International Journal of Sciences: Basic and Applied Research (IJSBAR)

ISSN 2307-4531
(Print & Online)

http://gssrr.org/index.php?journal=JournalOfBasicAndApplied

Classification and Identification of Volatile Organic Solvents based on Functional Groups using Electronic Nose

Tharaga Sharmilan[a], Duleesha Manohari[b], Indika Wanniarachchi[c], Sandya Kumari[d], Dakshika Wanniarachchi[e*]

[a]Department of Materials and Mechanical Technology, Faculty of Technology, University of Sri Jayewardenepura, Sri Lanka
[a,e]Instrument Centre, Faculty of Applied Sciences, University of Sri Jayewardenepura, Sri Lanka
[b]Department of Mathematics, Faculty of Applied Sciences, University of Sri Jayewardenepura, Sri Lanka
[c]Department of Physics, Faculty of Applied Sciences, University of Sri Jayewardenepura, Sri Lanka
[d]Department of Science and Technology, Faculty of Technology, University of Sri Jayewardenepura, Sri Lanka
[a]Email: tharagas@sjp.ac.lk, [b]Email: dakshikacw@sjp.ac.lk

Abstract

The Metal Oxide Semiconductor gas sensors based on SnO2 indicate cross sensitivity to many volatile organic compounds. Therefore, this study is focused on developing a methodology to distinguish organic solvents based on the functional groups present using an array of Metal Oxide Semiconductor gas sensors. Here, representative compounds for aliphatic, aromatic hydrocarbons, carbonyl groups, esters, alcohols and dichloromethane were used to evaluate gas sensors. Then data were analyzed using Principal Component Analysis and k-Nearest Neighbor methods. Finally, k-Nearest Neighbor best model was developed to predict the chemicals based on the sensor data. The overall results of this study sufficiently explain that the developed electronic nose system can distinguish the chemicals tested with Principal Component Analysis (96.6 percentage) and can predict with k-Nearest Neighbor (k=5) (90 percentage) the chemicals based on the sensor responses. These results demonstrate that the developed electronic nose can be used to classify and identify chemicals based in different functional groups.

----

* Corresponding author.

Keywords: e-nose; MOS gas sensors; carbonyl compounds; esters; chlorinated compounds; alcohols; PCA; kNN.

1. Introduction

Electronic Nose (e-Nose) is an electronic aroma detection system which is also known as an artificial olfaction system. It operates similar to how we smell. Different aroma enters in to the nasal cavity is bonded to one or more olfactory receptors in the nose which then sends signal to the brain. The data processed in the brain could distinguish different aroma based on the binding pattern of different olfactory receptors for a particular smell [1, 2]. Most of the recently developed e-nose systems use an array of sensors which represents the olfactory receptors in the human nose and an appropriate data processing method to identify the smell similar to what brain does [3, 4]. The e-nose systems are used in different applications [4, 5, 6, 7]. These are spread over many disciplines as well as industries such as food quality control [8, 9, 10, 11], agricultural field [6], bio medical applications [12, 13], medical diagnosis [15, 16, 17], security purposes [18] and in environment quality control fields [19, 20, 21]. The most crucial component in an e-nose is the sensor array which detects the incoming smell. Today there is a vast range of electronic gas sensors available such as Metal Oxide Semiconductor (MOS) gas sensors, catalytic field effect sensors (MOSFET), conducting polymer sensors, electrochemical sensors, quartz crystal microbalance and colorimetric sensors. The material properties and operational principle of each of these different sensor types have led to preference of a particular type of sensor for a given application. However, MOS gas sensors are widely employed in experimental and many commercially developed e-nose systems [13]. This could be due to the fact that availability in the market for a very low price and have many other advantages over other sensor types such as less maintenance, light weight and high durability [22]. The basic components of a MOS gas sensor includes a gas sensing metal oxide layer, gold sensing electrode, alumina substrate and a heater. There are two basic categories of metal oxides as n-type and p-type which differ whether the charge carrier type is electrons or holes respectively [23, 24]. Most of the commercially available gas sensors are n-type based on SnO2 [24]. The conductivity of the semiconducting metal oxide layer depends on the carrier concentration of the material (either n-type of p-type). The metal oxide layer has surface adsorbed O2 molecules which undergo successive reductions to form surface adsorbed O2- ions by taking electrons from the conduction band [24]. The presence of reducing gases such as CH4 react with O2- ions thus desorb from the SnO2 surface resulting more and more electrons taken from the conduction band. Consequently resistance of the material changes due to the change in charge carrier concentration of the material [24]. The difference in resistance compared to the base line (i.e. without the analyte) will be considered when detecting the amount of the gas present [22]. The performance of a metal oxide gas sensor material depends mainly on surface area, crystallinity grain size, presence of a catalyst and the operating temperature [23, 24, 25]. The selectivity of the SnO2 based gas sensors is achieved mainly from the temperature at which it operates [24]. The typical SnO2 based gas sensors operate at temperatures range 200-400 ^{0}C. The oxygen species found adsorbed to the surface of the metal oxide layer varies with the operating temperature [26, 27]. Thus by controlling the temperature, surface reactions can be controlled such that the material would be selective towards a particular type of gas [27]. In the case of alcohol detection, the presence of a catalytic layer of La2O3 enhances the performance of the SnO2 based gas sensor [26]. In general, SnO2 based gas sensors are not very specific. It is mentioned in the datasheets of the commercial gas sensors the sensor is capable of detecting multiple gases with varied sensitivity [28]. This may be an issue in the case a mixture of gases to be detected. Furthermore, the SnO2 based senor materials are disturbed when there is high moisture content is available [22]. The cross-sensitivity is an inherent problem related to the SnO2 based gas sensors. Therefore, these gas sensors cannot serve as a unique method to identify a particular gas. This may lead to ambiguity in identifying unknown gas mixtures. There have been many attempts taken to improve the selectivity factor either by decorating the base sensor material with a suitable catalytic oxide layer or in a non-invasive manner using multiple gas sensor signatures (using e-nose) using a suitable classification technique [27]. When developing an e-nose system it is important to know the response from various types of volatile compounds to the commercially available SnO2 based gas sensors. The commonly used statistical methods for e-nose data analysis are principal component analysis (PCA), variance analysis (ANOVA) and clustering methods [34]. In addition, artificial intelligence methods also carried out for the data analysis [34]. Even though a particular gas senor is marked such as `alcohol sensor', this could be responsive to many other volatile organic compounds as well. There has been many studies conducted for use of e-nose for volatile organic compound classifications [29] using household items. However, there is a need to evaluate whether an array of sensors would distinguish the chemical substances based on the functional groups present in the organic solvents. In this study a series of organic solvents were selected to represent alkanes (aliphatic and aromatic), aldehydes, ketones, esters, alcohols, chlorinated solvents and water as given in Table 1.

