Using Artificial Neural Network
to Predict a Variety of Pathogenic Microorganisms

Yu-Hsuan Liao,[1] Yu-Ning Yu,[2] Maysam F. Abbod,[3]
Chung-Hung Shih,[4,5*] and Jiann-Shing Shieh[1**]

[1]Department of Mechanical Engineering, Yuan Ze University, Chungli 32003, Taiwan
[2]Department of Computer Science, California State University, Dominguez Hill, Carson, CA 90746, US
[3]College of Engineering, Design and Physical Sciences, Brunel University London,
Uxbridge UB8 3PH, United Kingdom
[4]Division of Thoracic Medicine, Department of Internal Medicine, School of Medicine, College of Medicine,
Taipei Medical University, 11031, Taiwan
[5]Division of Pulmonary Medicine, Department of Internal Medicine,
Taipei Medical University Hospital, 11031, Taiwan

In this study, an electronic nose is used to record breathing data from healthy and pneumonia patients. The electronic nose records resistance data using a microarray of 11 sensors made of a metal oxide semiconductor. The recorded data are fed to an artificial neural network (ANN), which is used to train a model for the detection of infections. Initially, five patients' data are used to construct the ANN model. Then, another two patients' data are used to test the accuracy of the model. In this preliminary study, the ANN achieved good results, showing that it can be further developed into an efficient online pneumonia detection system in the near future.

1. Introduction

The intensive care unit (ICU) is an important part of medical treatment in hospitals and can also save the lives of critical patients. Generally, these patients have a poor survival rate except when they are correctly diagnosed. To reduce the mortality rate, patients have to receive many invasive examinations and treatments. Although these methods can keep a patient alive, they also allow bacteria to invade the patient's body. Nowadays, many patients easily get ventilator-associated pneumonia (VAP). The mortality rate of VAP[1] lies between 20 and 60%[2,3] and can be even higher. The whole detection process takes at least five days and requires the culture and identification of species of bacteria causing the disease. During this period, medical doctors cannot treat the patient owing to the lack of a diagnosis. Unfortunately, some patients might die without a true diagnosis while waiting for the laboratory outcome. Even for patients who survive, the bacteria continuously multiply during the waiting period, increasing the difficulty in treatment. Therefore, changing the standard operating procedure into online

pneumonia detection, without drawing blood or taking a chest X-ray or sputum culture, is more efficient and safer.

In general, nitrogen (N_2), oxygen (O_2), water vapor, and carbon dioxide (CO_2) are produced from breathing. A total of 1259 volatile organic compounds (VOCs)[4–8] can be produced by human metabolism activities. These VOCs circulate throughout the body in the blood. It is usually unclear whether empirical antibiotics are effective 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 to identify whether the patient has pneumonia without the need for a period of culture development, which can allow a physician to select suitable antibiotics in a short time and reduce the mortality rate.

Identifying whether a patient has pneumonia is the first step in this study. Although a culture of phlegm or blood can be used to identify the species of bacteria, at least five days are required to identify the type of bacterium and the consequences of medication allergy testing. This period of cultivation is crucial to patients, and patients should also be treated during this period. However, twenty bacteria can cause pneumonia.[9] Therefore, only the judgments of doctors can be used to treat patients when cultivating bacteria.

Because bacteria produce VOCs when metabolizing, it is important to analyze the spectrum of cultured pneumonia bacteria and collect VOCs from exhaled breathing gas. An electronic nose with 11 sensors[10] can provide a variety of resistances from exhaled VOCs. Using the 11 resistances as the input and the types of cultured bacteria identified in hospitals as the output, an artificial neural network (ANN) model[11–14] can be established. Therefore, the main goal of this study is to collect patients' breath gases data and build an ANN model to predict whether a patient has pneumonia in real time.

2.   Pneumonia Detection System

The development of a pneumonia detection system must be supported by a large amount of data. Data from 11 metal oxide semiconductor (MOS)-type sensors with materials such as tin oxide[15–17] and tungsten trioxide, and pneumonia VOC information have been converted into numerical data that represent the relationship between the resistance and the VOC. By comparing the varied resistance from each sensor, data should be selected carefully before training an ANN. Further details on how to build the system are given in the following sections.

