Analysis of feature selection with Probabilistic Neural Network (PNN) to classify sources influencing indoor air quality

Conference Paper *in* AIP Conference Proceedings · March 2017

DOI: 10.1063/1.4975275

7 authors, including:

Shaharil mad saad
Universiti Teknologi Malaysia
46 PUBLICATIONS **464** CITATIONS

ali yeon md shakaff
Universiti Malaysia Perlis
231 PUBLICATIONS **3,122** CITATIONS

Allan Melvin Andrew
Universiti Malaysia Perlis
50 PUBLICATIONS **327** CITATIONS

Ammar Zakaria
Universiti Malaysia Perlis
283 PUBLICATIONS **2,798** CITATIONS

Analysis of Feature Selection with Probabilistic Neural Network (PNN) to Classify Sources Influencing Indoor Air Quality

S. M. Saad[1, a)], A. Y. M. Shakaff[2], A. R. M. Saad[2], A.M. Yusof[2], A. M. Andrew[2], A. Zakaria[2] and A. H. Adom[2]

[1]*Faculty of Mechanical Engineering,*
Universiti Teknologi Malaysia (UTM), 81310 Skudai, Johor Bahru, Malaysia.
[2]*Center of Excellence for Advanced Sensor Technology (CEASTech),*
Universiti Malaysia Perlis (UniMAP), Taman Muhibbah, Jejawi, 02600 Arau, Perlis.

[a)] Corresponding author: shaharil@utm.my

Abstract. There are various sources influencing indoor air quality (IAQ) which could emit dangerous gases such as carbon monoxide (CO), carbon dioxide (CO_2), ozone (O_3) and particulate matter. These gases are usually safe for us to breathe in if they are emitted in safe quantity but if the amount of these gases exceeded the safe level, they might be hazardous to human being especially children and people with asthmatic problem. Therefore, a smart indoor air quality monitoring system (IAQMS) is needed that able to tell the occupants about which sources that trigger the indoor air pollution. In this project, an IAQMS that able to classify sources influencing IAQ has been developed. This IAQMS applies a classification method based on Probabilistic Neural Network (PNN). It is used to classify the sources of indoor air pollution based on five conditions: ambient air, human activity, presence of chemical products, presence of food and beverage, and presence of fragrance. In order to get good and best classification accuracy, an analysis of several feature selection based on data pre-processing method is done to discriminate among the sources. The output from each data pre-processing method has been used as the input for the neural network. The result shows that PNN analysis with the data pre-processing method give good classification accuracy of 99.89% and able to classify the sources influencing IAQ high classification rate.

INTRODUCTION

Most of the people spend their time in indoor environments unless they are travelling to other places, or performing outdoor activities. Therefore, it is no surprise when a study by the United States Environmental Protection Agency (EPA) found that Americans spend about ninety percent of their time in indoor environments [1]. Studies also showed that people's health is highly related to the quality of indoor environments, especially when the indoor air is polluted [2]. For those reasons, it is essential to ensure that the indoor environments are safe and comfortable to live in, including the air we breathe in. To achieve this, continuous IAQ monitoring system is proposed. Currently, researches on IAQ focused on sampling of the indoor air in order to monitor the IAQ in certain public places. Very limited researches have focused on continuous monitoring of IAQ.

Sources of outdoor pollution such as power plants, cars and other transportations are well-known to the public. But, less is known about indoor air pollutions. According to the US EPA, pollution in indoor air is much worse than outdoor air pollution, up to 10 times worse. The sources of indoor air pollutant that release gases or particle into the air are the primary cause of indoor air quality problems in home and office [3]. Indoor air may be polluted by various types of pollutants which may come from office machines, cleaning products, construction activities, carpets and furnishings, perfumes, cigarette smoke, water-damaged building materials, microbial growth, insects and outdoor pollutants. These pollutants emit dangerous gases such as CO, CO_2, O_3 and particle matter. Although these gases are usually safe for human, they could be hazardous to human being especially people with respiratory-related