Table 1: Chemicals used in this study and their respective groups

Alcohols	Carbonyl compounds	Hydrocarbons	Other solvents
Ethanol	Ethyl acetate	Toluene	Water
Methanol	Acetone	Hexane	Dichloromethane
iso-propyl alcohol	Iso-butylmethylketone		

The objective of this study is to investigate the effect of functional group on the sensitivity of sensor array and classify the chemical compounds using an e-Nose system. The results obtained were analyzed with two different classification techniques, principal component analysis and k-nearest neighbor analysis to predict the chemical of interest.

2. Materials and Methods

2.1. Materials

The following solvents were used as it is for e-nose sensor evaluation. Ethanol (Sigma aldrich, 96%), iso-propylalcohol (SRL,99.8%), methanol (SRL, 99.9%), ethyl acetate (SRL, 99.5%), acetone (Daejung, 99.5%), iso-butyl methyl ketone (SRL,99.5%), Toluene (SRL,99.9%), Hexane (SRL,95%) Dichloromethane (Sigma Aldrich 99.8%) and distilled water.

2.2. E-nose system

A custom built e-nose system is used in this study. An array of MOS gas Sensors have been used for the assessment of volatile compounds. A schematic diagram of the e-nose system developed is shown in Figure 1.

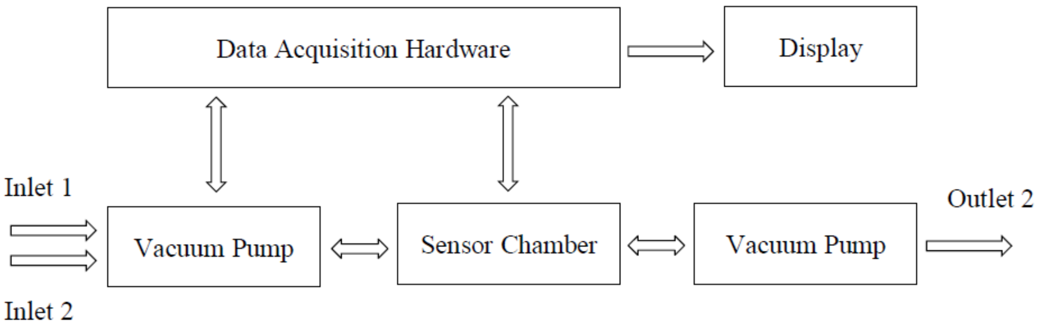

Figure 1: Schematic diagram of electronic nose system

Developed e-nose system contains three main parts, which are gas sensor chamber, data acquisition hardware and vacuum pumps. An air tight water proof case was used to build sensor chamber using a sensor array. In this study four gas sensors were used to develop sensor array. Transparent pipes were used to connect the sensor chamber and vacuum pumps in order to supply air to the sensor array. Arduino and related hardware were used to design the data acquisition system to record the sensor data when testing samples. Two air inlets were used to insert the reference air and aroma of the samples. Three vacuum pumps were used to draw the reference air and sample air in to the sensor chamber to be analyzed. Environment air was used as reference air in this study and to clean the sensor chamber between two sample measurements. Glass sample vials with an internal volume of 40 ml were used for the experiments.

Table 2: Sensors used in sensor chamber [28]

Sensor	Sensitive substances	Detection range
MQ2	LPG, iso-butane propane methane ,alcohol Hydrogen, smoke	200ppm-5000ppm (LPG and propane) 300ppm-5000ppm (butane) 5000ppm-20000ppm (methane) 300ppm-5000ppm (H2) 100ppm-2000ppm (Alcohol)
MQ3	Alcohol, benzene	0.05mg/L|10mg/L (Alcohol)
MQ4	Methane	300-10000 ppm (CH4)
MQ5	H2, LPG, CH4, CO, Alcohol	200-10000 ppm

The glass bottles were connected to the instrument by transparent pipes. The sensor chamber consists of four SnO2 based gas sensors (MQ-2, MQ-3, MQ-4, and MQ-5). These MQ series gas sensors contain a built in heater and an electro-chemical gas sensor. The sensitivity of the selected gas sensors are given in the Table 2.

