

An Artificial Olfactory System for Toxic Compounds Classification using Machine Learning Techniques

Carlos Cruz^a, Manuel Aleixandre^b, Daniel Matatagui^a, Mari Carmen Horrillo^{a*}

^a SENSAVAN, Instituto de Tecnologías Físicas y de la Información (ITEFI), CSIC, 28006 Madrid, Spain

^b Institute of Innovative Research, Tokyo Institute of Technology, Yokohama 226-8503, Japan
 carmen.horrillo.guemes@csic.es

The long-term exposure to nitrogen dioxide produces harmful effects for humans and any living being. Thus, in security applications, sensor arrays are required for detecting nitrogen dioxide by interfering gas classification. In this work, a compact and intelligent electronic nose (e-nose) based on a Shear-Horizontal Surface Acoustic Wave (SH-SAW) sensor array is proposed for sensing, classifying, and calibrating toxic chemicals. Different carbon-based nanostructured materials are deposited as sensitive layers providing excellent outcomes by mass and elastic changes in this type of sensors. The HS-SAW sensors achieve a high sensitivity, fast response, and reproducibility to different toxic gases such as nitrogen dioxide, carbon monoxide, ammonia, benzene and acetone. The gas flows were controlled by an automated system that consists of four mass flow controllers to obtain the desired concentrations.

The e-nose provides an efficient performance with supervised machine learning techniques. Outcomes indicate that Linear Discrimination Analysis (LDA) performs a 90% precise discrimination on test dataset and provides a clear discrimination of NO₂ with interfering toxic compounds. On the other hand, K-Nearest Neighbors (KNN) and Logistic Regression (LR) also achieve excellent classification scores (95% and 79% respectively). Decision surface for toxic compounds of different classification algorithms were also performed achieving good classification. An evaluation and comparison of the prediction methods: Partial Least Square (PLS), Artificial Neural Networks (ANNs) and cascade of ANNs are accomplished. The ANN cascade results show that this technique is an excellent candidate for an accurate prediction and classification of NO₂. Therefore, the designed and validated e-nose is a promising on-line tool of analysis for environmental applications.

1. Introduction

The performance of a gas sensor depends mainly on the proper use of sensing materials, the low noise and high accuracy of the signal acquisition system (Matatagui et al., 2019; Santos et al., 2012). In this work, the Shear Horizontal Surface Acoustic Wave (HS-SAW) sensors based on carbon sensing materials exhibit excellent sensitivity, response/recovery time, reproducibility, and long-term stability (Jha et al., 2009; de la O-Cuevas et al., 2021). In addition, data processing is an important factor, as the success of the Machine Learning (ML) process relies on it. It aims at extracting robust feature information from the dynamic response of the sensors, which can represent the unique "fingerprint" patterns for a particular gas. To ensure the effectiveness of the subsequent pattern recognition algorithm, ML techniques such as K-nearest neighbors (KNN), Partial Least Squares (PLS) or Artificial Neural Network (ANN) (Aleixandre et al. 2014; Yaqoob et al., 2021; Gutierrez-Osuna et al., 2002, Covington et al., 2021) have been widely used in the achievement of highly selective gas sensors. It is therefore of great importance that sensor signal processing (e.g., algorithms) is integrated into implementing electronic nose for realistic applications.

In this study, we develop a complete system of carbon-based HS-SAW sensor array together with signal processing units and ML algorithms in a smart and compact gas sensor system. The modular architecture provides a very self-contained and versatile platform, incorporating besides the ML capabilities for gas sensing.

2. Experimental setup

Figure 1 shows the experimental design and the processing abilities we have employed for this study. The experimental setup consisted of the e-nose (Figure 2a), the automatic gas line (Figure 2b) and the signal processing system to evaluate the discrimination and the prediction of the gas tested.

