TECHNICAL PAPER

Development of an E-nose system using machine learning methods to predict ventilator-associated pneumonia

Yu-Hsuan Liao[1] · Chung-Hung Shih[2,3] · Maysam F. Abbod[4] · Jiann-Shing Shieh[1] · Yu-Jen Hsiao[5]

Abstract

Currently, no device is available for the rapid screening of ventilator-associated pneumonia (VAP) at an early stage. Accordingly, we propose the design of an offline gas detection system to monitor and detect metabolites of pneumonia at an early stage. An electronic nose (e-nose) with 28 metal oxide semiconductor gas sensors was developed for predicting the presence of infection after patients have been intubated in the intensive care unit. The effectiveness of VAP identification was verified using clinical data. A total of 40 patients were included in this study, of whom 20 were infected with *Pseudomonas aeruginosa* and the remaining were uninfected. The results revealed that good accuracy rates of $0.9208\% \pm 0.0302\%$ and $0.8547\% \pm 0.0214\%$ were achieved by support vector machine and artificial neural network models, respectively. This study provides a simple, low-cost solution for the rapid screening of VAP at an early stage.

1 Introduction

Ventilator-associated pneumonia (VAP) is a common nosocomial infection occurring in ventilator-assisted patients in intensive care units (ICUs). Hospital-acquired pneumonia (HAP) (Self et al. 2013) is a major clinical problem associated with patients subjected to intubation and mechanical ventilation in ICUs (Woodhead et al. 2006; Chastre and Fagon 2002; Morehead and Pinto 2000; Koenig and Truwit 2006). The lower respiratory tract of a normal individual is usually a sterile environment; in intubated and mechanically ventilated patients, the incidence of bacterial infection can be as high as 80% (Fernández-Crehuet et al. 1997). VAP was estimated to account for 20–35% of the total cost burden imposed on hospitals. VAP increases hospitalization costs and may lead to death. Clinical sputum cultures, endotracheal aspirate analysis, and, less frequently, bronchoalveolar lavage are used to diagnose pneumonia; however, these techniques are poor in terms of sensitivity and positive predictive value for diagnosing VAP (Mandell et al. 2007). Reports of the diagnostic sensitivity and specificity of sputum cultures have varied extensively. Furthermore, sputum cultures can be difficult to interpret owing to the upper respiratory tract being colonized by bacteria, in contrast to the lower respiratory tract, which is typically sterile (Campbell and Forbes 2011). Therefore, numerous studies have been conducted on patients with other inflammatory pulmonary diseases, including VAP (Fens et al. 2011; Berkel et al. 2010; Phillips et al. 2010; Bos et al. 2014; van der Schee et al. 2015; Gao et al. 2016; Shafiek et al. 2015).

Exhaled breath contains gas-phase metabolites called volatile organic compounds (VOCs) that are produced by the host and bacteria. Different bacterial strains exhibit distinct patterns of VOCs in vitro and in animal models (Slupsky et al. 2009; Moens et al. 2006; Schöller et al. 1997; Dolch et al. 2012; Shestivska et al. 2012; Thorn et al. 2011). Recently, the breath of patients has been used to detect pneumonia infections (Nayeri et al. 2002; Corradi

✉ Chung-Hung Shih
chshih43@tmu.edu.tw

✉ Jiann-Shing Shieh
jsshieh@saturn.yzu.edu.tw

[1] Department of Mechanical Engineering, Yuan Ze University, Chungli 32003, Taiwan

[2] Division of Thoracic Medicine, Department of Internal Medicine, School of Medicine, College of Medicine, Taipei Medical University, Taipei City, Taiwan

[3] Division of Pulmonary Medicine, Department of Internal Medicine, Taipei Medical University Hospital, Taipei City, Taiwan

[4] College of Engineering, Design and Physical Sciences, Brunel University London, Uxbridge UB8 3PH, UK

[5] Department of Mechanical Engineering, Southern Taiwan University of Science and Technology, Tainan 710, Taiwan

et al. 2003); the smell from the breath of patients can help detect pneumonia more easily. However, conventional gas-analysis methods require the gas samples to be examined in a laboratory using a gas chromatography–mass spectrometry (GC–MS) analyzer (van den Velde et al. 2008; Zhu et al. 2010). Although GC–MS yields relatively accurate results, it is not suitable for monitoring in daily life because the instruments required are bulky and the procedure is a time-consuming process. Therefore, a device that can rapidly screen for VAP is lacking. Electronic noses (e-noses) are innovative devices for VOC sampling because they facilitate online recognition of complex VOC mixtures. A comparison of the predictive value of e-noses with that of the clinical pulmonary infection score indicated that e-noses could distinguish between infected and uninfected states (Hanson and Thaler 2005). The use of an e-nose to rapidly screen patients for pneumonia is a noninvasive technique that can help in the administering of antibiotic treatment in a timely manner. Therefore, e-noses constitute a suitable alternative for gas identification. An e-nose uses an array of sensors to acquire the specific response pattern of each type of gas and identifies this specific pattern for disease identification (Wilson and Baietto 2011; Pearce et al. 2006; Gardner and Bartlett 1992).

The use of an artificial neural network (ANN) has been proposed in the past (Sadrawi et al. 2015; Liu et al. 2015) as a suitable algorithm for modeling complex nonlinear and nonstationary relationships in medical research. An ANN can also be operated in real time when the weights have been trained and fixed, and it is fault tolerant. A support vector machine (SVM) is another machine learning method that is based on statistical theory (Chapelle and Vapnik 2000); SVMs are used for feature classification and regression. An SVM has many advantages when solving for a small number of samples; moreover, it can solve nonlinear and high-dimensional pattern recognition problems, and it is easy to use and implement in a real system owing to its simplicity (Gold and Sollich 2003; Anguita et al. 2003; Mustafa et al. 2009).

Herein, we propose the installation of a gas detection system in a ventilator to monitor and detect metabolites of pneumonia at an early stage. Unlike in our previous study (Liao et al. 2019), where we used equipment manufactured by Sensigent (i.e., Cyranose 320) to design a detection system through Matlab, in the present study, we independently designed our e-nose system. Specifically, we designed the sensor board, embedded hardware, and Android software for the system. Our data suggest that VOCs in exhaled breath can be used to discriminate intubated and mechanically ventilated patients with Hospital-acquired pneumonia (HAP) from ventilated patients without pneumonia in ICUs with high accuracy.

2 Materials and methods

The e-nose system designed in this study includes a sensor array panel with an air chamber and an e-nose circuit panel, in addition to have data collection, gas detection, and machine learning modules.

