

Asthma Identification Using Gas Sensors and Support Vector Machine

Hari Agus Sujono^{*1}, Muhammad Rivai², Muhammad Amin³

^{1,2} Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

³ Department of Pulmonology and Respiratory Medicine, Airlangga University, Surabaya, Indonesia

¹ Department of Electrical Engineering, Institut Teknologi Adhi Tama Surabaya, Surabaya, Indonesia

*Corresponding author, e-mail: hariagus.itats@yahoo.com¹, muhammad_rivai@ee.its.ac.id²

Abstract

The exhaled breath analysis is a procedure of measuring several types of gases that aim to identify various diseases in the human body. The purpose of this study is to analyze the gases contained in the exhaled breath in order to recognize healthy and asthma subjects with varying severity. An electronic nose consisting of seven gas sensors equipped with the Support Vector Machine classification method is used to analyze the gases to determine the patient's condition. Non-linear binary classification is used to identify healthy and asthma subjects, whereas the multiclass classification is applied to recognize the subjects of asthma with different severity. The result of this study showed that the system provided a low accuracy to distinguish the subjects of asthma with varying severity. This system can only differentiate between partially controlled and uncontrolled asthma subjects with good accuracy. However, this system can provide high sensitivity, specificity, and accuracy to distinguish between healthy and asthma subjects. The use of five gas sensors in the electronic nose system has the best accuracy in the classification results of 89.5%. The gases of carbon monoxide, nitric oxide, volatile organic compounds, hydrogen, and carbon dioxide contained in the exhaled breath are the dominant indications as biomarkers of asthma. The performance of electronic nose was highly dependent on the ability of sensor array to analyze gas type in the sample. Therefore, in further study we will employ the sensors having higher sensitivity to detect lower concentration of the marker gases.

Keywords: Asthma, Exhaled breath, Gas sensors, Support vector machine

Copyright © 2018 Universitas Ahmad Dahlan. All rights reserved.

1. Introduction

The exhaled breath contains many gases, mostly oxygen (O₂), carbon dioxide (CO₂), water vapor, nitric oxide (NO) and various volatile organic compounds (VOCs) [1]. These gases can be measured accurately using Gas Chromatography (GC) [2]. However, this technique is expensive, because the testing process takes a long time and requires interpretation from an expert [3]. An electronic nose can be used as an alternative to analyze the exhaled breath with low production costs, non-invasive, faster time in sample measurements, and portables [4,5]. The electronic nose has been used to diagnose kidney disease [6,7], diabetes [8], and lung cancer [9]. Biomarkers are physical symptoms of laboratory measurements that can serve as indicators of biological or pathophysiological processes or in response to therapeutic interventions [10].

Several studies have analyzed exhaled breath for asthma by using Cyranose 320 that contains 32 different polymer nanocomposite sensors. The analytical method used to analyze the sensor response is Principal Component Analysis (PCA). The results show that electronic nose can distinguish exhaled breath between healthy and asthma subjects, however, it is not good enough to distinguish the subjects of asthma with different severity [11]. The performance of the electronic nose depends on the features of the classification algorithm for exhaled breath. One of the classification methods that gets a lot of attention as state of the art in pattern classification is the Support Vector Machine (SVM) [12-18]. The SVM is an effective technique for quantitatively analyzing gas mixtures as this can solve the cross sensitivity problem of gas sensor array [19].

The purpose of this study is to analyze the gases contained in the exhaled breath in order to recognize healthy and asthma subjects with varying severity. The exhaled breath samples were taken from patients diagnosed with asthma and then analyze it in the laboratory.

An electronic nose consisting of seven gas sensors equipped with the SVM classification method is used to analyze the gases contained in the exhaled breath sample to determine the patient's condition. The non-linear binary classification is used to identify healthy and asthma subjects, whereas the multiclass classification is used to identify the subjects of asthma with different severity.

2. Research Method

2.1. Subjects and research design

The subjects are 30 patients diagnosed with asthma, and 30 healthy subjects volunteered in the study. All subjects are adults, not smokers, aged between 30-60 years, and do not suffer from acute or chronic illness. In current clinical practice, asthma is diagnosed and monitored from symptoms and physiological measurements using the Globe Initiative for Asthma (GINA) and Asthma Control Test (ACT). One of the standard characteristics of GINA is lung function examination by measuring Forced Expiratory Volume in the first second (FEV₁) or Peak Expiratory Flow (PEF) performed by forced expiratory maneuvers through standard procedures. From this measurement, the degree of asthma is divided into three classes, namely controlled, partly controlled, and uncontrolled asthma [20,21].

The classification of asthma subjects is based on standards with diagnostic results using ACT shown in Table 1. Each subject is asked to exhale the air collected in 1L Tedlar bag after breathing in and out for 5 minutes with clean room air. The exhaled breath is carried out at the asthma clinic at Dr. Soetomo Surabaya after getting approval from Hospital Research and Development. Figure 1 shows the research design performed for the classification of asthma.