2.1   Sensors

In this study, a CHS430 electronic nose [Taiwan Carbon Nanotube Technology Corporation (TCNT)][18–20] is used to collect breathing data to obtain bacteria data through the detection of patients' exhaled air. It consists of 11 MOS-type sensors, each of which is used to identify specific gases in real time. Typical responses of the sensors are shown in Fig. 1. The electronic nose is designed to identify whether a patient has pneumonia. The internal air sampling pump and advanced pattern recognition algorithm used to detect and recognize the chemical vapors
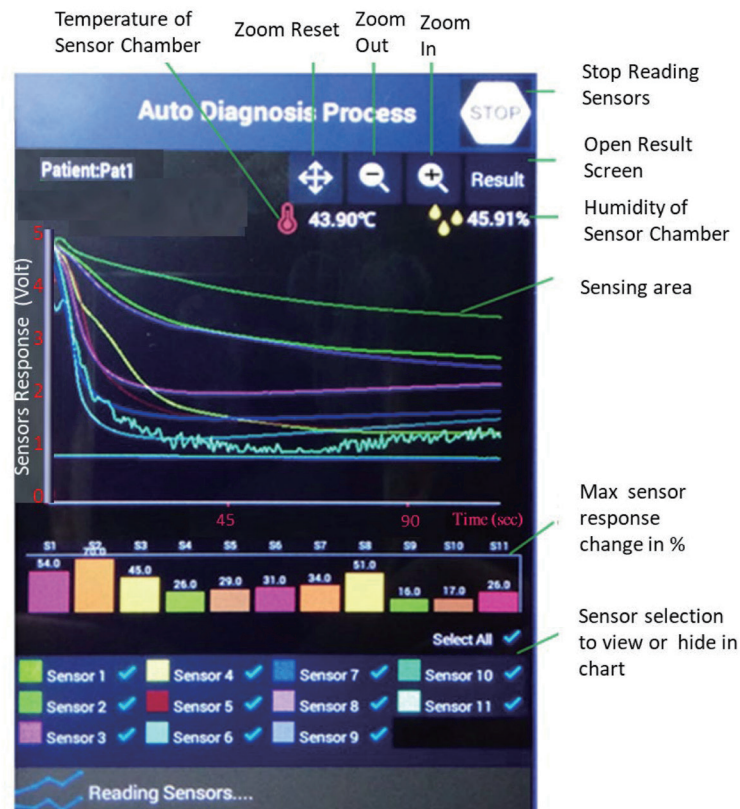
Fig. 1.　　(Color online) Reflection pattern of each sensor in array.

require careful design to maximize the sensor performance. Data are translated to numerical resistances that can be used to design the detection algorithm. Figure 1 shows the graphical user interface, which provides information to the user about the collected breathing data. It also shows the reflection pattern of each sensor array.

2.2　Data source

In this study, an electronic nose is used to collect the exhaled breathing data from patients who have already been infected with pneumonia on their first day at Taipei Medical University Hospital (TMUH). The study had full prior approval by an institutional review board and written informed consent was obtained from all the patients. The process of recording breathing data by the electronic nose was divided into several stages, as shown in Table 1, by controlling the sequence of pumps, solenoid valves, and sensors. Also, the process was labeled with the following flags: Flag 0: purging sample chamber prior to measuring baseline readings; Flag 1: drawing sample into sample chamber, Flag 2: resting phase between sample drawing and purging; Flag 3: purging sample chamber by pumping through sample intake hose; and Flag 4: turning off the power.

Table 1
Timing control sequence of pumps, solenoid valves, and sensors.

	Flag 0	Flag 1	Flag 2	Flag 3	Flag 4
Pump-2	ON	ON	OFF	ON	OFF
Pump-1	ON	ON	ON	OFF	OFF
Valve	OFF	ON	ON	OFF	OFF
Sensor	ON	ON	ON	ON	OFF

2.3 Data preprocessing

The steps of recording the breathing data by the electronic nose are labeled Flags 0–4. The electronic nose recorded one reading every 10 min. Since the resistances are used to identify the status of the gas, Flags 1 and 2 are the most important processes, which show the changes in resistance and saturation, as shown in Fig. 2. Data from Flags 1 and 2 are used for the analysis in this study. Therefore, for selecting data, Flags 1 and 2 were selected as the best sensing area of this process for ANN model training as the input data. Then, bacteria data were labeled as "infected" and "non-infected" patient data according to the type of cultured bacterium identified in hospitals using a binary system as the output of ANN model training. After testing the accuracy of the model, the data were normalized. To assess the sensitivity of the sensors, the mean absolute error (MAE) between infected and non-infected patients was calculated.