2.1 Sensor array panel

The sensor array is a key component of the e-nose system; it consists of 28 sensors. Table 1 lists the sensor manufacturer, target gas, and detection range. All 28 sensors are of the metal oxide semiconductor (MOS) type, such as TGS2600, TGS2602, TGS2603, TGS2610-C00, TGS2610-D00, TGS2611-C00, TGS2611-D00, TGS2612-D00, TGS2620, TGS3870, TGS2444, and TGS8100. MOS-type sensors have a simple structure, long life cycle, and low manufacturing cost; moreover, they are easy to miniaturize and are highly responsive. MOS-type sensors are composed of reactive materials such as Tin oxide (SnO_2), Zinc peroxide (ZnO_2) (Ji et al. 2010; Young et al. 2014; Liu et al. 2014; Chen et al. 2013), Tungsten trioxide (WO_3), Palladium (Pd), Carbon nanotube (CNT) (Young and Lin 2018) as a substrate and noble metals such as platinum, gold, tungsten, or palladium as a catalyst (Lin et al. 2015). The catalyst shortens the response time of the sensor to the chemical reaction equilibrium and expedites the response of the sensor.

2.2 E-nose circuit panel

Figure 1a shows the location of the 28 sensors (S1–S28) in a chamber sensor designed to enclose sensors of the same type. Figure 1b shows the block diagram of the sensor module comprising AD5144 analog potentiometers. The AD5144 potentiometers control by Inter-Integrated Circuit (I2C) bus and provide a nonvolatile solution for 256-position adjustment applications, and their resistance ranges from 10 to 100 kΩ; they ensure low resistor tolerance errors of ± 8%. The circuit controls the sensor resistance load and maintains the analog to digital converter (ADC) sampling voltage base in half maximum range to preserve each sensor's resistor as well as quality and health for long-term use. The central processing Unit (CPU) used is Freescale i.MX6, which has a single core running up to 1.0 GHz with 512 kB of L2 cache and 32-bit double data rate 3 (DDR3) with synchronous dynamic random access memory (SDRAM) support. Furthermore, it can integrate low voltage differential signaling (LVDS), a mobile industry processor interface (MIPI) display, MIPI camera port, and high definition multimedia interface (HDMI) v1.4.

Table 1 Specifications of the 28 sensors

SI No.	Sensor	Manufacturer	Target gas	Sensitivity (PPM)
1	TGS2600	Figaro Inc., Osaka, Japan	Hydrogen, ethanol	1–30
2	TGS2600	Figaro Inc., Osaka, Japan	Hydrogen, ethanol	1–30
3	TGS2600	Figaro Inc., Osaka, Japan	hydrogen, ethanol	1–30
4	TGS2602	Figaro Inc., Osaka, Japan	VOCs, ammonia, H2S	1–30
5	TGS2602	Figaro Inc., Osaka, Japan	VOCs, ammonia, H2S	1–30
6	TGS2602	Figaro Inc., Osaka, Japan	VOCs, ammonia, H2S	1–30
7	TGS2603	Figaro Inc., Osaka, Japan	Trimethylamine, methyl mercaptan	1–10
8	TGS2603	Figaro Inc., Osaka, Japan	Trimethylamine, methyl mercaptan	1–10
9	TGS2603	Figaro Inc., Osaka, Japan	Trimethylamine, methyl mercaptan	1–10
10	TGS2610-C00	Figaro Inc., Osaka, Japan	Butane, Propane	500–10,000
11	TGS2610-C00	Figaro Inc., Osaka, Japan	Butane, Propane	500–10,000
12	TGS2610-D00	Figaro Inc., Osaka, Japan	Butane, Propane	500–10,000
13	TGS2610-D00	Figaro Inc., Osaka, Japan	Butane, Propane	500–10,000
14	TGS2611-C00	Figaro Inc., Osaka, Japan	Methane	500–10,000
15	TGS2611-C00	Figaro Inc., Osaka, Japan	Methane	500–10,000
16	TGS2611-E00	Figaro Inc., Osaka, Japan	Methane	500–10,000
17	TGS2611-E00	Figaro Inc., Osaka, Japan	Methane	500–10,000
18	TGS2612-D00	Figaro Inc., Osaka, Japan	Methane, Propane, Butane	$1 \sim 25\%$ LEL
19	TGS2612-D00	Figaro Inc., Osaka, Japan	Methane, Propane, Butane	$1 \sim 25\%$ LEL
20	TGS2620	Figaro Inc., Osaka, Japan	Alcohol, Solvent vapors	50–5000
21	TGS2620	Figaro Inc., Osaka, Japan	Alcohol, Solvent vapors	50–5000
22	TGS2620	Figaro Inc., Osaka, Japan	Alcohol, Solvent vapors	50–5000
23	TGS3870	Figaro Inc., Osaka, Japan	Methane, CO	50–1000
24	TGS3870	Figaro Inc., Osaka, Japan	Methane, CO	50–1000
25	TGS3870	Figaro Inc., Osaka, Japan	Methane, CO	50–1000
26	TGS2444	Figaro Inc., Osaka, Japan	Ammonia	10–100
27	TGS8100	Figaro Inc., Osaka, Japan	Hydrogen, ethanol	1–30
28	TGS8100	Figaro Inc., Osaka, Japan	Hydrogen, ethanol	1–30

LEL lower explosive limit

The hardware framework of the e-nose system includes a valve, pump, gas chamber, temperature humidity control, sensor array, CPU mainboard, and control circuit (Fig. 2). The solenoid valve controls the inlet gas path and the pump initiates the gas flow, which is monitored using a flow sensor to monitor the flow ratio. A flow sensor from Sensirion is used to monitor the inlet gas operation in normal digital output. This facilitates long-term stability and precision even when the flow rate is slow. Furthermore, the gas chamber includes a sensor board that consists of 28 sensors and a potentiometer to attenuate the sensor output voltage and monitor the sensor resistance status. A high-precision 16-bit sigma–delta ADC with a low-noise programmable gain amplifier is used. The ADC includes TI LMP90080, which is a highly integrated multichannel, low-power 16-bit sensor analog front end chip. Finally, the e-nose system must be calibrated before use. The calibration procedure can be performed using standard gases, such as carbon monoxide gas or ethanol gas, to obtain the sensor array response curve. If this curve is within the tolerance range, the e-nose system can be used. However, the pump flow rate, temperature, and humidity of the chamber of the e-nose system are still needed to calibrate in regular intervals to make sure every parts of e-nose system within accurate values. Furthermore, even after the e-nose system is in use, we must still ensure that the pump flow rate, temperature, and humidity of the chamber are within a certain range, which can serve as a control system.