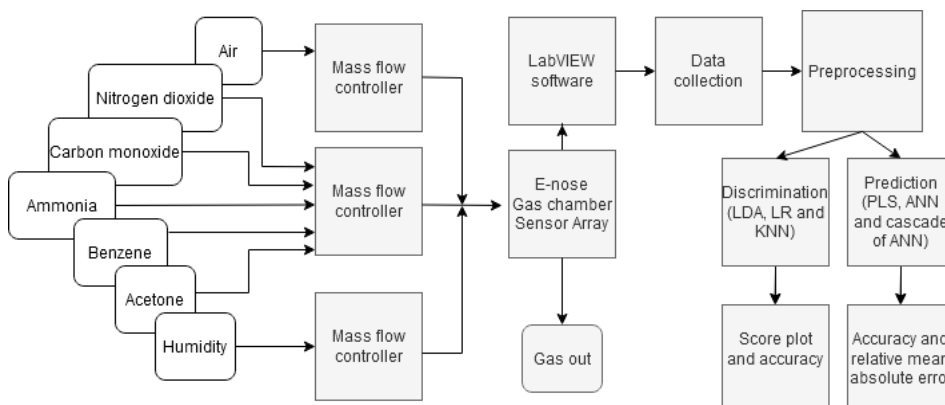


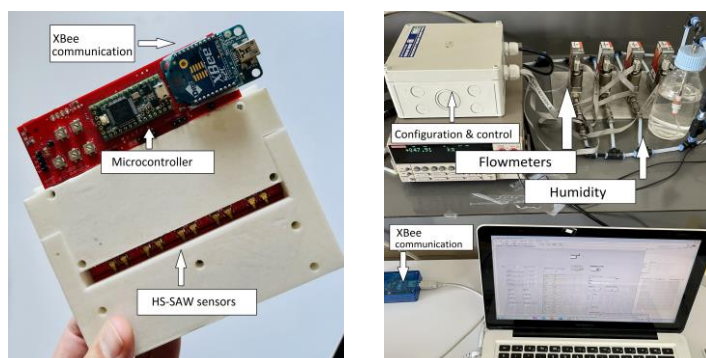
Figure 1: Experimental setup

2.1 Electronic nose

The system was designed using 1) a carbon-based sensor array, 2) a data acquisition stage, 3) a signal conditioning stage and 4) data transmission and software application stages (Figure 2a). The sensor array was developed with four carbon-based nanostructured materials, such as, mesoporous carbon (MC), reduced graphene oxide (rGO), graphene oxide (GO) and polydopamine/reduced graphene oxide (PDA/rGO). The sensors and the signal conditioning modules were mechanically adapted allowing an easy access to carry out changes and manipulation. The signal conditioning module feeds each sensor into a feedback loop that consists of two amplifier steps and a directional coupler. The output of the coupler was used to sample the oscillator frequency. A multiplexer selected one of the sensor-oscillator signals as a single output, which is mixed with the reference oscillator signal. This signal allows the operating frequency to be reduced and compensates for temperature and noise disturbances. In this way, a difference signal is obtained. The output signals are processed (filtered and amplified ones by the analogue-to-digital converter port of the teensy microcontroller), and the teensy module was used as a frequency counter (Matatagui et al., 2019).

2.2 Automatic gas line

An automated flow system controls the flowmeters and allows us to select the gases that enter into the gas cell and the different tested concentrations (Figure 2b). We have used synthetic air as carrier gas. More specifically, the gas control was performed by three Bronkhorst flowmeters. Their control and reading have been performed by two acquisition cards ADAM-4017 and ADAM-4024. The configuration and the reading of the flowmeters have been developed in a LabVIEW acquisition system that performs the pre-treatment and extraction of features of the gas measurements.



(a) (b)
Figure 2: SH-SAW e-nose (a) and implementation of the automatic gas line (b).

2.3 Processing system

The e-nose communication was performed by cable through UART/FIFO controllers and by wireless communication with XBee protocol. The latter method was also employed for the control of the gas line. The outcomes obtained have been processed using ML techniques with a PC using the LabVIEW and the Matlab software.