2.4 Sensitivity analysis of sensors

Seven patients' data were collected by the electronic nose. To determine the sensitivity of the sensors, we calculated the MAE between three infected and four non-infected patients. Figures 3(a)–3(d) show the relationship of three infected patients against each non-infected patient. It can be seen that sensors 9 and 10 do not distinguish between infected and non-infected patients. Hence, these two sensors were removed, and we constructed the model using the other nine sensors.

3. Machine Learning

After preprocessing the data, the varied resistances from 11 sensors of Flags 1 and 2 were selected as the input data of this process for ANN model training. Then, bacteria data according to the type of cultured bacterium identified in hospitals were labeled as infected and non-infected patient data using a binary system as the output of ANN model training. In this study, 0 was assigned to non-infected patients and 1 was assigned to infected patients. The normalization scale was applied to testing data. Then, these resistances from 11 sensors of patients' breath VOC gases were used as input values in the ANN's training, but here, we only used nine sensors because two sensors did not have any MAE differences between infection and non-infection according to Figs. 3(a)–3(d) results. The basic arithmetic unit of the ANN is used by artificial neurons of the structure of biological neurons imitated by the human brain. Millions of these biological neurons are connected by links and they interact with each other.
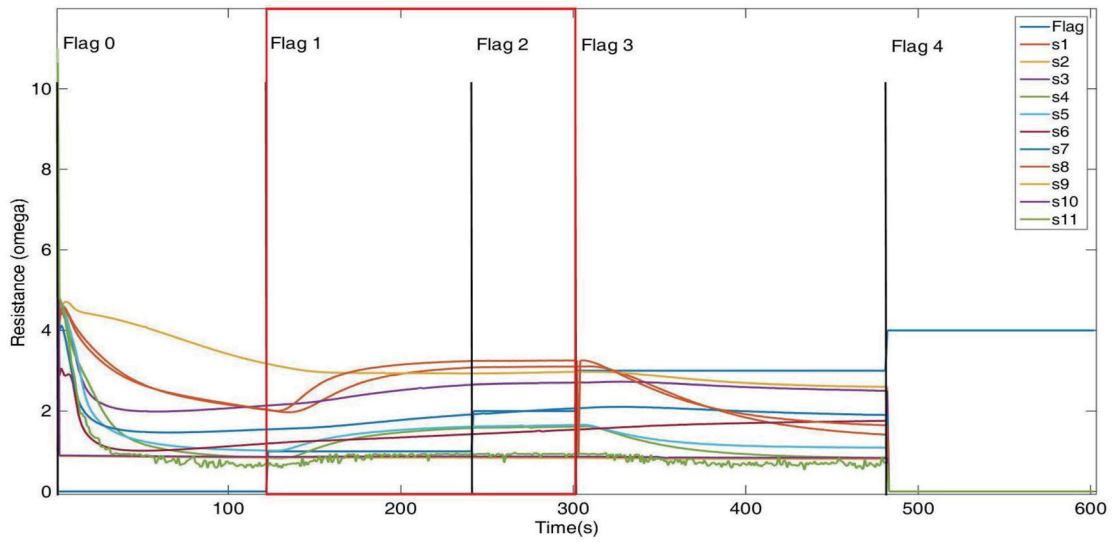
Fig. 2.   (Color online) Resistance response curves of infected patients' data.
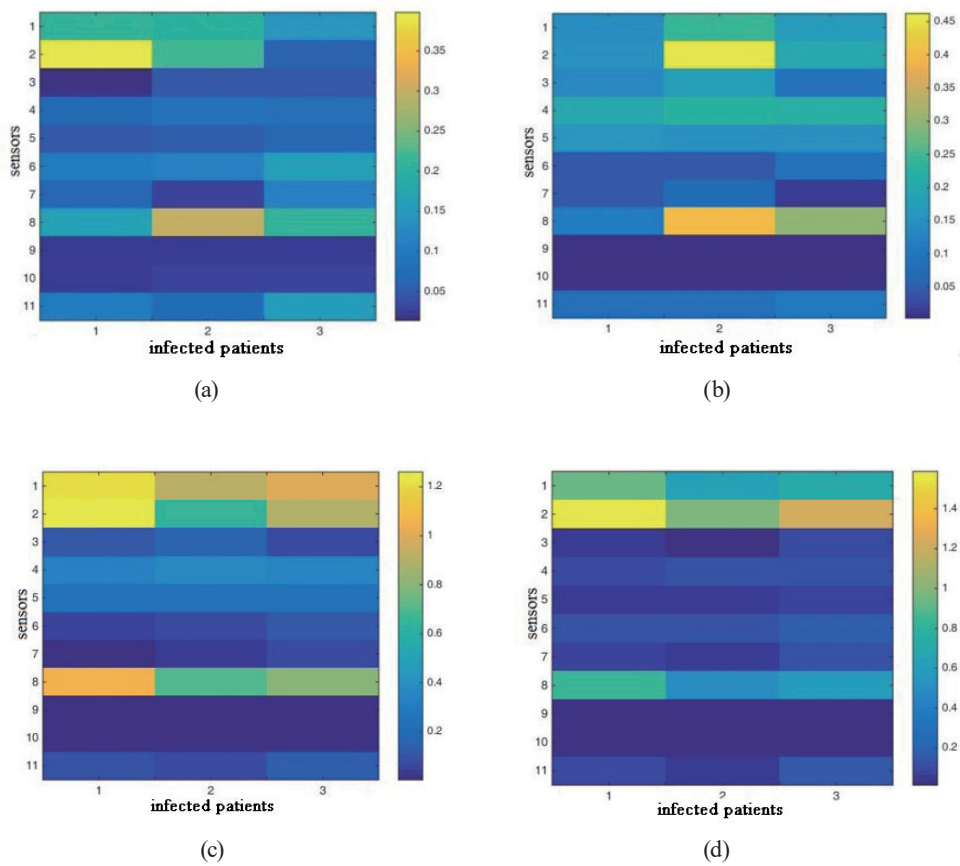

