chemistry proceedings

MDPI

Proceedings

Statistical Analysis for Selective Identifications of VOCs by Using Surface Functionalized MoS_2 Based Sensor Array †

Uttam Narendra Thakur [1,]*, Radha Bhardwaj [2], and Arnab Hazra[3,]*

[1] Dept. of Electrical & Electronics Engineering, Birla Institute of Technology and Science (BITS)-Pilani, Vidya Vihar
[2] Dept. of Electrical & Electronics Engineering, Birla Institute of Technology and Science (BITS)-Pilani, Vidya Vihar; arnabhazra2013@gmail.com
[3] Dept. of Electrical & Electronics Engineering, Birla Institute of Technology and Science (BITS)-Pilani, Vidya Vihar
* Correspondence: unthakur08@gmail.com (U.N.T); arnabhazra2013@gmail.com (A.H.)

Abstract: Disease diagnosis through breath analysis have attracted a significant attention in recent years due to its non-invasive nature, rapid testing ability and applicability for the patients of all ages. More than 1000 volatile organic component (VOC) exists in human breath, but only a selected VOCs are associated with specific diseases. Selective identifications of those disease marker VOCs by using array of multiple sensors is highly desirable in the current scenario. Not only the use of efficient sensors but also the use of suitable classification algorithms is essential for the selective and reliable detection of those disease markers in the complex breath. In the current study, we fabricated noble metals (Au Pd and Pt) nanoparticles functionalized MoS_2 based sensor array for the selective identifications of different VOCs. Four sensors i.e. pure MoS_2, Au/MoS_2, Pd/MoS_2 and Pt/MoS_2 were tested in the exposure different VOCs like acetone, benzene, ethanol, xylene, 2-propenol, methanol and toluene at 50°C. Initially, principal component analysis (PCA) and linear discriminant analysis (LDA) were used to discriminate those seven VOCs. As compared to the PCA, LDA was able to discriminate well among the seven VOCs. Four different machine learning algorithms like k-nearest neighbors (kNN), decision tree, random forest and multinomial logistic regression was used to identify those VOCs further. The classification accuracy of those seven VOCs by using KNN, decision tree, random forest and multinomial logistic regression were 97.14%, 92.43%, 84.1% and 98.97% respectively. These results authenticated that multinomial logistic regression performed best among all the four machine learning algorithms to discriminate and differentiate multiple VOCs popularly exists in human breath.

Keywords: Breath analysis; surface functionalized MoS_2; classification; discrimination

Published: 1 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

In the field of medical diagnostic and health care systems, breath analysis has gained a lot of interest for the non-invasive detection of diseases and monitoring health parameters [1,2]. More than 1000 volatile organic components (VOCs) are present in the exhaled breath, but only some of them are considered disease markers[3,4]. In this context, selective detection of the different VOCs using smart sensor systems has a high demand for efficient breath analysis. Selective detection can also be achieved by using suitable pattern recognition algorithms on sensor signals. For early detection of disease, the combination of a highly selective sensors and an effective machine learning algorithm is required. Diagnostic through breath is less time-consuming compared to the clinical process and, at the same time, it is cost-efficient as it does not require well-trained professionals and sensors are less costly[5,6].

Chemiresistive sensors typically recognize target VOC by changing its resistance depending upon the adsorption-desorption properties of the analyte to the detecting layer surface. An extensive variety of materials are used for VOC sensing, including thin metal films [7], metal oxides [8–10], polymers [11], etc. Accessible surface functionalization possibilities, high surface area to volume ratio, increased flexibility, high sensitivity, and tunable bandgap make two-dimensional molybdenum disulfide (MoS_2) an encouraging channel material to sense the VOC [12,13].

