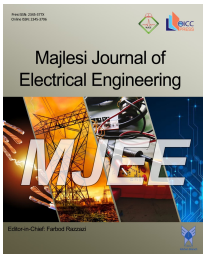

Volume 19, Issue 2, 192532 (1-12)

Majlesi Journal of Electrical Engineering (MJEE)

https://doi.org/10.57647/j.mjee.2025.1902.32

A deep learning approach for chronic obstructive pulmonary disease diagnosis from human exhaled breath gases

Nilakshi Maruti Mule[1,*], Dipti Durgesh Patil[2]

[1]*STES's Smt Kashibai Navale College of Engineering, Savitribai Phule Pune University, Pune, India.*
[2]*Department of Information Technology, MKSSS's Cummins College of Engineering for Women, Savitribai Phule Pune University, Pune, India.*

[*]**Corresponding author:** nilmule@gmail.com

Original Research

Received:
19 December 2024
Revised:
2 March 2025
Accepted:
24 April 2025
Published online:
1 June 2025

Abstract:

A proficient real-time decision support system has the potential to reduce the daily probability of acute exacerbation and loss of control for those suffering from chronic obstructive pulmonary disease (COPD). Applying statistical learning techniques to well-structured, medical E-nose data typically results in high accuracy. Volatile organic compounds or changes by disease processes can be measured in exhaled breath. This work elaborated on the integration of sensors into a sensor array, sampling methodologies, and an algorithm for data analysis. The clinical feasibility of the device was assessed in 40 COPD patients, 20 controls, 8 smokers, and 10 ambient air samples. The classification model utilizing Bi-Directional Long Short-Term Memory (Bi-LSTM) achieved an accuracy, sensitivity, specificity, and area under the curve of 99%, with recall, precision, and F1-score of 1 for COPD classification. The gas sensor array was non-invasive, economical, and provided a quick response. Research has shown that the VOC profiles of COPD patients differ from those of healthy controls, indicating that the E-nose system may serve as a viable diagnostic tool for COPD patients.

Keywords: Bidirectional long short term memory; Chronic obstructive pulmonary disease; Electronic-nose; Exhaled breath; Volatile organic compounds

1. Introduction

Chronic Obstructive Pulmonary Disease significantly burdens patients' daily lives and is a dangerous long-term lung ailment, gradually reducing lung airflow. Emphysema and chronic bronchitis are the two most notable pulmonary phenotypes in the COPD spectrum [1]. The primary causes of COPD are a history of smoking and additional risk factors, which include exposure to biomass, outdoor air pollution, occupational exposure to chemicals and vapors, and recurrent lower respiratory tract infections. The sixth most prevalent global cause of illness is COPD [2]. The most frequent COPD symptoms include breathing problems, tiredness, and a persistent cough. COPD increases the risk of other diseases. These include heart disease, lung cancer, pneumonia, flu-like lung infections, brittle bones, weak muscles, depression, and anxiety [3, 4]. The World Health Organization (WHO) documented a total of 3.23 million deaths from COPD globally in 2019, ranking it as the third most prevalent cause of mortality. The majority of mortality from COPD in individuals under the age of 70 is concentrated in low- and middle-income (LMIC) countries. Smoking is responsible for more than 70% of COPD cases in high-income nations. Smoking is responsible for around $30-40\%$ of COPD patients in LMIC [5, 6].

Thorough assessments are essential for the early diagnosis of COPD in clinical practice. The diagnostic instruments for early COPD are classified into four categories: risk factor identification, physiological assessments, imaging studies, and clinical or laboratory evaluations. Risk factors include tobacco use, childhood infections, respiratory illnesses, air pollution, and genetic predispositions. Physiological assessments encompass airway hyperresponsiveness, cardiovascular exercise testing, and lung clearance index measurement. Imaging studies encompass chest CT, parametric response mapping, hyperpolarized MRI, chronic bronchitis symp-

toms, and the 6-minute walk test in clinical characteristics [7, 8].

The exogenous and endogenous in origin are present in the breath in numbers that are yet unknown. Endogenous VOCs come from all parts of the airway because of metabolic processes in healthy and sick people and from the bloodstream through an interface between alveoli and capillaries. Exogenous VOCs can be changed by physical and biological airway mechanisms [9]. Physical exercise alters the composition of exhaled volatiles, increasing breath pH. Alterations in volatiles may account for an elevation in EBC pH [10]. Similar to physical exercise, modifications in exhaled gases during pregnancy should be considered when conducting breath tests on pregnant women with respiratory disorders [11].

These molecules' accurate detection and quantification are achieved through gas chromatography-mass spectrometry (GC-MS). High costs characterise GC-MS, the necessity for large equipment, and the requirement for specialized expertise. E-noses are cost-effective and user-friendly. These devices cannot identify and quantify molecules in gas mixtures; however, they can compare and discriminate gaseous samples based on the profiles of volatile substances [12–14].