3. Measurements

Different concentrations of toxic gases were the core of the experiments to obtain the sensor responses: Ammonia (NH_3), benzene (C_6H_6) and acetone ($\text{C}_3\text{H}_6\text{O}$) from 10 to 40 ppm; nitrogen dioxide (NO_2) from 0 to 1 ppm; and carbon monoxide (CO) from 1 to 6 ppm. The exposure time to the gases was 2 minutes. The recovery time in air, among exposures, was 20 minutes. Figure 3 shows the responses obtained by the different sensors for concentrations of $\text{C}_3\text{H}_6\text{O}$, C_6H_6 , and NH_3 20 ppm, NO_2 0.2 ppm and (CO) 2 ppm. Mixtures of gases of only two components were measured. NO_2 with each interfering gas, and with variate humidity (20 % and 40 % of relative humidity) for over 150 measurements.

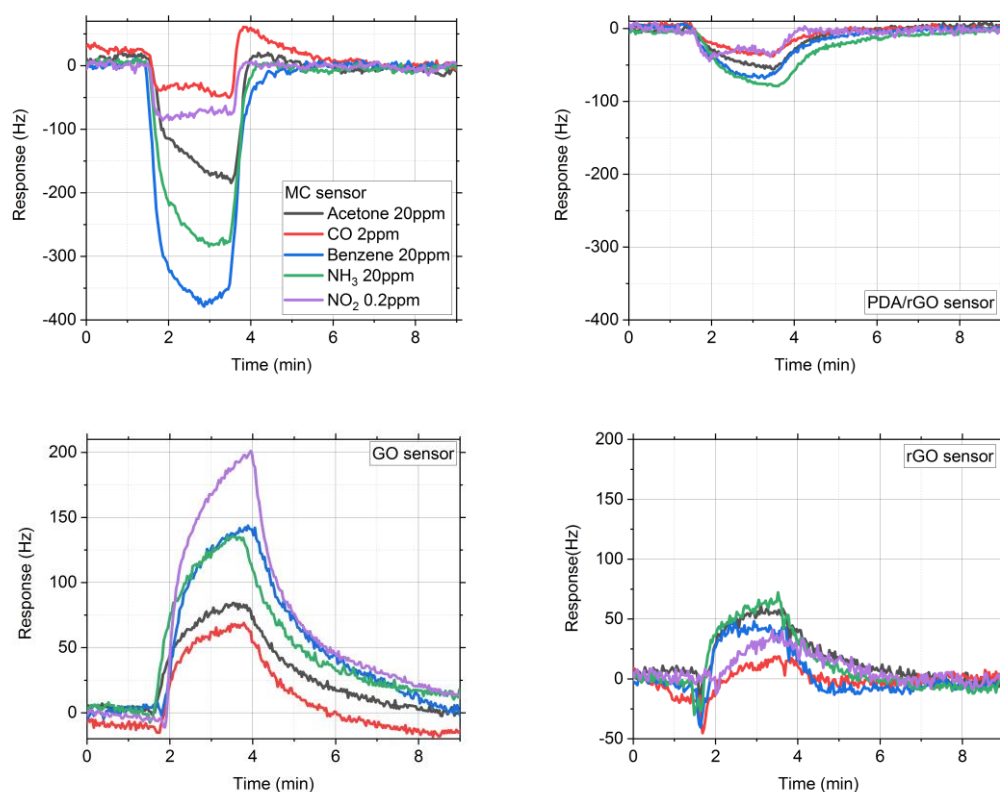


Figure 3: Response and recovery times of a SH-SAW sensor array: GO, rGO, MC and PDA/rGO for specific concentrations of benzene, acetone, ammonia, carbon monoxide and nitrogen dioxide.

4. Data Analysis

Six supervised ML techniques were implemented. One to discriminate the clustering analysis and five ones to validate for classification and prediction purposes, and in this way to determine the most efficient method.

