

Using artificial neural network to predict a variety of pathogenic microorganisms

Yu-Hsuan Liao¹, Yu-Ning Yu¹, Maysam F. Abbod², Chung-Hung Shih^{3,4}, Jiann-Shing Shieh¹

¹ Department of Mechanical Engineering, Yuan Ze University, Chungli 32003, Taiwan; comimic@gmail.com; wangzhchuang@126.com; jsshieh@saturn.yzu.edu.tw

² College of Engineering, Design and Physical Sciences, Brunel University London, Uxbridge UB8 3PH, United Kingdom; Maysam.Abbod@brunel.ac.uk

³ Division of Thoracic Medicine, Department of Internal Medicine, School of Medicine, College of Medicine, Taipei Medical University.

⁴ Division of Pulmonary Medicine, Department of Internal Medicine, Taipei Medical University Hospital, Taiwan; chshih43@tmu.edu.tw

* Correspondence: jsshieh@saturn.yzu.edu.tw; Tel.: +886-3-4648800 ext. 2470

*Correspondence: chshih43@tmu.edu.tw

Division of Pulmonary Medicine, Department of Internal Medicine, Taipei Medical University Hospital, Taiwan;

Abstract— **In this study, an electronic nose is used to record breathing data from healthy patients and pneumonia patients. The electronic nose records resistance data using a micro-array of 11 sensors. The recorded data is fed to Artificial Neural Network (ANN) which is used to train a model for detection of infections. The ANN has performed accurate classification which can be further developed to an online efficient pneumonia detection system. Initially, five patients' data are used to build up the ANN model. Then, another two patients' data are used to test the accuracy of the model. The results show that the model predicts the same outcome that is diagnosed by Taipei Medical University Hospital where samples have to be cultured to identify whether patients have pneumonia or not. In this preliminary study, ANN has achieved good results which can be further developed to an online efficient pneumonia detection system in the near future.**

Index terms- **Pneumonia, Artificial neural network, Electric nose**

1. Introduction

Intensive care unit (ICU) is an important part of medical treatment in the hospital, and it is also the place which can save critical patients. Generally, these patients have poor survival rate expect except when getting the true diagnosis. In order to reduce mortality rate, patients have to receive many invasive examinations and treatment. Although these methods can keep patient alive, it is also a way for bacteria invading into patient body. Nowadays, many patients are easily getting ventilator-associated pneumonia (VAP). Mortality rate of VAP [1] lies between 20% and 60% [2-3] and can be even higher. The whole detecting process takes at least five days for culture and identify which species of bacteria causing the diseases. During the period, medical doctor do not treat the patients without diagnosis. Unfortunately some patients might pass away without true diagnosis pending the period of laboratory outcome. By this time, the bacteria keeps multiplying and increase the difficulty of the treatment. Therefore, changing standard operating procedure into online pneumonia detection, without drawing blood, taking chest X-ray and sputum culture, is more efficient and safe.

According to previous study [4], nitrogen (N_2), oxygen (O_2), water vapor and carbon dioxide (CO_2) can be produced from breathing. There are almost 1259 kinds of volatile organic compounds (VOCs) [5-7] can be produced after the process of human

metabolism activities. These VOCs circulate through the body in the blood. It is usually unclear if the empirical antibiotics are effective or not in the first five days of admission. A fast detection system can be developed using a sensor system that detects bacterial species. The sensor data can be analyzed for identifying whether the patient has pneumonia or not without a need for the period of culture developments which can let physician select a suitable antibiotics in a short time and reduce the mortality rate.

At first, identifying whether the patient has pneumonia or not is the first step in this study. Although the culture of the phlegm and blood can identify the species of bacteria, it still need to take at least five days to know the kinds of bacteria and the consequence of the medication allergy testing. This period of cultivating is very crucial to patients, while patients should also take treatment in this period. However, there are dozens of bacteria can cause to pneumonia. Therefore, it only can depend on the judgments of doctors to treat the patients when cultivating bacteria.

Due to the fact that bacteria would produce VOC (volatile organic compounds) when metabolizing, hence it is important to analyze the spectrum of pneumonia cultured bacteria and collect VOC from exhaled breathing gas. Using 11 sensors electronic nose [8] can provide a variety of resistance from exhaled volatile compounds. Therefore, using the 11 resistances as input and the cultured bacteria type identified in the hospitals as output, can establish an ANN model [9-11] which can predict whether the patient has

pneumonia or not in real time.