This study employs a cost-effective e-nose system to analyze the VOC profiles of patients with COPD and healthy controls in the exhaled air. The device aims to differentiate COPD patients from healthy controls by analyzing variations in VOC profiles. The Bi-LSTM algorithm classifies subjects into two distinct groups.

This paper is structured as follows: Section 2 outlines the literature concerning the e-nose system. Section 4 outlines the materials and methods employed in this study. Section 6 presents the results and discussion regarding the sensor array responses and data analysis. Section 7 concludes the work.

2. Related work

Eight gas sensors were created by Binson et al. [15] to generate an e-nose system. Principal component analysis (PCA), linear discriminant analysis (LDA), and two-dimensionality reduction approaches were employed, and the ensemble learning methodology XGBoost was used to create the classification model. The system is tested on 30 COPD patients and 40 healthy controls to distinguish the breath samples. Better classification results were obtained by the PCA-XGBoost model, which had 90% specificity, 88.57% accuracy, and 86.67% sensitivity.

Binson et al. [16] focused on the sensor selection procedure, the configuration of sensors into an array, the design and execution of the circuit, the sampling technique, and an algorithm for data interpretation. The clinical feasibility of the system was assessed in 27 lung cancer patients, 22 individuals with COPD, and 39 healthy controls, comprising both smokers and non-smokers. The classification model utilizing a support vector machine (SVM) achieved an accuracy of 88.79%, sensitivity of 89.58%, and specificity of 88.23% for lung cancer and 78.70%, 72.50%, and 82.35% for COPD, respectively.

Cuastumal, Carlos A., et al. designed and created the SOE-I, a wireless electronic scent system for disease detection and monitoring using exhaled breath. The sensory system has been tested on patients with respiratory problems. A sensory system with eight metal oxide gas sensors has been used to condition the breath sample. The PCA method and linear discriminant analysis approach were used to categorize the data set, yielding a clear differentiation of the samples and 92% variance [17].

Further, the study by Maribel Rodriguez-Aguilar et al. aims to distinguish between healthy individuals, people with COPD who smoke (COPD-S), and people with COPD who breathe indoor air pollution (COPD-HAP). A cross-sectional study was done with 294 participants-88 with smoking-related COPD, 28 with HAP, and 178 healthy persons-. The cyranose320 electronic nose is used to analyze breath prints. The diagnostic power of the test has been assessed using receiver operating characteristic (ROC) curves, PCA, CDA, SVM, and group data. According to the findings, patients with COPD have a distinct breath pattern from that of healthy individuals, which accounts for 93.8% of the variability, 97.8% of correctly predicted outcomes, 100% of correctly classified outcomes, and 96.5% and 100%, respectively, of positive and negative predictive values [21]. The n = 364 physician-reported COPD diagnoses have been included by Job JMH van Bragt et al. The data analysis has used processing of the e-nose signal, a correction for ambient air, and statistics using PCA and LDA. LDA may distinguish between patients who have recently experienced an exacerbation and those who have not [22].

The exhaled breath profiles, as evaluated by GC-MS or E-nose, differ between stable disease and exacerbations, as per the theory proposed by P. van Velzen et al. These profiles can be used as biomarkers for COPD exacerbations. Using analysis of covariance ANCOVA, the symptom scores have been connected to the single-variable results of GC-MS and E-nose. Following PCA multivariate modelling, paired t-tests were carried out. Out of 68 patients, 16 breath samples were taken at three points, and 31 experienced an exacerbation. The classification of breath patterns for baseline vs. exacerbation and exacerbation vs. recovery was 71% and 78%, respectively [23].

Martin Thomas Gaugg et al. have determined the affected pathways by employing VOC profiles in breathed air. Using real-time secondary electrospray ionization high-resolution MS, they compared the exhaled breath of 26 frequent and 26 infrequent exacerbators matched for age, sex, and smoking history. Utilizing a Wilcoxon rank-sum test, the metabolites have been found. They have captured an ROC-AUC of 0.88 [24].

A. C. Hauschild et al. examined the breath samples of 35 healthy individuals in the control group and 84 volunteers with COPD or both COPD and bronchial carcinoma. They have standardized and incorporated various statistical learning methodologies to offer a thorough overview of their capacity for differentiating the patient groups. They have distinguished between COPD patients and healthy controls in MCC/IMS chromatograms. They ascertained that the accuracy of the COPD vs CG beast was 94%, and the accu-

racy of the CG vs COPD vs COPD + BC beast was 79% [25, 26]. Simone et al. [27] emphasize the necessity of normalization in technical aspects, such as sensor drift and statistical methods, as well as in sampling factors, including expiratory flow rate and patient-related variables like diet and smoking, to enhance the comparability of results.