Table 1. The ACT values

ACT Score		
≥ 25	20 - 24	< 20
Controlled	Partly-Controlled	Uncontrolled

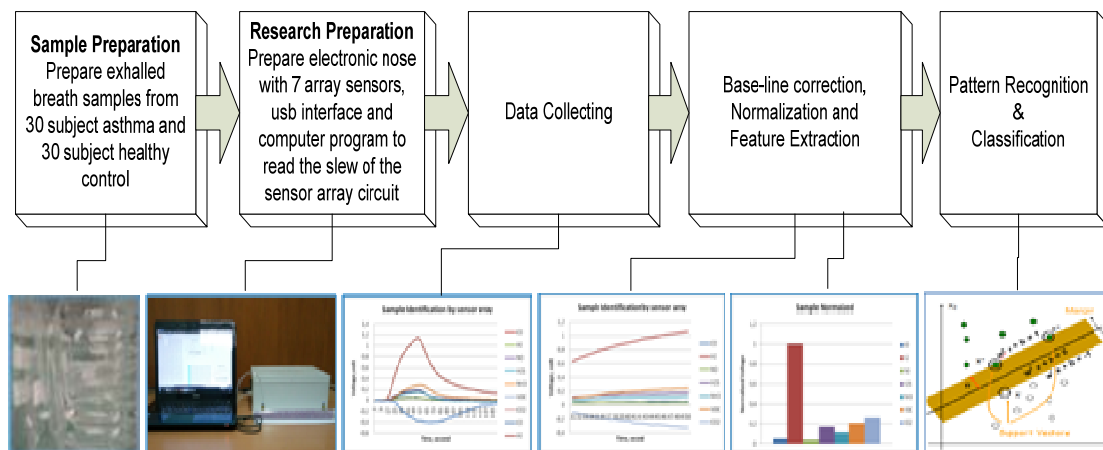


Figure 1. The research design for asthma identification using gas sensor

2.2. Electronic nose system

The main part of the electronic nose system is a 240 ml chamber consisting of a gas sensor array. Each sensor has sensitivity and selectivity to certain gases flowing through the chamber. The sensor resistance will change from fresh air to gas sample, R_S . The sensor output voltage is determined by equation (1).

$$V_{OUT} = \frac{R_L}{R_L + R_S} V_{CC} \quad (1)$$

where R_L is the load resistance, R_S is sensing resistance, V_{CC} is the constant voltage applied to the sensor and V_{OUT} is the transient output voltage. Each sensor response is an analog signal, which is then filtered, amplified and converted to digital form and then sends it to the computer every second. Pumps are activated to flow fresh air from silica gel or gas from the exhaled breath bag into the sensor chamber. The flow rate is maintained at 100 ml/min. Before carrying the gas sample, the sensor chamber is dried with fresh air to clean the sensor from residual gas. The humidity rate in the fresh air is reduced by flowing the air through a tube containing silica gel material.

The sampling process of the exhaled breath is carried out in the certain time order. In the period of 1st to 15th seconds, the sensor response is in a based line time, t_b . During this time, the valve is OFF so that fresh air can flow from B to C. In the period of 16th to 55th seconds, the exhaled breath is injected to the sensor chamber. This process is called reaction stage time, t_r . At this moment, the valve is ON so that the gas in tedlar bag is sucked from A to C. In the period of 56th to 150th seconds, the injection is stopped and sensor array is cleaned, and this process is called as purge stage time, t_p . At this stage, the valve is OFF again. After the 150th seconds, the valve remains OFF until further sampling took place. The block diagram of the electronic nose system is shown in Figure 2, while the photo setup of the apparatus and its computer interface is shown in Figure 3. The sensor response during sampling process is shown in Figure 4. The portion of the sensor response at the reaction stage over a period of 30th to 49th seconds is a feature extraction which can be considered to represent the overall sensor response for the classification process.

