Electronic Nose-Based Breath Analysis for Early COPD

Pongsaphat Chantaramanee[1], Prompriang Prakopkit[1]
Natpassorn Laonet[1*] and Kanokrat Palanupap[2]

[1]Princess Chulabhorn Science High School Phitsanulok, Makham Sung, Mueang Phitsanulok, Phitsanulok 65000, Thailand

[2]Department of Internal Medicine, Mae Sot Hospital, Mae Sot, Mae Sot, Tak 63110, Thailand

**e-mail: pongsaphat.cha@pccpl.ac.th*

Abstract

Chronic Obstructive Pulmonary Disease (COPD) is a chronic respiratory disease characterized by persistent inflammation and lung tissue damage. According to the World Health Organization (WHO), COPD causes approximately 3 million deaths annually. Early diagnosis remains challenging because symptoms are often mild and non-specific, while spirometry, the current standard diagnostic method, is limited by its cost, inconvenience, and invasiveness. To address these limitations, this study developed a preliminary COPD screening device based on an electronic nose (E-nose) integrated with a real-time mobile application. Volatile organic compounds (VOCs) in exhaled breath collected from a total of 60 participants at Mae Sot Hospital, including COPD patients and healthy participants, were analyzed using MQ gas sensors, and seven machine learning models were developed and compared using Weka and Orange Data Mining. Breath from COPD patients showed higher levels of acetone, decane, and p-xylene, but lower levels of benzene and cyclopentane, than that of healthy individuals. Among the eight models, XGBoost achieved the highest performance, with an accuracy of 96.67%, sensitivity of 97.06%, specificity of 96.15%, and precision of 97.06%, followed by Random Forest (93.33% accuracy). These findings demonstrate that the proposed E-nose system has strong potential as a rapid, non-invasive preliminary screening tool for COPD, improving access to early detection in community healthcare settings while reducing the burden on the healthcare system.

keywords : Chronic Obstructive Pulmonary Disease (COPD), Electronic Nose, Machine learning, Volatile Organic Compounds (VOCs)

Introduction

Chronic obstructive pulmonary disease (COPD) is a progressive respiratory disorder characterized by persistent airflow limitation, chronic inflammation, and damage to lung tissue. Common symptoms include shortness of breath, chronic cough, sputum production, fatigue, and reduced exercise tolerance. As the disease advances, these symptoms may increasingly restrict daily activities and reduce overall quality of life. Long-term tobacco exposure remains the principal risk factor, although occupational chemicals, airborne particles, household smoke, and

environmental pollution may also contribute to disease development.

One of the major difficulties in COPD management is that the condition often remains undetected during its early stages. Initial symptoms may be mild, intermittent, or mistaken for normal aging or temporary respiratory illness. As a result, many patients seek medical attention only after a substantial decline in lung function has already occurred. Earlier identification is therefore important because it may allow patients to receive timely clinical assessment, reduce exposure to risk factors, and begin appropriate treatment before the disease becomes more severe.

Spirometry is widely used to assess airflow obstruction and remains essential for confirming COPD. However, accurate testing requires calibrated equipment, trained personnel, standardized procedures, and sufficient cooperation from the patient. These requirements may limit access in small hospitals, community clinics, and remote areas. A rapid preliminary screening method could therefore help identify individuals who may require further pulmonary evaluation and support more efficient use of limited healthcare resources.

Exhaled breath contains volatile organic compounds (VOCs) produced through metabolic activity, inflammation, and oxidative stress. Changes in health status may alter the pattern of these compounds. An electronic nose uses an array of gas sensors to capture a combined breath signal, which can be interpreted as a characteristic breath profile. Because these sensor responses are complex and often overlap, machine-learning algorithms can assist in identifying patterns that distinguish individuals with COPD from those without the disease.

Accordingly, this study aimed to compare VOC-related breath measurements between COPD and non-COPD groups, develop an electronic nose for preliminary screening, and evaluate several machine-learning classifiers. The best-performing model was then integrated with an iOS application for real-time result display. The proposed system was designed as a screening aid to support early referral and was not intended to replace spirometry or formal clinical diagnosis.

