Identification of Human Pathogen in Nutrient Culture Media using an Electronic Nose

Syahida Amani Zulkifli[1], Che Wan Sharifah Robiah Mohamad[1], Abu Hassan Abdullah[1, 2]
[1] *School of Mechatronic Engineering, Universiti Malaysia Perlis (UniMAP), 02600 Arau, Perlis, Malaysia*
[2] *Centre of Excellence for Advanced Sensor Technology, (CEASTech), Universiti Malaysia Perlis, 02600 Jejawi, Perlis, Malaysia*
syahidaamanizulkifli@yahoo.com

***Abstract*—This paper present human pathogen bacteria for early screening using electronic nose. An electronic nose (E-nose) designed for mimicking the mammalian olfactory system to recognize gases and odors. Electronic nose used for detecting different bacteria such as Pseudomonas aeruginosa and Shigella cultured on media agar. In addition, the data from the electronic nose (E-nose) is processed using a statistical method which is principal component analysis (PCA) and existing classification method which is K-Nearest Neighbor Method (KNN). The study shows the capability of electronic nose (E-nose) for early screening for bacterial infection in the human stomach.**

***Index Terms*—Accuracy; Classification; Electronic Nose; Volatile Organic Compound.**

I. INTRODUCTION

Volatile compound is produced by all organisms as part of their metabolism and certain species of bacteria are often recognized as having a characteristics smell in vitro when isolated on recovery media [1]. Investigation of the volatile compound is continued for many years and lead to a determination of the group of substance characteristics for a particular disease. Chemical analysis of bacterial culture includes analysis of bacterial metabolism, bacterial cell walls composition and fatty acids profiling have been introduced as bacterial differentiation and detection methods [2]. Early detection and classification of pathogenic microorganisms are crucial for rapid treatment initiation and improved clinical outcome [3]. One non-invasive which is gaining interested in the analysis of odor measurement and volatile organic compound for medical use is gas sensor array which is electronic nose (E-nose).

There is a various number of the researcher that has been reported working on diagnosis with many types of disease by using Electronic Nose. An electronic nose is a system that consists of different types of electronic, chemical gas related sensor array to partially specify, in addition to appropriate statistical methods to enable the recognition of complex odors [4].

In the last decade, compact gas sensors have been integrated into an instrument which can provide a fast and non-instructive measurement of gaseous agents [5].

The growth of bacteria on organic matter generates volatile organic compounds and fixed gases that can be used as an indicator of its presence and identify characteristics of microorganisms concerned [6]. Therefore, a most effective device for odor measurement is the electronic nose. The electronic nose device was predominantly conducted with polymer-type with sensors 32 sensors in the sensor array [7].

In this paper, an approach for clustering bacteria with the electronic nose is presented. With the advancement of electronic nose, a small size array for detection of more chemicals with lower cost, higher predictive, accuracy and portability becomes more desired objective and as instrumental analysis [8]. The methods evaluate the suitability of the sample for classification by representing the output from principal component analysis and k- nearest neighbor (KNN).

II. DATA PROCESSING TECHNIQUE

This study demonstrates different data analysis techniques such as feature reduction and pattern recognition classification in order to determine the performance of electronic nose in investigating bacteria species in the human stomach.

A. Feature Reduction

The most feature reduction technique for the electronic nose is mainly Principal Component Analysis (PCA). PCA is a statistical method where it is transformed multidimensional data into coordinates that maximize the variance while minimizing the correlation in the data set [9]. The first principal component has the largest possible variance. In this study, two principal components for instance first principal component and second principal component have been taken on the dataset of an electronic nose. Thus, it produces graphical data in order to identify the pattern for clustering, similarity and differences from electronic nose data set.