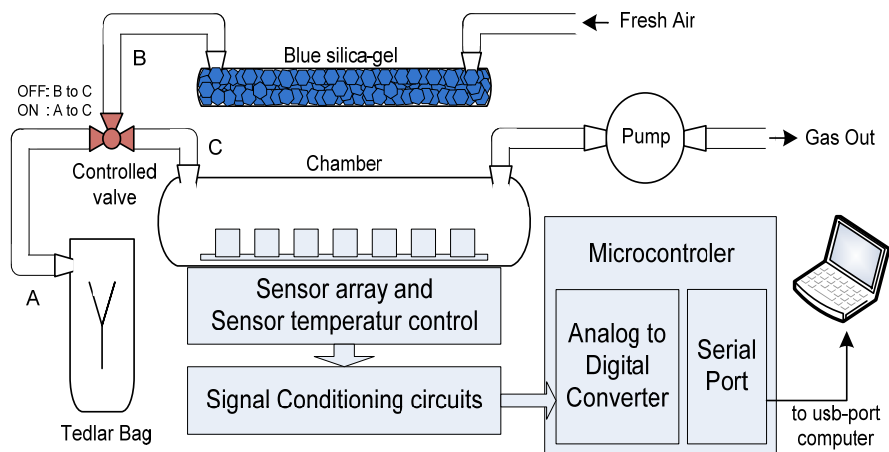


Figure 2. The electronic nose system

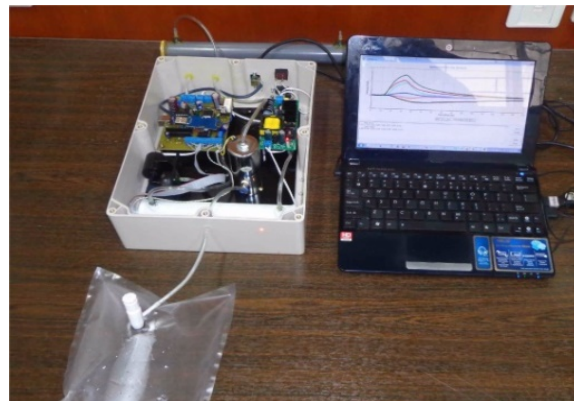


Figure 3. The exhaled breath analysis using the electronic nose

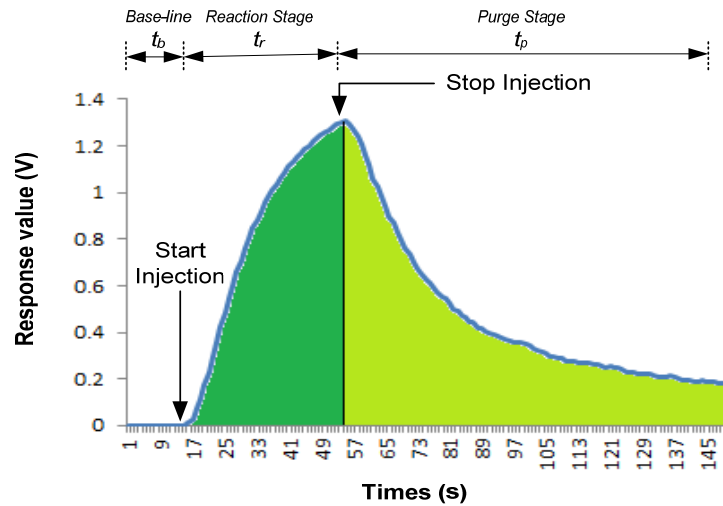


Figure 4. The time response of a gas sensor

Table 2. The gas sensors used in the electronic nose system

Sensors	Gases	Sensitivities (ppm)
MQ7	CO	10 – 10000
MQ8	H ₂	100 – 1000
MQ131	NO _x	0.01 – 2
MQ136	H ₂ S	1 – 100
MQ137	NH ₃	5– 200
MQ138	VOCs	10 - 1000
TGS4161	CO ₂	350 – 10000

The sensor array used on the electronic nose consists of seven gas sensors which its selectivity and sensitivity are shown in Table 2. Each sensor is intended to detect gases in exhaled breath that indicate the presence of asthma in the subject. In some studies have showed that asthma patients had higher concentration values for NO [22], and carbon monoxide (CO) [23]. There is also a relationship between the VOC in exhaled breath subject with lung disease [24]. There are several factors associated with exhaled breath by asthma patients including an increase in hydrogen (H₂) due to digestive system, CO₂ affected by exposure to air pollution, ammonia (NH₃) due to acid-base status in the airway of asthmatics [25], and hydrogen sulfide (H₂S) as harmful hydrocarbon elements in the human body [26].

2.3. Data Analysis

In the sampling process, the sensor array measures the level of gases in the exhaled breath and then sends it to the computer for analysis. The data package of the sampling process on a subject is a data set as matrix [150x7] consisting of a response of seven sensors for 150 seconds. Only the signal response at the reaction stage is used for analysis. Data analysis consists of three stages, namely pre-processing, feature extraction and classification. In the pre-processing includes baseline correction and normalization. The baseline correction is conducted by subtracting each sensor signal response with the average signal response at the baseline, $B_{i,s}(t_b)$. The result is the pre-processed signal response defined at equation (2).

$$R_{i,s}^P = R_{i,s} - \frac{1}{N_b} \sum_{t_b=1}^N B_{i,s}(t_b) \quad (2)$$

where $i = 1, 2, \dots, N_i$ (N_i is the number of sample subjects), $s = 1, 2, \dots, N_s$ (N_s is the number of sensors), $t_b = 1, 2, \dots, N_b$ (N_b is the maximum time period on the base line stage). Normalization of the data is desired to reduce a pattern variation due to variations in the vapor concentration [27]. The normalized data is expressed by equation (3).

$$R_{i,s}^{PN}(t_r) = \frac{R_{i,s}^P(t_r)}{(R_{i,s}^P(t_r))_{MAX}} \quad (3)$$

Output signal response for each sensor at the reaction stage after being processed by equation (2) and (3), $R_{i,s}^{PN}(t_r)$ is a feature extraction signal for use on classification process. In this process, a dataset [20x7] is derived from the $R_{i,s}^{PN}(t_r)$ on the signal response at the sampling periode of 30th - 49th seconds.

SVM is a machine learning method that works on the principle of Structural Risk Minimization (SRM) with the aim of finding the best hyperplane that separates the two classes in the input space. The method proposed in 1992 by Vladimir N. Vapnik, Bernhard E. Boser and Isabelle M. Guyon has been widely applied especially in the field of bioinformatics [28]. There are two types of classification conducted by the SVM, namely non-linear binary classification and non-linear multi-class classification. For multi-class classification employs one-against-one approach and Gaussian radial basis function kernel or RBF kernel.

The SVM concept can be explained simply as a search for the best hyperplane that serves as a separator of two classes in the input space. Figure 5 shows some patterns of members of two classes: +1 and -1. The pattern incorporated in class -1 is symbolized by green (circle), while the pattern in class +1 is symbolized by red (box). The classification problem can be translated by finding the line or hyperplane that separates the two groups. Various alternative discrimination boundaries are shown in Figure 5(a). The best separator hyperplane between the two classes can be found by measuring the hyperplane's margins and searching for the maximum point.

Margin is the distance between the hyperplane and the closest pattern of each class. The closest pattern is called a support vector. The solid line in Figure 5(b) shows the best hyperplane, which is located at the center of the two classes, while the green and red dots that are in the black circle are the support vectors. The effort to locate the hyperplane is at the heart of the learning process in SVM.

The available data is denoted as $\bar{x}_i \in \mathcal{R}^d$, whereas each label is denoted $y_i \in \{-1, +1\}$ for $i = 1, 2, \dots, l$. It is assumed that both classes -1 and +1 can be perfectly separated by d dimensioned hyperplane, defined as: $\bar{w} \cdot \bar{x} + b = 0$. The \bar{x}_i pattern that belongs to class -1 (negative sample) can be formulated as a pattern that satisfies inequality $\bar{w} \cdot \bar{x}_i + b \leq -1$. While the \bar{x}_i pattern that belongs to the class +1 (positive sample) can be formulated as a pattern that satisfies inequality $\bar{w} \cdot \bar{x}_i + b \geq +1$. The greatest margin can be found by maximizing the distance value between the hyperplane and its nearest point, $\frac{1}{\|\bar{w}\|}$. This can be formulated as a Quadratic Programming (QP) problem, which is to find the minimal point of equation shown in equation (4), and (5).