2. The pneumonia detection system

The developments of pneumonia detection system need to be supported by large amount of data. Data from the 11 sensors, and pneumonia VOC information have been transferred into numeral data that represent the resistance value and VOC relationship. Comparing the varieties from each sensor, data should be selected carefully before training the ANN [12-13]. Further details on how to build the system are given in the following sections.

2.1 Sensors

In this study, CHS430 (a kind of electric nose which is used to collect the breathing data in the study) is used to gather bacteria data by detecting patients exhaled air as shown in Fig. 1. CHS430 is a kind of electric nose [14-16] that manufactured by Taiwan Carbon Nanotube Technology Corporation (TCNT). CHS430 consists 11 sensors, each sensor is used to identify different gases. The CHS430 is designed to identify whether the patient has infection pneumonia or not. Internal air sampling pump and advanced pattern recognition algorithm to detect and recognize the chemical vapors requires carefully designed algorithm in order to get the best sensors performance. Data can be translated to numerical resistance format in order to be used to design the detection algorithm. Fig. 1 shows the graphical user interface which provide

information to the user about the collected breathing data, also it can show the reflection pattern of each sensor array.

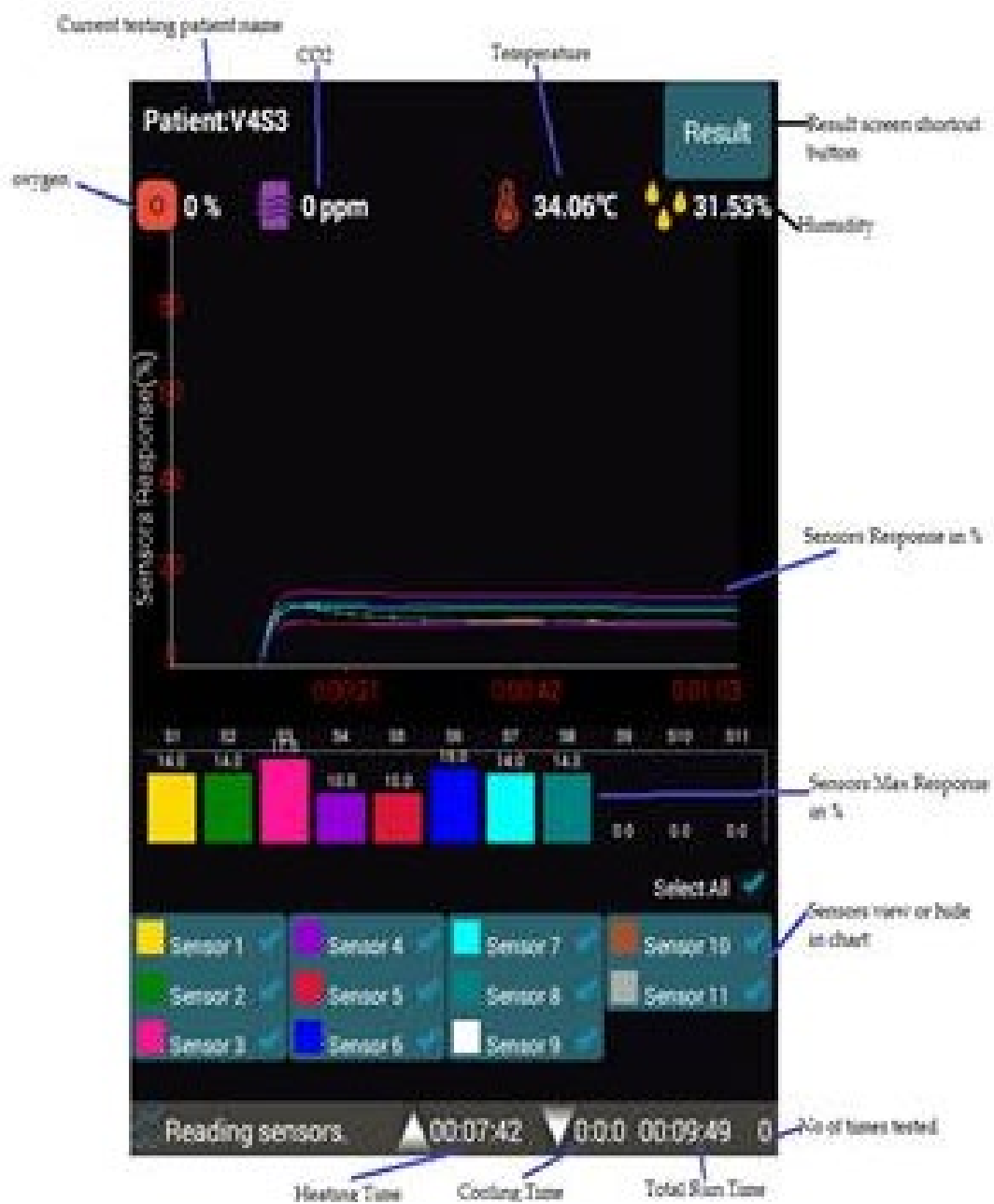


Fig. 1 The reflection pattern of each sensor array

2.2 Data Source