Fig. 3.   (Color online) MAE between three infected and four non-infected patients (a) against first non-infected patient, (b) against second non-infected patient, (c) against third non-infected patient, and (d) against fourth non-infected patient.

Also, the neurons can take input data and perform simple operations on the data. The result of these operations is passed to other neurons. Moreover, each link is associated with weight. ANNs are capable of learning due to altering weights. After the input signal is multiplied by the weight, afferent artificial neurons and the weights represent the stimulation intensity of the input signal to neurons. The output at each neuron is called its activation value. Then, these output values connect with the next layer of the ANN. By connecting layer by layer, a neural network is completed as shown in Fig. 4. In theory, an ANN is composed of an input layer, a hidden layer, and an output layer.

First, five patients' data (two infected patients and three non-infected patients) were used to train an ANN model, whereas the result of sampling phlegm and blood was the target of the ANN model as shown in Fig. 5. Hence, out of the seven cases, four cases were considered for training, one case was considered for validation, and the other two cases were considered for testing.
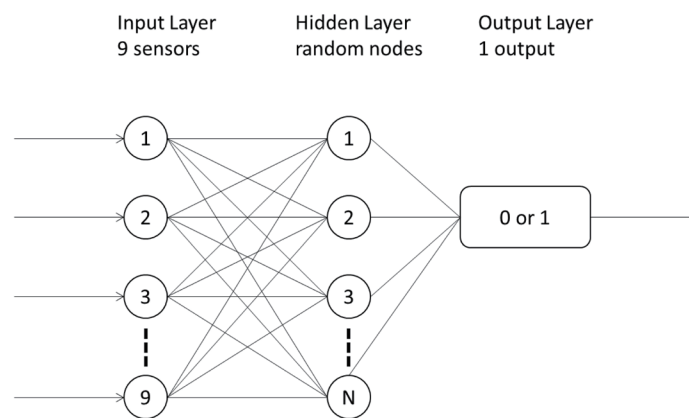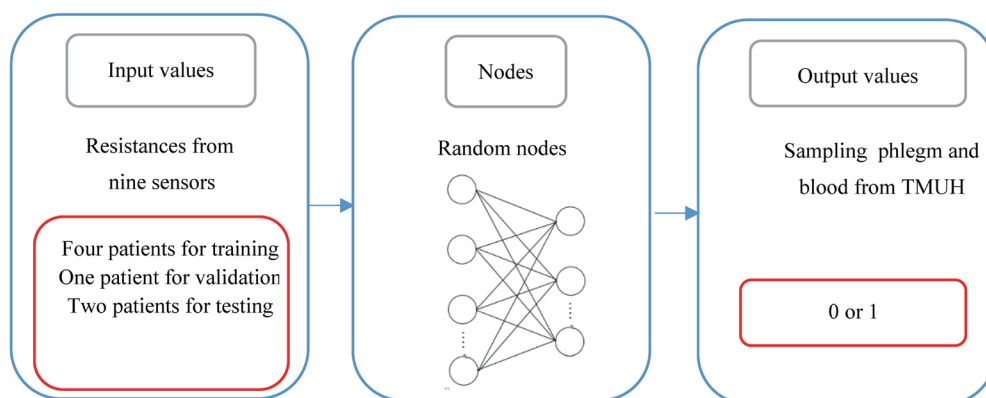

Fig. 4.   Structure of ANN.

Fig. 5.   (Color online) Structure of ANN used to predict a variety of pathogenic microorganisms.

4.   Results