2.3 Breath samples and collection method

The single inclusion criterion for the patients to be part of the study was that they had to be on mechanical ventilation for more than 24 h Cardiopulmonary surgery patients and

(a)
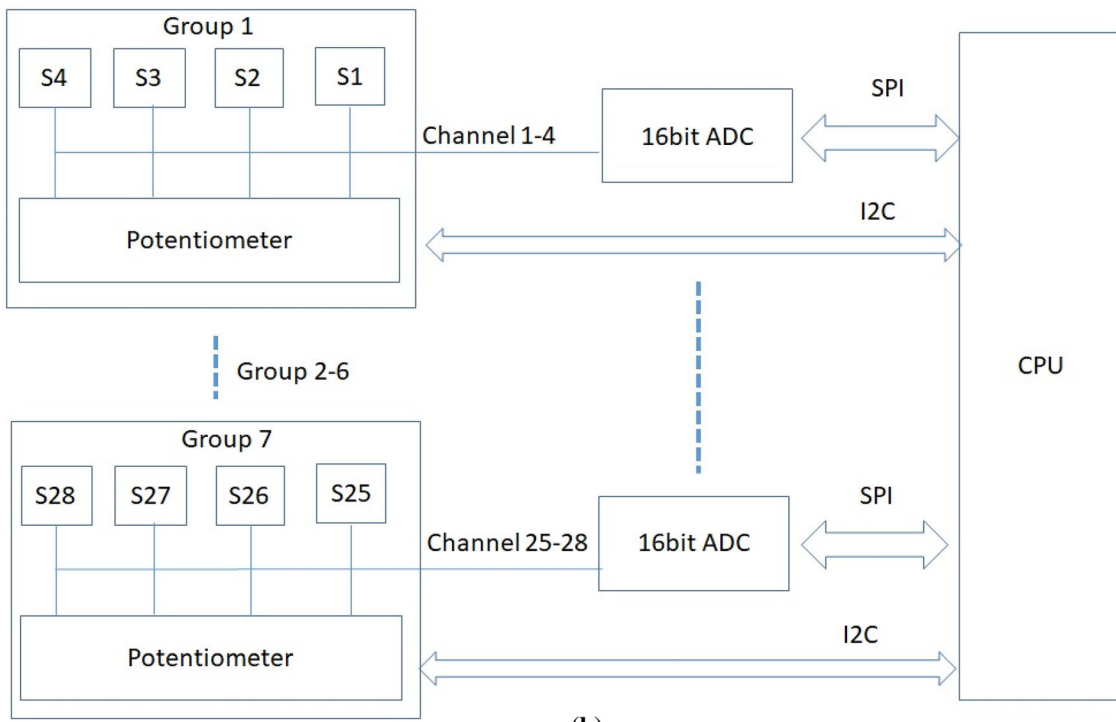
(b)

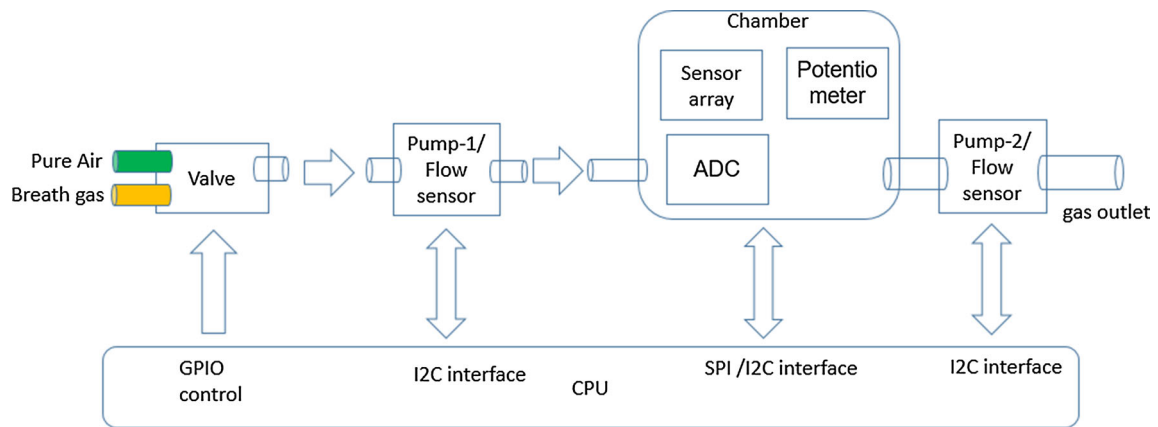Fig. 1 E-Nose circuit panel block diagram **a** Sensor placement, **b** Sensor module block diagram
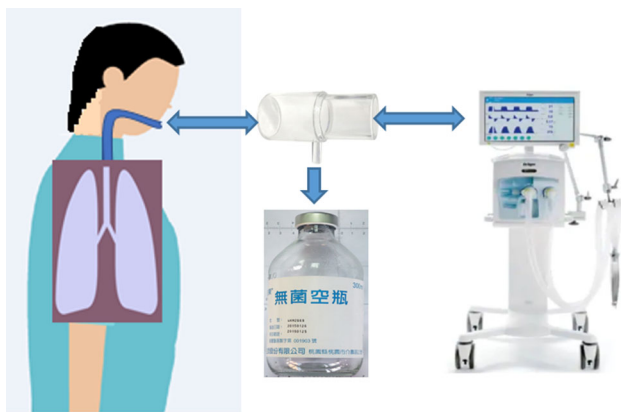
Fig. 2 Framework of e-nose system hardware

Fig. 3 Setup for collecting exhaled breath samples from VAP patients

Fig. 4 Setup for e-Nose breath sample reading

patients who previously received mechanical ventilation during the same hospitalization were excluded. The exhaled gas was sampled within 24 h after the patient was admitted into the ICU. The sampling method is illustrated in Fig. 3.

In this study, the bacterium *Pseudomonas aeruginosa* (*P. aeruginosa*) was collected in a sterile vacuum bottle connected to a respirator. The e-nose system equipped with 28 sensor arrays was then used offline to screen the patterns of respiratory gases in the sterile bottle, as shown in Fig. 4.

The sampling equipment was approved by Taiwan Food and Drug Administration (TFDA) product code 001903. The sterile bottle was provided by Xindong Co., Ltd., and the medical equipment intravenous (IV) set was used to connect the two sides. The sterile vacuum bottle was connected to the other side of the respirator. To ensure patient safety, the sampling process was performed by a medical professional and the sampling time was limited to 10 s.