Methodology

The research was conducted in three main stages: (1) identifying volatile organic compound patterns in exhaled breath associated with chronic obstructive pulmonary disease, (2) developing an electronic nose-based preliminary COPD screening device with an application-based result display, and (3) evaluating the performance of the completed screening system.

Identification of COPD-Related Volatile Organic Compounds in Exhaled Breath

The electronic nose enclosure was designed using Autodesk Fusion 360. Acrylic sheets were then cut, shaped, and assembled to form the device housing. Five gas sensors—MQ-135, MQ-138, MiCS-5524, MQ-3, and MQ-7—were installed inside the chamber in an array configuration. This arrangement allowed

the system to capture a combined sensor-response pattern rather than relying on a single gas measurement.

An ESP32 microcontroller was used to collect and process the sensor signals, while an LCD screen displayed operational information. A DHT-22 sensor was also incorporated to monitor temperature and relative humidity, enabling the environmental conditions affecting the MQ sensors to be observed and controlled more effectively. Finally, tubing was connected between the breath-collection bag and the sensor chamber so that the exhaled-breath sample could be introduced into the device for analysis.

Development of the Electronic Nose-Based COPD Screening System and Application

Breath-related gas measurements obtained from participants with COPD and those without COPD were used to develop classification models. The measured variables included acetone, decane, p-xylene, benzene, and cyclopentane.

A set of machine-learning models was developed using Weka and Orange Data Mining, including Support Vector Machine, Random Forest, XGBoost, k-Nearest Neighbors, Neural Network, Decision Tree, and Gradient Boosting. Model performance was evaluated through 10-fold cross-validation. In this process, the dataset was divided into ten subsets; nine subsets were used for training, while the remaining subset was used for testing. The procedure was repeated until each subset had served as the test set once, providing a more reliable estimate of model performance.

The best-performing model was then selected for integration into the electronic nose system. An iOS application was developed using Xcode to receive the processed classification output and display the screening result in real time.

Performance Evaluation of the Electronic Nose and Application

The completed electronic nose and application were tested with both individuals diagnosed with COPD and individuals without COPD. The obtained results were analyzed using SPSS.

Differences between the two groups were examined using the independent t-test at a significance level of $\alpha = 0.05$. The corresponding p-values were calculated to determine whether the observed differences were statistically significant.

The overall performance of the screening system was evaluated using four principal metrics: accuracy, sensitivity, specificity, and precision. Accuracy represented the proportion of all cases classified correctly. Sensitivity reflected the system's ability to identify individuals with COPD, whereas specificity indicated its ability to correctly classify individuals without the disease. Precision represented the proportion of positive COPD predictions that were correct.

Result and Discussion

Identification of Volatile Organic Compounds in Exhaled Breath for COPD Classification

Following the development of the electronic nose enclosure using Autodesk

Fusion 360, the gas sensors were arranged in an array inside the chamber.

The system was used to compare VOC-related measurements in exhaled breath between participants with COPD and those without COPD. Five compounds were identified, as shown in Table 1.

Table 1. Comparison of VOCs-Related from Breath Between the COPD and Non-COPD Groups

VOCs	COPD group (n=34)	Non-COPD group (n=26)	p-value
Acetone	4.71 ± 0.32	2.55 ± 0.42	< 0.001
Decane	2.45 ± 0.44	1.07 ± 0.25	< 0.001
p-Xylene	0.62 ± 0.41	1.02 ± 0.38	0.0002
Benzene	0.33 ± 0.19	0.70 ± 0.21	< 0.001
Cyclopentane	0.23 ± 0.15	0.65 ± 0.16	< 0.001

As shown in Table 1, the results were consistent with those reported by Tian et al. (2025) for Acetone and Decane, both of which were significantly higher in the COPD group (n=34) than in the non-COPD group (n=26) ($p < 0.001$). In contrast, p-xylene, benzene, and cyclopentane were significantly lower in the COPD group.

