et al., 2019b), where the authors have explored the Convolutional Neural Networks (CNN) in different settings.

The present study focuses on validating if the rapid detection approach is suitable to be applied in diverse E-Nose settings (five different databases) (RODRIGUEZ GAMBOA et al., 2019d). Additionally, to test DL techniques such as Convolutional Neural Networks (CNN) against a more classical method like SVM for classification tasks in E-Nose using the proposed approach.

### 5.3.2 Materials and methods

In this work, we used five E-Nose databases that include fifteen datasets to test our approach. The tested databases correspond to different systems used in several settings and with distinct experimental setups, guaranteeing varied conditions. We intend to make this work as a reference for further researches, encouraging the research community to perform more studies or analysis by using E-Noses with numerous databases as well as making public the collected databases.

*5.3.2.1 Databases*

*5.3.2.1.1 Database 1: Electronic nose dataset for the detection of wine spoilage thresholds*

This public database consists of time series collected through an electronic nose for a wine quality control application focused on spoilage thresholds. This database has two datasets, one of them composed of only wines (three-class classification problem), and the other comprises wines and ethanol (four classes). The database contains 235 recorded measurements of wines divided into three groups, labeled as high quality (HQ), average quality (AQ), and low quality (LQ), in addition to 65 ethanol measurements. The time series acquired at 18.5 Hz of sampling frequency during 180 seconds correspond to 3330 data points per sensor. Each file in the dataset has eight columns: relative humidity (%), temperature (°C), and the resistance readings in kΩ of the six MOX gas sensors: MQ-3, MQ-4, MQ-6, MQ- 3, MQ-4, MQ-6. More details are available in (RODRIGUEZ GAMBOA et al., 2019d, 2019e).

*5.3.2.1.2 Database 2: Electronic Nose for Quality Control of Colombian Coffee through the Detection of Defects in "Cup Tests"*

This dataset consists of time-series recorded by an electronic nose used for the coffee quality control to detect defects in the grain (RODRIGUEZ GAMBOA, JUAN CARLOS; DURAN ACEVEDO, 2009; RODRÍGUEZ; DURÁN; REYES, 2010). The dataset contains 58 measurements of coffee divided into three groups and labeled as high quality (HQ), average quality (AQ), and low quality (LQ), inducing a three classes classification problem. The time series acquired at 1 Hz of sampling frequency during 300 seconds correspond to 2400 data points for each measurement. Where, each file in the dataset has eight columns with the resistance readings in kΩ of the gas sensors: SP-12A, SP-31, TGS-813, TGS-842, SP-AQ3, TGS-823, ST-31, TGS-800.

*5.3.2.1.3 Database 3: Gas sensor arrays in open sampling settings Data Set*

The authors compiled an extensive database through a chemical detection platform for detecting potentially hazardous substances at different concentrations, composed of nine portable sensor array modules (72 metal-oxide chemical sensors in a wind tunnel facility.) Each module had eight MOX gas sensors (manufactured by Figaro Inc.) and positioned at six different line locations normal to the wind direction. Thus, creating thereby a total number of

54 measurement locations, uniformly distributed for a total of 18000 different measurements. We split this database into six datasets (each dataset corresponds to one line location: L1, L2, …, L6.) Compounds, such as acetone, acetaldehyde, ammonia, butanol (butyl-alcohol), ethylene, methane, methanol, carbon monoxide, benzene, and toluene (ten classes) were measured to generate the database (VERGARA et al., 2013).

*5.3.2.1.4 Database 4: Gas sensor array exposed to turbulent gas mixtures Data Set*

This dataset was obtained using the same wind tunnel mentioned in section 2.1.3, but the wind tunnel was adapted from the previous setup to include two independent gas sources. Besides, only one module (eight MOX gas sensors array) was used in a fixed location in the wind tunnel. The sensors array was exposed to binary mixtures of ethylene with either methane or carbon monoxide. Volatile Organic Compounds (VOCs) were released at four different rates to induce different concentration levels in the module vicinity. Each configuration was repeated six times, for a total of 180 measurements. See (FONOLLOSA et al., 2014, 2015) for additional details. In this work, we split the dataset to generate a four-class classification problem, including the followings categories (high ethylene concentration, medium ethylene concentration, low ethylene concentration, and without ethylene.) Hence, this is a challenging problem because the measurements were performed using two interfering gases (Methane, carbon monoxide) at different concentrations, and all groups of measurements include binary mixtures of ethylene with combinations of the mentioned interfering VOC.

*5.3.2.1.5 Database 5: Twin gas sensor arrays Data Set*

This database comprises the recordings of five twin eight gas sensor detection units. This database has five datasets (B1, B2, …, B5) where each dataset corresponds to the measurements of one twin system (authors followed the same measuring experimental protocol in the five twin units). Every single day, a different unit was tested, which included the presentation of 40 different gas conditions, presented in random order, exposing each unit to 10 concentration levels of Ethanol, Methane, Ethylene, and Carbon Monoxide (four classes). The conductivity of each sensor for 600 s in each experiment was acquired by using a sample rate of 100 Hz. The authors tested the detections platforms for 22 days, but only 16 days, the measurements were collected. Hence, the complete dataset comprises 640 records (FONOLLOSA et al., 2016).

*5.3.2.2 Deep Learning Models*

The models generated were implemented using the *Python* programming language. In this study, three DL architecture implementations types were used for the classification tasks. The first type was a set of three simple DL models named SniffNets in (ARAUJO, I. C. S. ; SILVA, A. J. ; GAMBOA, 2019). The SniffNets were implemented employing the machine learning framework: Keras (BUITINCK et al., 2013). The second architecture implementation type was a DL model to perform meta-learning, adjusting the connections between different computing cells by differentiable search to obtain the best graph configuration while training. The authors called that methodology as Differentiable Architecture Search (DARTS) (LIU; SIMONYAN; YANG, 2018). The DARTS implementation has been made available by its authors, but we adapted it so that the model could fit the shape of the target data. This implementation was created using the PyTorch library (ADAM PASZKE, SAM GROSS, SOUMITH CHINTALA, GREGORY CHANAN, EDWARD YANG, ZACHARY DEVITO, ZEMING LIN, ALBAN DESMAISON, LUCA ANTIGA, 2017). Finally, the third model corresponds to a simple Deep MLP model with only fully connected layers based on the model used in (RODRIGUEZ GAMBOA et al., 2019d).

The input format used for the models with convolutional layers was a feature matrix with dimensions $R_f$ x $C_f$, where $R_f$ corresponds to the rows (represents the time interval) and the columns $C_f$ that corresponds to the gas sensors used to detect the specimens.