The model was created using the ANN toolbox within MATLAB. At the beginning, one hidden layer with random neurons was set. Figure 6 shows the training error for the five patients' data. In the figure, the blue dots show the error, while the green line shows the separation between Flags 1 and 2. The three additional red lines stand for 0.8, 0.5, and 0.2. Since Flag 2 is more stable than Flag 1, a reference to Flag 2 is made to identify whether the patient is infected. Furthermore, Fig. 6 shows the training output values between 0 and 1 of the ANN model from patients 1 to 5; those of patients 1 and 3 are lower than 0.2 after the green line, whereas that of patient 2 is close to 0.2. For these three patients, it can be judged that they do not have pneumonia. On the other hand, those of patients 4 and 5 are higher than 0.8 with respect to Flag 2; hence, these two patients are infected.

After validation and training the ANN model, the data of two other patients (one non-infected patient and one infected patient) were used to test the accuracy of the model. Figure 7 shows the testing output values between 0 and 1 of the ANN model from patients 6 and 7. The predicted value of the first patient (i.e., patient 6) is less than 0.2; hence, this patient is clearly not infected. However, the predicted values of the second patient (i.e., patient 7) show that this patient is infected with pneumonia. The results show that the model predicts the same result as medical doctors in TMUH. After the green line in patients 6 and 7 shown in Fig. 7, Flag 2 is stable to know whether the patient has pneumonia.
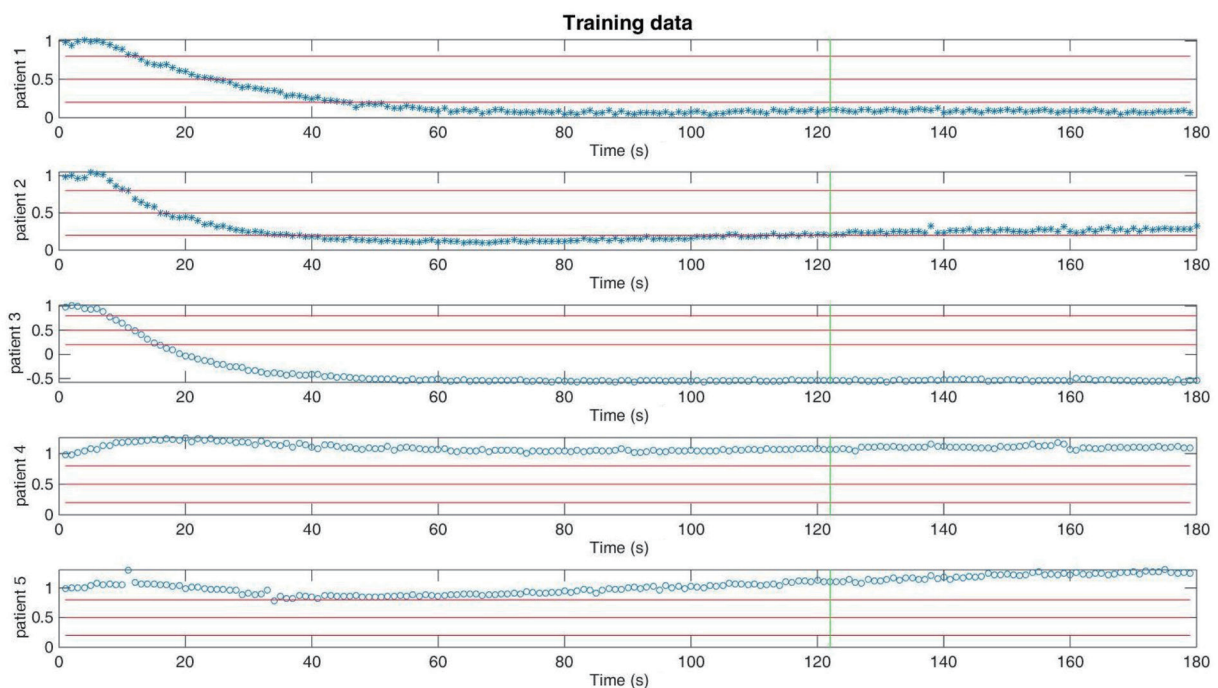