2.3. Operation of e-nose

The e-nose system developed in this study operates as sniffing cycles. One sniffing cycle contains events according the following sequence. First sensor chamber cleaning then sniffing process and odor-lock followed by sensor chamber cleaning. The next cycle begins soon after the cleaning cycle and the instrument continue to collect the data until the process is terminated by the user. The data collection is conducted for 10 consecutive sniffing cycles for each chemical. Experimental conditions have been maintained for the classification of chemicals is given below: Air flow rate = 12 ml/s, Amount of each chemicals = 9 ml, Temperature = 30 ^{0}C. One sniffing cycle time was limited to 3 minutes: 1 minute for cleaning, 1 minute for sniffing and odor-lock process and 1 minute for cleaning process and for each sample, data collected for three successive sniffing cycles. This time duration was programmed through the software. The sample air inlet was closed while sensor chamber was cleaning. Environment air was used for the cleaning of the sensor chamber. A continuous gas flow was maintained in the sensor chamber except during odor-lock process. Data obtained from the sensor array were stored in the micro Secure Digital (SD) card to conduct the data analysis.

2.4. Comparison of sensor raw values

MQ series gas sensors are electro-chemical gas sensors which respond to an incoming gas as change in resistance. The response to a particular gas is recorded after analog to digital conversion as the sensor raw value. Since the gases that are analyzed in this paper are mostly not included in the calibration plot of $\frac{\boldsymbol{Rs}}{\boldsymbol{R0}}$ (Ro-Resistance of Sensor in the environmental conditions, i.e. without gas to be analyzed and Rs-Resistance of Sensor) vs. concentration (ppm) by the manufacturer, we will be using sensor raw value instead of ppm values. The sensor raw values for the environmental air is recorded and given in supplementary materials (Table S1: Sensor raw values for the environmental air). The sensitivity of gas sensors for each chemical was compared considering a sniffing cycle to obtain the relative response.

2.5. Data Analysis

Table 3: Chemical labels used in data analysis

Chemicals	Label
iso-Butyl Methyl Ketone	0
Ethyl Acetate	1
Dichloro Methane	2
Ethanol	3
Acetone	4
iso-Propyl Alcohol	5
Water	6
Hexane	7
Toluene	8
Methanol	9

Data analysis have been done to distinguish the chemical tested by sensors and to predict the chemicals based on the sensor responses given for unknown. Machine learning algorithms are the best solution to accomplish these objectives. But, Machine learning algorithm will be too slow due to high dimensionality of input data. Therefore, after the data preprocessing step, Principal Component Analysis (PCA) was initially used to reduce the dimensionality and visualize the data. All 10 chemicals were labeled from zero to nine to perform the data analysis as given in Table 3.

2.5.1. Principal Component Analysis

It is one of the feature extraction techniques and mostly used for the dimensionality reduction process. Most of the data scientists use Jupiter notebook as it is online source in their data analysis. Therefore, Jupiter note book was used to analyze the data set in this study. This data set should be scalable when performing PCA. Therefore, data were standardized on to unit scale for the optimal performance of the Machine learning algorithm. 70% of data set has been used to make the model. 30% were used to test the model. In the first step pre-processed data were high (99 row data points). It has four dimension projections. It is difficult to visualize the data. Explained variance ratio was found to provide amount of variance each principal component have after doing dimensionality reduction. Then visualization of data have been plotted to identify the distinguished chemicals by sensors.

2.5.2. K-Nearest Neighbor

k-Nearest Neighbor (k-NN) algorithm was used for the prediction purposes as machine learning algorithm at the first step of this study. k is the number of closet point to training any data point. It is a simple algorithm to solve both classification and regression analysis. Python language and scikit library is used to implement kNN algorithm. Prediction of chemicals was done in two ways: with PCA and without doing PCA. Then results were compared. In the kNN process, first scalar transformation has been done. At first kNN model is produced using default neighbor value of classifier. Then, training and test data set have been loaded and the value of k was chosen. Range of values of k was taken from 1 to 20 and accuracy of kNN model based on different values of k was plotted. Less difference between testing accuracy and training accuracy was selected to find the best number of neighbors in each case (kNN with PCA, kNN without PCA). Then model was fitted according to the number of neighbors and accuracy of the model on test data and train data was checked. Then prediction was done using the 30% of the test data sets and compared the results in each case with actual test target values. At this stage, chemical can be predicted based on the sensor responses given.

Then, confusion matrix was used get a better idea about performance of classification model. It is a table (table 4) with different combinations of actual and predicted class.

Table 4: Confusion Matrix

	Predicted Class (P)	Predicted Class (N)
Actual Class (P)	True Positive (TP)	False Negative (FN)
Actual Class (N)	False Positive (FP)	True Negative (TN)

Finally, classification report was analyzed to measure the quality of predictions of trained model by the kNN algorithm. The numbers of true and false predictions are used predict the measures of the classification report. The measured performance was interpreted in terms of Precision, Recall and F-measure (Eq. 1, 2, 3).

$$\text{Precision} = TP/\ (TP+FP) \qquad (1)$$

$$\text{Recall}=TP/\ (TP+FN) \qquad (2)$$

$$\text{F-Measure}=\ (2x\text{Recall}x\text{Precision})\ /\ (\text{Recall} + \text{Precision}) \qquad (3)$$

$$\text{Classification Rate/ Accuracy}=\ (TP+TN)/\ (TP+TN+FP+FN) \qquad (4)$$

Definitions of the terms given in Table 4 and Equations 1,2,3 and 4 are given below:

Positive (P) : Observation is positive

Negative (N) : Observation is not positive

True Positive (TP) : Observation is positive, and is predicted to be positive

False Positive (FP) : Observation is negative, but is predicted positive

False Negative (FN) : Observation is positive, but is predicted negative

True Negative (TN) : Observation is negative, and is predicted to be negative

Precision: It is the ability of a classifier not to label an instance positive that is actually negative

Recall: It is the ability of a classifier to find all positive instances

F-Measure: It is a weighted harmonic mean of precision and recall

Classification Rate/Accuracy: the number of correctly classified patterns to the total number of patterns

3. Results and Discussion

3.1. Comparison of Sensor responses for different chemical classes

A typical sniffing cycle consists of one minute cleaning of the sensor chamber, one minute odor lock and one minute cleaning to bring the sensor raw values back to the starting point. The following Figure 2 illustrates the response of each gas sensor towards different chemicals during one sniffing cycle. According to these figures, an elevated sensor raw value is obtained for each sensor MQx during odor lock period, which ultimately decreased during cleaning process.