Pattern recognition algorithm also plays an essential role in the detection of VOC. A suitable classifier is required to achieve an effective classification rate in VOC sensing based on the sensor data. Different algorithms like partial least squares discriminant analysis [14], canonical discriminant analysis [15], k-nearest neighbor [4,16], Discriminant Function Analysis [17], support vector machine [18], random forest [19], logistic regression [20], etc. were reported in the literature. In some of the reported literature, different types of neural network classifier were used [21–24].

In the current study, we have used principal component analysis (PCA) and linear discriminant analysis (LDA) to visualize the data in lesser dimensions compared to the original extent. Also, four different supervised algorithms, k-nearest neighbour (kNN), decision tree, random forest, and multinomial logistic regression, were implemented to identify the best-suited algorithm based on their performance.

2. Material and methods

2.1.Preparation of MoS_2 and noble metal nanoparticles solutions

All materials MoS_2 (Sigma Aldrich), gold (III) chloride ($AuCl_3$, 99 %, Sigma Aldrich), palladium chloride ($PdCl_2$, 60%, Molychem) and chloroplatinic acid ($H_4PtCl_6xH_2O$, 40 %, Molychem) were analytical grade and used without further any purification. 0.2 Wt% MoS_2 solution was prepared in deionized water and stirred for 1.5 h at room temperature to maintain homogeneity. And similarly, 0.1 MM aqueous solutions of noble metal nanoparticles (Au,Pd,Pt) were prepared by adding corresponding metal salts in deionized water with continuous stirring and dropwise diluted HCl was also added to get stable and uniform nanoparticles at room temperature.

Au, Pd and Pt nanoparticle loaded MoS_2 samples were prepared by spray coating technique. Firstly, MoS_2 solution was spray coated on washed SiO_2/Si substrate and dried at room temperature. And in final step, nanoparticle solutions was spray coated on previously deposited MoS_2 and dried at room temperature.

A thermal annealing was performed for 4 h at 250 °C to provide crystallization and thermal stability in all 4 samples (MoS_2, Au-MOS_2, Pd-MoS_2 and Pt-MoS_2).

2.2. Fabrication of Sensors

Au source and drain electrodes of 150 nm thickness were deposited on all four samples by using electron bean evaporation unit. Sensors was then placed into a sensor holder and further sensing performance was studied.

The sensor holder was placed in glass sealed sensing chamber of size 650 ml on a heating plate. The sensing performance of prepared sensors was examined by static mode sensing setup where, VOCs were injected by using micro syringes (Hamilton micro syringe) and sensor was recovered by flowing 450 SCCM synthetic air by using mass flow controller. The amount of injected VOC was calculated by using formula: C (ppm) = 2.46 × (V1D/VM) × 103, where D (gm/mL), M (gm/mol) and V (Lit) represent density of the VOC, molecular weight of the VOC and volume of vaporization chamber respectively[13,25,26]. 7 different VOCs, i.e. acetone, 2-propanol, benzene, ethanol, methanol, toluene, and xylene were tested during the study. Sensing performance was recorded by using Keithley 6487 source meter applying 1 V constant bias. The sensitivity of the sensor was calculated by formula; Ra-Rv/Ra×100 where Ra and Rv were the resistances of the sensor in the air and in target VOC.

To read the generated output of sensors stored in CSV file a python script was used. All the algorithms, analysis, and plotting were performed on Python 3.7 and Jupiter notebook as a platform.

3. Results and discussion

3.1.VOC sensing

As a reference ambient, synthetic air was used to perform the gas sensing measurements of four different sensors: pure MoS_2, Au- MoS_2, Pd- MoS_2 and Pt- MoS_2. Figure 1 shows the change in resistance (MΩ) with respect to time at 50^{o}C. In the presence of VOCs, as the exposer time increases, the resistance offered by the sensor is decreasing. This decrease in resistance confirms that the sensor is n-type property. In the presence of seven distinct VOCs, i.e. acetone, 2-propanol, benzene, ethanol, methanol, toluene, and xylene Four different sensors, i.e. pure MoS_2, Au- MoS_2, Pd- MoS_2 and Pt- $MoS_{2,}$ were observed and stored for further processing of data.