*5.3.2.2.1 SniffNets*

We generated three models with different architecture implementations based on (ARAUJO, I. C. S. ; SILVA, A. J. ; GAMBOA, 2019), adapting the architectures to test the proposed rapid detection approach (RODRIGUEZ GAMBOA et al., 2019d). In the three models, we used the *softmax* as the activation function for the output layer. The code used in this study is available at (S. ARAUJO, 2019a)

The first model is a convolutional network named Sniff ConvNet in this document, which consists of two layers that apply a bi-dimensional convolution (Conv2D), followed by two fully connected (FC) layers. We used the activation function *ReLU* in both Conv2D and FC layers. The second model is a residual network named Sniff ResNet composed of two residual blocks, each one with two Conv2D layers. In each block, the first convolutional layer has a skip connection joined to the second convolutional layer output. Two FC layers follow the two residual blocks. Like as the Sniff ConvNet model, we used the activation function *ReLU* in the Conv2D and FC layers. The third model is a fusion neural network called Sniff Multinose. In this case, we adopted a different approach, where the feature matrix has a shape $R_f$ X $C_f$. We split the feature matrix by columns $C_f$, and each column was used as an input of a Multilayer Perceptron (MLP) model. Then, we concatenated the outputs of all MLP models and utilized them as inputs of another MLP network to complete the classification model.

*5.3.2.2.2 DARTS: Differentiable Architecture Search*

Searching for optimal neural network architectures is a task that can be both difficult and time consuming for the researchers. Seeking a way to automate this task, a study was proposed to perform this search using differentiation (LIU; SIMONYAN; YANG, 2018). The key reason behind the use of differentiation is that the search space is continuous, and the algorithm performs the architecture optimization over the validation set performance. The algorithm performs the search in a network considered as a directed acyclic graph. Each node xi represents the output of a subnetwork in the chart. For example, xi can be a feature vector from a fully connected Multilayer perceptron or a feature map from a convolutional layer. Let O be the set of all the arcs(i,j) being an operation between the i-th node to the j-th node pondered by a factor α(i,j). The arc(i, j) represents the connection between nodes xi and xj. In which this connection is the o(i,j) operation which inputs are Xi and outputs Xj . After the initialization of a set of candidate operations between all nodes (i, j) of the graph. The search task is then performed by first computing the gradient of the loss function with respect to the factors α(i,j) and then concerning the weights of the model. Thus, after computing the minimal loss concerning the α and the weights in the arcs between (i,j), the algorithm determines the optimal architecture according to the values of α (LIU; SIMONYAN; YANG, 2018). The code used in this study is available at (S. ARAUJO, 2019b).

*5.3.2.2.3 Deep MLP model*

We also used a Deep MLP model presented in (RODRIGUEZ GAMBOA et al., 2019d). The configuration of the model consists of eight layers each with *ReLU* as the activation function except for the output layer, in which we used *softmax*. In this model, the input layer was configured to have 100 neurons and all the hidden layers to have 30 neurons. The code used in this study is available at (S. ARAUJO, 2019a).

*5.3.2.2.4 Training Configurations*

We trained the three sets of DL models until to reach 20 epochs by using the Stochastic Gradient Descent (SGD) algorithm for optimization with a learning rate of 0.001 and a momentum of 0.9. Besides, we used the loss function called categorical cross-entropy and the holdout cross-validation method.

*5.3.2.2.5 Configurations for the SVM model*

The SVM model available in the scikit-learn library was used (RAUL GARRETA, 2013). Furthermore, we defined the following parameters to optimize the model: A Radial Bayes Function (RBF) as the kernel, the regularization parameter C as 10, and the other settings as the default value. Given a dataset $D$ with vectors of $n$ features, the value computed for the *gamma* parameter is $(n \bullet variance(D_{flat}))^{-1}$. Where $variance(D_{flat})$ is the variance over the flattened dataset. The algorithm computed the *gamma* value over the normalized data, using standardization or *z-score normalization*.

**5.3.3 Results and Discussion**

According to the rapid detection approach, the rising window protocol was applied to find the early portion with the best validation accuracy (less error rate) in each dataset. We used different methods: the DARTS search model architecture, a deep MLP, three DL models called Sniff (ConvNet, Resnet, and Multinose), and SVM, to validate the proposed approach and determine if it could be applied independently to the classification method. The results of the experiments were summarized in **Table 1** to compare the performance on those windows (chosen windows and the last windows). The results let infer that the models generated with initials windows are similar or even outperform in the majority of cases the models obtained using the complete information of the measurements. The accuracy of the test data in the chosen windows is depicted in **Fig. 1**, to facilitate the visualization.

The chosen windows (with the best accuracy on test data) for each dataset are depicted in **Fig. 2**. It is important to remark that the first-window (w1) corresponds to the first 10% of measurement data, the second-window (w2) has 20% of the information, continuing until the ten-window (w10) that has the 100% of measurements data. The results suggest that, usually, using 30% or even less of information lets obtain suitable models. The necessary time to get the appropriate models for each dataset and classification method is detailed in **Table 2**. Those times are only a reference to know which methods work faster in the training process when is adopted the rapid detection approach.

The results showed similar accuracy for each dataset comparing the tested classification methods. The main difference is presented in the size of the best window. Comparing the Sniff models (gray bars in the figures), SniffMultinose reached the best-combined performance (15 datasets.) The DARTS algorithm (red bars in the figures) generates models that usually are similar to the models created by the Sniff architectures. Still, it is the method that needs more time in the training process to generate reliable models.

Analyzing the window size with the best accuracy on the test data, we conclude that based on the tested classification methods (DL techniques and SVM), the rapid detection approach is a reliable option to apply in electronic nose applications. Their results let to validate this methodology to be used in E-Nose datasets. Besides, it is relevant to remark that in the majority of cases, the SVM (blue bars in the figures) generates models that use an early portion of information with minor size **Fig. 2**, allowing to make forecasts in less time regarding the tested architectures. Additionally, the training time is quite less, reducing the computational cost. Therefore, the mentioned findings suggest that in the electronic nose field not worth it to use Convolutional Neural Networks (CNN) and deep learning techniques for classification tasks. These techniques increase the necessary time to generate reliable models and generally not reach much better results.