$$\min_{\bar{w}} \lambda(w) = \frac{1}{2} \|\bar{w}\|^2 \quad (4)$$

$$y_i(\bar{w} \cdot \bar{x}_i + b) - 1 \geq 0, \quad \forall i \quad (5)$$

This problem can be solved by various computational techniques, including Lagrange Multiplier.

$$L(\bar{w}, b, \beta) = \frac{1}{2} \|\bar{w}\|^2 - \sum_{i=1}^l \beta_i (y_i((\bar{w} \cdot \bar{x}_i + b) - 1)) , \quad (i = 1, 2, \dots, l) \quad (6)$$

β_i is a Lagrange multiplier that is either 0 or positive $\beta_i \geq 0$. The optimal value of equation (6) can be calculated by minimizing L to \bar{w} and b and maximizing L to β_i . Considering the nature at the optimum point of slope $L = 0$, equation (6) can be modified as a maximization problem containing β_i only, as indicated by (7).

Maximize:

$$L(\bar{\beta}) = \sum_{i=0}^l \beta_i - \frac{1}{2} \sum_{i,j=1}^l \beta_i \beta_j y_i y_j \bar{x}_i \cdot \bar{x}_j \quad (7)$$

Subject to

$$\beta_i \geq 0 \text{ (for } i = 1, 2, \dots, l \text{)}; \sum \beta_i y_i = 0$$

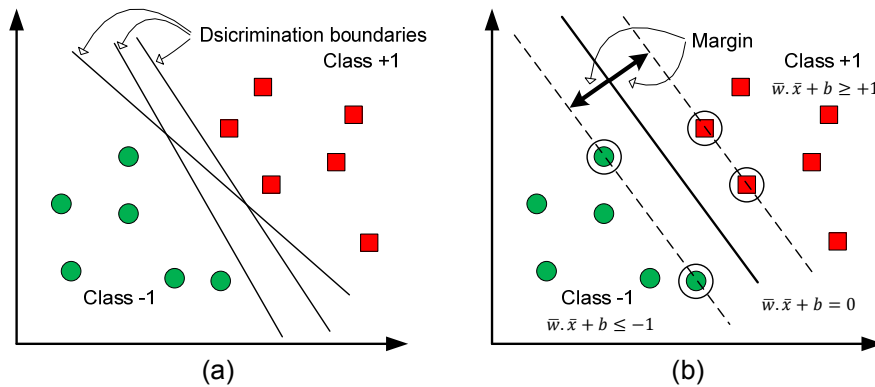


Figure 5. The SVM finds the best hyperplane that separates the two classes

From the results of this calculation, it will be obtained β_i which mostly positive value. Data correlated with positive β_i is called a support vector. In general, problems in real world domains are rarely linearly separate but mostly non-linear. To solve the non-linear problems, the SVM is modified by entering the Kernel function. In non-linear SVM, the data x is mapped by the function $\Phi(\bar{x})$ to a higher-dimensional vector space. In this new vector space, the hyperplane that separates the two classes can be constructed. To classify non-linear data, the SVM formula must be modified. Therefore, the two limiting fields of (5) have to be changed so that they are more flexible for certain conditions with the addition of the variable ξ_i ($\xi_i \geq 0$) shown in the equation (8).

$$y_i(\bar{x}_i \cdot \bar{w} + b) \geq 1 - \xi_i, \quad \forall_i \tag{8}$$

Thus the equation (4) is changed to:

$$\min_{\bar{w}} \lambda(w) = \frac{1}{2} \|\bar{w}\|^2 + C \sum_{i=1}^l \xi_i \tag{9}$$

C is chosen to control the tradeoff between margin and misclassification. A large C value means it will give a larger penalty for the misclassification.

Another method of solving the non-linear data problems in SVM is by mapping data to a higher dimension space (feature space) [29], with data in that space can be separated linearly by using the transformation of $\Phi: \mathcal{R}^d \rightarrow H$. Thus the training algorithm depends on the data through the dot product in H (e.g. $\Phi(x_i) \cdot \Phi(x_j)$). If there is a kernel K function, such as $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$, thus in the training algorithm requires only the kernel K function without having to know the exact Φ transformation. By transforming $x_k \rightarrow \Phi(x_k)$, then the value w becomes $w = \sum_{i=1}^l \beta_i y_i \Phi(x_i)$ and the learning function becomes:

$$h(x_d) = \sum_{i=1}^l \beta_i y_i \Phi(x_i) \cdot \Phi(x_d) + b \tag{10}$$

The feature space usually has a higher dimension resulting a feature space that may be very large. To solve this problem, then it uses the "kernel trick" $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$, for which the equation (10) becomes:

$$h(x_d) = \sum_{i=1}^l \beta_i y_i K(x_i, x_j) + b \tag{11}$$

with x_i is the support vector. There are some kernel functions that are commonly used in the SVM, one of which is used in this study is the RBF kernel. In particular, it is commonly used in

support vector machine classification [30]. The RBF kernel on two samples x_i and x_j represented as feature vectors in some input space is defined in equation (12)

$$K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / 2\sigma^2) \tag{12}$$

The classification reliability is indicated by the sensitivity and specificity resulted by either positive (diseased) or negative (healthy). In the classification, the number of disease subjects identified as the disease subjects are denoted by t_p (true positive), the disease subjects identified as the healthy subjects are denoted by f_p (false positive), the healthy subjects identified as the disease subject are denoted by f_n (false negative) and the healthy subjects identified as the healthy subjects are denoted by t_n (true negative). The sensitivity, specificity, and accuracy for two and three classes are defined as in Table 3 and 4, respectively.