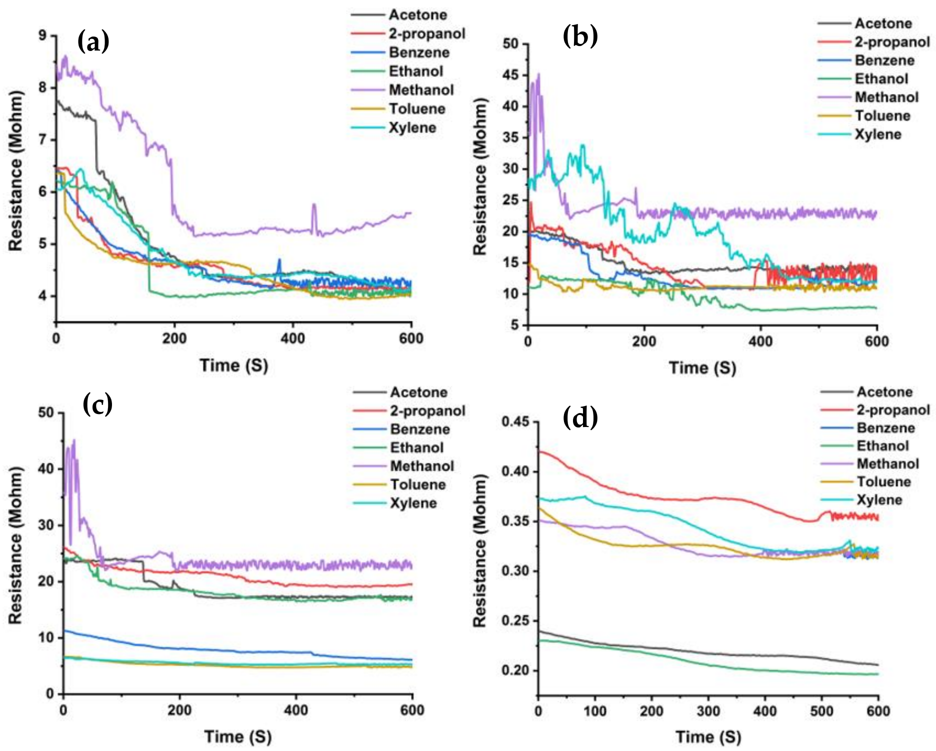

Figure 1. Change in resistance offered by sensors (a) MoS_2 (b) Au- MoS_2 (c) Pd- MoS_2 (d) Pt- MoS_2 with respect to time in presence of 7 VOCs.

3.2. Data analysis

Figure 2 describes the influence of volatile organic components (VOCs) on the outcomes of two-dimensionality reduction techniques: principal component analysis and linear discriminant analysis. The measurement parameters were kept constant during the experiment. Operating temperature was 50^{o}C, response was taken up to 600 sec. and the sample concentration was 100 ppm.

The response obtained by the four different sensors for seven different VOCs was used for principal component analysis (PCA). The three-dimensional plot between the first principal component (PC1), second principal component (PC2), and third principal component (PC3) is represented in figure 2. As we have four independent variables (sensor responses), the maximum principal component obtained was four. Therefore, in this analysis, we have considered only the first three principal components contributing the most to the explained variance. The total explained variance was 93.58%, in which PC1 contributes 52.52%, PC2 contributes 30.91%, and PC3 contributes 10.14%. All seven VOCs have their compact cluster, and they have separated, but the separation between the cluster of acetone/2-propanol and benzene/toluene is quite less that increases the possibility of the misclassification.