11th Asian Conference on Chemical Sensors
AIP Conf. Proc. 1808, 020042-1–020042-7; doi: 10.1063/1.4975275
Published by AIP Publishing. 978-0-7354-1476-1/$30.00

problem and children, if their amount exceeded certain limits as proposed by the US EPA. Other physical parameters like indoor temperature, relative humidity and ventilation issues might affect how individual respond to indoor environment as well. Therefore, it is important to identify and control the source of indoor pollutions to prevent or resolve building related illness.

In order to identify the sources influencing IAQ, an indoor air quality monitoring system (IAQMS) was developed. The system of IAQMS is divided into three parts: sensor module cloud (SMC), base station and service-oriented client. SMC contains an array of sensor modules. It is used to measure the indoor air parameters like Nitrogen Dioxide (NO_2), Carbon Dioxide (CO_2), Ozone (O_3), Carbon Monoxide (CO), Oxygen (O_2), Volatile Organic Compounds (VOCs) and Particulate Matter (PM_{10}) along with temperature and humidity. The captured data then is transmitted the captured data to base station through wireless network.

This IAQMS applies a classification method based on Probabilistic Neural Network (PNN). It is used to classify the sources of indoor air pollution based on five conditions: ambient air, human activity, presence of chemical products, presence of food and beverage, and presence of fragrance. The PNN classifier is a non-linear pattern recognition algorithm based on Bayesian classification and classical estimators for probability density function. It uses exponential activation function instead of sigmoidal activation function. The computational time of PNN is less compared to other conventional classifier like Multilayer Perceptron (MLP) classifier [3]. For each algorithm in data pre-processing method, a separate PNN model was developed. The purpose of this classification is to find which algorithm provides the best classification results that can be simple enough to be implemented to the real time system for in situ air quality monitoring.

EXPERIMENT SETUP AND DATA COLLECTION

There are certainly a lot of activities and conditions which could trigger IAP. However, for the purpose of this study, the sources of IAP are limited to 5 conditions only because they are commonly present in indoor environment: ambient air, human activity, presence of chemical products, presence of food and beverage, and presence of fragrance [7][8][9][10]. The first condition of sources of indoor air pollution is the ambient air. Ambient air refers to the air that normally exists in indoor environment without the presence of other sources of indoor air pollution. Ambient air could pollute the indoor air if it carries excessive dust from carpet and furniture or if it carries too much ozone from office machines[7]. The second condition of sources of IAP is human activity. Human activities like smoking cigarette and burning fire woods release poisonous gases like CO, CO_2, and PM at a high concentration than ambient air does which could harm human's health[8][10]. For the purpose of this study, cigarette smoking has been chosen as the proxy for human activity.

The third condition of sources of indoor air pollution is the presence of chemical products or substances. Chemical products like chemical cleaning agents which are usually used in homes and offices may release VOCs at poisonous level. Excessive level of VOCs may lead to respiratory-related diseases such as lung cancer[7]. Thus, for the purpose of this study, chemical cleaning product will be used as the proxy for the presence of chemical. The fourth condition for sources of indoor air pollution is the presence of food and beverage. Cooking activity or certain food and beverage emit VOCs which could lead to uncomfortable smell inside a building[9]. VOCs are known to have led to eye irritation, headache and nausea to certain people. Therefore, rotten fish which has strong smell and high level of VOCs is chosen as the proxy to food and beverage.

Finally, presence of fragrance is the fifth condition from the sources of indoor air pollution. Fragrance like air fresheners and perfumes usually deliver pleasant smell. But, excessive use of perfumes may cause annoyance and headache to certain people. In addition, air fresheners usually emit high amount of VOCs which may cause irritation and discomfort to certain people[7]. For this project, air freshener is used to substitute for the presence of fragrance. Table 1 summarized the 5 conditions for sources of indoor air pollution and their proxy that have been used in this study.