Data were collected from selected patients with VAP infection and those without infection. The present research received approval from the Institutional Review Board at
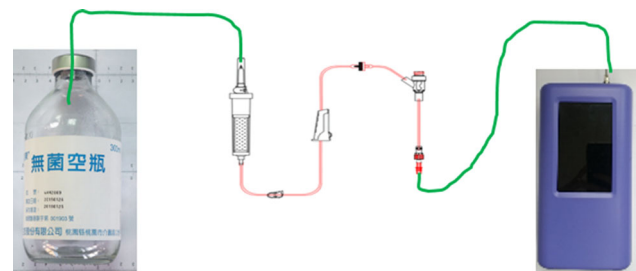
Taipei Medical University Hospital (TMUH), Taiwan. Written informed consent was obtained from the 40 patients participating in the study. Of the 40 patients, 20 contracted pneumonia at TMUH and the remaining did not have pneumonia (Table 2). Patients in the ICU are usually those that have undergone neurosurgery without any evidence of infection or lung disease. Raw data of 120 samples per patient were collected to obtain 4800 raw data points. To avoid bacterial variation caused by treatment implementation, respiratory gases from patients not injected with antibiotics were collected to constitute an experimental database.

2.4 Detection process

Figure 5 shows the architecture of the e-nose system. The system architecture shows the overall system flow in relation to external auxiliary systems.

The detection process is automatically controlled by the e-nose system. Figure 6 shows the timing sequence of the motor solenoid valve and sensor control. The detection procedure is divided into four phases, each defined as a flag:

Flag 0—Baseline phase (600 s): Each sensor baseline is recorded, which is used in data analysis.

Table 2 Patients with infection *P. aeruginosa* and without infection

No.	Sex	Age	WBC	PLT	Seg	CRP	Sputum	No.	Sex	Age	WBC	PLT	Seg	CRP	Sputum
1	Male	77	11.25	216	59.0	7.08	*P. aeruginosa*	1	Female	86	16.49	262	85.1	2.86	X
2	Male	33	7.93	98	87.4	N/A	*P. aeruginosa*	2	Female	81	10.31	239	79.6	1.87	X
3	Female	67	19.35	474	83.4	17.74	*P. aeruginosa*	3	Male	73	16.49	262	85.1	2.86	X
4	Female	90	15.19	505	82.6	N/A	*P. aeruginosa*	4	Male	69	14.55	483	82.1	4.91	X
5	Male	61	7.87	371	69.7	1.76	*P. aeruginosa*	5	Female	93	7.81	235	73	N/A	X
6	Male	83	10.21	251	82.1	4.2	*P. aeruginosa*	6	Female	81	10.46	290	79.4	0.81	X
7	Male	55	6.00	164	73.4	24.31	*P. aeruginosa*	7	Female	81	13.75	259	80.5	5.69	X
8	Female	87	11.60	315	64.9	1.42	*P. aeruginosa*	8	Male	83	11.31	208	81.4	1.18	X
9	Male	85	10.99	266	81.6	N/A	*P. aeruginosa*	9	Male	76	7.81	235	73	N/A	X
10	Male	46	15.92	374	89.1	16.42	*P. aeruginosa*	10	Male	79	13.75	259	80.5	5.69	X
11	Female	91	8.28	141	84.8	8.60	*P. aeruginosa*	11	Male	56	2.59	195	72.2	31.03	X
12	Female	51	12.10	65	89.3	0.41	*P. aeruginosa*	12	Male	63	7.98	15	89.2	N/A	X
13	Female	88	13.90	145	89.0	N/A	*P. aeruginosa*	13	Male	53	17.45	145	95.2	N/A	X
14	Female	64	11.13	93	58.5	3.97	*P. aeruginosa*	14	Male	47	7.38	126	80.6	N/A	X
15	Female	87	2.46	139	37.6	6.90	*P. aeruginosa*	15	Female	66	1.30	87	70	N/A	X
16	Female	83	6.75	77	75.7	4.87	*P. aeruginosa*	16	Female	66	7.67	66	91.9	7.98	X
17	Female	78	5.89	135	85.7	15.46	*P. aeruginosa*	17	Female	74	11.60	315	64.9	1.42	X
18	Male	94	6.75	77	75.7	4.87	*P. aeruginosa*	18	Male	85	11.92	249	79.7	N/A	X
19	Male	91	58.50	144	84.6	N/A	*P. aeruginosa*	19	Male	68	13.57	370	86	N/A	X
20	Male	64	11.54	146	94.1	9.52	*P. aeruginosa*	20	Male	65	5.14	40	76.4	18.09	X

P. aeruginosa: *Pseudomonas aeruginosa*, *WBC* White Blood Cell (10^{-3}/ul); *PLT* platelet (10^3/ui), *Seg* segmented eutrophils (%), *CRP* C-reactive protein (mg/L), *N/A* not available, *X* non-infected patient

Fig. 5 System architecture
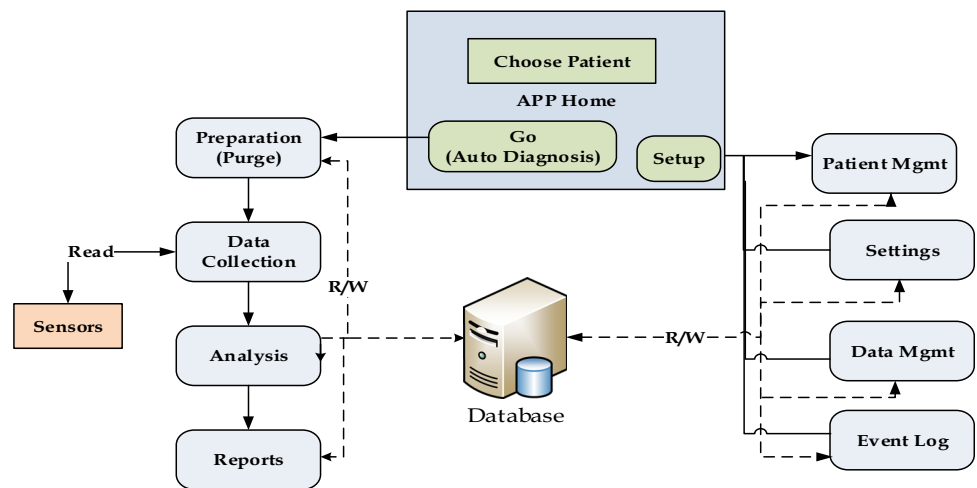
Flag 1—Breath gas injection phase (30 s): The pump is turned on. A breath gas sample is drawn from the airbag to the air chamber at constant speed. The sensor signal begins to respond to the injected target expiratory gas.

Flag 2—Sensor reaction phase (30 s): The pump is turned off. The sensor continues to react with components in the breath gas. The response of the MOS sensor reaches its maximum value.

Flag 3—Cleaning phase (600 s): The pump is turned on again and draws clean air into the air chamber for 600 s.

The sensor response gradually returns to the baseline. After the response becomes stable, the device is ready to measure the next sample.