4.1 Linear Discrimination Analysis (LDA)

LDA is used to maximise separation among gases and minimise variance. Its use is focused on the type of point and/or feature and/or subspace that offers the most discrimination to separate the data. Thus, LDA reduces the degree of over-fitting due to dimensionality in non-regularised models. Figure 4 shows the results obtained for the NO_2 and interferences. The discrimination accuracy was 90%.

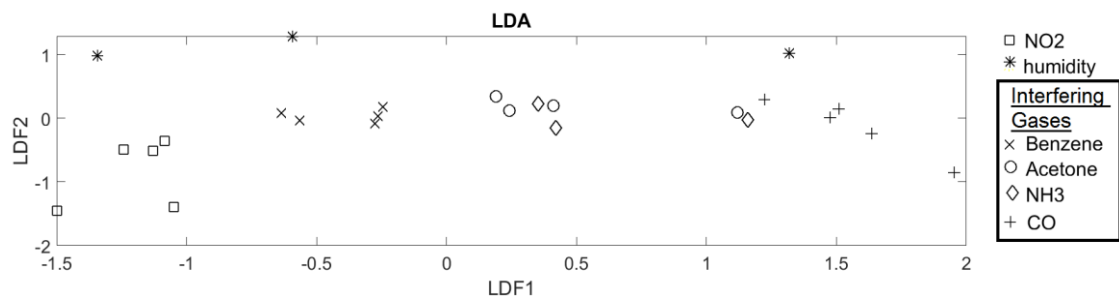


Figure.4: LDA reached an accuracy of around 90% for NO₂ and interferences on the given test data and classes.

4.2 K-nearest neighbor (KNN)

KNN is a subset within ML techniques based on the biological neural networks of the human brain. It starts with an untrained network and establishes a training pattern in the input layer. Signals are then fed through the network and the output is determined at the last layer. This technique takes many labelled points and uses them to learn how to label other ones. Figure 5 shows a grid of points spanning the entire space within some bounds of the sensor responses. KNN correctly classified gases with an accuracy of 95%. However, there is not a clear classification for acetone and benzene over a 3-nearest neighbour classifier training. Figure 5 shows how to plot the decision surface for aive Bayes and Classification tree, which are compared with LDA and KNN classification algorithms.

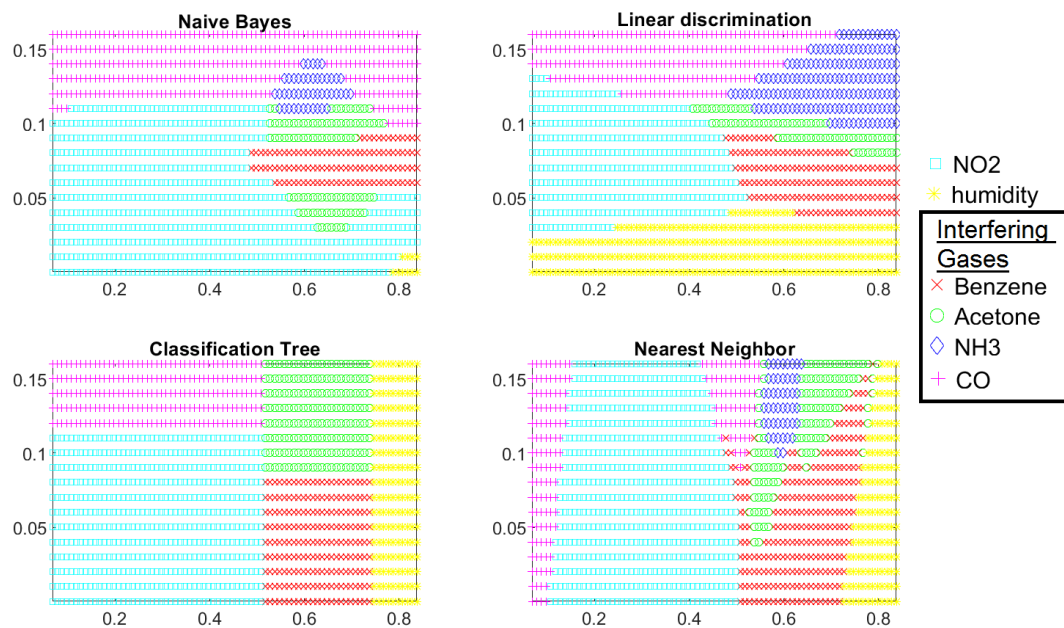


Figure 5: Classification algorithms generate a decision-making rule visualized in the form of a decision surface. 3 repeated measurements are performed from each of six classes.