Table 3. Definition of Sensitivity, Specivicity and Accuracy for two classes

		Predicted Result		Sensitivity	Specivicity	Accuracy
		Positive	Negative			
Actual condition	Positive	t_p	f_n	$\frac{t_p}{t_p + f_n}$	$\frac{t_n}{t_n + f_p}$	$\frac{t_p + t_n}{t_p + f_p + t_n + f_n}$
	Negative	f_p	t_n			

Table 4. Definition of Accuracy for three classes

		Predicted Result			Accuracy
		A	B	C	
Actual Condition	A	t_A	f_{AB}	f_{AC}	$T = t_A + t_B + t_C$ $Acc = \frac{T}{T + f_{AB} + f_{AC} + f_{BA} + f_{BC} + f_{CA} + f_{CB}}$
	B	f_{BA}	t_B	f_{BC}	
	C	f_{CA}	f_{CB}	t_C	

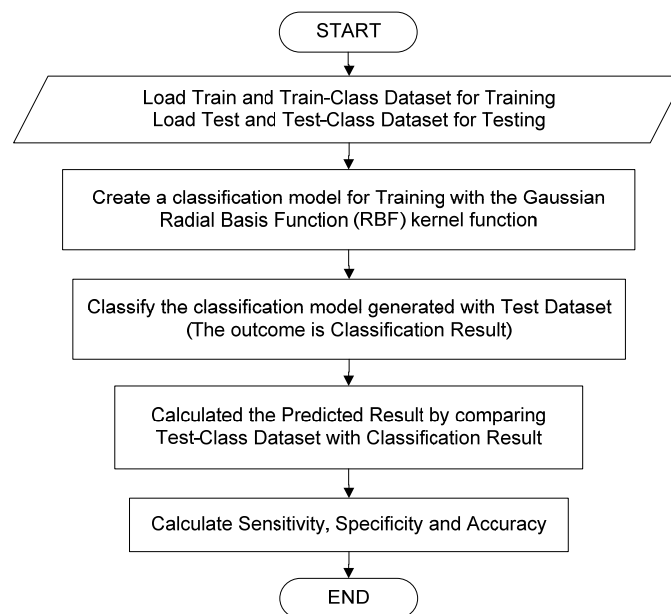


Figure 6. The SVM Algorithm for classification

3. Results and Analysis

The electronic nose has been tested to detect and identify exhaled breath samples from subjects with healthy and asthma subjects and among asthma subjects with different

severity using the procedure described in the method. Figure 7 shows the response of the seven sensors to the four exhaled breath samples during a periode of 150 seconds. Figure 8 (a) shows the average response for the two categories. It is possible to obtain a combination of sensors that clearly indicate each of the two categories. Sensor responses of CO, NO, H₂S, and CO₂ are smaller for healthy subjects, but not for H₂, NH₃, and VOCs. Figure 8 (b) shows the sensor responses for the four categories. We have conducted five types of classifications.

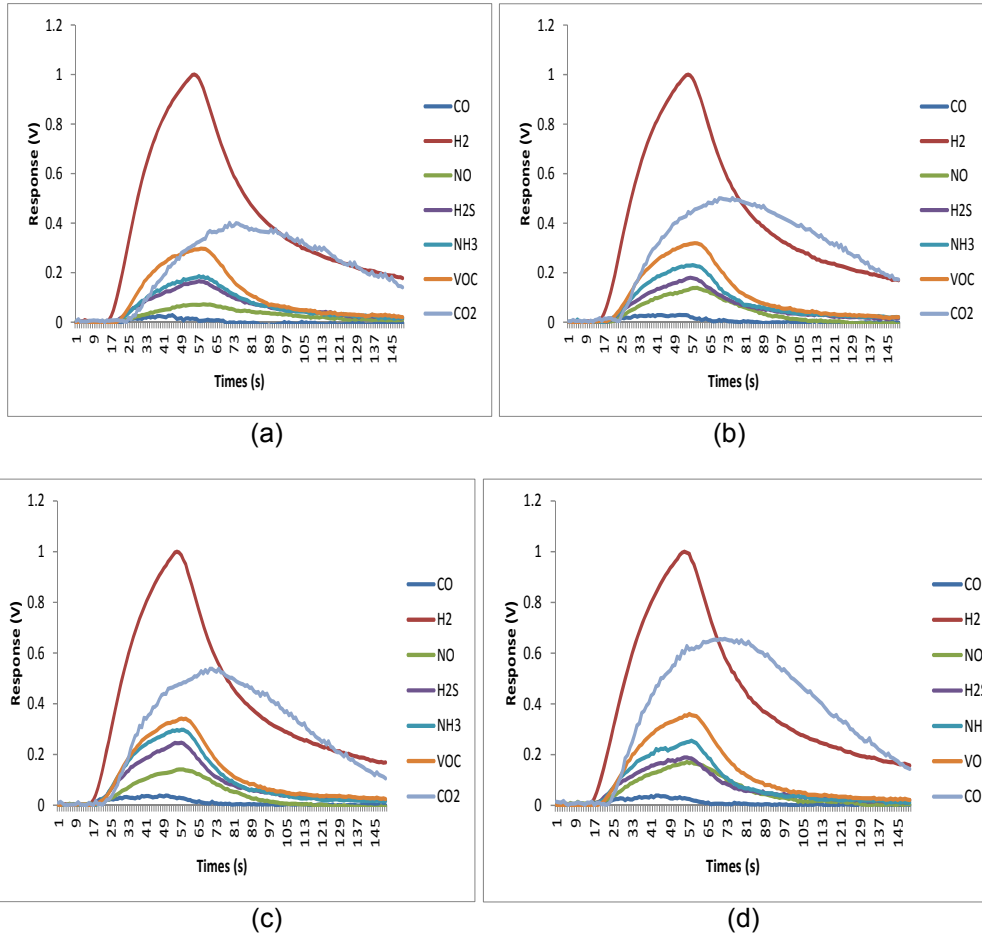


Figure 7. The sensor responses to (a) healthy (b) controlled asthma (c) partly controlled asthma, and (d) uncontrolled asthma