Development of the Electronic Nose-Based COPD Screening System and Application

Gas measurements obtained from the COPD and non-COPD groups, including acetone, decane, p-xylene, benzene, and cyclopentane, were processed using Weka and Orange Data Mining to develop classification models. The performance of each model is presented in Table 2.

Table 2. Performance of the Machine-Learning Classification Models

Model	Acc (%)	Spec	Rec	Prec	F1	AUC	MCC
SVM	91.67	0.88	0.94	0.91	0.93	0.91	0.83
Random Forest	93.33	0.92	0.94	0.94	0.94	0.93	0.86
XGBoost	**96.67**	**0.96**	**0.97**	**0.97**	**0.97**	**0.96**	**0.93**
KNN	85.00	0.80	0.88	0.85	0.87	0.86	0.70
Neural Network	71.67	0.70	0.73	0.75	0.74	0.71	0.43
Decision Tree	81.67	0.76	0.85	0.82	0.84	0.81	0.62
Gradient Boosting	90.00	0.85	0.94	0.89	0.91	0.90	0.80

As shown in Table 2, the models produced different classification results when evaluated using k-fold cross-validation. XGBoost achieved the highest accuracy at 96.67%, followed by Random Forest at 93.33% and SVM at 91.67%. The Neural Network model showed the lowest accuracy at 71.67%.

Performance Evaluation of the Electronic Nose and Application

XGBoost was selected for integration into the electronic nose because it achieved the highest classification accuracy and could operate efficiently with the system's data-transmission process. The screening result was processed and displayed through an iOS application developed using Xcode. The performance of the completed system is shown in Table 3.

Table 3. Perfomance of the Integrated Electronic Nose Screening

Performance metric	Value (%)
Accuracy	96.67
Sensitivity	97.06
Specificity	96.15
Precision	97.06

As presented in Table 3, the integrated electronic nose and application successfully distinguished between participants with COPD (n=34) and without COPD (n=26). The system achieved an accuracy of 96.67%, sensitivity of 97.06%, specificity of 96.15%, and precision of 97.06%.

Conclusions

This study developed an electronic nose system that combines exhaled-breath sensing, machine learning, and real-time mobile reporting for preliminary COPD screening. Among the total sample size of 60 participants, the sensor array detected significant differences between the two study groups: acetone and decane were higher in the COPD group, whereas p-xylene, benzene, and cyclopentane were lower.

Among the seven evaluated classifiers, XGBoost showed the strongest and most balanced performance. It achieved an accuracy of 96.67%, with high specificity, recall, precision, F1-score, AUC, and MCC. When incorporated into the prototype, the model produced a sensitivity of 97.06%, specificity of 96.15%, and precision of 97.06%. The iOS application also displayed the screening result successfully in real time.

The prototype illustrates how accessible sensor technology and data-driven analysis can be combined to support early referral in settings where formal pulmonary testing is not immediately available. Its most appropriate role is as a preliminary screening aid, not as a stand-alone diagnostic instrument. With improved environmental control, standardized breath sampling, and validation in an independent population, the system may contribute to wider access to respiratory screening while reducing unnecessary pressure on specialized healthcare services.

Acknowledgments

The authors sincerely thank Natpassorn Laonet for her guidance, valuable advice, and continued support throughout the project. We are also grateful to Kanokrat Palanupap, M.D., Department of Internal Medicine, Mae Sot Hospital, for her medical consultation, constructive recommendations, and for sharing practical clinical perspectives on respiratory disease screening. Appreciation is extended to Princess Chulabhorn Science High School Phitsanulok for providing funding, facilities, and opportunities that made this study possible.

References

Simalaotao, P., and Saenrat, C. (2020). *A comparison of classification performance for factors supporting learning through electronic instructional media*. King Mongkut's

University of Technology North Bangkok. [In Thai]

Srinonchat, C., and Chansri, C. (2017). *Improving the odor-classification performance of an electronic nose using artificial neural networks*. Faculty of Engineering, Rajamangala University of Technology Thanyaburi. [In Thai]

Tian, J., and Zhang, Q. (2025). Exhaled volatile organic compounds as novel biomarkers for early detection of COPD, asthma, and PRISm: A cross-sectional study. *BMC*.