Fig. 6.   (Color online) Training output values between 0 and 1 of ANN model from 5 patients.
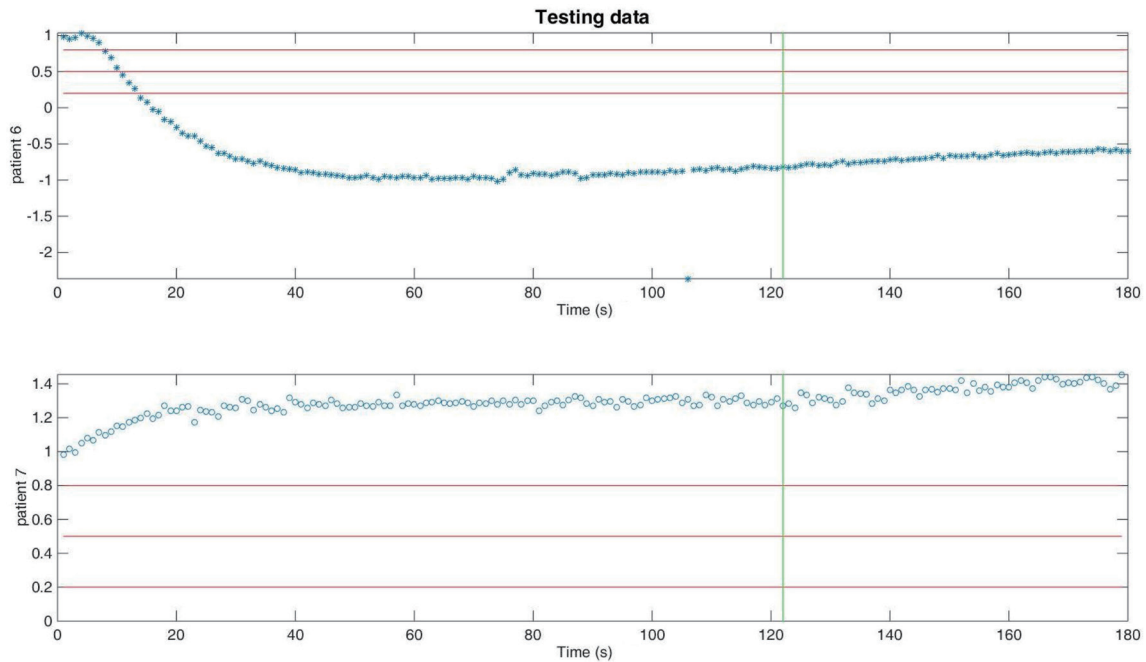
Fig. 7.   (Color online) Testing output values between 0 and 1 of ANN model from 2 patients.

Table 2
Averages calculated using Flag 2.