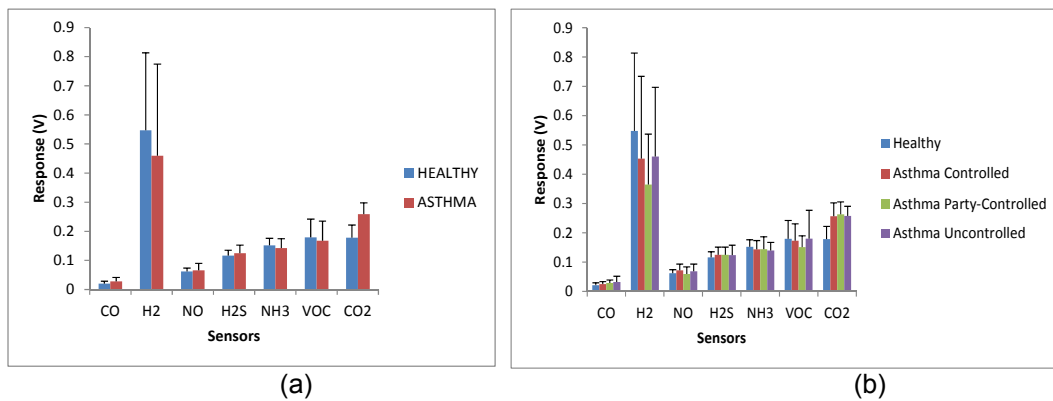


Figure 8. The sensor responses to (a) two categories and, (b) four categories

3.1. Classification of Healthy and Asthma using seven sensors

In this study, we used the electronic nose to distinguish between exhaled breath samples of healthy and asthma subjects using non-linear binary SVM classification. The composition of the subject database is shown in Table 5. The exhaled breath samples of subjects were collected from patients who have been diagnosed with early diagnosis standard using ACT. We randomly collected 20 samples for the training sets from each category. The remaining 10 samples from each category were used for the test set. In this study, we aim to test the ability of the electronic nose to distinguish each category of healthy and asthma subjects using binary classification.

Table 5. The composition of the subject database for two categories

Categories of subjects	Number	Age
Healthy	30	32 – 60
Asthma	30	30 – 58

Table 6 shows the results of classification for two categories. For the 10 asthma subjects as the test set, the system could correctly identify 10.0 samples as asthma and incorrectly 0.0 sample as healthy. For the 10 healthy subjects as the test set, the system could correctly identify 7.8 samples as healthy and incorrectly 2.2 samples as asthma. In this identification results, the sensitivity, specificity, and accuracy are 100, 78.0, and 89%, respectively.

Table 6. The non-linear binary SVM classification result of two categories using seven sensors

Actual Condition	Asthma	Positive	Predicted Result		Sensitivity	Specificity	Accuracy
			Positive	Negative			
	Healthy	Negative	2.2	7.8	100	78	89

3.2. Classification of Healthy and Asthma with Different Severity

We have collected three categories of asthma subjects separated by severity, each consisting of ten samples. Ten samples for healthy subjects were taken randomly from 30 samples. The composition of the subject database is shown in Table 7. We randomly collected six samples for the training sets from each category. The remaining four samples from each category were used for the test set. Table 8 shows the result of classification for four categories. CA subject could be distinguished from healthy subject with 82% accuracy, 90% sensitivity and 75% specificity. PA subject could be differentiated from healthy subject with 87.5% accuracy, 75% sensitivity and 100% specificity. UA subject could be distinguished from healthy subject with 66% accuracy, 65% sensitivity and 67.5% specificity.

Table 7. The composition of the subject database for four categories

Categories of subjects	Number
Healthy	10
Controlled asthma (CA)	10
Partly-controlled asthma (PA)	10
Uncontrolled asthma (UA)	10

Table 8. The non-linear binary SVM classification result of four categories

Actual Condition	CA	Healthy	Positive	Negative	Sensitivity	Specificity	Accuracy
	PA	Healthy	3.0	1.0	75	100	87.5
	UA	Healthy	2.6	1.4	65	67.5	66.25

3.3. Classification of Three Categories of Asthma with Different Severity

In this study, we have tested the electronic nose to distinguish three different categories of asthma subjects with different severity of CA, PA and UA using the non-linear multi class SVM classification. Table 9 shows the results of non-linear multi class SVM classification for this study. For the four CA subjects as the test set, the system could correctly identify 0.9 as CA, incorrectly 1.1 samples as PA, incorrectly 2.0 samples as UA. For the four PA subjects as the test set, the system could correctly identify 2.2 samples as PA, incorrectly 1.8 samples as CA, and incorrectly 0.0 samples as UA. For the four UA subjects as the test set, the system could correctly identify 1.6 samples as UA, incorrectly 1.0 sample as CA, and incorrectly 1.4 samples as PA. In this identification, the accuracy is 39%. The result of identification shows that by only providing accuracy value, the system cannot distinguish among categories. Therefore, we have employed the binary classification for these three categories. Table 10 shows the non-linear binary SVM classification for three categories of asthma subjects. For the four CA samples as the test set, the system could correctly identify 0.4 samples as CA, and incorrectly 3.6 samples as PA. This indicates that the sensitivity is bad (i.e. 10%). For the four PA samples as the test set, the system could correctly identify 3.3 samples as PA, and incorrectly 0.7 samples as CA.