The responses of the sensors (S1–S28) in the four phases are illustrated in Fig. 7. The response remains stable during the baseline phase (Flag 0) and changes during the injection phases (i.e., Flags 1 & 2). Each sensor reaches its maximum response at least once during the sensor reaction phase (Flag 2). Finally, during the purge phase (Flag 3), the

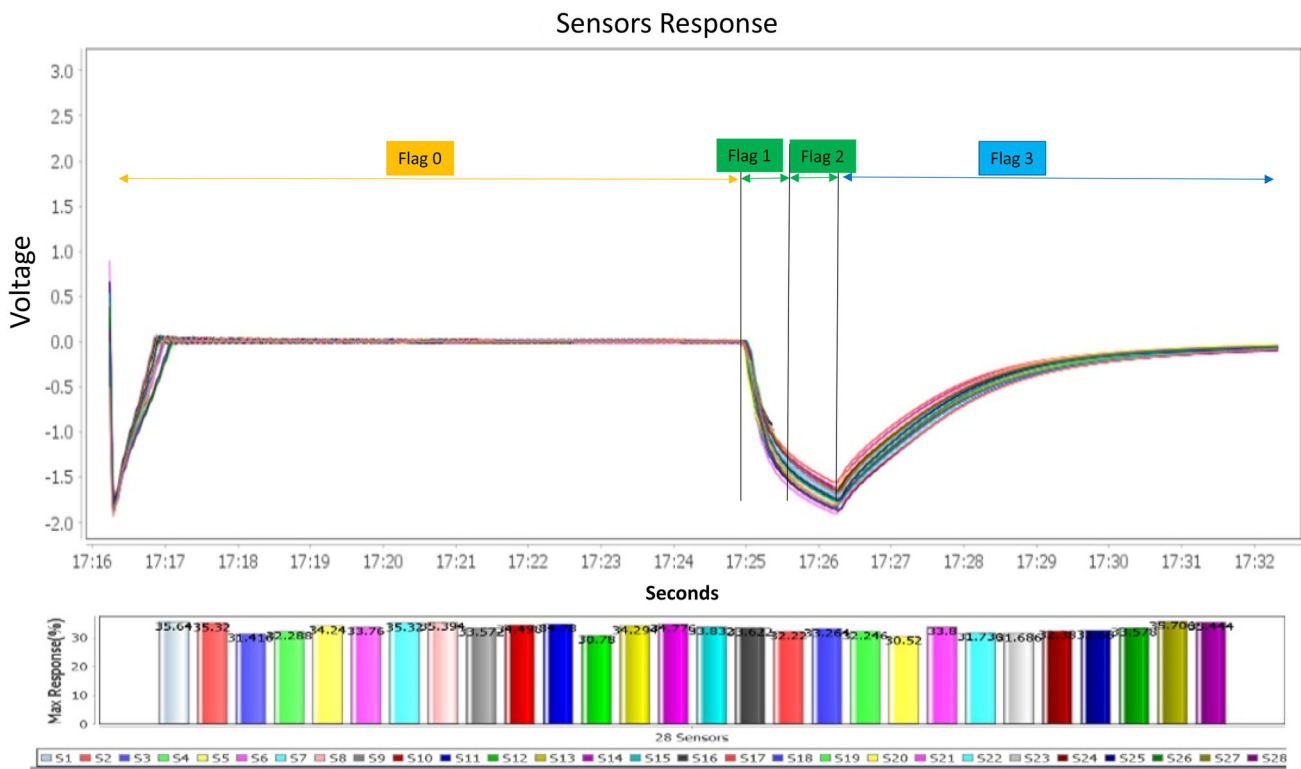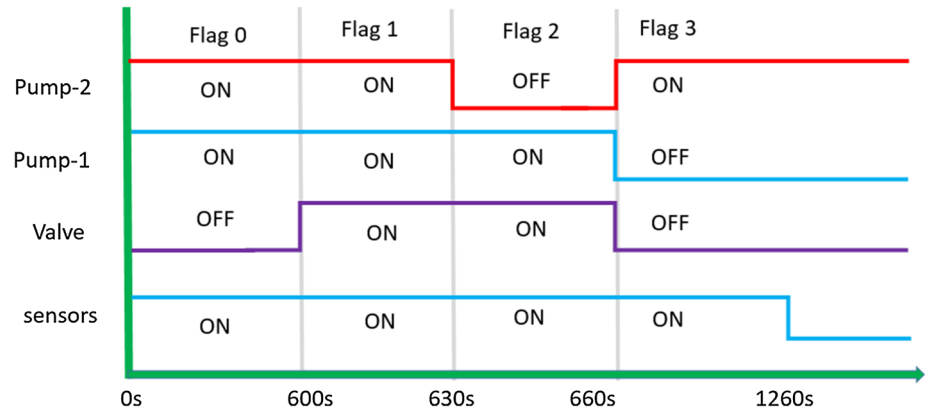**Fig. 6** Timing sequence of motor solenoid valve and sensor control

Fig. 7 Response curves of 28 sensors

sensor response gradually returns to the baseline when clean air is injected into the chamber. After the entire process, the digitized breath sample is represented by 28 response curves. Each response curve has 120 data points. The data samples are then used for further analysis.

2.5 Machine learning methods

2.5.1 SVM architecture

The SVM technique was introduced by Vapnik, and in recent years, it has been further developed. Several studies have reported that an SVM is usually able to achieve higher classification accuracy than other existing classification algorithms. In the past decade, the SVM has emerged as an important learning technique for solving classification and regression problems in various fields. A novel kernel function called the radial basis polynomial kernel (RBPK), which combines the characteristics of the radial basis function (RBF) kernel and the polynomial kernel, is defined as

$$K(x_i x_j) = \exp\left(\frac{\left|x_i x_j + c\right|^d}{\sigma^2}\right) \quad \text{where } c > 0, \ d > 0 \quad (1)$$

The RBPK function has good predictive ability from the polynomial kernel function and good learning ability from the RBF kernel function.

2.5.2 ANN architecture

The ANN algorithm is a machine learning algorithm that mimics human neural networks, and it is used for prediction, clustering, and pattern recognition based on past and current training data. An ANN allows for the inclusion of a large number of variables, and each variable can be assigned a different weight to produce an output that is very close to the predicted result. A typical classification network consists of an input layer, a hidden layer, and an output layer. In addition, the data are usually divided into training, validation, and testing sets. In previous studies (Liu et al. 2015), nonlinear and nonstationary medical data have been used with a backpropagation network having four layers: an input layer, two hidden layers with 17 and 10 neuron nodes each, and an output layer. Recently, deep learning has been extended to contain more layers; for example, CifarNet contains 5 layers, AlexNet contains 8 layers, and VGGNet contains 15–19 layers (Liu et al. 2019). For our study, since the dataset was of only 40 patients, we selected only four layers: an input layer, two hidden layers with 30 neuron nodes each, and an output layer.