Analytics		Model 1	Model 2	Model 3
Non-infected	Patient 1	−0.69	0.16	0.004
Infected	Patient 2	1.37	1.05	1.03

To verify the consistency of these data, cross-validation was performed. This preliminary study only recorded seven patients' breathing data: four non-infected patients and three infected patients. In each model, two cases were used (non-infected and infected) to test the accuracy of the model; as a result, three cross-validation models were generated. Table 2 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. In these results, values larger than 0.8 mean that the patient is infected, whereas values less than 0.2 mean that the patient is not infected. Table 2 shows that each model achieved good performance for each testing data. The accuracy of each model is 100%.

5.   Conclusion

In this study, an electronic nose constructed with 11 MOS-type sensors was used to collect patients' breathing VOC gas data and build an ANN model to predict whether a patient has

pneumonia in real time. After the 5 patients were trained in this ANN model, the other 2 patients were tested in this preliminary study with good prediction accuracy. However, we will increase the amount of patients' data in the future. The proposed algorithm needs to be comprehensively evaluated on a wider database. This may help to enhance the efficiency of the proposed algorithm to ensure that the model has good performance for detecting the bacteria with which the patient is infected.

Acknowledgments

References

1 J. Chastre and J. Y. Fagon: AJRCCM. **165** (2002) 7. https://doi.org/10.1164/ajrccm.165.7.2105078
2 J. Y. Fagon, J. Chastre, A. J. Hance, P. Montravers, A. Novara, and C. Gibert: Am. J. Med. **94** (1993) 281. https://doi.org/10.10156/0002-9343(93)90060-3 .
3 N. Bercault and T. Boulain: Crit. Care Med. **29** (2001) 2303. https://doi.org/10.1097/00003246-200112000-00012
4 M. Phillips: Anal. Biochem. **247** (1997) 272. https://doi.org/10.1006/abio.1997.2069
5 C. M. Slupsky, A. Cheypesh, D. V. Chao, H. Fu, K. N. Rankin, T. J. Marrie, and P. Lacy: J. Proteome Res. **8** (2009) 3029. https://doi.org/10.1021/pr900103y
6 M. Shirasu and K. Touhara: J. Biochem. **150** (2011) 257. https://doi.org/10.1093/jb/mvr090
7 A. W. Boots, J. B. N. van Berkel, J. W. Dallinga, A. Smolinska, E. Wouters, and F. van Schooten: J. Breath Res. **6** (2012) 2. https://doi.org/10.1088/1752-7155/6/2/027108
8 K. D. G. van de Kant, L. J. T. M. van der Sande, Q. Jöbsis, O. C. P. van Schayck, and E. Dompeling: Respir. Res. **13** (2012) 117. https://doi.org/10.1186/1465-9921-13-117
9 C. L. Wu, S. C. Ku, K. Y. Yang, W. F. Fang, C. Y. Tu, C. W. Chen, K. H. Hsu, W. C. Fan, M. C. Lin, W. Chen, C. Y. Ou, and C. J. Yu: J. Formosan Med. Assoc. **112** (2013) 31. https://doi.org/10.1016/j.jfma.2011.09.028
10 J. B. Change and V. Subramanian: IEEE Spectrum. **45** (2008) 50. https://doi.org/10.1109/MSPEC.2008.4457857
11 M. F. Abbod, K. Y. Cheng, X. R. Cui, S. J. Huang, Y. Y. Han, and J. S. Shieh: Biomed. Signal Process. Control **6** (2011) 414. http://dx.doi.org/10.1016/j.bspc.2011.01.002
12 X. Cui, C. W. Lin, M. F. Abbod, Q. Liu, and J. S. Shieh: J. Taiwan Inst. Chem. Eng. **43** (2012) 15. https://doi.org/10.1016/j.jtice.2011.07.005
13 X. R. Cui, M. Abbod, Q. Liu, J. S. Shieh, T. Y. Chao, C. Y. Hsieh, and Y. C. Young: J. Nutr. Health Aging **15** (2011) 341. https://link.springer.com/article/10.1007/s12603-010-0260-1
14 Y. H. Liao, Z. C. Wang, F. G. Zhang, M. F. Abood, C. H. Shieh, and J. S. Shieh: Sensors **19** (2019) 8, 1866. https://doi.org/10.3390/s19081866
15 A. D. Lam, S. D. Prior, S. T. Shen, S. J. Young, and L. W. Ji: Engineering Innovation and Design (CRC Press, London, 2019). https://doi.org/10.1201/9780429019777
16 A. D. Lam, S. D. Prior, S. T. Shen, S. J. Young, and L. W. Ji: Smart Science, Design & Technology (CRC Press, London, 2020). https://doi.org/10.1201/9780429058127
17 Y. H. Liao, Y. J. Hsiao, Y. Nagarjuna, and J. S. Shieh: Microsyst. Technol. (2020). https://doi.org/10.1007/s00542-019-04732-5
18 J. W. Gardner, and P. N. Bartlett: Sens. Actuators, B **18** (1994) 210. https://doi.org/10.1016/0925-4005(94)87085-3
19 S. L. Ferreira, A. M. dos Santos, G. R. de Souza, and W. L. Polito: Energy **33** (2008) 1801. https://doi.org/10.1016/j.energy.2008.08.002
20 A. Jonsson, F. Winquist, J. Schnürer, H. Sundgren, and I. Lundström: Int. J. Food Microbiol. **35** (1997) 187. https://doi.org/10.1016/S0168-1605(96)01218-4