TABLE 1. Sources of indoor air pollutants

Condition	**Proxy**	**Brand**
Human Activity	Cigarette	Marlboro
Chemical Present	Cleaning Agent	Dettol
Fragrance Present	Air Freshness	Ambi Pur Lavender
Food & Beverage Present	Rotten Fish	Mackerel
Ambient Air	Ambient Air	Ambient

Once all the 5 conditions of sources of indoor air pollution have been identified, an experiment simulating the 5 conditions was set up for data collection purposes. The experiment was conducted in medium-size room of 4.5m x 2.4m x 2.6m which is equipped with an air-conditioner located at the center of the room at a height of 2.2m from the floor. The sensor module which is used to collect the data on indoor air was installed hanging up to the wall of the room with 1.1 meter height above the ground, a position considered as breathing zone for the occupants [11]. The sensor module was powered up using an adaptor 7.5V and was programmed to send the data to the base station every 1 minute. The data collection was conducted over 16 days between 9.00 a.m. and 5.00 p.m. with the room Temp set at 22°C. Every day, after each experiment, the air in the room was purged out by opening windows to clean the air. The process of data collection for all 5 conditions is starts from day 1 to day 16 (February 2, 2014 until February 17, 2014).

FEATURE SELECTION

One of the most important parts of intelligent system is its ability to extract useful information that is less redundant than the original one to aid fast processing on pattern recognition or classification. Before doing the data analysis, the sensor output must be processed to free itself from the drift effect, the intensity dependence and possibly from non-linearities [12]. This is called feature extraction. There are many ways or method to do the feature extraction but in this study the feature extraction can be done based on data pre-processing method. Data pre-processing is a procedure that involves on extracting certain significant characteristics from the sensor response curves or transient response in order to produce a set of numerical data or feature for further processing [13]. Choosing the correct pre-processing technique is important because it may induce the success of subsequent analysis and affect the performance of pattern recognition [14]. Most data pre-processing techniques are basically derived from a typical sensor response as shown in Fig. 1 shows when the sensor is exposed to a certain odour. *Vo* is a measured value in clean ambient air or initial value called baseline while *Vs* is response value to odour or smell. Basically, data pre-processing techniques can be divided into three major categories: baseline manipulation, normalization and compression. The study selects 5 data pre-processing techniques which are frequently used in odour pattern recognition as summarized in Table 2. Raw data is also chosen as one of the features.