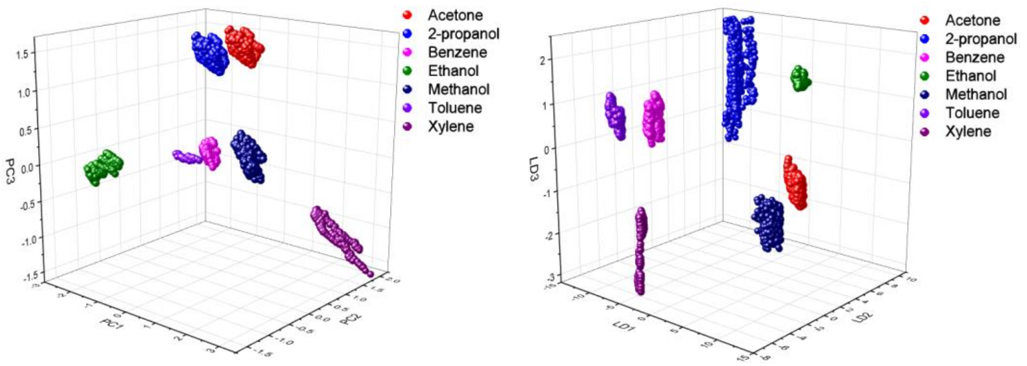

Figure 2. Scatter plot from the exposer of 4 sensors to seven VOCs in (**a**) PCA (**b**) LDA.

Taking account of the problem of discrimination among the different VOCs, linear discrimination analysis was performed, too. In linear discriminant analysis (LDA), the same sensor response vector was used. Figure 2(b) shows that the employment of the classifier allows the discrimination of all the seven VOCs. Thus, LDA is highly efficient for investigating the VOCs based on the sensor response. A three-dimensional plot is shown in figure 2(b), which clearly depicts the performance of LDA on the raw data (sensor response vector). The different VOCs are densely clustered within their groups, and they are well separated from each other. So there is a significantly less probability of misclassification among the VOCs. 2-propanol is slightly more stretched along the axis of the second linear discriminant function (LD2), and xylene is along the third discriminant function (LD3). The three discriminant function, LD1, LD2, and LD3 contributes 71.22%, 27.42% and 1.21% respectively, the total resultant explained variance for the classifier becomes 99.85%.

3.3. VOC identification

The previously discussed LDA and PCA plot gives only the visual representation of the separation of VOCs based on the sensor response. The goal of the sensor setup is to design a generalized model based on the known data during the training phase and tries predict the class when an unknown data sample is encountered.

The supervised algorithm was performed in the current work to determine the VOCs; four different machine learning algorithms like k-nearest neighbour (kNN), decision tree, random forest, and multinomial logistic regression were used to identify those seven VOCs. The normalized sensor response was feed to the algorithms, and the whole data set was divided into training testing data with 70% and 30%, respectively. The data set consists of 4200 measurements of each sensor, with each class containing 600 data vectors and seven classes. So, 2940 vectors were used to train the model, and the remaining 1260 vectors were used to test the model. For identification of VOCs, above 84% was the classification accuracy for every classifier with an accuracy of 97.14%, 92.43%, 84.1%, 98.97% for kNN, decision tree, random forest, and multinomial logistic regression, respectively. A confusion matrix is used to calculate the classification accuracy, and the confusion matrix furnishes the observation into what components were mistakenly classified. Figure 4(a) shows the confusion matrix of kNN where 11 samples of toluene were classified as xylene and 10 samples of benzene was wrongly predicted as ethanol. Figure 4(b) is a representation of the confusion matrix obtained from the decision tree algorithm. The confusion matrix of the random forest and multinomial logistic regression are shown in Figures 4(c) and 4(d), respectively. In multinomial logistic regression, only 12 benzene samples were identified as acetone, and one sample of xylene was identified as toluene.

with an accuracy of 97.14%, 92.43%, 84.1%, 98.97% for kNN, decision tree, random forest, and multinomial logistic regression, respectively.