B. Data Classification: K-Nearest Neighbor Method (KNN)

The K-Nearest Neighbor method is called a lazy learning algorithm. This is because the method does not build models but collect the closest k records that can be found in training data set which have high similarity and put into groups. The purpose of KNN is for estimation of a fixed number of observations. KNN is chosen because of its simplicity and flexibility to corporate for different data types. As for classification, it functions to select most frequent neighbor. Before using this method, a training set and a test sample to know the value of k. Nutrient agar is recommended for the cultivation of non-fastidious microorganisms [12].
Three types of the common used with KNN:

i. Euclidean Distance:

$$\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \quad (1)$$

ii. Minkowski Distance:

$$(\sqrt{\sum_{i=1}^{k}(x_i - y_i)^q})^{1/q} \quad (2)$$

iii. Manhattan Distance:

$$\sum_{i=1}^{k}|x_i - y_i| \quad (3)$$

The pair (x_i, y_i) is the training set given to solve the estimated $y(x)$ from new input which is x. The index of integers,i, and k is the prediction of how many neighbours that are going to be used. The distance is calculated by the formula between the instances.

Important features of KNN are simple to be implemented. It is also lazy learning due to the calculated distance of its neighbor. The KNN also has small error ration robustness. In addition, it is capable of working with little information. Lastly, the KNN is an instance-based learning method for classification purposes. The most used formula is the Euclidean distance because it is easy to be calculated [10].

III. Methodology

A. Electronic Nose System

The sensor array in this study is constructed due to the high sensitivity and quick response of the sensors to metabolites. The electronic nose system consists of a few sensors which are TGS2620, TGS826, TGS2180, TGS2600, TGS2201, TGS2610, TGS4161, TGS2611 and SHT77 (temperature and humidity). The acquired analog signals from sensors are converted into digital signals by the built-in Analogue to Digital Converter (ADC) which is stored in the computer for further processing. Figure 1 shows the electronic nose hardware that used for data collection.

Figure 1: Electronic nose hardware

B. Sample Preparation

The bacterial samples used for this study are common pathogens in the human stomach as Table 1. Then, a cotton swab is used to streaking the bacteria over the plate. Bacterial culture streaking on the plate allows bacteria to grow, particularly in the culture medium in a controlled laboratory condition. The bacteria were spread across the agar plate and it is needed to incubate at 37˚C for 24 hours. The petri dish was incubated in an inverted position as to decrease the rate of evaporation hence results in proper microbial growth. The optimum incubation for bacteria growth is 37˚C for 24 hours for their proper growth. After 24 hours, this agar was observed either the colonies of bacteria present or not. If not, the incubation plate was extended another 24 hours to allow slower growth of bacteria. By referring Figure 2, it is an example of the bacteria growth on the plate which is *Pseudomonas aeruginosa* bacterium.

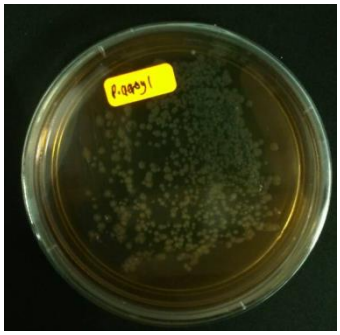

Figure 2: Bacteria grow on the plate (Pseudomonas aeruginosa bacterium)

C. Sampling Procedure

The sampling procedure is collected after incubation within the range 18 to 24 hours. The electronic nose uses a static headspace technique for sampling process. It will sniff the samples without changing the composition and properties of the bacteria. There are three stages of sampling process: the baseline stage (purge cycle), the response stage (sniff cycle) and the recovery stage. For the "Purge Cycle", ambient air supplies the chamber inlet to clean the sensors. After completion of the "Purge Cycle", the system will be set to an idle state to enable the sensors to return to their baseline values. The "Sniff Cycle" follows, and the indoor air sample was supplied to sensor arrays through the chamber inlet. Figure 1 shows the electronic nose setup for the sampling procedure.