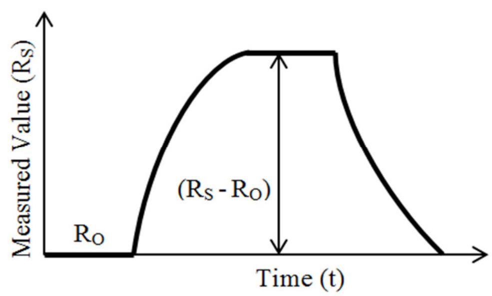


FIGURE 1. Typical sensor response

TABLE 2. Data pre-processing techniques selected

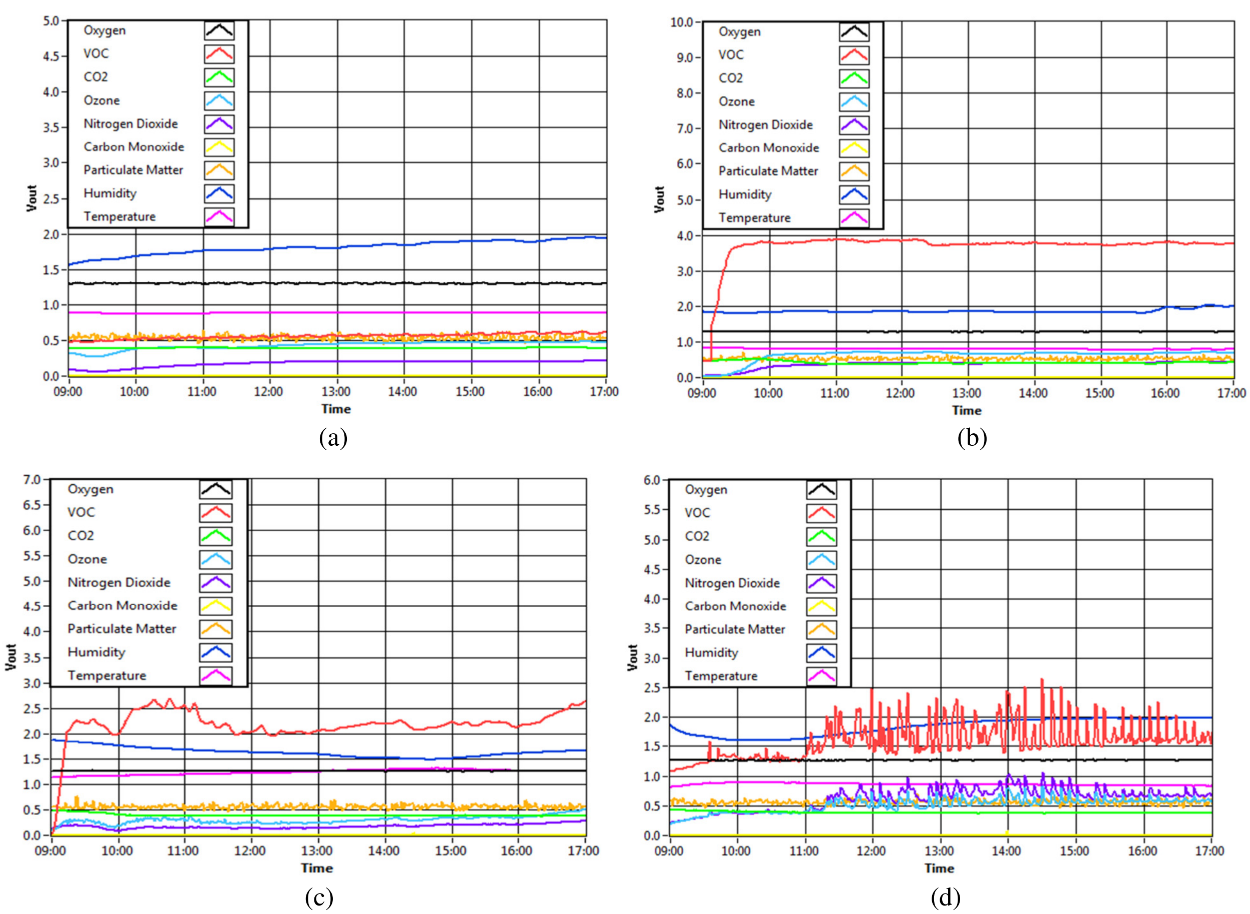
No	Technique	Abbreviation	Formula	References
1	RAW	RW	$X_{ij} = V_{ij}$	[15]
2	Differential	DIFF	$X_{ij} = V_{ij} - V_{bj}$	[14][15] [16]
3	Relative	REL	$X_{ij} = \frac{V_{ij}}{V_{bj}}$	[14][15][16]
4	Fractional	FRACT	$X_{ij} = \frac{V_{ij} - V_{bj}}{V_{bj}}$	[14][15][16]
5	Sensor Normalization	SN	$X_{ij} = \frac{V_{ij} - V_{ij}^{min}}{V_{ij}^{max} - V_{ij}^{min}}$	[14][15][16]
6	Vector Array Normalization	VAN	$X_{ij} = \frac{V_{ij}}{\sqrt{\sum_{q=1}^{N}(V_{ij})^2}}$	[14][15][16]