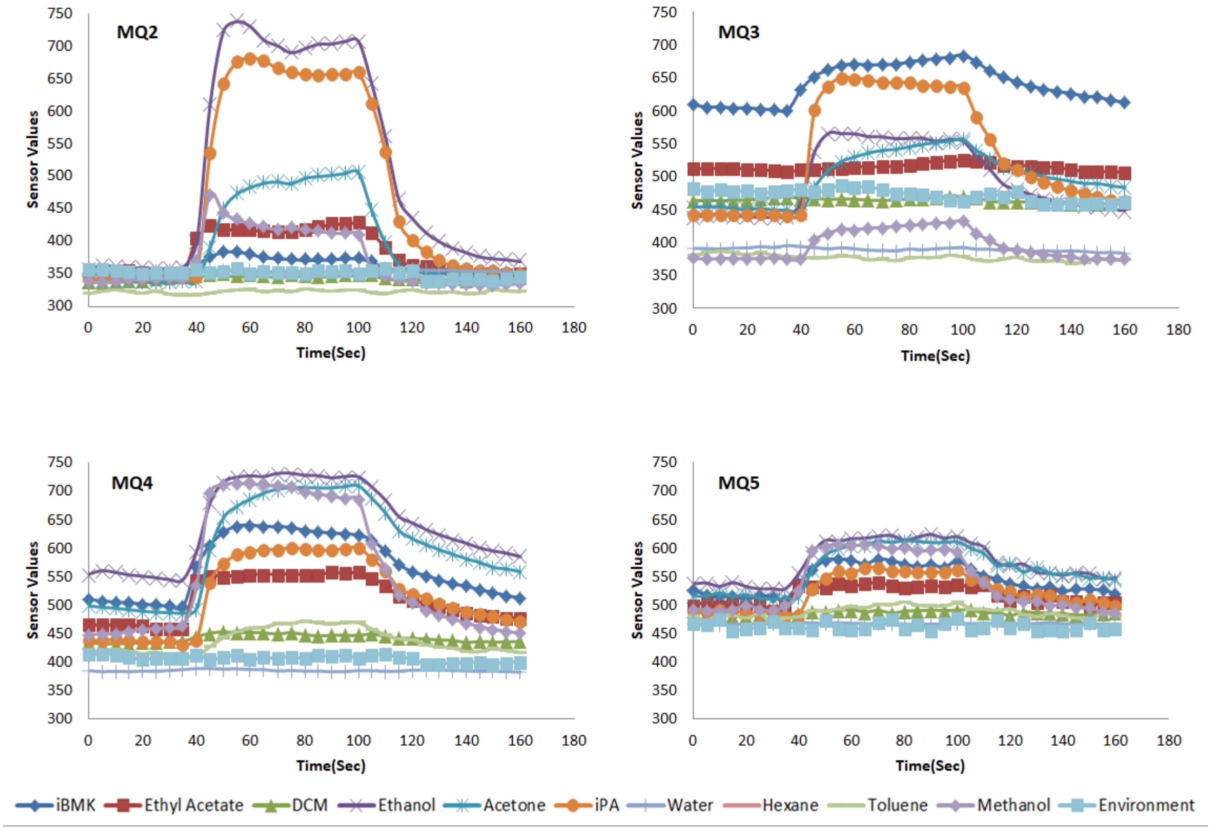

Figure 2: Response of MQ 2-5 towards different chemicals

The MQ series gas sensors are based on SnO2 materials. When exposed to the gas that is to be analyzed, some of the surface adsorbed reactive oxygen species react with the gas. As a result resistance of the material is changed which is read as change in analog voltage [30]. The sensor responses are therefore increased when there is a gas present. MQ2 sensor is generally used to detect methane, butane and LPG. According to the results highest response is observed for ethanol and iso-propyl alcohol (iPA). MQ3 is most responsive for iso-butylmethylketone (iBMK), iPA, ethyl acetate and acetone in addition to ethanol. MQ4 and MQ5 are most sensitive for ethanol and acetone. In general alcohols and carbonyl compounds have higher sensitivity towards these sensors compared to hydrocarbons (hexane and toluene) and dicholormethane. MQ3 is an alcohol detector which causes dehydrogenation and dehydration [31, 32] of alcohol and acetone to achieve higher selectivity of the sensor over other gases. However, proper categorization of the chemicals with different functional groups requires classification techniques such as PCA and kNN which are discussed in detail below.

3.2. Classification using PCA and kNN

In this PCA model, first principal component contains 79.1 % of the variance and the second principal component contains 17.5% of the variance. Together two components contain 96.6% of the information. It means scikit-learn chose the minimum number of PCs such that 96.6% of the variance is retained. Therefore raw matrix is reduced to 2 principal components. Ten samples of each class projected in PCA model. Dimensionality reduced PCA model results are shown in Figure 3.