In this study, electronic nose is used to collect the exhaled breathing data. These breathing data are collected from patients who have already infected pneumonia in the first day. The electronic nose collects the breathing data from seven patients who are staying in the Taipei Medical University Hospital (TMUH). While the electronic nose records the breathing data, the process was divided into several stages and labelled as flags which are Flag 0: Purging sample chamber prior to measuring baseline readings, Flag 1: Drawing sample into sample chamber, Flag 2: Resting phase between sample draw and purge, Flag 3: Purging sample chamber by pumping through sample intake hose, and Flag 4: Turn off the power.

2.3 Data pre-processing

For selecting data, an ideal phase of flags 1 and 2 has been selected for training. The steps of recording the breathing data by the electronic nose are Flags 0-4. The electronic nose recorded one reading every ten minutes. Since the resistance values are used to identify the status of the gas, Flags 1 and 2 are the most important processes that shows the changing of the resistance values and saturation, as shown in Fig. 2. Therefore, data from Flag 1 and Flag 2 are used for the analysis in this study. Then bacteria data had all be labeled as infection and the rest non-infection patient data were

labeled as non-infection using binary system. After testing the accuracy of the model, the data has been normalized. For assessing the sensitivity of the sensors, the mean absolute error (MAE) between infection patients and non-infection patients are calculated.

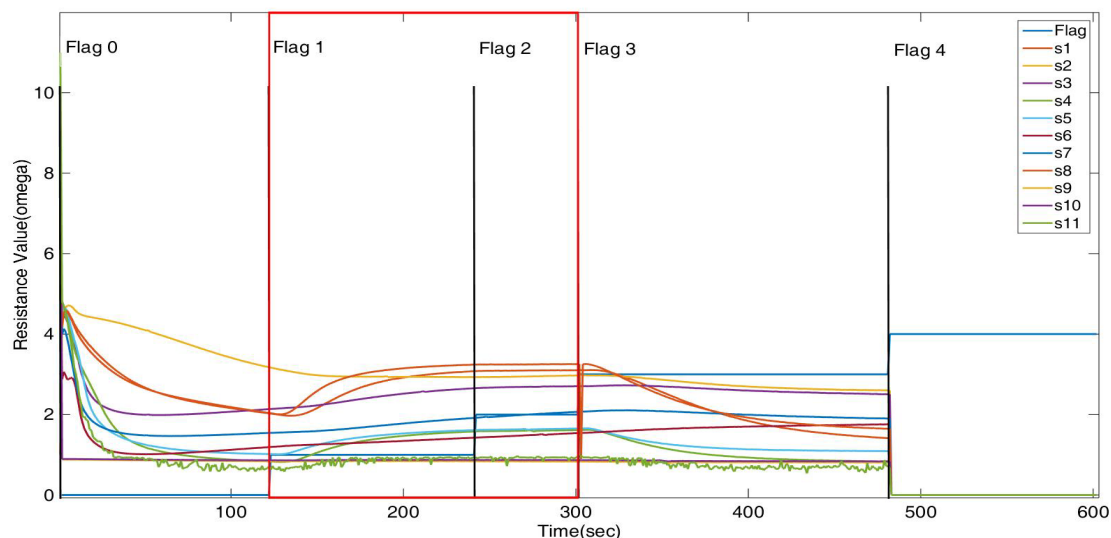


Fig. 2 Resistance respond curve of infection patients' data.