RESULTS AND DISCUSSIONS

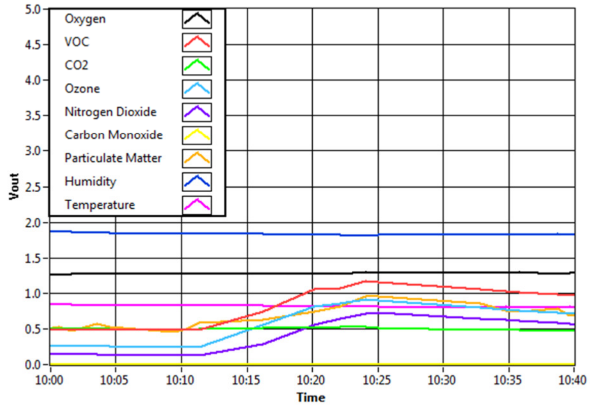
Sensor Response

In this section, sensors' response towards the 5 different conditions of sources of indoor air pollution- ambient air, human activity, chemical presence, fragrance product (air freshener) and foods and beverages (rotten fish) is discussed. Fig. 2a shows that the sensors gave a relatively steady reading throughout the time. The sensors' response was as expected since there was no substance which could interrupt the ambient air concentration. On the other hand, in Fig. 2b, with the presence of chemical substance which is represented by chemical cleaning product, it can be observed that certain gas sensors such as VOCs, NO_2 and O_3 reacted differently as compared to ambient environment. The reading of VOCs gas sensor particularly, raised sharply when the chemical was present in the room. Similar situation could be observed with the presence of food and beverages which is represented by rotten fish as shown in Fig. 2c. The graph for VOCs increased dramatically when the smell was first introduced and then remained at the peak. The graph for other gases did not change much.

Fig. 2d represents the response of the sensors when the automatic air freshener released fragrance into the room every 15 minutes. The fragrance of the air freshener however, vaporized quickly into the air after it was released. Thus, these changes of high and low concentration of fragrance in the air could be observed from the disturbed graph. Meanwhile, for the last condition, 30-minutes data were recorded instead of 8 hours because the effects of cigarette smoke only last for 30 minutes. Fig. 2e illustrates the effect of the cigarette smoking activity on the sensor in the room. Notably, in all graphs, different set of gas sensors reacted differently towards different conditions. In the following sections, the raw data collected went through pattern recognition procedures.



(a) (b) (c) (d)



(e)

FIGURE 2. (**a**) Ambient environment; (**b**) Chemical presence; (**c**) Food and beverages;(**d**) Fragrance presence; (**e**) Human activity;

PNN Analysis

In this section, the result of developed PNN models along with its classification performance of six feature databases, namely, RW, DIFF, REL, FRACT, SN and VAN are discussed. All the PNN classifier models have 9 input neurons and one output neuron. The PNN classifier is simple compared to MLP classifier because it only contain one variables to be adjusted called spread factor, σ. The spread factor was not fixed and was adjusted until the desired performance was achieved. For PNN classifier, the range of spread factor value is from 0 to 2 but for this study only 7 values was chosen which is 0.01, 0.05, 0.10, 0.50, 1.00, 1.50 and 2.00. For each feature, a separate neural network model was developed to classify the eight different methods. Each PNN model used two datasets; training dataset with 2880 training samples (60% of 4800) and was tested with the remaining 1920 samples (40% of 4800). The detailed parameter for PNN training is given in Table 3.

TABLE 3. Parameters for PNN training

Training Parameter for PNN	Value
Sample • Number of samples used for training: 2880 • Number of samples used for testing: 1920	4800
Input Neurons	9
Spread Factor	Flexible
Output neurons	5
Testing Tolerance	0

The classification results for the RW database is shown in Table 4. From Table 4, it is observed that the maximum classification accuracies for the RW features network model is 99.39% with spread factor is 0.05.