**Table 1.** Summary of experiments, showing the test data accuracy of the best window (b-win) and the accuracy of the last window (l-win), i.e., using all information after the gas injection for all tested datasets. For each dataset (rows), the best accuracy is highlighted.

| Dataset / Method | Sniff-ConvNet | | Sinff-Resnet | | Sniff-Multinose | | DARTS | | MLP | | SVM | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | b-win | l-win | b-win | l-win | b-win | l-win | b-win | l-win | b-win | l-win | b-win | l-win |
| Wines | 97.5 | 97.5 | 97.5 | 97.5 | 99.1 | 93.2 | 100 | 87 | 100 | 95.7 | 100 | 100 |
| | w3 | w10 | w3 | w10 | w4 | w10 | w1 | w10 | w4 | w10 | w2 | w10 |
| Wines & ethanol | 98 | 95.3 | 94 | 94 | 96.7 | 94.7 | 98.3 | 95 | 100 | 95 | 100 | 98.3 |
| | w2 | w10 | w10 | w10 | w4 | w10 | w9 | w10 | w2 | w10 | w1 | w10 |
| Coffee | 79.3 | 65.5 | 79.3 | 44.8 | 89.7 | 69 | 92 | 67 | 91.6 | 66.6 | 100 | 100 |
| | w1 | w10 | w8 | w10 | w3 | w10 | w3 | w10 | w5 | w10 | w1 | w10 |
| Wind tunnelL1 | 79.6 | 79.6 | 85.4 | 69.9 | 92.2 | 91.3 | 96.77 | 95.16 | 82.9 | 73.2 | 90.2 | 73.2 |
| | w2 | w10 | w2 | w10 | w5 | w10 | w5 | w10 | w5 | w10 | w2 | w10 |
| Wind tunnelL2 | 95.8 | 83.2 | 97.9 | 81 | 97.9 | 92.6 | 98.4 | 98.4 | 92.1 | 78.9 | 97.4 | 89.5 |
| | w4 | w10 | w3 | w10 | w2 | w10 | w7 | w10 | w6 | w10 | w6 | w10 |
| Wind tunnelL3 | 96.1 | 90.3 | 92.2 | 87.4 | 99 | 94.2 | 98.4 | 98.4 | 90.2 | 92.7 | 100 | 97.6 |
| | w6 | w10 | w7 | w10 | w4 | w10 | w7 | w10 | w3 | w10 | w5 | w10 |
| Wind tunnelL4 | 91.2 | 95.1 | 96.1 | 81.4 | 100 | 96.1 | 98.4 | 98.4 | 95.1 | 87.8 | 97.6 | 95.1 |
| | w9 | w10 | w6 | w10 | w9 | w10 | w7 | w10 | w7 | w10 | w2 | w10 |
| Wind tunnelL5 | 94.2 | 91.3 | 89.3 | 85.4 | 93.2 | 98 | 98.4 | 98.4 | 90.2 | 87.8 | 95.1 | 80.5 |
| | w7 | w10 | w5 | w10 | w6 | w10 | w7 | w10 | w4 | w10 | w5 | w10 |
| Wind tunnelL6 | 95.7 | 88.2 | 89.2 | 90.3 | 100 | 95.7 | 98.4 | 98.4 | 91.9 | 89.2 | 94.6 | 94.6 |
| | w4 | w10 | w6 | w10 | w2 | w10 | w7 | w10 | w8 | w10 | w4 | w10 |
| Turbulent gas mixtures | 60 | 52.2 | 53.3 | 34.4 | 57.7 | 30 | 73.3 | 60.1 | 72.2 | 72.2 | 77.7 | 86.1 |
| | w3 | w10 | w4 | w10 | w7 | w10 | w3 | w10 | w10 | w10 | w4 | w10 |
| Twin gas sensorB 1 | 100 | 97.5 | 97.5 | 97.5 | 97.5 | 97.5 | 100 | 93 | 100 | 100 | 100 | 100 |
| | w2 | w10 | w7 | w10 | w10 | w10 | w9 | w10 | w3 | w10 | w3 | w10 |
| Twin gas sensorB 2 | 97.5 | 91.2 | 96.2 | 100 | 100 | 80 | 96.9 | 92.1 | 100 | 90.6 | 100 | 96.9 |
| | w3 | w10 | w6 | w10 | w1 | w10 | w10 | w10 | w1 | w10 | w1 | w10 |
| Twin gas sensorB 3 | 100 | 92.5 | 100 | 90 | 98.7 | 88.7 | 96.9 | 94.4 | 100 | 93.7 | 100 | 100 |
| | w8 | w10 | w8 | w10 | w2 | w10 | w10 | w10 | w3 | w10 | w2 | w10 |
| Twin gas sensorB 4 | 92.5 | 52.5 | 100 | 92.5 | 100 | 90 | 87.5 | 77.6 | 100 | 62.5 | 100 | 100 |
| | w3 | w10 | w8 | w10 | w3 | w10 | w9 | w10 | w5 | w10 | w2 | w10 |
| Twin gas sensorB 5 | 100 | 77.5 | 100 | 92.5 | 100 | 87.5 | 100 | 80.1 | 100 | 87.5 | 100 | 100 |
| | w3 | w10 | w9 | w10 | w1 | w10 | w10 | w10 | w1 | w10 | w1 | w10 |