4.3 Logistic regression (LR)

LR achieves very good performances with linearly two separable classes. A 79% accuracy in the gas classification has been achieved to determine if a new sample fits into the interfering category.

4.4 Partial least squares (PLS)

PLS regression is a quick, efficient, and optimal method based on the standard mathematical approach for fitting a linear regression. The algorithm returns the relative mean absolute error (RMAE) metric for its evaluation and response loadings of the gas responses.

4.5 Artificial neural network (ANN)

ANN is a highly capable ML technique to perform nonlinear and complex tasks with a high accuracy degree. The ANNs were structured into ensembles consisting of 10 feed-forward networks evaluated with the same trained dataset of the normalized sensor responses. The output of the ensembles was the robust mean (removing the lowest and highest outputs of the prediction). The ANN ensembles make the possibility of deviations is less likely for a robust prediction.

4.6 Cascade of ANN

Cascade of ANN is a useful tool to improve the learning since new information is added to an already-trained network. We built and trained three ANN ensembles (NO_2 , interfering gases and humidity). Initially, we trained two ensembles for the NO_2 and the interfering gases. The process was repeated with a third ensemble for only NO_2 . Two additional entries were introduced into this step, which corresponded to the results of the two gases previously obtained. Next, we trained a new ensemble with the humidity responses. The lowest error to predict concentrations was the obtained one from this ensemble.

RMAE is a valuable precision metric if large errors with high consequences are obtained. This score describes the average of the differences among predicted and observed values. Table 1 summarizes the results of this type of Error for the newest ML techniques used. The ANN technique showed a clear improvement over PLS. The ANN cascade showed many better accuracy levels than PLS and ANNs. Figure 6 shows the results of the ANN cascade. However, the training and processing times increase considerably (179 s) if compared with PLS (9 s) and ANN (16 s) as the need to train and the evaluation of several networks increase their complexity with the number of interferences.

Table 1: Metric of the error prediction obtained for the different gases measured through PLS, ANNs, and ANN Cascade.

	RMAE for NO_2 (%)	RMAE for Interferings (%)	RMAE for Humidity (%)	Time (s)
PLS	13.12	17.90	2.01	9
ANN	9.92	15.15	0.30	16
Cascade of ANN	7.62	11.16	0.04	188

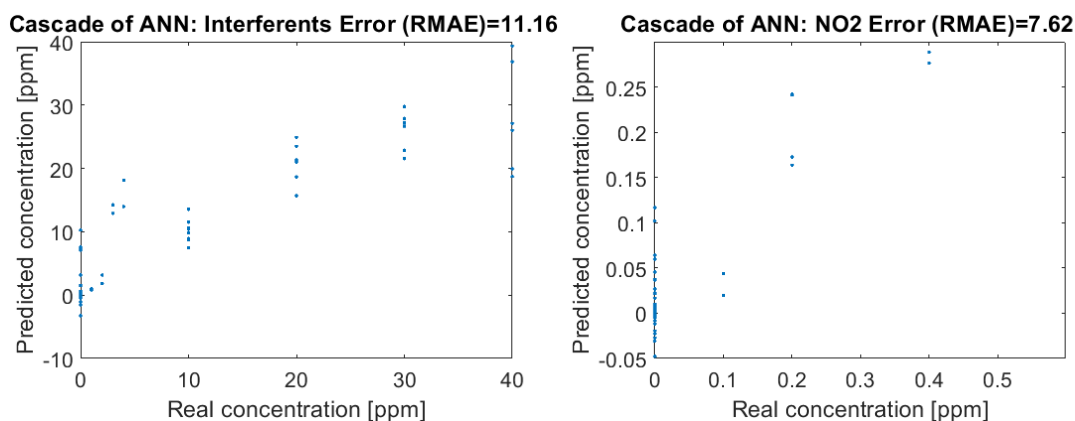


Figure 6: Prediction of the interferences and NO_2 using the cascade of ANN ensemble.