2.5.3 Evaluating the predictive performance of machine learning methods

The receiver operating characteristic (ROC) curve analysis is a statistical method that has been widely used in medical diagnosis for evaluating the performance of classifiers (Ottenbacher et al. 2001; Yeh et al. 2009). Furthermore, the area under the ROC curve (AUC) is the main performance index of a classification system. An AUC of 1.0 typically implies perfect discrimination; however, this is an ideal case and would not occur in the real world. An AUC of 0.5 is equivalent to a random model. An AUC of less than 0.5 implies that an alternative method would be required to build the model, as it would mean that the classifier does not fit the real situation. Therefore, in this study, the ROC curve and AUC analysis were used to evaluate the predictive performance of the machine learning methods (i.e., the SVM and ANN). Moreover, the sensitivity (SEN), accuracy (ACC), and positive predictive value (PPV) of the methods were calculated using Eqs. (2)–(4) as follows:

$$SEN = \frac{TP}{TP + FN}, \tag{2}$$

$$ACC = \frac{TP + TN}{TP + FP + TN + FN}, \tag{3}$$

$$PPV = \frac{TP}{TP + FP} \tag{4}$$

where TP is a true positive classification, FN is a false negative classification, TN is a true negative classification, and FP is a false positive classification.

3 Data analysis and results

3.1 Data preprocessing

Data preprocessing has a major effect on the performance of e-nose systems. A good signal preprocessing method not only reduces noise but also reduces recognition complexity and errors, thereby improving system identification performance. Accordingly, in this study, we implemented a hardware first-order low-pass RC filter. A nonlinear log–sigmoid activation function was used, which is defined in Eq. (5), where a is the output of the nonlinear log–sigmoid function and n is the input of the nonlinear activation function. The training and testing data are normalized in the range of 0–1

$$a = f(n) = \frac{1}{1 + e^{-n}}. \tag{5}$$

3.2 Predictive ability of ANN and SVM ROC models

The concept of cross-validation is to generate k-fold cross-validations of equal size. One data fold is used as the validation set, and the other folds are used for training to generate a neural network model. This process is repeated k times, and for each repetition, the validation sets are selected from the k-fold; the sets are never used twice. This process generates a k number of neural networks that are averaged to generate the final prediction.

To estimate the generalized prediction abilities of the ANN model, testing data were applied to the model, and the outputs were used to generate the ROC curve. Table 3 summarizes the ROC curve analysis results for the ANN and SVM models. To evaluate the performance of the models, we employed fivefold cross-validation. The results revealed that the ANN model had good VAP recognition rates, and the average ACC, SEN, and PPV values were 0.8547 ± 0.0214 (mean $\pm$ SD), 0.6779 ± 0.1133, and 0.6408 ± 0.1583, respectively. The results also showed that the SVM model exhibited high VAP recognition rates, and the average ACC, SEN, and PPV values were 0.9208 ± 0.0302, 0.8439 ± 0.0303, and 0.8145 ± 0.0660, respectively. Therefore, the SVM model performed better than the ANN model in terms of detecting VAP. The optimized SVM model shown in Fig. 8 was established with the ideal combination of parameters of $C = 10$, $\sigma = 0.001$, to achieve better performance in VAP detection.

The ROC curves (Fig. 9) for the ANN and SVM models were analyzed. The average AUC obtained from fivefold cross-validation tests for the ANN model was 0.868 ± 0.04868 (mean $\pm$ SD). In addition, the average

Table 3 ROC curve analysis results by test dataset for ANN and SVM models

Dataset	Model type	AUC	ACC	SEN	PPV
At best threshold					
Average	ANN Model	0.868 ± 0.04868	0.8547 ± 0.0214	0.6779 ± 0.1133	0.6408 ± 0.1583
Average	SVM Model	0.926 ± 0.0167	0.9208 ± 0.0302	0.8439 ± 0.0303	0.8145 ± 0.0660

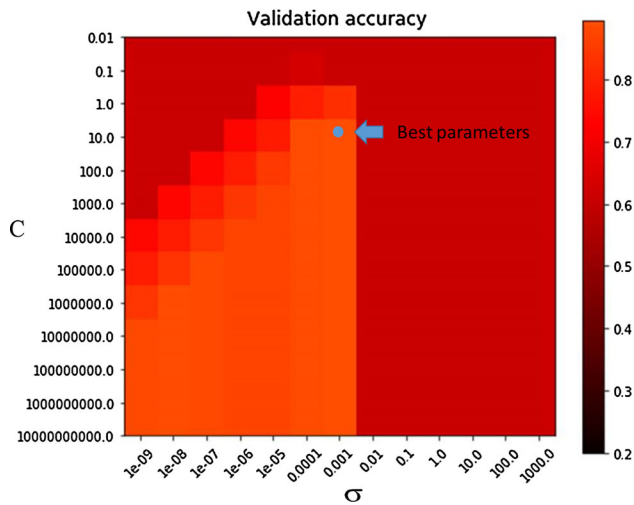*AUC* area under the curve, *ACC* accuracy, *SEN* sensitivity, *PPV* positive predictive value

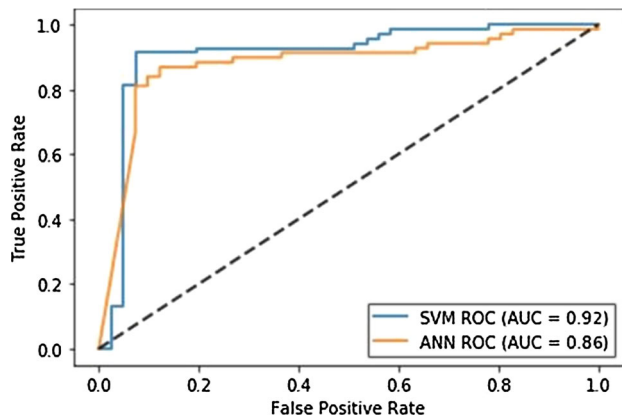**Fig. 8** SVM search results for choosing the best parameters

Fig. 9 ANN and SVM average ROC curve

AUC obtained from fivefold cross-validation tests for the SVM model was 0.926 ± 0.0167. The high mean values and the low SD values of SVM model confirm the uniformity and reliability of the results.

4 Discussion

In this study, an e-nose system consisting of 28 gas sensors was developed and tested on 40 patients. Our experience of each type of sensor consists of at least 3 sensors which could get good stability and performance. Furthermore,

sensors need filter to use target gas to screen and define sensor quality. Moreover, each sensor module needs aging over 168 h to ensure stability.

The proposed system has a high degree of automation, has a relatively low environmental impact, has low artificial effects, improves the repeatability and reliability of tests, and can be integrated with or customized for various test methods to meet different needs. In addition, the advancement of system software functions along with improvements in identification technology will improve the performance of the system.