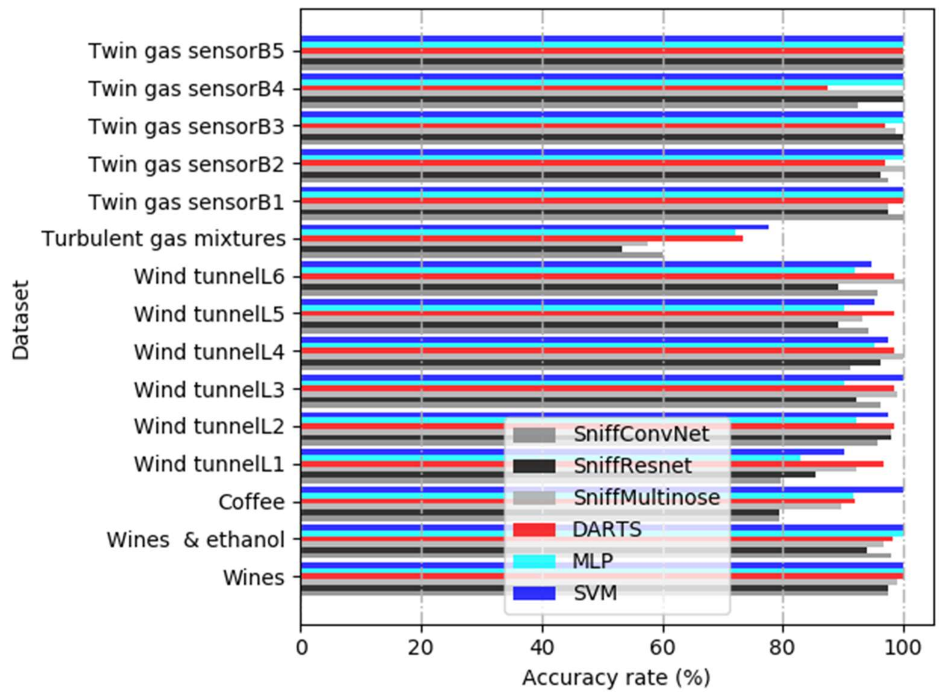


**Fig. 1.** The classification accuracy rate of the test data over the 15 datasets for the tested methods.

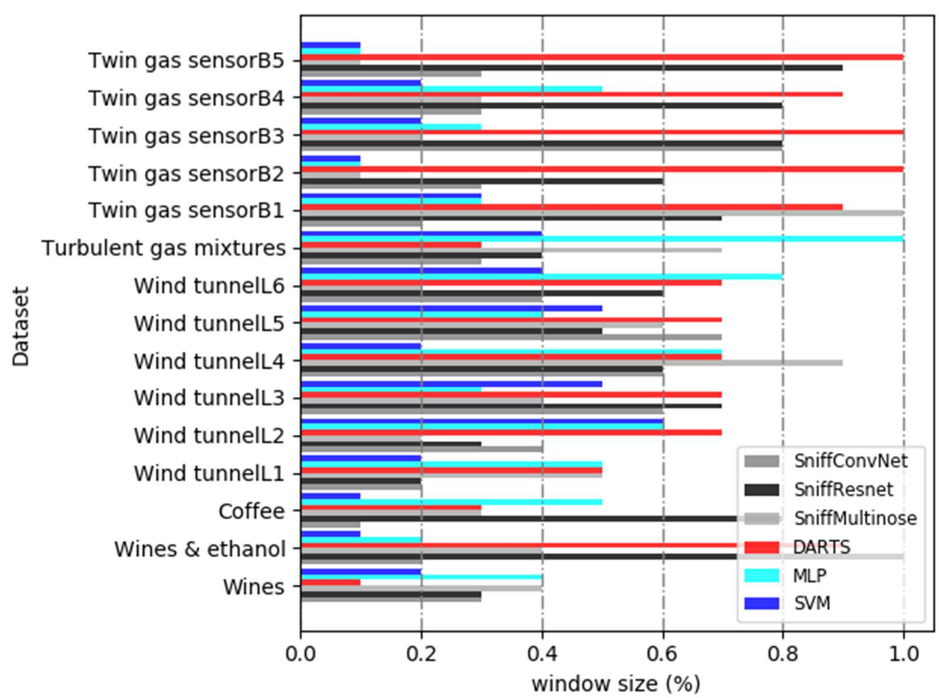


**Fig. 2.** The window size in percentage regarding the complete measurements over the 15 datasets for the tested methods.

**Table 2.** Training time in seconds for the tested classifications methods on the 15 datasets, only for reference.

| Dataset / Method | Sniff-ConvNet | Sinff-Resnet | Sniff-Multinose | DARTS | MLP | SVM |
|---|---|---|---|---|---|---|
| **Wines** | 1475.2 | 148.93 | 83.55 | 5940 | 23.63 | 0.98 |
| **Wines & ethanol** | 1885.66 | 191.22 | 114.2 | 7380 | 33.28 | 1.65 |
| **Coffee** | 237.16 | 134.23 | 39.74 | 960 | 99.9 | 0.0375 |
| **Wind tunnelL1** | 509.87 | 143.09 | 228.52 | 3360 | 65.676 | 6.22 |
| **Wind tunnelL2** | 492.57 | 143.75 | 228.69 | 3120 | 71.746 | 3.44 |
| **Wind tunnelL3** | 527.76 | 162.31 | 256.22 | 3360 | 82.245 | 5.65 |
| **Wind tunnelL4** | 546.04 | 173.22 | 169.41 | 3300 | 90.31 | 5.82 |
| **Wind tunnelL5** | 560.77 | 184.49 | 201.69 | 3300 | 99.32 | 6.17 |
| **Wind tunnelL6** | 545.27 | 185.97 | 285.59 | 2820 | 104.273 | 5.11 |
| **Turbulent gas mixtures** | 3119.55 | 105.13 | 115.94 | 3120 | 55.734 | 2.83 |
| **Twin gas sensorB1** | 5137.12 | 453.2 | 105.99 | 3300 | 28.692 | 1.1 |
| **Twin gas sensorB2** | 5035.33 | 79.2 | 124.11 | 3360 | 24.93 | 1.2 |
| **Twin gas sensorB3** | 5027.64 | 80.42 | 110.2 | 3360 | 31.236 | 1.1 |
| **Twin gas sensorB4** | 2588.93 | 67.41 | 68.47 | 1800 | 32.4921 | 0.4 |
| **Twin gas sensorB5** | 2574.06 | 79.88 | 84.79 | 1860 | 40.126 | 0.4 |

### 5.3.4 Conclusions

In this research, we validated the rapid detection approach (RODRIGUEZ GAMBOA et al., 2019d) in several datasets with diverse electronic nose settings, showing that it is suitable to be used in this field, with better or similar accuracy compared against a conventional approach that needs the complete information of the measurements.

The investigation allowed finding that in the majority of cases is possible to obtain a reliable forecast using only the first 30% (even less) of the measure after the gas injection started. Therefore, subsequent investigations could focus on generating models using only this portion of the gas sensors signals, which entails reducing the time to produce models and make the forecasts (accelerating response).

In this work, we validated the proposed approach using several classification methods; the SVM algorithm and three different DL architecture approach: (i) the Differentiable Architecture Search (DARTS) algorithm, (ii) three deep learning models based on SniffNets, and (iii) a Deep MLP. Although deep learning models are useful when there is a large volume of data, and it can automatically identify patterns. The results showed that using SVM models in the majority of cases, the results are similar or even better and were consistent concerning the early portion of signals needed to make reliable forecasts. Therefore, SVM still is an excellent option in the electronic nose field and could be used to apply the rapid detection approach, as well, the tested deep learning techniques. Still, SVM needs less time for the training process against the other tested classifications methods.

## Acknowledgments

We gratefully acknowledge the support of the following Brazilian agencies: CAPES, CNPq, and FACEPE, who promoted and financed our research project.

## References

[1] P.-F. Qi, Q.-H. Meng, M. Zeng, A CNN-based simplified data processing method for electronic noses, en: 2017 ISOCS/IEEE Int. Symp. Olfaction Electron. Nose, IEEE, 2017: pp. 1-3. https://doi.org/10.1109/ISOEN.2017.7968887.