Table 9. Non-linear multi class SVM classification result

		Predicted Result			Accuracy
		CA	PA	UA	
Actual Condition	CA	0.9	1.1	2.0	39%
	PA	1.8	2.2	0	
	UA	1.0	1.4	1.6	

Table 10. Non-linear Binary SVM classification result

			Predicted Result		Sensitivity	Specificity	Accuracy
			Positive	Negative			
Actual Condition	CA	Positive	0.4	3.6	10.0%	82.5%	46,25%
	PA	Negative	0.7	3.3			
	CA	Positive	1.6	2.4	40%	75%	57,5%
	UA	Negative	1.0	3.0			
	PA	Positive	4.0	0.0	100%	57.5	78,75%
	UA	Negative	1.7	2.3			

This indicates that the specificity is good (i.e. 82.5%). Similarly, to distinguish between CA and UA subjects, the system's ability shows poor sensitivity (i.e. 40%) and quite good specificity (i.e. 75%). Finally, to differentiate between PA and UA subjects, the system's capability shows good accuracy (i.e. 78.75%).

3.4. Classification of Healthy and Asthma using five sensors

In this study, we used the response from five sensors, i.e. CO, NO, VOC, H₂ and CO₂ sensors, to differentiate between healthy and asthma subjects using non-linear binary SVM classification. Table 11 shows the classification results obtaining the values of the sensitivity, specificity, and accuracy are 100, 79, and 89.5%, respectively.

Table 11. The non-linear binary SVM classification result of two categories using five sensors

			Predicted Result		Sensitivity	Specificity	Accuracy
			Positive	Negative			
Actual Condition	Asthma	Positive	10.0	0.0	100	79	89.5
	Healthy	Negative	2.1	7.9			

3.5. Classification of Healthy and Asthma using three sensors

In this study, we used the response from three sensors only, i.e. CO, NO and VOC sensors, to differentiate between healthy and asthma subjects. Table 12 shows the classification results obtaining the values of the sensitivity, specificity, and accuracy are 90, 79, and 84.5%, respectively.

Table 12. The non-linear binary SVM classification result of two categories using three sensors

Actual Condition	Asthma	Positive	Predicted Result		Sensitivity	Specificity	Accuracy
			Positive	Negative			
Healthy		Negative	9.0	1.0	90	79	84.5
		Positive	2.1	7.9			

The ability of the electronic nose to identify between healthy and asthma subjects using various number of sensors is shown in Figure 9. It concludes that the use of five gas sensors in the electronic nose system has the best accuracy in the classification results. This suggests that the gases of CO, NO, VOC, H₂, and CO₂ are the dominant indications as markers of asthma.

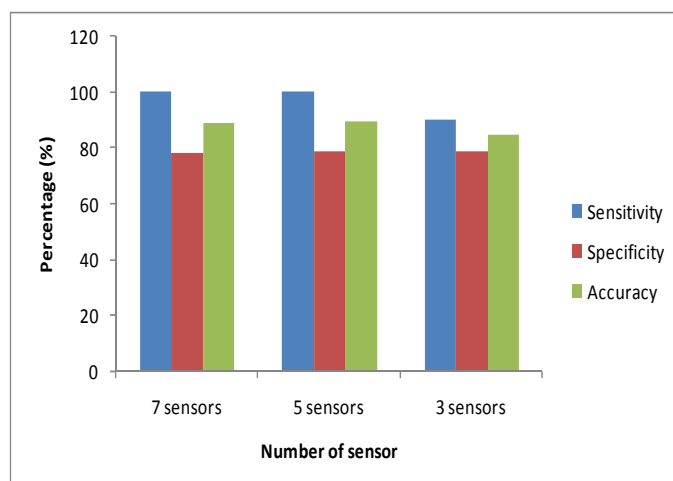


Figure 9. The classification results of the electronic nose for various number of sensors

4. Conclusion

In this research, the exhaled breath identification system for healthy and asthma subjects with different severity was attained by means of electronic nose. The electronic nose system has performed well by identifying the gases contained in exhaled breath. The system produces signal response processed by three stages, namely pre-processing, feature extraction, and classification. The feature extraction and classification was performed by SVM method, which was an effective technique to analyze gas mixtures quantitatively because of its ability to solve the cross sensitivity problems from the gas sensor array. The result of this study showed that the system using non-linear multi class SVM provided quite low accuracy to differentiate among asthma subjects with several degrees of severity. By using non-linear binary SVM, this system can distinguish between the partially controlled and uncontrolled asthma subjects with the accuracy of 78.8%. However, the system provided high rate of sensitivity, specificity, and accuracy to recognize between healthy and asthma subjects. The use of five gas sensors in the electronic nose system has the best accuracy in the classification results of 89.5%. This study suggests that the gases of CO, NO, VOC, H₂, and CO₂ contained in the exhaled breath can be used as biomarkers of asthma.

Acknowledgement

This research was carried out with financial aid support from the Ministry of Research, Technology and Higher Education of the Republic of Indonesia (Kemendiknas RI).

References

- [1] Yan K, Zhang D, Wu D, Wei H, Lu G. Design of a Breath Analysis System For Diabetes Screening and Blood Glucose Level Prediction. *IEEE Trans. Biomed. Eng.* 2014; 61: 2787-2795.