5 Conclusion

In this study, machine learning was used to predict whether patients were infected with major bacteria *P. aeruginosa* in the ICU. The ANN and SVM models were trained and tested separately. ROC curve analysis showed that the ANN and SVM models had a high pneumonia identification rate. The SVM model was more accurate than the ANN model. Because only 40 patients were included in this study, the ANN model may not have had sufficient data for training. If we were to include more data in the future, the ANN model would outperform the SVM model because the nonlinear mapping for the ANN is higher than that for the SVM. The designed system utilizes a low-cost microarray e-nose that is easy to install into a mechanical ventilation system in the ICU to facilitate the collection of breath samples from VAP patients. In the future, data collected from multiple hospitals can be tested not only with the traditional ANN (or shallow learning) but also with deep learning (e.g., convolutional neural networks) (Esteva et al. 2017; Rajpurkar et al. 2017; De Fauw et al. 2018), which can be applied to VAP detection systems.

References

Anguita D, Ridella S, Rivieccio F (2003) Hyperparameter design criteria for support vector classifiers. Neuro Comput 55(1):109–134. https://doi.org/10.1016/S0925-2312(03)00430-2

Berkel JJBN, Dallinga JW, Moller GM, Godschalk RWL, Moonen EJ, Wouters EFM, van Schooten FJ (2010) A profile of volatile organic compounds in breath discriminates COPD patients from controls. Respir Med 104:557–563. https://doi.org/10.1016/j.rmed.2009.10.018

Bos LDJ, Weda H, Wang Y, Knobel HH, Nijsen TME, Vink TJ, Zwinderman AH, Sterk PJ, Schultz MJ (2014) Exhaled breath metabolomics as a noninvasive diagnostic tool for acute respiratory distress syndrome. Eur Respir J 44:188–197. https://doi.org/10.1183/09031936.00005614

Campbell S, Forbes BA (2011) The clinical microbiology laboratory in the diagnosis of lower respiratory tract infections. J Clin Microbiol 49(S9): S30–S33. https://doi.org/10.1128/JCM.00789-11 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3185862/

Chapelle O, Vapnik V (2000) Model selection for support vector machines. In: Advances in neural information processing systems, vol 12. MIT Press, Cambridge, pp 230–236. http://papers.nips.cc/paper/1663-model-selection-for-support-vector-machines.pdf

Chastre J, Fagon JY (2002) Ventilator-associated pneumonia. Am J Resp Crit Care Med 165:867–903. https://doi.org/10.1164/ajrccm.165.7.2105078

Chen TP, Young SJ, Chang SJ, Hsiao CH, Wu SL (2013) Photoelectrical and low-frequency noise characteristics of ZnO nanorod photodetectors prepared on flexible substrate. IEEE Trans Electron Dev 60(1):229–234. https://doi.org/10.1109/TED.2012.2228655

Corradi M, Pesci A, Casana R, Alinovi R, Goldoni M, Vettori MV, Cuomo A (2003) Nitrate in exhaled breath condensate of patients with different airway diseases. Nitric Oxide 8(1):26–30. https://doi.org/10.1016/S1089-8603(02)00128-3

De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, Askham H, Glorot X, O'Donoghue B, Visentin D, et al (2018) Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med 24:1342–1350. https://doi.org/10.1038/s41591-018-0107-6 https://www.nature.com/articles/s41591-018-0107-6

Dolch ME, Hornuss C, Klocke C, Praun S, Villinger J, Denzer W, Schelling G, Schubert S (2012) Volatile organic compound analysis by ion-molecule reaction mass spectrometry for Gram-positive bacteria differentiation. Eur J Clin Microbiol Infect Dis 31:3007–3013. https://doi.org/10.1007/s10096-012-1654-2

Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542:15–118. https://doi.org/10.1038/nature21056

Fens N, de Nijs SB, Peters S, Dekker T, Knobel HH, Vink TJ, Willard NP, Zwinderman AH, Krouwelse FH, Janssen HG et al (2011) Exhaled air molecular profiling in relation to inflammatory subtype and activity in COPD. Eur Respir J 38:1301–1309. https://doi.org/10.1183/09031936.00032911

Fernández-Crehuet R, Díaz-Molina C, de Irala J, Martínez-Concha D, Salcedo-Leal I, Masa-Calles J (1997) Nosocomial infection in an intensive-care unit: identification of risk factors. Infect Control Hosp Epidemiol 18(12):825–830. https://doi.org/10.1017/S0195941700086550

Gao J, Zou Y, Wang Y, Wang F, Lang L, Wang P, Zhou Y, Ying K (2016) Breath analysis for noninvasively differentiating Acinetobacter baumannii ventilator-associated pneumonia from its respiratory tract colonization of ventilated patients. J Breath Res 10(2):027102. https://doi.org/10.1088/1752-7155/10/2/027102

Gardner JW, Bartlett PN (1992) Sensors and sensory systems for an electronic nose. Springer, New York

Gold C, Sollich P (2003) Model selection for support vector machine classification. Neurocomputing 55:221–249. https://doi.org/10.1016/S0925-2312(03)00375-8

Hanson CW, Thaler ER (2005) Electronic nose prediction of a clinical pneumonia score: biosensors and microbes. Anesthesiology 102(1):63–68. https://doi.org/10.1097/00000542-200501000-00013

Ji LW, Wu CZ, Lin CM, Meen TH, Lam KT, Peng SM, Youg SJ, Liu CH (2010) Characteristic improvements of ZnO-based metal–semiconductor–metal photodetector on flexible substrate with ZnO cap layer. Jpn J Appl Phys 49(5R):052201. https://doi.org/10.1143/JJAP.49.052201

Koenig SM, Truwit JD (2006) Ventilator-associated pneumonia: diagnosis, treatment, prevention. Clin Microbiol Rev 19(4):637–657. https://doi.org/10.1128/CMR.00051-05

Liao YH, Wang ZC, Zhang FG, Abbod MF, Shih CH, Shieh JS (2019) Machine learning methods applied to predict ventilator-associated pneumonia with *Pseudomonas aeruginosa* infection via sensor array of electronic nose in intensive care unit. Sensors. 19(8):E1866. https://doi.org/10.3390/s19081866

Lin ZD, Young SJ, Chang SJ (2015) Carbon nanotube thin films functionalized via loading of au nanoclusters for flexible gas sensors devices. IEEE Trans Electron Devices 63(1):476–480. https://doi.org/10.1109/TED.2015.2504105

Liu YH, Young SJ, Hsiao CH, Ji LW, Meen TH, Water W, Chang SJ (2014) Visible-blind photodetectors with Mg-doped ZnO nanorods. IEEE Photon Technol Lett 26(7):645–648. https://doi.org/10.1109/LPT.2014.2301845