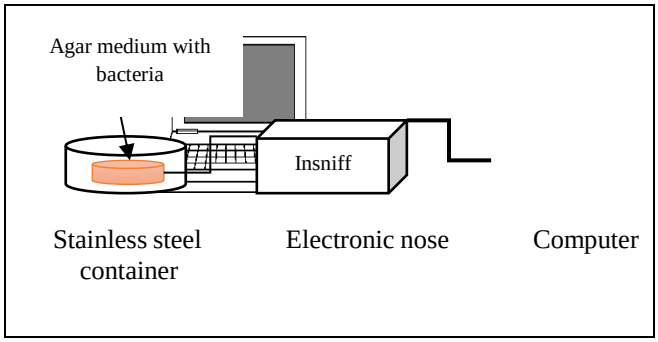

Figure 3: Electronic nose setup for sampling procedure

For data collection, each sample of bacteria and also control agar need to sniff 5 times. So, the data collected for each type of samples is 1000 data as shown in Table 1.

Table 1
Bacteria Species and Collection of Data

Bacteria species	Time of sniff	Data collected
Pseudomonas aeruginosa	5 times	1000
Shigella	5 times	1000
Control (blank agar)	5 times	1000

IV. Result And Discussion

A. Principal Component Analysis (PCA)

The electronic nose data were processed off-line using MATLAB software version 2015. Principal component analysis (PCA) graph has been plotted by using MATLAB R2015a software as to study and obtain the important volatile compounds that involve in differentiating among samples.

Since, the main objective of this research is to evaluate the relationship between different bacteria species in nutrient agar media. Experimentally, control (blank) agar tend to produce odor itself after undergoing incubation as it is affected by surrounding temperature. Thus, PCA helps to classify the three classes into a group of the cluster and to identify the volatile compounds that produced from the samples.

Figure 2 shows PCA plot of both bacteria sample (*Pseudomonas aeruginosa* and *Shigella*) in nutrient agar (referred as NA) where *Pseudomonas aeruginosa* bacteria (blue label), *Shigella* bacteria (black label) and blank Nutrient agar (red label). It is shown a scatter plot of a principal component represents about 98.61% of total variance in two different groups of agar. The graph shows clear separation of different bacteria species.

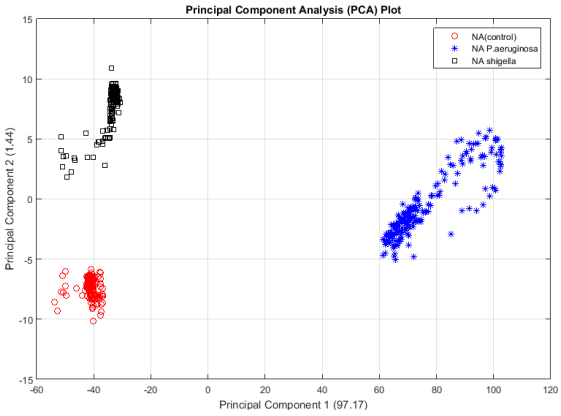

Figure 4: PCA plot for *Pseudomonas aeruginosa* and *Shigella* bacterium in nutrient agar media (refer)

V. Classification Using K-Nearest Neighbor Method (KNN)

Data classification method is used by K-Nearest Neighbor method (KNN) because it is a non-parametric classifier. The results are shown by the confusion matrix for each bacteria. The data are also trained to be able to show the possibility of testing the data. Also, Euclidean data is used to determine the possible k in classification.

The confusion matrix in Table 2 consists of training matrix and also the test matrix. The samples were divided into two groups, 70% for the training and the other 30% for the testing. Based on the table, the train is done to exert the potential of the method to be classified. From the sample used, the training matrix showed a high percentage overall which meant the possibility of the data is used for classification. The classification rate of bacteria using electronic nose showed a percentage that can be understood that it is 99.0% and 56.7% for testing data in its class.