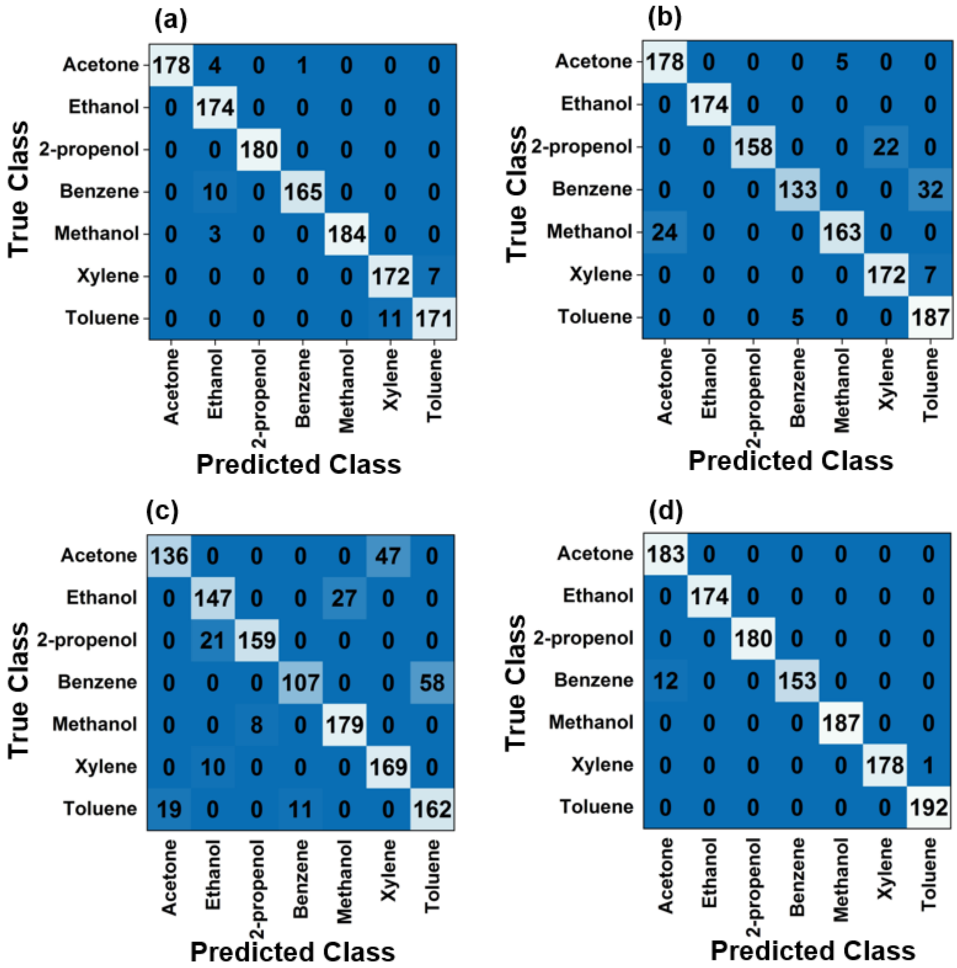

Figure 4. Confusion matrix of (**a**) k-nearest neighbour, (**b**) decision tree, (**c**) random forest, and (**d**) multinomial logistic regression

4. Conclusions

The capability of surface-functionalized MoS_2 sensor to distinguish between the various VOCs was appraised by PCA and LDA, in which LDA laid out the excellent separation between VOCs. Further, to evaluate the effectiveness of sensor output to identify the VOCs, four different machine learning (supervised) based classification algorithms were implemented, and among them, k-nearest neighbour and multinomial logistic regression performed outstandingly with an accuracy of 97.14% and 98.97%, respectively. Thus, high selectivity and accuracy authenticate that the system discriminates and differentiates multiple VOCs popularly exists in human breath.