[2] Y.-J. Liu, Q.-H. Meng, X.-N. Zhang, Data Processing for Multiple Electronic Noses Using Sensor Response Visualization, IEEE Sens. J. 18 (2018) 9360-9369. https://doi.org/10.1109/JSEN.2018.2871599.

[3] Y.-J. Liu, M. Zeng, Q.-H. Meng, Electronic nose using a bio-inspired neural network modeled on mammalian olfactory system for Chinese liquor classification, Rev. Sci. Instrum. 90 (2019) 025001. https://doi.org/10.1063/1.5064540.

[4] M. Längkvist, S. Coradeschi, A. Loutfi, J. Rayappan, M. Längkvist, S. Coradeschi, A. Loutfi, J.B.B. Rayappan, Fast Classification of Meat Spoilage Markers Using Nanostructured ZnO Thin Films and Unsupervised Feature Learning, Sensors. 13 (2013) 1578-1592. https://doi.org/10.3390/s130201578.

[5] P. Peng, X. Zhao, X. Pan, W. Ye, P. Peng, X. Zhao, X. Pan, W. Ye, Gas Classification Using Deep Convolutional Neural Networks, Sensors. 18 (2018) 157. https://doi.org/10.3390/s18010157.

[6] G. Wei, G. Li, J. Zhao, A. He, G. Wei, G. Li, J. Zhao, A. He, Development of a LeNet-5 Gas Identification CNN Structure for Electronic Noses, Sensors. 19 (2019) 217. https://doi.org/10.3390/s19010217.

[7] J.C. Rodriguez Gamboa, E.S. Albarracin E, A.J. da Silva, L. L. de Andrade Lima, T.A. E. Ferreira, Wine quality rapid detection using a compact electronic nose system: Application focused on spoilage thresholds by acetic acid, LWT. 108 (2019) 377-384. https://doi.org/10.1016/J.LWT.2019.03.074.

[8] S.K. Jha, R.D.S. Yadava, K. Hayashi, N. Patel, Recognition and sensing of organic compounds using analytical methods, chemical sensors, and pattern recognition approaches, Chemom. Intell. Lab. Syst. 185 (2019) 18-31. https://doi.org/10.1016/J.CHEMOLAB.2018.12.008.

[9] J.C. Rodriguez Gamboa, E.S. Albarracin E., A.J. da Silva, T.A. Tiago, Electronic nose dataset for detection of wine spoilage thresholds, Data Br. 25 (2019) 104202. https://doi.org/10.1016/j.dib.2019.104202.

[10] C.M. Rodriguez Gamboa, Juan Carlos; Duran Acevedo, Dataset: Electronic Nose for Quality Control of Colombian Coffee through the Detection of Defects in "Cup Tests", (2009). https://doi.org/10.17632/7spd6fpvyk.1.

[11] J. Rodríguez, C. Durán, A. Reyes, Electronic nose for quality control of Colombian coffee through the detection of defects in «Cup Tests»., Sensors (Basel). 10 (2010) 36-46. https://doi.org/10.3390/s100100036.

[12] A. Vergara, J. Fonollosa, J. Mahiques, M. Trincavelli, N. Rulkov, R. Huerta, On the performance of gas sensor arrays in open sampling systems using Inhibitory Support Vector Machines, Sensors Actuators B Chem. 185 (2013) 462-477. https://doi.org/10.1016/J.SNB.2013.05.027.

[13] J. Fonollosa, I. Rodríguez-Luján, M. Trincavelli, A. Vergara, R. Huerta, Chemical Discrimination in Turbulent Gas Mixtures with MOX Sensors Validated by Gas Chromatography-Mass Spectrometry, Sensors. 14 (2014) 19336-19353. https://doi.org/10.3390/s141019336.

[14] J. Fonollosa, I. Rodríguez-Luján, M. Trincavelli, R. Huerta, Data set from chemical sensor array exposed to turbulent gas mixtures, Data Br. 3 (2015) 216-220. https://doi.org/10.1016/J.DIB.2015.02.022.

[15] J. Fonollosa, L. Fernández, A. Gutiérrez-Gálvez, R. Huerta, S. Marco, Calibration transfer and drift counteraction in chemical sensor arrays using Direct Standardization, Sensors Actuators B Chem. 236 (2016) 1044-1053. https://doi.org/10.1016/J.SNB.2016.05.089.

[16] J.C.R. ARAUJO, I. C. S. ; SILVA, A. J. ; GAMBOA, Modelos de deep learning para classificação de gases detectados por matrizes de sensores nariz artificial, (2019).

[17] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, G. Varoquaux, API design for machine learning software: experiences from the scikit-learn project, CoRR. abs/1309.0 (2013). http://arxiv.org/abs/1309.0238.

[18] H. Liu, K. Simonyan, Y. Yang, DARTS Differentiable Architecture Search, CoRR. abs/1806.0 (2018). http://arxiv.org/abs/1806.09055.

[19] A.L. Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, Automatic differentiation in PyTorch, en: NIPS 2017 Work. Autodiff Decis. Progr. Chairs, 2017.

[20] I.C. S. Araujo, SniffNets: Project of the Deep Learning Course, GitHub. (2019). https://github.com/IsmaelCesar/SniffNets (accedido 23 de marzo de 2020).

[21] I.C. S. Araujo, DARTS: Differentiable architecture search for convolutional and recurrent networks, GitHub. (2019). https://github.com/IsmaelCesar/darts (accedido 23 de marzo de 2020).

[22] G.M. Raul Garreta, Learning scikit-learn: Machine Learning in Python, Packt Publishing, 2013. https://books.google.com.co/books?hl=en&lr=&id=OOotAgAAQBAJ&oi=fnd&pg=PT7&dq=Scikit-learn:+Machine+learning+in+Python&ots=45w0uHfShb&sig=zDjylHPGtxoixTfn7CrbVApmax0#v=onepage&q=Scikit-learn%3A Machine learning in Python&f=false (accedido 4 de febrero de 2020).

# CHAPTER 6

## 6. Bibliographical Production

In this section are listed the papers submitted and/or published during the doctoral studies at the *Universidade Federal Rural de Pernambuco* (UFRPE) in the post-graduate program of Biometria e Estatística Aplicada.

RODRIGUEZ GAMBOA, J. C.; ALBARRACIN E., E. S.; DA SILVA, ADENILTON J.; E. FERREIRA, T. A. Electronic nose dataset for detection of wine spoilage thresholds. Data in Brief (RODRIGUEZ GAMBOA et al., 2019b). Related to chapters two and three in this thesis.