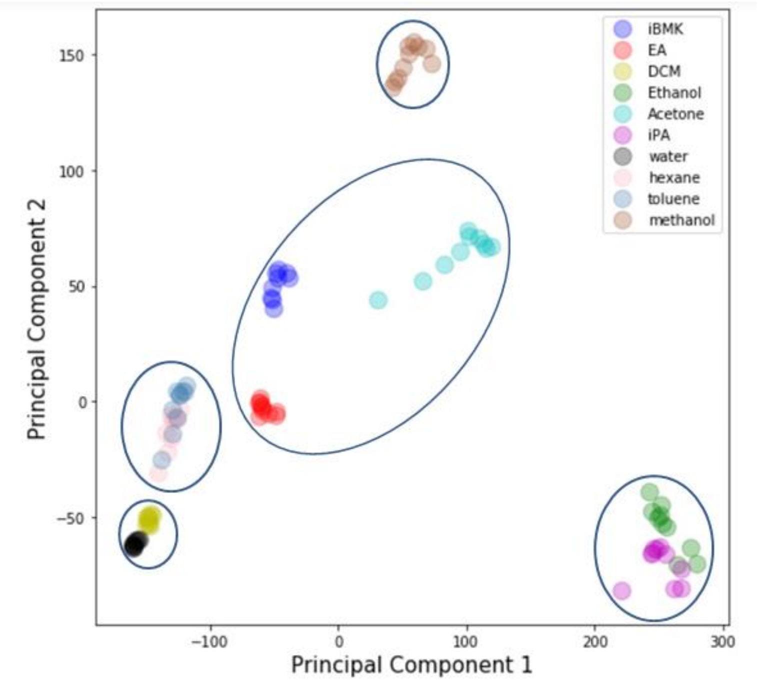

Figure 3: Two dimensional representation of the classification problem

It can be seen that all the chemical groups (hydrocarbons, other polar groups, carbonyl compounds and alcohols) are distinguished correctly. Carbonyl compounds are shown in middle of the image. Since iso-butylmethylketone and acetone belonging to ketone category, these are slightly separated from ester group (Ethyl Acetate). Ethanol and iso-propyl alcohol are classified correctly from carbonyl compounds. In the alcohol group, methanol has only deviated from ethanol and iso-propyl alcohol. Hexane and toluene are hydrocarbon molecules which do not have any polarity. These are clustered together. Furthermore, dichloromethane and water are classified together as these have given low sensor scores with all four gas sensor types. It is because of the variation in the sensor responses in the presence of different organic compounds, it can be stated that e-Nose system can be able to discriminate the different substances. In order to evaluate the accuracy of identifying each of these chemicals using the sensor array, kNN classification was conducted. Then kNN classification process was carried out after PCA process and without doing PCA process to find the best kNN model for this study. First, kNN classification was done with the PCA process. In this process, a range of values of k were taken from 1 to 20 and accuracy of kNN model based on different values of k was plotted. The Plot is given in Figure 4. The accuracies related to each of the k values is given in table S2 (Table S2: Accuracies of kNN (k1 to k20) with doing PCA).

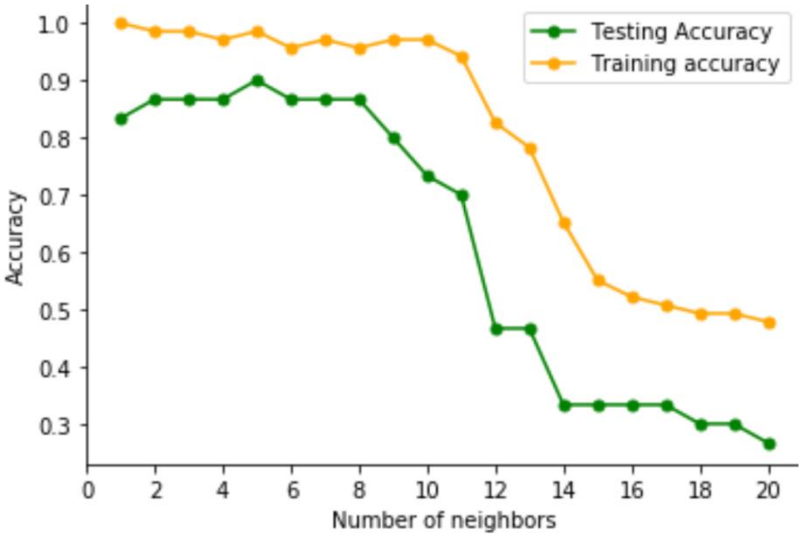

Figure 4: Plot of kNN varying number of neighbors with PCA

In Figure 4, less error between testing accuracy and training accuracy was selected as k=5 when performing kNN with PCA Process. When the model was run with k=5, training accuracy and test accuracy received respectively as 98.5% and 90%. But when default model (k value of default model is also 5) was run, training and test accuracy were 98.5% and 90%. When model was fitted at k=5, under fitting percentage is same as the default model. However, test accuracy was slightly lesser than training accuracy in both models (k=5). Therefore, this model is under fitted. As a comparison, the kNN classification was conducted without the dimensionality reduction (i.e. done without doing PCA process). Similar to the previous process, a range of values of k were taken from 1 to 20 and accuracy of kNN model based on different values of k was plotted to find the best k value for the model. The plot was given in Figure 5 and accuracies related to each of the k values is given in supporting information (Table S3:Accuracies of kNN (k1 to k20) without doing PCA).