- [2] K Lamote, P Brinkman, L Vandermeersch. Breath Analysis by Gas Chromatography-Mass Spectrometry and Electronic Nose to Screen for Pleural Mesothelioma: A Cross-Sectional Case-Control Study. *Oncotarget*. 2017; 8(53): 91593-91602.
- [3] M Brattoli, E Cisternino, PR Dambruoso. Gas Chromatography Analysis with Olfactometric Detection (GC-O) as a Useful Methodology for Chemical Characterization of Odorous Compounds. *Sensors*. 2013; 13(12): 16759-16800.
- [4] S Chen, Y Wang, S Choi. Applications and Technology of Electronic Nose for Clinical Diagnosis. *Open Journal of Applied Biosensor*. 2013; 2: 39-50.
- [5] M Rivai, D Purwanto, H Juwono, HA Sujono. Electronic Nose using Gas Chromatography Column and Quartz Crystal Microbalance. *TELKOMNIKA Telecommunication Computing Electronics and Control*. 2011; 9(2): 319-326.
- [6] T Saidi, O Zaim, M Moufid. 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 and Actuators B: Chemical*. 2018; 257: 178-188.
- [7] D Guo, D Zhang. A Novel Breath Analysis System Based on Electronic Olfaction. *IEEE Transactions on Biomedical Engineering*. 2010; 57(11): 2753-2763.
- [8] HM Saraoğlu, AO Selvi, MA Ebeoğlu, C Taşaltın. Electronic Nose System Based on Quartz Crystal Microbalance Sensor for Blood Glucose and HbA1c Levels from Exhaled Breath Odor. *IEEE Sensors Journal*. 2013; 11: 4229-4235.
- [9] Tırzite M, Bukovskis M, Strazda G, Jurka N, Taivans I. Detection Of Lung Cancer in Exhaled Breath with an Electronic Nose Using Support Vector Machine Analysis. *J Breath Res*. 2017; 11(3): 036009.
- [10] Z Diamant, JD Boot. Biomarkers in Asthma and Allergic Rhinitis. *Journal Pulmonary Pharmacology & Therapeutics*. 2010; 23: 468-481.
- [11] S Dragonieri, R Schot, B Mertens, SL Cessie. An Electronic Nose in The Discrimination of Patients with Asthma and Controls. *J. Allergy Clin. Immunol*. 2007; 120: 856-862.
- [12] P Borah, D Gupta. Review: Support Vector Machines in Pattern Recognition. *International Journal of Engineering and Technology*. 2017; 9(38): 43-48.
- [13] H Xisheng, Z Wang, C Jin, Y Zhen, X Xue. A Simplified Multi-Class Support Vector Machine with Reduced Dual Optimization. *Elsevier on Pattern Recognition Letters*. 2012; 33: 71-82.
- [14] Mulyati, WA Kusuma, M Nurilmala. Identification of Tuna and Mackerel Based on DNA Barcodes using Support Vector Machine. *TELKOMNIKA Telecommunication Computing Electronics and Control*. 2016; 14(2): 778-783.
- [15] B Gunawan, M Rivai, H Juwono. Characterization of Polymeric Chemi-resistors for Gas Sensor. *TELKOMNIKA Telecommunication Computing Electronics and Control*. 2012; 10(2): 275-280.
- [16] MS Bhatt, TP Patalia. Indian Monuments Classification using Support Vector Machine. *International Journal of Electrical and Computer Engineering*. 2017; 7(4): 1952-1963.
- [17] SVS Prasad, TS Savithri, IVM Krishna. Comparison of Accuracy Measures for RS Image Classification using SVM and ANN Classifiers. *International Journal of Electrical and Computer Engineering (IJECE)*. 2017; 7(3): 1180-1187.
- [18] TG Ngo, QT Ngo, DD Nguyen. Image Retrieval with Relevance Feedback using SVM Active Learning. *International Journal of Electrical and Computer Engineering (IJECE)*. 2016; 6(6): 3238-3246.
- [19] X Wang, J Chang, Z Yu. *Gas Quantitative Analysis by Combining Kernel Independent Component Analysis and Least Squares Support Vector Machine*. Research Gate. International Conference on Electrical and Control Engineering. 2011.
- [20] Global Initiative for Asthma (GINA). Global Strategy Report for Asthma Management and Prevention. 2016. Available at www.ginasthma.com
- [21] Boulet LP, Fitzgerald JM, Levy ML, Cruz AA, Pedersen S, Haahtela T, Bateman ED. Asthma Guidelines Implementation: A Guide to The Translation of GINA Guidelines Into Improved Care. *Eur Respir J*. 2012; 39(5): 1220-9.
- [22] Harnan SE, Essat M, Gomersall T, Tappenden P, Pavord I, Everard M, Lawson R. Exhaled Nitric Oxide in The Diagnosis of Asthma in Adults: A Systematic Review. *Clin Exp Allergy*. 2017; 47(3): 410-429.
- [23] M Shameem, Q Hussain, A Ahmad. Increased Levels of Exhaled Carbon Monoxide and Their Correlation with Airflow Obstruction in Asthma and COPD. *European Respiratory Journal*. 2012: 1054.
- [24] Dallinga JW, Robroeks CM, van Berkel JJ, Moonen EJ, Godschalk RW, Jöbsis Q, Dompeling E, Wouters EF, van Schooten FJ. Volatile Organic Compounds in Exhaled Breath as a Diagnostic Tool for Asthma in Children. *Clin Exp Allergy*. 2010; 40(1): 68-76.
- [25] MMT Lozowska, Z Zietkowski. Inflammatory Markers and Acid-Base Equilibrium in Exhaled Breath Condensate of Stable and Unstable Asthma Patients. *Int Arch Allergy Immunol*. 2012; 159: 121-129.
- [26] Zhang J, Wang X, Chen Y, Yao W. Correlation Between Levels of Exhaled Hydrogen Sulfide and Airway Inflammatory Phenotype in Patients with Chronic Persistent Asthma. *Respirology*. 2014; 19(8): 1165-1169.

-
- [27] HA Sujono, M Rivai. Vapor Identification System Using Quartz Resonator Sensor Array and Support Vector Machine. *ARPN Journal of Engineering and Applied Sciences*. 2014; 9(12): 2426-2430.
- [28] R Akbani, T Korkmaz. Applications of Support Vector Machines in Bioinformatics and Network Security. Book edited by Yagang Zhang, February 1, 2010.
- [29] C Shen, H Li. On The Dual Formulation of Boosting Algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* 2010; 12(30): 2216-2231.
- [30] YW Chang, CJ Hsieh, KW Chang, M Ringgaard, CJ Lin. Training and Testing Low-Degree Polynomial Data Mappings Via Linear SVM. *J. Machine Learning Research*. 2010; 11: 1471-1490.