Table 2
Matrix Train and Test of K-Nearest Neighbour (KNN) for the Samples

Sample	Matrix- Train			(%)	Matrix-Test			(%)
Shigella bacterium	70	0	0	100	15	0	0	50
Pseudomonas aeruginosa bacterium	0	68	0	97.2	0	30	0	100
Control agar	0	0	70	100	0	0	6	20

In data analysis, it is necessary to understand the clustering method chosen. In this case, it would be Principal Component Analysis or PCA. Before doing so, it is needed to establish smoother data by using the Savitzky-Golay filter. After smoothing the data, the data collected must be subtracted from the ambiance to pull out the pure data of air present in the room. It is a must in any location as it is one of the important steps before clustering process. Firstly, the percentage of PC1 and PC2 can determine the dimensional plot that is going to be used. Above 80%, the 2D plot can be used. So, by obtaining the result of the PCA, data is reducing dimensionally and the clustering of data is possible. The possibility of clustering gain can help in classifying the data.

VI. Conclusion

This study focused on the ability of the electronic nose to discriminate different bacteria commonly found in the human stomach. From the result obtained, the electronic nose can be proved that it is able to cluster different types of bacteria. The overall results presented show that the bacteria can be accurately discriminated using PCA and classifier such as KNN. The PCA method was chosen because multiple collections of data can be obtained and interpreted. The PCA is also able to represent data in a simpler and reduced form. For the data classification, the K-Nearest Neighbor (KNN) method is also simpler in order to classify the clusters of data. The method can be label as the simplest machine learning method for classification of data. Also, the method is supported by the flexible and simple incorporate different data types.

Acknowledgment

This work is supported by the Fundamental Research Grant Scheme (FRGS) grant 9003-00512 by the Ministry of Higher Education, Malaysia. The author would like to thank all the research fellows involved in this project especially from School of Mechatronic Engineering, Universiti Malaysia Perlis (UniMAP), Arau, Malaysia.

References

[1] R. M. S. Thorn, D. M. Reynolds, and J. Greenman, "Multivariate analysis of bacterial volatile compound profiles for discrimination between selected species and strains in vitro," J. Microbiol. Methods, vol. 84, no. 2, pp. 258–264, 2011.

[2] M. Sohrabi, L. Zhang, K. Zhang, A. Ahmetagic, and M. Q. Wei, "Volatile Organic Compounds as Novel Markers for the Detection of Bacterial Infections."

[3] N. M. Zetola, C. Modongo, K. Mathlagela, E. Sepako, O. Matsiri, T. Tamuhla, B. Mbongwe, E. Martinelli, G. Sirugo, R. Paolesse, and C. Di Natale, "Identification of a large pool of microorganisms with an array of porphyrin based gas sensors," Sensors (Switzerland), vol. 16, no. 4, pp. 1–14, 2016.

[4] M. Ghasemi-Varnamkhasti, S. S. Mohtasebi, M. Siadat, and S. Balasubramanian, "Meat quality assessment by electronic nose (Machine Olfaction Technology)," Sensors, vol. 9, no. 8, pp. 6058–6083, 2009.

[5] M. Trincavelli, S. Coradeschi, A. Loutfi, B. Soderquist, and P. Thunberg, "Direct identification of bacteria in blood culture samples using an electronic nose," IEEE Trans. Biomed. Eng., 2010.

[6] N. Yusuf, M. I. Omar, and A. Zakaria, "Diagnosis of Bacteria for Diabetic Foot Infection using Electronic Nose Technology," pp. 114–118, 2013.

[7] A. D. Wilson, "Advances in Electronic - Nose Technologies for the Detection of Volatile Biomarker Metabolites in the Human Breath," 2015.

[8] L. Zhang, F. Tian, and G. Pei, "A novel sensor selection using pattern recognition in electronic nose," Measurement, vol. 54, pp. 31–39, 2014.
[9] M. B. Banerjee and T. Nandy, "Multivariate Preprocessing Techniques Towards Optimising Response of Fused Sensor from Electronic Nose and Electronic Tongue," pp. 949–954, 2016.
[10] M. Sharma, "International Journal of Advanced Research in Computer Science and Software Engineering Generalized K-Nearest Neighbour Algorithm- A Predicting Tool," vol. 3, no. 11, pp. 1–4, 2013.