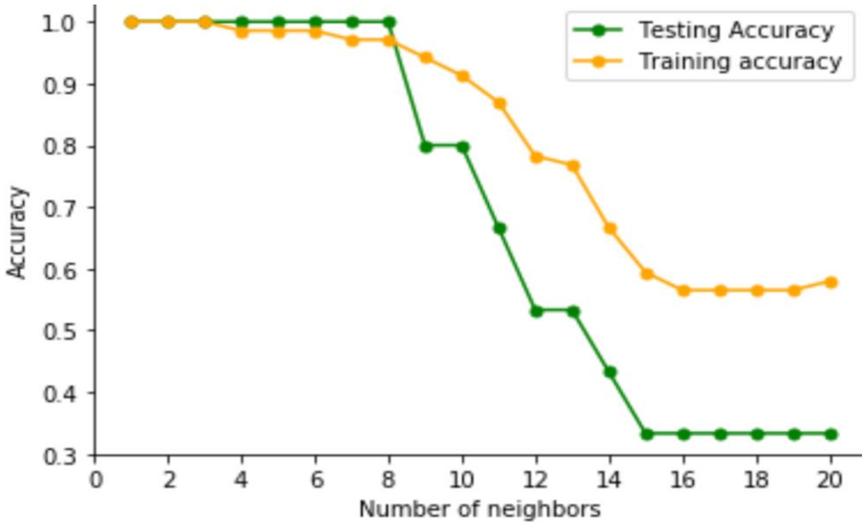

Figure 5: Plot of kNN varying number of neighbors without PCA

In Figure 5, when default model (k=5) was run, training and test accuracy were 98.5% and 100%. Test accuracy was higher than training accuracy in the default model. When the model is not able to get a sufficiently low error value on the training data, it is under fitted. Therefore this default model is under fitted. Under fitted model is not a good model [33] and it will have poor performance on the training data set. In another instance there can be training accuracy is higher than the testing accuracy yet the gap between two curves are high. Under fitting and over fitting can lead to poor model performance. Therefore good fit can be selected at the point between under fitting and over fitting. Making training error small and gap between training and test error small are two important conditions to select the good fit. Therefore k=10 is selected as good fit point because training accuracy and testing accuracy are 91.3% and 80% respectively with error 11.3%. Comparison of Training and test accuracy for the best fit models with PCA and without PCA are given in below Table 5.

Table 5: Comparison of train and test accuracy

	K=10 without PCA	K=5 with PCA
Train Accuracy	91.3%	98.5%
Test Accuracy	80%	90%

According to the Table 5, it can be stated that k=5 is best value for the kNN model with doing PCA process (Train=98.5%, Test=90%). However, if kNN classification was carried out without dimensionality reduction, training accuracy and test accuracy of good fit point are 91.3% and 80%. Confusion matrix was further analyzed to find the best model and to explain the performance of that best model. Confusion matrix of k=5 with PCA process is given below Table 6.

Table 6: Confusion matrix of k=5 kNN model with doing PCA process

	0	1	2	3	4	5	6	7	8	9	All
0	5	0	0	0	0	0	0	0	0	0	5
1	0	2	0	0	0	0	0	0	0	0	2
2	0	0	0	0	0	0	3	0	0	0	3
3	0	0	0	2	0	0	0	0	0	0	2
4	0	0	0	0	3	0	0	0	0	0	3
5	0	0	0	0	0	3	0	0	0	0	3
6	0	0	0	0	0	0	1	0	0	0	1
7	0	0	0	0	0	0	0	5	0	0	5
8	0	0	0	0	0	0	0	0	2	0	2
9	0	0	0	0	0	0	0	0	0	4	4
All	5	2	0	2	3	3	4	5	2	4	30

Table 7: Confusion matrix of k=10 kNN model without doing PCA process

	0	1	2	3	4	5	6	7	8	9	All
0	0	5	0	0	0	0	0	0	0	0	5
1	0	2	0	0	0	0	0	0	0	0	2
2	0	0	3	0	0	0	0	0	0	0	3
3	0	0	0	2	0	0	0	0	0	0	2
4	0	0	0	0	3	0	0	0	0	0	3
5	0	0	0	0	0	3	0	0	0	0	3
6	0	0	0	0	0	0	1	0	0	0	1
7	0	0	0	0	0	0	0	4	1	0	5
8	0	0	0	0	0	0	0	0	2	0	2
9	0	0	0	0	0	0	0	0	0	4	4
All	0	7	3	2	3	3	1	4	3	4	30

Diagonal elements of confusion matrix show the number of correct classifications for each class and off diagonal elements provides misclassifications. According to the Table 6; iBMK, EA, ethanol, acetone, iPA, water, hexane, toluene and methanol are classified correctly and DCM is misclassified as water but none of water is misclassified as DCM. According to the equation 4, classification accuracy is obtained as 90%. Confusion matrix of k=10 kNN model without PCA process is given below Table 7.

Table 8: Classification report of kNN model (k=5) with doing PCA process

	Precision	Sensitivity	F1-score	Support
0	1.00	1.00	1.00	5
1	1.00	1.00	1.00	2
2	0.00	0.00	0.00	3
3	1.00	1.00	1.00	2
4	1.00	1.00	1.00	3
5	1.00	1.00	1.00	3
6	0.25	1.00	0.40	1
7	1.00	1.00	1.00	5
8	1.00	1.00	1.00	2
9	1.00	1.00	1.00	4
Micro avg	0.90	0.90	0.90	30
Macro avg	0.82	0.90	0.84	30
Weighted avg	0.88	0.90	0.88	30

According to the table 7; EA, DCM, ethanol, acetone, iPA, water, toluene and methanol are classified correctly and iBMK is misclassified as EA but none of EA is misclassified as iBMK and 80% of Hexane is correctly predicted and 20% is misclassified as toluene. None of toluene is misclassified as hexane. According to the equation 4, classification accuracy is obtained as 80%. When comparing Table 6 and Table 7, default kNN (k=5) model with doing PCA process is showing good model (table 7). Classification report is further analyzed to find the best model. The classification reports of them are given in below tables 8 and 9. Model with k=5 kNN algorithm with PCA process is better at recall, precision and F1-score (Table 8 and 9). Overall, the model using k=5 Nearest Neighbor Algorithm performed with principal component analysis is best model in this study.