References

1. S. Dragonieri, G. Pennazza, P. Carratu, and O. Resta, "Electronic Nose Technology in Respiratory Diseases," *Lung*, vol. 195, no. 2, pp. 157–165, 2017, doi: 10.1007/s00408-017-9987-3.
2. P. Sahatiya, A. Kadu, H. Gupta, P. Thanga Gomathi, and S. Badhulika, "Flexible, Disposable Cellulose-Paper-Based MoS 2 /Cu 2 S Hybrid for Wireless Environmental Monitoring and Multifunctional Sensing of Chemical Stimuli," *ACS Appl. Mater. Interfaces*, vol. 10, no. 10, pp. 9048–9059, Mar. 2018, doi: 10.1021/acsami.8b00245.

3. F. S. Cikach and R. A. Dweik, “Cardiovascular Biomarkers in Exhaled Breath,” *Prog. Cardiovasc. Dis.*, vol. 55, no. 1, pp. 34–43, 2012, doi: 10.1016/j.pcad.2012.05.005.
4. J. Pereira *et al.*, “Breath analysis as a potential and non-invasive frontier in disease diagnosis: An overview,” *Metabolites*, vol. 5, no. 1, pp. 3–55, 2015, doi: 10.3390/metabo5010003.
5. K. Arshak, E. Moore, G. M. Lyons, J. Harris, and S. Clifford, “A review of gas sensors employed in electronic nose applications,” *Sens. Rev.*, vol. 24, no. 2, pp. 181–198, 2004, doi: 10.1108/02602280410525977.
6. V. H. Tran *et al.*, “Breath analysis of lung cancer patients using an electronic nose detection system,” *IEEE Sens. J.*, vol. 10, no. 9, pp. 1514–1518, 2010, doi: 10.1109/JSEN.2009.2038356.
7. A. Pundt, “Hydrogen in Nano-sized Metals,” *Adv. Eng. Mater.*, vol. 6, no. 12, pp. 11–21, Feb. 2004, doi: 10.1002/adem.200300557.
8. A. Hazra, “Amplified Methanol Sensitivity in Reduced Graphene Oxide FET Using Appropriate Gate Electrostatic,” *IEEE Trans. Electron Devices*, vol. 67, no. 11, pp. 5111–5118, Nov. 2020, doi: 10.1109/TED.2020.3025743.
9. T. Gakhar and A. Hazra, “Oxygen vacancy modulation of titania nanotubes by cathodic polarization and chemical reduction routes for efficient detection of volatile organic compounds,” *Nanoscale*, vol. 12, no. 16, pp. 9082–9093, 2020, doi: 10.1039/c9nr10795a.
10. E. Kanazawa *et al.*, “Metal oxide semiconductor N2O sensor for medical use,” *Sensors Actuators B Chem.*, vol. 77, no. 1–2, pp. 72–77, Jun. 2001, doi: 10.1016/S0925-4005(01)00675-X.
11. H. Bai and G. Shi, “Gas Sensors Based on Conducting Polymers,” *Sensors*, vol. 7, no. 3, pp. 267–307, Mar. 2007, doi: 10.3390/s7030267.
12. V. Selamneni, H. Raghavan, A. Hazra, and P. Sahatiya, “MoS 2 /Paper Decorated with Metal Nanoparticles (Au, Pt, and Pd) Based Plasmonic-Enhanced Broadband (Visible-NIR) Flexible Photodetectors,” *Adv. Mater. Interfaces*, vol. 8, no. 6, p. 2001988, Mar. 2021, doi: 10.1002/admi.202001988.
13. P. Bindra and A. Hazra, “Selective detection of organic vapors using TiO2 nanotubes based single sensor at room temperature,” *Sensors Actuators B Chem.*, vol. 290, pp. 684–690, Jul. 2019, doi: 10.1016/j.snb.2019.03.115.
14. C. Di Natale *et al.*, “Lung cancer identification by the analysis of breath by means of an array of non-selective gas sensors,” *Biosens. Bioelectron.*, vol. 18, no. 10, pp. 1209–1218, Sep. 2003, doi: 10.1016/S0956-5663(03)00086-1.