TABLE 4. Performance of PNN for RW feature

PNN Model	Spread Factor	Classification Accuracy (%)
1	0.01	99.25
2	0.05	99.39
3	0.10	98.92
4	0.50	97.11
5	1.00	96.86
6	1.50	92.70
7	2.00	82.57

The same procedure was repeated for the other features for PNN which are DIFF, REL, FRACT, SN and VAN. The classification rate for these features is shown in Table 5 and Fig. 3 along with the classification rate for RW feature. It can be seen that all features obtained remarkably high classification accuracy of above 99% except for SN feature. SN feature gained the lowest classification accuracy at 95.07% (spread factor is 0.01). The highest accuracy is gained by VAN feature again with 99.89% (spread factor is 0.01) classification accuracy. The result showed that the test classifications highly correlate with the sample data. The PNN model was able to classify the source that influenced IAQ and the classification rate was 99.89%.

TABLE 5. The best performance of PNN for each feature

PNN Model	Spread Factor	Classification Accuracy (%)
RW	0.05	99.39
DIFF	0.01	99.31
REL	0.01	99.40
FRACT	0.01	99.33
SN	0.01	95.07
VAN	0.01	99.89

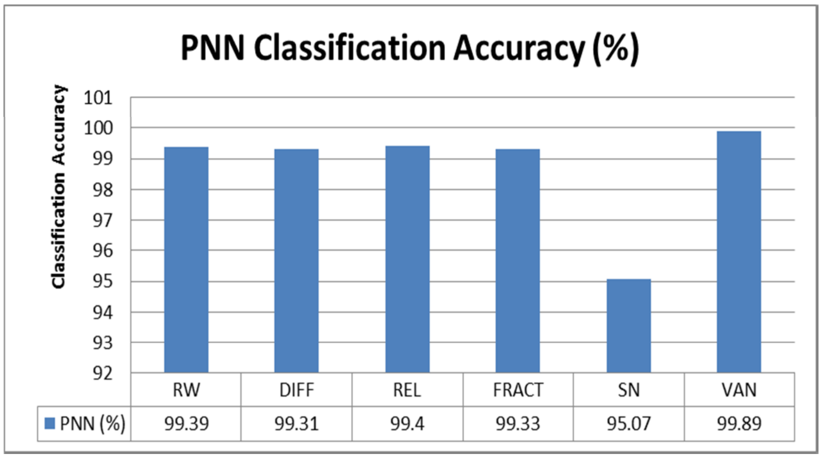


FIGURE 3. PNN classification rate for each feature

CONCLUSION

In this paper, an indoor air quality monitoring system (IAQMS) has been developed with the additional function of classifying sources influencing the IAQ based on five different environments such as ambient air, chemical presence, fragrance presence, foods and beverages, and human activity. This IAQMS consists of three parts: sensor module cloud, base station and service-oriented client. The sensor module cloud (SMC) contains collections of sensor modules that measure the air quality data and transmit the captured data to base station through wireless network. Each sensor modules includes an integrated sensor array that can measure indoor air parameters like Nitrogen Dioxide (NO_2), Carbon Dioxide (CO_2), Ozone (O_3), Carbon Monoxide (CO), Oxygen (O_2), Volatile Organic Compounds (VOCs) and Particulate Matter (PM_{10}) along with temperature and humidity. This IAQMS applies a classification method based on Probabilistic Neural Network (PNN) as classification method. The PNN classifier is simple compare to other neural network classifier like Multilayer Perceptron (MLP) classifier because it only contain one variables to be adjusted called spread factor, σ. In order to get good and best classification accuracy, an analysis of several feature selection based on data pre-processing method is done to discriminate among of sources. The output from each data pre-processing method has been used as the input for the neural network. The result shows that the PNN model with data pre-processing method based on VAN feature give a good classification accuracy again with 99.89% (spread factor is 0.01) classification accuracy. The result showed that the test classifications highly correlate with the sample data. Overall, it can be concluded that the system delivered a high classification rate based feature selection and PNN analysis.