Horvath et al. [28] want to establish technical standards and guidelines for sample collecting and analytical methods while emphasizing future research goals. The current document evaluates recent breakthroughs and technological advancements in EBC and FeNO. The authors do not aim to offer clinical recommendations regarding disease diagnosis and management. Binson et al. [40] created an electronic nose system with five metal-oxide semiconductor gas sensors. The study seeks to distinguish COPD and lung cancer from control subjects. The gadget underwent testing with 199 volunteers, comprising 93 controls, 55 individuals with COPD, and 51 lung cancer patients. During the training and validation phase, the ensemble learning technique XGBoost was employed, achieving a classification accuracy of 79.31% for lung cancer and 76.67% for COPD.

Rosdiana et al. [41] developed an electronic nose system for the diagnosis of tuberculosis (TB) and chronic obstructive pulmonary disease (COPD) via the artificial neural network (ANN) approach. The study seeks to ascertain the role of the electronic nose in identifying tuberculosis and chronic obstructive pulmonary disease, as well as to evaluate the sensor response in the design of the electronic nose utilizing Arduino ATmega for illness detection. The data is analyzed quantitatively. Following data analysis, the ANN testing method was conducted to compare COPD and TB using the pattern graph. Tuberculosis exhibits a pattern of (2 2 1 1 0 1), while Chronic Obstructive Pulmonary Disease displays a pattern of (2 1 1 1 0 1). Anant et al. [42] sought to assess the efficacy of exhaled breath profiles by e-nose in differentiating COPD and pulmonary TB from lung cancer. The study analyzed data from 80 healthy individuals: 70 patients with COPD, 60 patients with pulmonary tuberculosis, and 113 patients with lung cancer. R programming was employed to differentiate the algorithms. The classification accuracy for distinguishing lung cancer from COPD and TB was 70% and 80%, respectively.

Using metabolic instruments, Binson et al. [43] developed breath fingerprints in human exhalation to promptly diagnose lung cancer, COPD, and asthma. The study assessed breath samples from 218 individuals, comprising 48 lung cancer patients, 52 COPD patients, 55 asthma patients, and 63 healthy controls. The KPCA-XGBoost model achieved an accuracy of 91.74%, sensitivity of 90.57%, and specificity of 92.65% for lung cancer prediction; 89.84%, 88.14%, and 91.30% for COPD prediction; and 70.66%, 68.75%, and 72.41% for asthma prediction, respectively. Binson et al. [44] differentiate lung cancer and COPD from healthy controls with a metal oxide semiconductor sensor-based electronic nasal system. The clinical feasibility of the system was assessed in 32 lung cancer patients, 38 individuals with COPD, and 72 healthy controls, comprising both smokers and non-smokers. Supervised classification algorithms were employed to categorize exhaled breath samples and quantify the concentration of VOCs inside them. In distinguishing lung cancer from controls, the k-nearest neighbours achieved an accuracy of 91.3%, sensitivity of 84.4%, and specificity of 94.4%. For COPD discrimination, the support vector machine yielded superior results with an accuracy of 90.9%, sensitivity of 81.6%, and specificity of 95.8%. Table 1 shows the summary of COPD detection.

3. Analytical platforms

The volatiles in exhaled breath samples can be examined using various analytical spectrometric methods. Numerous factors must be considered when selecting an analytical process. The benefits and drawbacks of a specific analytical

Table 1. Summary of the COPD detection.

Ref.	Comparison	Sample	Sample size	Analytical platform	Statistical approach	Biomarkers	Accuracy
[18]	COPD vs. Healthy Smokers	End-Tidal Breath	20-Copd, 6-Healthy smokers	GC-DMS	PLS-DA, ANOVA, Kruskal-Wallis	VOCs Patterns	84%
[19]	COPD vs. Asthma vs. Non-smoking control vs. Smoking Control	Expiratory vital capacity	30-Copd, 20-Asthma, 20-Non-smoking control, 20-Smoking	E-nose	PCA	VOCs Patterns	96%
[20]	COPD vs. Smokers vs. Non-Smokers	—-	52-Healthy non-smokers, 52-COPD ex-smokers, 49- Healthy smokers, 37-Smokers with COPD	TD-GC-MS	ANOVA, PCA, LDA, Newman-Keuls test	Linear Alkanes, Alcohols, Aldehydes, Aromatic Compound	89.4%-COPD-ExSmokers, 82.6%-COPD-smokers

Continued of Table 1.