Liu Q, Chen YF, Fan SZ, Abbod MF, Shieh JS (2015) EEG signals analysis using multiscale entropy for depth of anesthesia monitoring during surgery through artificial neural networks. Comput Math Methods Med 215:1–16. https://doi.org/10.1155/2015/232381

Liu Q, Cai JF, Fan SZ, Abbod MF, Shieh JS, Kung YC, Lin LS (2019) Spectrum analysis of EEG signals using CNN to model patient's consciousness level based on anesthesiologists' experience". IEEE Access 7:53731–53742. https://doi.org/10.1109/ACCESS.2019.2912273

Mandell LA, Wunderink RG, Anzueto A et al (2007) Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. Clin Infect Dis 44(S2):S27–S72. https://doi.org/10.1086/511159

Moens M, Smet A, Naudts B, Verhoeven J, Ieven M, Jorens P, Geise HJ, Blockhuys F (2006) Fast identification of ten clinically important micro-organisms using an electronic nose. Lett Appl Microbiol 42(2):121–126. https://doi.org/10.1111/j.1472-765X.2005.01822.x

Morehead RS, Pinto SJ (2000) Ventilator-associated pneumonia. Arch Intern Med 160(13):1926–1936. https://doi.org/10.1001/archinte.160.13.1926

Mustafa NBA, Syed KA, Zaipatimah A, Yit WB, Aidil AZA, Zainul AMS (2009) Agricultural produce sorting and grading using support vector machines and fuzzy logic. IEEE Int Conf Signal Image Process Appl 174:391–396. https://doi.org/10.1109/ICSIPA.2009.5478684

Nayeri F, Millinger E, Nilsson I, Zetterström O, Brudin L, Forsberg P (2002) Exhaled breath condensate and serum levels of hepatocyte growth factor in pneumonia. Resp Med 96:115–119. https://doi.org/10.1053/rmed.2001.1225

Ottenbacher K, Smith P, Illig S, Linn R, Fiedler R, Granger C (2001) Comparison of logistic regression and neural networks to predict rehospitalization in patients with stroke. J Clin Epidemiol 54(11):1159–1165. https://doi.org/10.1016/S0895-4356(01)00395-X

Pearce TC, Schiffman SS, Nagle HT, Gardner JW (2006) Handbook of machine olfaction: electronic nose technology. Wiley, Hoboken

Phillips M, Basa-Dalay V, Bothamley G, Cataneo RN, Lam PK, Natividad MPR, Schmitt P, Wai J (2010) Breath biomarkers of active pulmonary tuberculosis. Tuberculosis 90(2):145–151. https://doi.org/10.1016/j.tube.2010.01.003

Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, Ding D, Bagul A, Langlotz C, Shpanskaya K, et al (2017) CheXNet: radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv:1711.05225v1. https://arxiv.org/abs/1711.05225

Sadrawi M, Fan SZ, Abbod MF, Jen KK, Shieh JS (2015) Computational depth of anesthesia via multiple vital signs based on artificial neural networks. Biomed Res Int 2015:1–13. https://doi.org/10.1155/2015/536863

Schöller C, Molin S, Wilkins K (1997) Volatile metabolites from some Gram-negative bacteria. Chemosphere 35:1487–1495. https://doi.org/10.1016/S0045-6535(97)00209-9

Self WH, Courtney DM, McNaughton CD, Wunderink RG, Kline JA (2013) High discordance of chest X-ray and computed tomography for detection of pulmonary opacities in ED patients: implications for diagnosing pneumonia. Am J Emerg Med 31:401–405. https://doi.org/10.1016/j.ajem.2012.08.041

Shafiek H, Fiorentino F, Merino JL, López C, Oliver A, Segura J, de Paul I, Sibila O, Agustí A, Cosío BG (2015) Using the electronic nose to identify airway infection during COPD exacerbations. PLoS ONE 10(9):e0135199. https://doi.org/10.1371/journal.pone.0135199

Shestivska V, Španěl P, Dryahina K, Sovová K, Smith D, Musílek M, Nemec A (2012) Variability in the concentrations of volatile metabolites emitted by genotypically different strains of *Pseudomonas aeruginosa*. J Appl Microbiol 113:701–713. https://doi.org/10.1111/j.1365-2672.2012.05370.x

Slupsky CM, Cheypesh A, Chao DV, Fu H, Rankin KN, Marrie TJ, Lacy P (2009) Streptococcus pneumoniae and Staphylococcus aureus pneumonia induce distinct metabolic responses. J Proteome Res 8(6):3029–3036. https://doi.org/10.1021/pr900103y

Thorn RMS, Reynolds DM, Greenman J (2011) Multivariate analysis of bacterial volatile compound profiles for discrimination between selected species and strains in vitro. J Microbiol Methods 84:258–264. https://doi.org/10.1016/j.mimet.2010.12.001

van den Velde S, Nevens F, van Hee P, van Steenberghe D, Quirynen M (2008) GC-MS analysis of breath odor compounds in liver patients. J Chromatogr B Analyt Technol Biomed Life Sci 875(2):344–348. https://doi.org/10.1016/j.jchromb.2008.08.031

van der Schee MP, Paff T, Brinkman P, van Aalderen WMC, Haarman EG, Sterk PJ (2015) Breathomics in lung disease. CHEST J 147(1):224–231. https://doi.org/10.1378/chest.14-0781

Wilson AD, Baietto M (2011) Advances in electronic-nose technologies developed for biomedical applications. Sensors 11(1):1105–1176. https://doi.org/10.3390/s110101105

Woodhead M, Welch CA, Harrison DA, Bellingan G, Ayres JG (2006) Community-acquire pneumonia in the intensive care unit: secondary analysis of 17,869 cases in the ICNARC case mix programme database. Crit Care. https://doi.org/10.1186/cc4927

Yeh JR, Fan SF, Shieh JS (2009) Human heartbeat analysis using a modified algorithm of detrended fluctuation analysis based on empirical mode decomposition. Med Eng Phys 31(1):92–100. https://doi.org/10.1016/j.medengphy.2008.04.011

Young SJ, Lin ZD (2018) Ethanol gas sensors based on multi-wall carbon nanotubes on oxidized Si substrate. Microsyst Technol 24(1):55–58. https://doi.org/10.1007/s00542-016-3154-2

Young SJ, Liu YH, Hsiao CH, Chang SJ, Wang BC, Kao TH, Tsai KS, Wu SL (2014) ZnO-based ultraviolet photodetectors with novel nanosheet structures. IEEE Trans Nanotechnol 13(2):238–244. https://doi.org/10.1109/TNANO.2014.2298335

Zhu J, Bean HD, Kuo YM, Hill JE (2010) Fast detection of volatile organic compounds from bacterial cultures by secondary electrospray ionization-mass spectrometry. J Clin Microbiol 48:4426–4431. https://doi.org/10.1128/JCM.00392-10