Table 9: Classification report of kNN model (k=10) without PCA process

	Precision	Sensitivity	F1-score	Support
0	1.00	1.00	1.00	5
1	1.00	1.00	1.00	2
2	0.00	0.00	0.00	3
3	1.00	1.00	1.00	2
4	1.00	1.00	1.00	3
5	1.00	1.00	1.00	3
6	0.25	1.00	0.40	1
7	1.00	1.00	1.00	5
8	1.00	1.00	1.00	2
9	1.00	1.00	1.00	4
Micro avg	0.80	0.80	0.80	30
Macro avg	0.80	0.88	0.81	30
Weighted avg	0.76	0.80	0.76	30

4. Conclusions

The commercially available gas sensors show poor selectivity. However, selectivity can be improved if detection can be combined with responses from several gas sensors. Therefore, in this study the combination of several gas sensors were used to classify chemical compounds with different functional groups using a custom built e-nose system. The comparison of sensor raw value indicates that the sensors MQ 2-5 could be sensitive to many volatile organic compounds than given in the data sheet. The possible classification of sensor responses to different chemicals was carried out with PCA and kNN methods. In PCA, two dimensionality reductions were done to visualize the data. It is because of 96.6% of the variation in the sensor responses in the presence of different organic compounds, it can be stated that e-Nose system can be able to distinguish the different substances. First k-Nearest Neighbor model was classified after dimensionality reduction process. The best kNN (k=5) model is obtained with 90% classification accuracy. But 80% classification accuracy is obtained when performing kNN model (k=10) without dimensionality reduction. All 10 chemicals have nearly high precision value in both cases (88%, 76%). Therefore, performing prediction of chemicals based on the sensor responses is best for kNN after conducting dimensionality reduction. The best k value for the model is 5 (k=5). The best model is able to accurately predict all chemicals except DCM. It can be concluded that sensors MQ2-5 in the electronic nose system can classify different chemicals with different functional groups.

5. Author Contributions

T.S. and D.W. conceived and designed the experiments; I.W. helped to develop the data acquisition and control systems; T.S. performed the experiments; T.S. and D.M. analyzed the data; T.S. and D.M. were responsible for writing original draft preparation; D.W., S.K. and I.W. were responsible for supervision, writing review and editing

6. Funding

This research was funded by National Research Council of Sri Lanka, grant number NRC 17-038 (http://www.nrc.gov.lk/).

Acknowledgements

Authors would like to thank Instrument Center and Department of Physics staffs for their supporting to conduct the experiments

7. Abbreviations

The following abbreviations are used in this manuscript:

PCA	Principal Component algorithm
KNN	K-Nearest Neighbor
MOS	Metal Oxide Sensors
iPA	iso Propyl Alcohol
DCM	Dichloromethane
iBMK	iso Butyl Methyl Ketone

References

[1]. Hurot, Charlotte, Natale, S., Arnaud, B., and Hou,Y. Bio-Inspired Strategies for Improving the Selectivity and Sensitivity of Arti_cial Noses: A Review. In Sensors., 2020, 20(6),1803.

[2]. Hu, Wenwen, Wan, L., Jian,Y., Ren,C., Jin,K., Su,X.,Bai,X., Haick,H., Yao,M., and Wu,W. Electronic noses: from advanced materials to sensors aided with data processing. In Advanced Materials

Technologies ,2019, 4(2),1800488.

[3]. Xu, Sai, Sun,X., Lu,H., and Zhang,Q. Detection of type, blended ratio, and mixed ratio of pu'er tea by using electronic nose and visible/near infrared spectrometer. In Sensors, 2019,19(10), 2359.

[4]. Wu, Zhiyuan,Wang,H.,Wang,X., Zheng,H., Chen,Z., and Meng,C. Development of Electronic Nose for Qualitative and Quantitative Monitoring of Volatile Flammable Liquids. In Sensors, 2020, 20(7), 1817.

[5]. Mahmoudi, Esmaeil. Electronic nose technology and its applications. In Sensors and Transducers,2009, 107(8), 17.

[6]. Wilson, Alphus D. Diverse applications of electronic-nose technologies in agriculture and forestry. In Sensors,2013,13(2),2295-2348.

[7]. Win,Tin,D. The electronic nose{a big part of our future. In AU JT,2005, 9(1), 1-8.

[8]. Dutta, Ritaban, K. R. Kashwan, M. Bhuyan, Evor L. Hines, and J. W. Gardner. Electronic nose based tea quality standardization. In Neural Networks, 2003,16(5-6),847-853. https://doi.org/10.1016/S0893-6080(03)00092-3

[9]. Zhou, B., Wang, J., and Qi, J. Identification of different wheat seeds by electronic nose. In International Agrophysics, 2012, 26(4), 413{418. https://doi.org/10.2478/v10247-012-0058-y

[10]. Ghasemi-Varnamkhasti, M., Mohtasebi, S. S., Siadat, M., and Balasubramanian, S. Meat quality assessment by electronic nose (Machine Olfaction Technology). In Sensors., 2003, 9(8), 6058{6083. https://doi.org/10.3390/s90806058

[11]. Xu, M., Wang, J., and Gu, S. Rapid identi_cation of tea quality by E-nose and computer vision combining with a synergetic data fusion strategy. In Journal of Food Engineering, 2019, 241, 10-17. https://doi.org/10.1016/j.jfoodeng.2018.07.020

[12]. Wilson, Alphus, D., and Baietto, M. Advances in electronic-nose technologies developed for bio medical applications. In Sensors, 2011, 11(1), 1105{1176. https://doi.org/10.3390/s110101105

[13]. Wilson, Alphus D., and Baietto, M. Advances in electronic-nose technologies developed for biomedical applications. In Sensors,2009, 9(7), 5099-5148.