About the Authors

Yu-Hsuan Liao received his B.S. degree from the Department of Electronic Engineering, National Chin-Yi University of Technology, in 1998 and his M.S. degree from the Department of Electronic Engineering, Chien Hsin University of Science and Technology in 2007. He is currently working toward a Ph.D. degree in mechanical engineering from Yuan Ze University, Taiwan. His research interests include MEMS sensors, electronic noses, and signal processing analysis techniques. (comimic@gmail.com)

Yu-Ning Yu completed her B.S. and M.S. degrees in mechanical engineering from Yuan Ze University, Taiwan, in 2015 and 2016, respectively. Currently, she is a master student and a research assistant in computer science of California State University, Dominguez Hill. Her interests are artificial intelligence in biomedical engineering and the sensor network simulation of data transitions. (memory539@kimo.com)

Maysam Abbod received his B.Sc. degree in electronic engineering from the University of Technology, Baghdad, in 1987 and his Ph.D. degree in automatic control and systems engineering from the University of Sheffield, UK, in 1992. He is currently a reader at the Department of Electronic and Computer Engineering, Brunel University London, UK. His research interests are focused on artificial intelligence applications, particularly on biomedical systems, automation and control, diagnosis and prognosis, optimization, and scheduling. Dr. Abbod has published over 160 papers in peer-reviewed international journals. He also serves as an editor for Electronics (MDPI), Sensors (MDPI), and Engineering Applications of Artificial Intelligence (Elsevier). (Maysam.Abbod@brunel.ac.uk)

Chung-Hung Shih received his M.D. degree from Kaohsiung Medical University, Kaohsiung, Taiwan, in 1983 and his Ph.D. degree from the Graduate Institute of Clinical Medicine, National Yang-Ming University, Taipei, Taiwan, in 1990. He is currently an associate professor of medicine at Taipei Medical University. He is a certified chest physician and specializes in critical care medicine and respiratory therapy. (chshih43@tmu.edu.tw)

Jiann-Shing Shieh received his B.S. and M.S. degrees in chemical engineering from National Cheng Kung University, Taiwan, in 1983 and 1986, respectively, and his Ph.D. degree in automatic control and systems engineering from the University of Sheffield, UK, in 1995. He is currently a professor with the Department of Mechanical Engineering and a joint professor with the Graduate School of Biotechnology and Bioengineering, and also serves as the dean of the College of Engineering, Yuan Ze University, Taiwan. His research interests are focused on biomedical engineering, particularly bio-signal processing of ECG, BP, EEG, SPO2, center of pressure (COP) position signals, artificial intelligent analysis and control, medical automation, pain model and control, critical care medicine monitoring and control, dynamic cerebral autoregulation research, and brain death index research. Professor Shieh has published over 130 papers in peer-reviewed international journals. He also serves as a section editor of the Journal of Clinical Medicine and an academic editor of the Journal of Healthcare Engineering. (jsshieh@saturn.yzu.edu.tw)