ACKNOWLEDGMENTS

The equipment used in this project was pre-developed by Universiti Malaysia Perlis (UniMAP) for previous studies. The authors would like to offer a special thanks to the members of the Centre of Excellence for Advanced Sensor Technology (CEATech), UniMAP, Malaysia for their critical advice and warm cooperation. The authors also would like to acknowledge the financial sponsorship provided by Malaysian Technical University Network (MTUN) grant and the Malaysian Ministry of Higher Education (MOHE) under myBrain15 scheme. Lastly, the authors would like to thanks to Universiti Teknologi Malaysia (UTM) for supporting this research.

REFERENCES

1. EPA, "Buildings and their impact on the environment: A statistical summary," *U.S. Environmental Protection Agency Green Building Workgroup*, 2009. [Online]. Available: http://www.epa.gov/greenbuilding/pubs/gbstatpdf. [Accessed: 26-Nov-2014].
2. ASHRAE, "Ventilation for acceptable indoor air quality," 1999.
3. EPA, "Building air quality - a guide for building owners and facility managers," 1991.
4. J. Gonzalez-Jimenez, J. G. Monroy, and J. L. Blanco, "The multi-chamber electronic nose-an improved olfaction sensor for mobile robotics," *Sensors*, vol. 11, no. 6, pp. 6145–6164, 2011.
5. M. Budde, M. Busse, and M. Beigl, "Investigating the use of commodity dust sensors for the embedded measurement of particulate matter," in *Networked Sensing Systems (INSS), 2012 Ninth International Conference on Networked Sensing*, 2012, pp. 1–4.
6. C. Technologies, "HSM-20G Humidity Sensor User's manual," 2009.
7. EPA, "Indoor air: An introduction to indoor air quality (IAQ)," *US Environmental Protection Agency*, 2014. [Online]. Available: http://www.epa.gov/iaq/ia-intro.html. [Accessed: 27-Nov-2014].
8. R. B. G Invernizzi, A Ruprecht, R Mazza, E Rossetti, A Sasco, S Nardini, "Particulate matter from tobacco versus diesel car exhaust: an educational perspective," 2004.
9. K. Meena, "Indoor air pollution: sources, health effects and mitigation strategies," 2009.
10. Loomis, D., Grosse, Y., Lauby-Secretan, B., El Ghissassi, F., Bouvard, V., Benbrahim-Tallaa., "The carcinogenicity of outdoor air pollution," *Lancet Oncol.*, vol. 14, pp. 1262–1263, 2013.
11. DOSH, "Industry code of practice on indoor air quality," Kuala Lumpur, 2010.
12. A. C. Romain, J. Nicolas, V. Wiertz, J. Maternova, and P. Andre, "Use of a simple tin oxide sensor array to identify five malodours collected in the field," *Sensors Actuators, B Chem.*, vol. 62, pp. 73–79, 2000.
13. C. Distante, M. Leo, P. Siciliano, and K. C. Persaud, "On the study of feature extraction methods for an electronic nose," *Sensors Actuators, B Chem.*, vol. 87, no. 2, pp. 274–288, 2002.
14. R. Gutierrez-Osuna and H. T. Nagle, "A method for evaluating data-preprocessing techniques for odor classification with an array of gas sensors," *IEEE Trans. Syst. Man, Cybern. Part B Cybern.*, vol. 29, no. 5, pp. 626–632, 1999.
15. J. Nicolas, a. C. Romain, V. Wiertz, J. Maternova, and P. André, "Using the classification model of an electronic nose to assign unknown malodours to environmental sources and to monitor them continuously," *Sensors Actuators, B Chem.*, vol. 69, pp. 366–371, 2000.
16. J. W. Gardner and P. N. Bartlett, "A brief history of electronic noses," *Sensors Actuators B Chem.*, vol. 18, no. 1–3, pp. 210–211, 1994.