[14]. Pavlou, A., Turner,A.P.F., and Magan,N. Recognition of anaerobic bacterial isolates in vitro using electronic nose technology. In Letters in Applied Microbiology,2002, 35(5), 366-369.

[15]. McWilliams, Annette,Beigi,P.,Srinidhi,A., Lam,S., and Calum E. MacAulay. Sex and smoking status effects on the early detection of early lung cancer in high-risk smokers using an electronic nose. In IEEE Transactions on Biomedical Engineering, 2015, 62(8),2044-2054.

[16]. Guntner, Andreas T., Koren ,V., Chikkadi, K., Righettoni,M., and Sotiris E. Pratsinis.E-nose sensing of low-ppb formaldehyde in gas mixtures at high relative humidity for breath screening of lung cancer.In ACS Sensors, 2016, 1(5),528-535.

[17]. Hockstein, Neil G., Erica R. Thaler, Torigian,D.Wallace T. M. Jr, De_enderfer,O. and C. William H.Diagnosis of pneumonia with an electronic nose: correlation of vapor signature with chest computed tomography scan _ndings. In The Laryngoscope ,2004,114 (10), 1701-1705.

[18]. Pobkrut, Theerapat,Eamsa-Ard,T., and Kerdcharoen,T. Sensor drone for aerial odor mapping for agriculture and security services. In 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)IEEE,2016,1-5

[19]. Wilson, Alphus D. Review of electronic-nose technologies and algorithms to detect hazardous chemicals in the environment.In Procedia Technology, 2012,453-463.

[20]. De Vito, S., E. Massera, G. Burrasca, A. Di Girolamo, M. Miglietta, G. Di Francia, and D. Della Sala. TinyNose: Developing a wireless e-nose platform for distributed air quality monitoring applications. In SENSORS, 2008, IEEE,701-704.

[21]. Ta_stan, Mehmet, and G□okozan, H. Real-time monitoring of indoor air quality with internet of things-based E-nose. In Applied Sciences,2019,9(16),3435.

[22]. Fine, George, F., Leon, M. Cavanagh, Afonja, A.,and Binions,R. Metal oxide semi-conductor gas sensors in environmental monitoring. In sensors,2010, 10(6) ,5469-5502.

[23]. Korotcenkov, Ghenadii. Handbook of gas sensor materials. In Conventional Approaches, 2013,1, Springer-Verlag.

[24]. Wetchakun, K., T. Samerjai, N. Tamaekong, C. Liewhiran, C. Siriwong, V. Kruefu, A. Wisitsoraat, A. Tuantranont, and S. Phanichphant. Semiconducting metal oxides as sensors for environmentally hazardous gases.In Sensors and Actuators B: Chemical, 2011, 160(1),580- 591.

[25]. Eranna, G., B. C. Joshi, D. P. Runthala, and R. P. Gupta. Oxide materials for development of integrated gas sensors|a comprehensive review.In Critical Reviews in Solid State and Materials Sciences,2004, 29(3-4), 111-188.

[26]. [26] Korotcenkov, Ghenadii. Metal oxides for solid-state gas sensors: What determines our choice. In Materials Science and Engineering,2007,139(1), 1-23.

[27]. Liu, Xiao, Cheng,S.,Liu,H., Hu,S., Zhang,D., and Ning,H. A survey on gas sensing technology. In Sensors,2012, 12(7),9635-9665.

[28]. TECHNICAL DATA MQ-2,3,4 and 5 GAS SENSOR. available in https://www.pololu. com/file/0J309/MQ2.pdf, https://www.pololu.com/file/0J310/MQ3.pdf, https://www. pololu.com/file/0J311/MQ4.pdf, https://www.parallax.com/sites/default/files/downloads/ 605-00009-MQ-5-Datasheet.pdf

[29]. Wu, Zhiyuan,Wang,H.,Wang,X.,Zheng,H.,Chen,Z., and Meng,C. Development of Electronic Nose for Qualitative and Quantitative Monitoring of Volatile Flammable Liquids. In Sensors, 2020,20(7), 1817

[30]. Park, Yun,S.,Kim,Y., Kim,T., Eom,T.H., Kim,S.Y., and Jang,H.W., Chemoresistive materials for electronic nose: Progress, perspectives, and challenges. In InfoMat,2019, 1(3), 289-316.

[31]. Wang, Chengxiang, Yin,L., Zhang,L., Xiang,D., and Gao,R. Metal oxide gas sensors: sensitivity and inuencing factors. In Sensors,2010,10(3), 2088-2106.

[32]. Kanan, So_an, M., Oussama M. El-Kadri, Imad A. Abu-Yousef, and Marsha C. Kanan. Semiconducting metal oxide based sensors for selective gas pollutant detection. In Sensors,2009,9(10), 8158-8196.

[33]. AWS. Amazon Machine Learning: Developer Guide. https://docs.aws.amazon.com/ machine-learning/latest/dg/model-fit-underfitting-vs-overfitting.html, 2020.

[34]. S. M. Scott, D. James, and Z. Ali, "Data analysis for electronic nose systems," Microchim. Acta, vol. 156, no. 3–4, pp. 183–207, 2006.