RODRIGUEZ GAMBOA, J. C.; ALBARRACIN E., E. S.; DA SILVA, ADENILTON J.; LEITE, L.; E. FERREIRA, T. A. Wine quality rapid detection using a compact electronic nose system: application focused on spoilage thresholds by acetic acid. LWT Food Science and Technology (RODRIGUEZ GAMBOA et al., 2019a). Related to chapter four in this thesis.

ARAUJO, I. C. S.; SILVA, A. J.; GAMBOA, J. C. R. Modelos de deep learning para classificação de gases detectados por matrizes de sensores nariz artificial. Bracis, Salvador - Brasil, 2019 (ARAUJO; GAMBOA; SILVA, 2020). Related to chapter five in this thesis.

RODRIGUEZ GAMBOA, J. C.; DA SILVA, ADENILTON J.; S. ARAUJO, ISMAEL C.; ALBARRACIN E., E. S.; DURAN A., CRISTHIAN M. Validation of the rapid detection approach for enhancing the electronic nose systems performance, using different deep learning models and support vector machines (RODRIGUEZ GAMBOA et al., 2020). This research article compiles the latest results of my doctoral research, and it will be submitted to a recognized scientific journal to publish. We published a preprint version at ArXiv.org. Related to chapter five in this thesis.

ALBARRACIN E., E. S.; RODRIGUEZ GAMBOA, J. C.; M.MARQUES, ELAINE C.; STOSIC, TATIJANA. Complexity analysis of Brazilian agriculture and energy market. Physica A: Statistical Mechanics and its Applications (ALBARRACÍN E. et al., 2019). Related to the proposed future work in this thesis.

# CHAPTER 7

## 7. Conclusion

As long as this doctoral research work was conducted, it was developed an E-Nose system (called O-NOSE.) That device was used to obtain a database for the detection of wine spoilage thresholds, using an array of six metal-oxide-semiconductor (MOS) gas sensors, which includes an automatic module for the adequacy of the gas mixture and sampling system, using air as the conveying gas. The aim of this part of work focused on developing a system with excellent performance and portability, more details in the section: “Experimental Design, Materials, and Methods” (RODRIGUEZ GAMBOA et al., 2019b). It was generated classifiers models using the methods: Support Vector Machine (SVM), and Multilayer Perceptron (MLP) neural network to perform the trials. For additional details, refer to the “Results” section (RODRIGUEZ GAMBOA et al., 2019a).

As a result of this thesis, it was proposed a novel approach for the electronic nose systems (RODRIGUEZ GAMBOA et al., 2019a), treating an early portion of the raw signals (while the measurement process is still running and without applying preprocessing techniques.) The mentioned approach was tested in a wine quality application, and we achieved excellent results against the traditional methodology, using a deep Multilayer Perceptron (MLP) classifier with the raw data of an E-Nose composed by an array of six gas sensors. Then, it was obtained outperformance regarding the traditional method using the rapid detection approach (up to 63 times faster). The proposed methodology focused on reducing the necessary time for making the forecast, accelerating the response time. The results and the explanation of the approach were embodied in a research paper (RODRIGUEZ GAMBOA et al., 2019a), and the collected database and details were published in the data paper (RODRIGUEZ GAMBOA et al., 2019b), the database is available[2].

[2] https://data.mendeley.com/datasets/vpc887d53s/3

Subsequently, we focused on to validate if the rapid detection approach proposed for the electronic nose is suitable to be applied in diverse E-Nose settings (five different databases containing 15 datasets.) Additionally, to prove whether it is worth using Deep Learning models for classification tasks in this field. Consequently, the rapid detection approach could be extended to several E-Nose environments, and the DL approaches are suitable to make forecasts using this approach. However, the SVM still is an excellent option to classify samples in electronic nose applications, and the results even show outperformance against the deep learning architectures tested in several experimental settings (RODRIGUEZ GAMBOA et al., 2020).

As future work, the method to find the optimal early portion with the best performance could be optimized, performing the search with a defined criterion that allows choosing the best slice. Regarding the mentioned issue, it was done some experiments involving the sample entropy and cross-sample entropy, seeking quantifies the information contained in each portion and correlates it with the results. The preliminary results showed coherence and correlation regarding the results, but more trials must be realized.

# References

ADAM PASZKE, SAM GROSS, SOUMITH CHINTALA, GREGORY CHANAN, EDWARD YANG, ZACHARY DEVITO, ZEMING LIN, ALBAN DESMAISON, LUCA ANTIGA, A. L. Automatic differentiation in PyTorch. NIPS 2017 Workshop Autodiff Decision Program Chairs. Anais...2017

ALBARRACÍN E., E. S. et al. Complexity analysis of Brazilian agriculture and energy market. Physica A: Statistical Mechanics and its Applications, v. 523, p. 933-941, jun. 2019.

ALEIXANDRE, M. et al. Quantification of wine mixtures with an electronic nose and a human panel. Frontiers in Bioengineering and Biotechnology, v. 6, n. FEB, p. 14, 12 feb. 2018.

ARAUJO, I. C. S. ; SILVA, A. J. ; GAMBOA, J. C. R. Modelos de deep learning para classificação de gases detectados por matrizes de sensores nariz artificialSalvador - Brasil, 2019.

ARAUJO, I.; GAMBOA, J.; SILVA, A. Modelos de deep learning para classificação de gases detectados por matrizes de sensores nariz artificial. Anais do Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2019). Anais...Sociedade Brasileira de Computação - SBC, 17 ene. 2020Disponível em: <https://sol.sbc.org.br/index.php/eniac/article/view/9339>. Acesso em: 28 mar. 2020

BERNA, A. Metal Oxide Sensors for Electronic Noses and Their Application to Food Analysis. Sensors, v. 10, n. 4, p. 3882-3910, abr. 2010.

BUITINCK, L. et al. API design for machine learning software: experiences from the scikit-learn project. CoRR, v. abs/1309.0, 2013.

BURLACHENKO, J. et al. Sample handling for electronic nose technology: State of the art and future trends. TrAC Trends in Analytical Chemistry, v. 82, p. 222-236, 1 sep. 2016.

CRETIN, B. N.; DUBOURDIEU, D.; MARCHAL, A. Influence of ethanol content on sweetness and bitterness perception in dry wines. LWT - Food Science and Technology, v. 87, p. 61-66, 1 ene. 2018.

DURÁN-ACEVEDO, C. M. et al. Exhaled breath analysis for gastric cancer diagnosis in Colombian patients. Oncotarget, v. 9, n. 48, p. 28805-28817, 22 jun. 2018.