15. N. Fens *et al.*, “External validation of exhaled breath profiling using an electronic nose in the discrimination of asthma with fixed airways obstruction and chronic obstructive pulmonary disease,” *Clin. Exp. Allergy*, vol. 41, no. 10, pp. 1371–1378, 2011, doi: 10.1111/j.1365-2222.2011.03800.x.
16. Z. Haddi *et al.*, “E-Nose and e-Tongue combination for improved recognition of fruit juice samples,” *Food Chem.*, vol. 150, pp. 246–253, 2014, doi: 10.1016/j.foodchem.2013.10.105.
17. M. Ghasemi-Varnamkhasti, A. Mohammad-Razdari, S. H. Yoosefian, Z. Izadi, and M. Siadat, “Aging discrimination of French cheese types based on the optimization of an electronic nose using multivariate computational approaches combined with response surface method (RSM),” *Lwt*, vol. 111, pp. 85–98, 2019, doi: 10.1016/j.lwt.2019.04.099.
18. T. Saidi, O. Zaim, M. Moufid, N. El Bari, R. Ionescu, and B. Bouchikhi, “Exhaled breath analysis using electronic nose and gas chromatography–mass spectrometry for non-invasive diagnosis of chronic kidney disease, diabetes mellitus and healthy subjects,” *Sensors Actuators, B Chem.*, vol. 257, pp. 178–188, 2018, doi: 10.1016/j.snb.2017.10.178.
19. E. C. Nallon, V. P. Schnee, C. Bright, M. P. Polcha, and Q. Li, “Chemical Discrimination with an Unmodified Graphene Chemical Sensor,” *ACS Sensors*, vol. 1, no. 1, pp. 26–31, 2016, doi: 10.1021/acssensors.5b00029.
20. D. Poli *et al.*, “Exhaled volatile organic compounds in patients with non-small cell lung cancer: Cross sectional and nested short-term follow-up study,” *Respir. Res.*, vol. 6, no. 1, p. 71, Dec. 2005, doi: 10.1186/1465-9921-6-71.
21. A. K. Srivastava, “Detection of volatile organic compounds (VOCs) using SnO2 gas-sensor array and artificial neural network,” *Sensors Actuators, B Chem.*, vol. 96, no. 1–2, pp. 24–37, 2003, doi: 10.1016/S0925-4005(03)00477-5.
22. J. Fu, G. Li, Y. Qin, and W. J. Freeman, “A pattern recognition method for electronic noses based on an olfactory neural network,” *Sensors Actuators, B Chem.*, vol. 125, no. 2, pp. 489–497, 2007, doi: 10.1016/j.snb.2007.02.058.
23. R. Dutta, D. Morgan, N. Baker, J. W. Gardner, and E. L. Hines, “Identification of Staphylococcus aureus infections in hospital environment: Electronic nose based approach,” *Sensors Actuators, B Chem.*, vol. 109, no. 2, pp. 355–362, 2005, doi: 10.1016/j.snb.2005.01.013.
24. P. Montuschi *et al.*, “Diagnostic performance of an electronic nose, fractional exhaled nitric oxide, and lung function testing in asthma,” *Chest*, vol. 137, no. 4, pp. 790–796, 2010, doi: 10.1378/chest.09-1836.
25. R. Bhardwaj, V. Selamneni, U. N. Thakur, P. Sahatiya, and A. Hazra, “Detection and discrimination of volatile organic compounds by noble metal nanoparticle functionalized MoS2coated biodegradable paper sensors,” *New J. Chem.*, vol. 44, no. 38, pp. 16613–16625, 2020, doi: 10.1039/d0nj03491f.
26. P. Bindra and A. Hazra, “Electroless deposition of Pd/Pt nanoparticles on electrochemically grown TiO 2 nanotubes for ppb level sensing of ethanol at room temperature,” *Analyst*, vol. 146, no. 6, pp. 1880–1891, 2021, doi: 10.1039/D0AN01757D.