2.4 Sensitivity analysis of sensors

There are seven patients' data which are collected by the electronic nose. The study had full prior approval by an institutional review board and written informed consent was obtained from all the patients. At first, the sensitivity of the sensors to be investigated, hence the MAE between infection patients and non-infection patients is calculated. Fig. 3 shows the relationship with infection patients and non-infection patients. From the result, it can be seen that sensors 9 and 10 do not have any reflection between infection and non-infection. Hence these two sensors were taken off, and the

model is built using the other nine sensors.

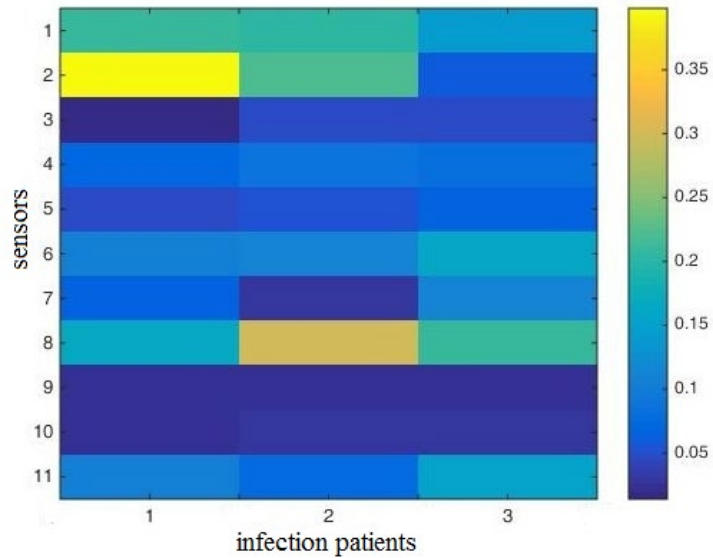


Fig. 3 MAE between infection patients and first non-infection patient

3. Machine Learning

After performing pre-processing of the data, the data has all been labeled as infection and the rest non-infection patient data are labeled as non-infection. In this study, patients who are non-infection have been labelled as 0, and those who have been identified as infection are labelled as 1. Then, these resistance values from nine sensors of VOCs are input values in ANN's training.

At the beginning, five patients' data (two patients are with infection and three patients are without infection) are used to build up the ANN model, results from the hospital laboratory test were used as the target for the ANN model. The other 2 patient's data out of the 7 are used for testing.

4. Results

The model is created using ANN toolbox within MATLAB. At the beginning, one hidden layer with random neurons have been set. Fig. 4 shows the training error for the 5 patient's data. In the figure the blue dots show the error while the green line shows the separation between Flags 1 and 2. Three additional red lines are standing for 0.8, 0.5 and 0.2, respectively. Since Flag 2 is more stable than Flag 1, reference to Flag 2 is made to identify whether the patient is infection or non-infection. Furthermore, in the figure, patients 1 and 3 are lower than 0.2 after the green line, while patient 2 is close to 0.2. For these three patients, it can judged that the patients do not have pneumonia. On the other hand, patients 4 and 5 are higher than 0.8 with respect to Flag 2, hence these two patients have been infection.