EL BARBRI, N. et al. Electronic Nose Based on Metal Oxide Semiconductor Sensors as an Alternative Technique for the Spoilage Classification of Red Meat. Sensors, v. 8, n. 1, p. 142-156, 21 ene. 2008.

FONOLLOSA, J. et al. Chemical Discrimination in Turbulent Gas Mixtures with MOX Sensors Validated by Gas Chromatography-Mass Spectrometry. Sensors, v. 14, n. 10, p. 19336-19353, 16 oct. 2014.

FONOLLOSA, J. et al. Data set from chemical sensor array exposed to turbulent gas mixtures. Data in Brief, v. 3, p. 216-220, 1 jun. 2015.

FONOLLOSA, J. et al. Calibration transfer and drift counteraction in chemical sensor arrays using Direct Standardization. Sensors and Actuators B: Chemical, v. 236, p. 1044-1053, 29 nov. 2016.

GARDNER, J. W.; BARTLETT, P. N. A brief history of electronic noses. Sensors and Actuators B: Chemical, v. 18, n. 1-3, p. 210-211, mar. 1994.

GUERRINI, L. et al. Smelling, Seeing, Tasting - Old Senses for New Sensing. ACS Nano, v. 11, n. 6, p. 5217-5222, 2017.

HINES, E. L.; LLOBET, E.; GARDNER, J. W. Electronic noses: a review of signal processing techniques. IEE Proceedings - Circuits, Devices and Systems, v. 146, n. 6, p. 297, 1999.

HU, W. et al. Electronic Noses: From Advanced Materials to Sensors Aided with Data Processing. Advanced Materials Technologies, v. 4, n. 2, p. 1800488, 14 dic. 2018.

JASINSKI, G. et al. An electronic nose based on the semiconducting and electrochemical gas sensors. EMPC 2017 - 21st European Microelectronics and Packaging Conference and Exhibition. Anais...Institute of Electrical and Electronics Engineers Inc., 24 abr. 2018Disponível em: <https://ieeexplore.ieee.org/document/8346895/>. Acesso em: 28 mar. 2020

JHA, S. K. et al. Recognition and sensing of organic compounds using analytical methods, chemical sensors, and pattern recognition approaches. Chemometrics and Intelligent Laboratory Systems, v. 185, p. 18-31, 15 feb. 2019.

LÄNGKVIST, M. et al. Fast Classification of Meat Spoilage Markers Using Nanostructured ZnO Thin Films and Unsupervised Feature Learning. Sensors, v. 13, n. 2, p. 1578-1592, 25 ene. 2013a.

LÄNGKVIST, M. et al. Fast Classification of Meat Spoilage Markers Using Nanostructured ZnO Thin Films and Unsupervised Feature Learning. Sensors, v. 13, n. 2, p. 1578-1592, 25 ene. 2013b.

LEFFINGWELL, J. Olfaction–Update No. 5. Leffingwell Reports, v. 2, n. 1, p. 1-34, 2002.

LIU, H.; SIMONYAN, K.; YANG, Y. DARTS Differentiable Architecture Search. CoRR, v. abs/1806.0, 2018.

LIU, Y.-J.; MENG, Q.-H.; ZHANG, X.-N. Data Processing for Multiple Electronic Noses Using Sensor Response Visualization. IEEE Sensors Journal, v. 18, n. 22, p. 9360-9369, 15 nov. 2018a.

LIU, Y.-J.; MENG, Q.-H.; ZHANG, X.-N. Data Processing for Multiple Electronic Noses Using Sensor Response Visualization. IEEE Sensors Journal, v. 18, n. 22, p. 9360-9369, 15 nov. 2018b.

LIU, Y.-J.; ZENG, M.; MENG, Q.-H. Electronic nose using a bio-inspired neural network modeled on mammalian olfactory system for Chinese liquor classification. Review of Scientific Instruments, v. 90, n. 2, p. 025001, 5 feb. 2019a.

LIU, Y.-J.; ZENG, M.; MENG, Q.-H. Electronic nose using a bio-inspired neural network modeled on mammalian olfactory system for Chinese liquor classification. Review of Scientific Instruments, v. 90, n. 2, p. 025001, 5 feb. 2019b.

LOUTFI, A. et al. Electronic noses for food quality: A review. Journal of Food Engineering, v. 144, p. 103-111, 1 ene. 2015.

LOZANO, J.; SANTOS, J. P.; HORRILLO, M. C. Wine Applications With Electronic Noses. En: Electronic Noses and Tongues in Food Science. [s.l.] Elsevier Inc., 2016. p. 137-148.

MARTÍ, M. P. et al. Electronic noses in the quality control of alcoholic beverages. TrAC - Trends in Analytical Chemistry, v. 24, n. 1, p. 57-66, ene. 2005.

MARTINS, N. et al. An ancient winemaking technology: Exploring the volatile composition of amphora wines. LWT, v. 96, p. 288-295, 1 oct. 2018.

ODOTECH; OFF-SITE, I. Odotech White Paper How Does an Electronic Nose Work ? n. January, 2013.

PASK, G. M. et al. Specialized odorant receptors in social insects that detect cuticular hydrocarbon cues and candidate pheromones. Nature Communications, v. 8, n. 1, p. 297, 17 dic. 2017.

PENG, P. et al. Gas Classification Using Deep Convolutional Neural Networks. Sensors, v. 18, n. 2, p. 157, 8 ene. 2018a.

PENG, P. et al. Gas Classification Using Deep Convolutional Neural Networks. Sensors, v. 18, n. 2, p. 157, 8 ene. 2018b.

PERESTRELO, R.; RODRIGUEZ, E.; CÂMARA, J. S. Impact of storage time and temperature on furanic derivatives formation in wines using microextraction by packed sorbent tandem with ultrahigh pressure liquid chromatography. LWT, v. 76, p. 40-47, 1 mar. 2017.

PERIS, M.; ESCUDER-GILABERT, L. Electronic noses and tongues to assess food authenticity and adulterationTrends in Food Science and TechnologyElsevier Ltd, , 1 dic. 2016.

QI, P.-F.; MENG, Q.-H.; ZENG, M. A CNN-based simplified data processing method for electronic noses. 2017 ISOCS/IEEE International Symposium on Olfaction and Electronic Nose (ISOEN). Anais...IEEE, may 2017aDisponível em: <http://ieeexplore.ieee.org/document/7968887/>. Acesso em: 28 oct. 2019

QI, P.-F.; MENG, Q.-H.; ZENG, M. A CNN-based simplified data processing method for electronic noses. 2017 ISOCS/IEEE International Symposium on Olfaction and Electronic Nose (ISOEN). Anais...IEEE, may 2017bDisponível em: <http://ieeexplore.ieee.org/document/7968887/>. Acesso em: 9 may. 2019

RAUL GARRETA, G. M. Learning scikit-learn: Machine Learning in Python. [s.l.] Packt Publishing, 2013.