We have presented an innovative neural network structure that uses the parts that are easiest to regress and to assist the prediction of the other parts. Since the measurement space has no priori information from the combination of gases, the improvement must be due to the added information from the prediction at each step. An increasing in the number of neurons does not optimize the result. In addition, an increase of step number in the ANN cascade would increase the complexity of the system and would make a training more complex and difficult by increasing the processing time required. By contrast, a higher statistic will help to reduce the prediction error.

5. Conclusions

In this work, an electronic nose has been presented to classify, discriminate NO₂ with respect to interferences and humidity. After sensors' responses treatment, final results showed that LDA, KNN are excellent candidates with an accuracy of more than 90%. In addition, three different methods have been used for NO₂, interferences and moisture prediction: PLS, ANN and cascade of ANN. All the ML techniques provided excellent results. For instance, the error is reduced in the ANN implementation if compared to the PLS technique. The cascade of ANN ensemble can improve the prediction error by more than 50% for NO₂ and interferences compared to the results obtained for a simple ANN network. This is probably since the networks optimize their results more efficiently with only one class at a time due to a smaller number of possible combinations. Once gases exhibit a well-differentiated response and their prediction is calculated, this information is better incorporated into the network as an independent input. As a result, LDA is an excellent method for discriminating NO₂ and the cascade of ANN ensemble is a novel method for gas prediction through a portable electronic nose.

Acknowledgments

Funding: Spanish Ministry of Science and Innovation for financing the project RTI2018-095856-B-C22 (AEI/FEDER).

References

- Aleixandre M., Matatagui D., Santos J. P., Horrillo M. Carmen., 2014, Cascade of Artificial Neural Network committees for the calibration of small gas commercial sensors for NO₂, NH₃ and CO, *SENSORS*, IEEE, pp. 1803-1806, doi: 10.1109/ICSENS.2014.6985376.
- Covington A., Marco S., Persaud K. C., Schiffman S. S., Nagle H T., 2021, Artificial Olfaction in the 21st Century, in *IEEE Sensors Journal*, vol. 21, no. 11, pp. 12969-12990, doi: 10.1109/JSEN.2021.307641.
- de la O-Cuevas E., Alvarez-Venicio V., Badillo-Ramírez I., Islas S. R., del Pilar Carreón-Castro M., Saniger J. M., 2021, Graphenic substrates as modifiers of the emission and vibrational responses of interacting molecules: The case of BODIPY dyes. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 246, 119020.
- Gutierrez-Osuna R., 2002, Pattern analysis for machine olfaction: A review. *IEEE Sensors journal*, 2(3), 189-202.
- Matatagui D., Bahos F. A., Gràcia I., Horrillo M. Carmen., 2019. Portable low-cost electronic nose based on surface acoustic wave sensors for the detection of BTX vapors in air. *Sensors*, 19(24), 5406.
- Jha S. K., Yadava R. D. S., 2009, Preprocessing of SAW Sensor Array Data and Pattern Recognition, in *IEEE Sensors Journal*, vol. 9, no. 10, pp. 1202-1208.
- Santos, J. P., Aleixandre, M., & Cruz, C., 2012. Hand held electronic nose for VOC detection. *Chemical Engineering*, 30.
- Yaqoob U., Younis M. I., 2021, Chemical gas sensors: Recent developments, challenges, and the potential of machine learning—a review. *Sensors*, 21(8), 2877.