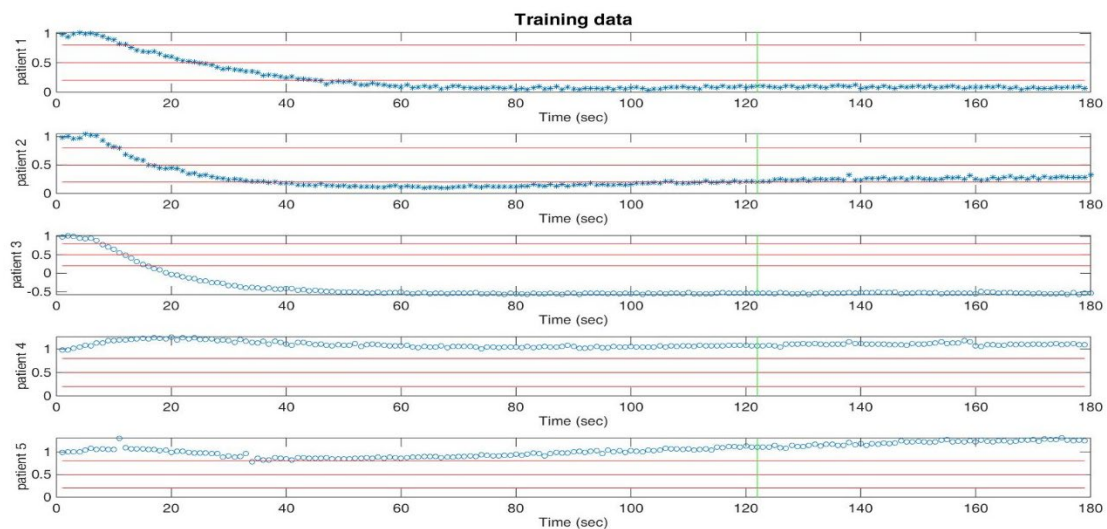


Fig. 4 The accurately of the model

After testing training ANN model another two patients' data (one is non-infection patient, and the other is infection patient) are used to test the accuracy of the model. Fig. 5 shows the testing data results, the predict value of the first patient (i.e., patient 6) are less than 0.2, hence this patient is clearly non-infection case. However, the predicted values of 2nd patient (i.e., patient 7) show this patient has been infected with pneumonia. The results show that the model predicts the same result which were obtained from TMUH. After the green line, Flag 2 is quite stable to know whether the patient have pneumonia or not.

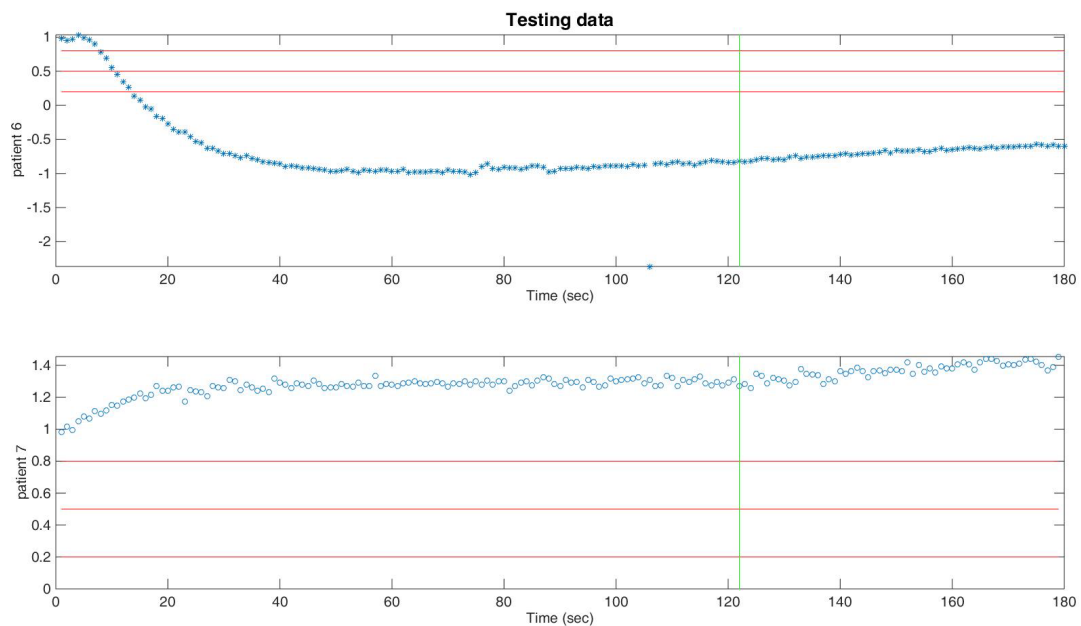


Fig. 5 The results of testing data

In order to make sure about the consistency of these data, cross validation is performed to prove it. This preliminary study only records seven patients' breathing

data, there are four patients who are non-infected, and three patients who are infected.

In each model, two cases are used (one is non-infection, the other one is infection) to test the accuracy of the model, as a result, three cross validation models were generated.

Table 1 shows the result of each model for the same database. Since Flag 2 is more stable than Flag 1, the average from Flag 2 is calculated to identify whether the patient is infected or non-infected. In these results, values larger than 0.8 means the patient has infection; whereas values less than 0.2 means the patient has no infection. Table 1 shows that each model has achieved good performance. The accuracy of each model is 100%.

Table 1 The average values which are calculated using Flag 2.

		model 1	model 2	model 3
non-infection	Patient 1	-0.69	0.16	0.004
infection	Patient 2	1.37	1.05	1.03

5. Conclusion and Future Work

In this study, electronic nose is used to collect patients' breathing data. According to the characteristic of substance when patient breathing, sensor data were used to identify whether the patient have infection VAP or not infection using ANN which has demonstrated good prediction accuracy.