RODRÍGUEZ-GAMBOA, J. C.; ALBARRACÍN-ESTRADA, E. S.; DELGADO-TREJOS, E. Quality Control Through Electronic Nose System. En: ELDIN, A. B. (Ed.). . Modern Approaches To Quality Control. Rijeka, Croatia: InTech, 2011. p. 505-522.

RODRÍGUEZ-MÉNDEZ, M. L. et al. Electronic noses and tongues in wine industryFrontiers in Bioengineering and BiotechnologyFrontiers Media S.A., , 25 oct. 2016.

RODRIGUEZ GAMBOA, JUAN CARLOS; DURAN ACEVEDO, C. M. Dataset: Electronic Nose for Quality Control of Colombian Coffee through the Detection of Defects in “Cup Tests”. Disponível em: <https://data.mendeley.com/datasets/7spd6fpvyk/1>. Acesso em: 20 mar. 2020.

RODRIGUEZ GAMBOA, J. C. et al. Wine quality rapid detection using a compact electronic nose system: Application focused on spoilage thresholds by acetic acid. LWT, v. 108, 2019a.

RODRIGUEZ GAMBOA, J. C. et al. Electronic nose dataset for detection of wine spoilage thresholds. Data in Brief, 2019b.

RODRIGUEZ GAMBOA, J. C. et al. Mendeley Data - Electronic nose dataset for detection of wine spoilage thresholds. Disponível em: <https://data.mendeley.com/datasets/vpc887d53s/3>. Acesso em: 28 mar. 2020c.

RODRIGUEZ GAMBOA, J. C. et al. Wine quality rapid detection using a compact electronic nose system: Application focused on spoilage thresholds by acetic acid. LWT, v. 108, p. 377-384, 1 jul. 2019d.

RODRIGUEZ GAMBOA, J. C. et al. Electronic nose dataset for detection of wine spoilage thresholds. Data in Brief, v. 25, p. 104202, 1 ago. 2019e.

RODRIGUEZ GAMBOA, J. C. et al. Validation of the rapid detection approach for enhancing the electronic nose systems performance, using different deep learning models and support vector machines. 4 may 2020.

RODRÍGUEZ, J.; DURÁN, C.; REYES, A. Electronic nose for quality control of Colombian coffee through the detection of defects in «Cup Tests». Sensors (Basel, Switzerland), v. 10, n. 1, p. 36-46, ene. 2010.

S. ARAUJO, I. C. SniffNets: Project of the Deep Learning Course. Disponível em: <https://github.com/IsmaelCesar/SniffNets>. Acesso em: 23 mar. 2020a.

S. ARAUJO, I. C. DARTS: Differentiable architecture search for convolutional and recurrent networks. Disponível em: <https://github.com/IsmaelCesar/darts>. Acesso em: 23 mar. 2020b.

SÁENZ-NAVAJAS, M. P. et al. Sensory-active compounds influencing wine experts’ and consumers’ perception of red wine intrinsic quality. LWT - Food Science and Technology, v. 60, n. 1, p. 400-411, 1 ene. 2015.

STUPAK, M. et al. Rapid approach for the determination of alcoholic strength and overall quality check of various spirit drinks and wines using GC–MS. Food Control, v. 80, p. 307-313, 1 oct. 2017.

VAZALLO-VALLEUMBROCIO, G. et al. Commercial enological tannins: Characterization and their relative impact on the phenolic and sensory composition of Carménère wine during bottle aging. LWT - Food Science and Technology, v. 83, p. 172-183, 15 sep. 2017.

VERGARA, A. et al. On the performance of gas sensor arrays in open sampling systems using Inhibitory Support Vector Machines. Sensors and Actuators B: Chemical, v. 185, p. 462-477, 1 ago. 2013.

WEI, G. et al. Development of a LeNet-5 Gas Identification CNN Structure for Electronic Noses. Sensors (Basel, Switzerland), v. 19, n. 1, 8 ene. 2019a.

WEI, G. et al. Development of a LeNet-5 Gas Identification CNN Structure for Electronic Noses. Sensors, v. 19, n. 1, p. 217, 8 ene. 2019b.

WILKENS, W. F.; HARTMAN, J. D. An electronic analog for the olfactory processes. Annals of the New York Academy of Sciences, v. 116, n. 2 Recent Advanc, p. 608-612, jul. 1964.

ZHANG, Q. et al. «Sensory analysis» of Chinese vinegars using an electronic nose. Sensors and Actuators, B: Chemical, v. 128, n. 2, p. 586-593, 15 ene. 2008.

ZHAO, Z. et al. Vortex-assisted dispersive liquid-liquid microextraction for the analysis of major Aspergillus and Penicillium mycotoxins in rice wine by liquid chromatography-tandem mass spectrometry. Food Control, v. 73, p. 862-868, 1 mar. 2017.

刘云翔刘天伟姚文斌杜駉骏郑力刘续博李晓丹. CN105954466A - Electronic nose acquisition system and method for identifying quality of edible spice - Google Patents, 2016. Disponível em: <https://patents.google.com/patent/CN105954466A/en?oq=CN105954466A>. Acesso em: 29 mar. 2020

刘步中陈亮李朝林徐耀庄海军杨永徐通. CN104713985B - 一种基于无线传感器的矿井环境有害气体检测系统 - Google Patents, 2016. Disponível em: <https://patents.google.com/patent/CN104713985B/zh?oq=CN104713985A>. Acesso em: 29 mar. 2020

王俊姜水. CN105606653A - 一种无线电子鼻系统及监测运输过程中果蔬品质的方法 - Google Patents, 2016. Disponível em: <https://patents.google.com/patent/CN105606653A/zh?oq=CN105606653A>. Acesso em: 29 mar. 2020

田逢春谢鑫张健吕博黄扬帆陈建军杨先一廖海林陈小娟. CN104820072B - 基于云计算的电子鼻空气质量监测系统的监测方法 - Google Patents, 2016. Disponível em: <https://patents.google.com/patent/CN104820072B/zh?oq=CN104820072A>. Acesso em: 29 mar. 2020

郑丽敏杨璐任发政朱虹田立军詹小琳王智凝. CN105181890A - 一种嵌入式电子鼻检测系统及方法 - Google Patents. Disponível em: <https://patents.google.com/patent/CN105181890A/zh?oq=CN105181890A>. Acesso em: 29 mar. 2020.