For future works, the study attempts to increase patients' data. The proposed algorithm needs to be comprehensively evaluated in a wider database. This will help to enhance the proposed algorithm for efficiency purposes. Currently, this study suggests that model which is built by machine learning has good performance and accuracy. Further development and more evaluation are required by increasing the database to make sure that the model has good performance for detecting different types of bacteria.

Acknowledgment

This research is financially supported by the Taiwan Carbon Nanotube Technology Corporation.

References

- [1] J. Chastre and J.-Y. Fagon, Ventilator-associated pneumonia, *American journal of respiratory and critical care medicine* 165 (2002) 867-903.
- [2] J.-Y. Fagon, J. Chastre, A. J. Hance, P. Montravers, A. Novara, and C. Gibert, Nosocomial pneumonia in ventilated patients: a cohort study evaluating

- attributable mortality and hospital stay, *The American journal of medicine* 94 (1993) 281-288.
- [3] N. Bercault and T. Boulain, Mortality rate attributable to ventilator-associated nosocomial pneumonia in an adult intensive care unit: a prospective case-control study, *Critical care medicine* 29 (2001) 2303-2309.
- [4] J. N. Labows, K. J. McGinley, G. Webster, and J. Leyden, Headspace analysis of volatile metabolites of *Pseudomonas aeruginosa* and related species by gas chromatography-mass spectrometry, *Journal of Clinical Microbiology* 12 (1980) 521-526.
- [5] M. Shirasu and K. Touhara, The scent of disease: volatile organic compounds of the human body related to disease and disorder, *Journal of biochemistry* 150 (2011) 257-266.
- [6] A. W. Boots, J. J. van Berkel, J. W. Dallinga, A. Smolinska, E. F. Wouters, and F. J. van Schooten, The versatile use of exhaled volatile organic compounds in human health and disease, *Journal of breath research* 6 (2012) 027108.
- [7] K. D. Van de Kant, L. J. van der Sande, Q. Jöbsis, O. C. van Schayck, and E. Dompeling, Clinical use of exhaled volatile organic compounds in pulmonary diseases: a systematic, *Respiratory research* 13 (2012) 1.
- [8] J. B. Chang and V. Subramanian, Electronic noses sniff success, *Spectrum, IEEE* 45 (2008) 50-56.
- [9] M. F. Abbod, K.-Y. Cheng, X.-R. Cui, S.-J. Huang, Y.-Y. Han, and J.-S. Shieh, Ensembled neural networks for brain death prediction for patients with severe head injury, *Biomedical Signal Processing and Control* 6 (2011) 414-421.
- [10] X. Cui, C.-W. Lin, M. F. Abbod, Q. Liu, and J.-S. Shieh, Diffuse large B-cell lymphoma classification using linguistic analysis and ensembled artificial neural networks, *Journal of the Taiwan Institute of Chemical Engineers* 43 (2012) 15-23.
- [11] X. Cui, M. Abbod, Q. Liu, J.-S. Shieh, T. Chao, C. Hsieh, et al., Ensembled

artificial neural networks to predict the fitness score for body composition analysis, *The journal of nutrition, health & aging* 15 (2011) 341-348.

- [12] M. F. Abbod, K.-Y. Cheng, X.-R. Cui, S.-J. Huang, Y.-Y. Han, and J.-S. Shieh, Ensembled neural networks for brain death prediction for patients with severe head injury, *Biomedical Signal Processing and Control* 6 (2011) 414-421.
- [13] X. Cui, C.-W. Lin, M. F. Abbod, Q. Liu, and J.-S. Shieh, Diffuse large B-cell lymphoma classification using linguistic analysis and ensembled artificial neural networks, *Journal of the Taiwan Institute of Chemical Engineers* 43 (2012) 15-23.
- [14] J. W. Gardner and P. N. Bartlett, A brief history of electronic noses, *Sensors and Actuators B: Chemical* 18 (1994) 210-211.
- [15] S. Ferreira, A. Dos Santos, G. De Souza, and W. Polito, Analysis of the emissions of volatile organic compounds from the compression ignition engine fueled by diesel–biodiesel blend and diesel oil using gas chromatography, *Energy* 33 (2008) 1801-1806.
- [16] A. Jonsson, F. Winqvist, J. Schnürer, H. Sundgren, and I. Lundström, Electronic nose for microbial quality classification of grains, *International Journal of Food Microbiology* 35 (1997) 187-193.