Ref.	Comparison	Sample	Sample size	Analytical platform	Statistical approach	Biomarkers	Accuracy
[29]	COPD vs. Healthy	Alveolar	118-COPD, 63-Healthy	GC-MS	J48, JRIP, PART	Isoprene, Benzene, Toluene, Hexanal, Benzaldehyde, Nonadecane	73.4%
[30]	COPD vs. non-smoking controls vs. smoking controls	Mixed	25-COPD, 25-non-smoking controls, 11-smoking controls	MS	Mann- Whitney U test, KNN.	Acetone	—-
[31]	COPD vs Lung Cancer	—-	682-COPD, 211-Lung Cancer	E-Nose	PCA, LDA	VOCs Pattern	83%
[32]	COPD vs. Control	—-	22-COPD, 14-Controls	SESI-HRMS	PCA, Whitney-Mann U test	Amino acids and related compounds	89%
[33]	COPD vs. Healthy	Alveolar Air	57-COPD, 100-Healthy	GC-MS	Logistic Regression	Hexanal, Heptanal, Nonanal, Propanoic acid, Nonanoic acid	—-
[33]	COPD vs. Healthy	Alveolar Air	57-COPD, 100-Healthy	GC-MS	Logistic Regression	Hexanal, Heptanal, Nonanal, Propanoic acid, Nonanoic acid	—-
[34]	COPD vs. Healthy	End-tidal breath	45-COPD, 23-Healthy smokers, 28-Healthy non-smokers	MCC/IMS	Kolmogorov-Smirnov test, Kruskal Wallis test	VOCs Patterns	70%
[35]	AECOPD vs. COPD vs. Healthy	Alveolar Air	30-COPD, 24-Healthy	TD-GC-toF-MS	ANOVA, Random Forest	Cyclohexane, n-butane, 4-heptanone, 2-pentanone, n-heptane, Methyl propyl sulfide	92%
[36]	COPD vs. with & without AAT deficiency vs. Healthy	Tidal Breath	10- with COPD, 23-without AAT, 10-Healthy	E-nose	LDA, Mann-Whitney U test, Wilcoxon signed rank test	VOCs Pattern	—-
[37]	COPD vs. Healthy	Mixed	50-COPD, 29-Healthy	GC-MS	SVM, Random Forest	Isoprene, C 16 hydrocarbon,4, 7-dimethyl undecane, 2,6- dimethyl-heptane, 4-methyl-octane, Hexadecane, 3, 7-dimethyl, 1, 3, 6-octatriene, 2, 4, 6-trimethyl-decane, Hexanal, Terpineol, Benzonitrile, Octadecane, Undecane	92%
[38]	COPD vs Asthma	—-	60- Asthma, 21-Fixed asthma, 39-Classic asthma, 40-COPD	E-nose	P.C.A., C.D.A.	Carbon Monoxide	88%-COPD vs. Fixed Asthma, 83%- COPD vs. Classic Asthma
[39]	COPD vs. Healthy	—-	23-COPD, 33-Healthy	FGC E-nose	PCA, CDA, Mann Whitney U test, Kolmogorov-Smirnov test	Hydrocarbons, Alcohols, Aldehydes or Ketones	—-

approach must be considered, and it is necessary to check whether offline or online sampling is required. The section below describes the main analytical tools used to study breath volatiles.

3.1 GC-MS based instrumentation

The most effective method for offline measurements is GC-MS. It has a high sensitivity that can occasionally fall into the ppb range and has great potential in identifying and quantifying unidentified components from complicated biological matrices. Samples must be collected in specific bags or absorbent materials and transferred to the laboratories for GC-MS analysis. As a result, samples must be held for days or weeks before examination [45].

3.2 PTR-MS and SESI-MS based instrumentation

While breath samples have been analyzed offline using PTR-MS and SESI-MS, their effectiveness is most significant when utilized online. Accurate identification of the volatiles depends on the success of real-time analysis to detect abrupt changes in volatile concentrations. However, storing the samples is unnecessary because the tests may be performed in the patient's presence. It prevents errors in storage and breath sample deterioration [46].

PTR-MS and SESI-MS demand competent operators as GC-MS requires them. There are fewer PTR-MS studies than GC-MS research because PTR-MS instruments are expensive. A SESI-MS is inexpensive than a GC-MS but hasn't been utilized much [47].

3.3 SIFT-MS based instrumentation

Selected Ion Flow Tube Mass Spectrometry (SIFT-MS) is designed to allow direct and real-time quantification of metabolites in humid breath. This method does not require collecting samples into bags or onto a trap. It might preconcentrate trace VOCs or may damage the breath sample. With a fair degree of precision, SIFT-MS has been used to identify individual metabolites linked to specific clinical and physiological states and to directly identify and quantify many VOCs in a single breath exhalation from a patient or volunteer. The sample volume is less than one part per billion, a detection threshold [48].

3.4 IMS based instrumentation

Ion mobility spectrometry is an analytical tool that can analyze VOCs in real-time or close to it. It analyses on its own or in conjunction with GC columns offering preseparation. The instrumentation costs are significantly lower than the above-described mass spectrometric methods. The size and power requirements are notably reduced because no hoover system is needed. IMS, specifically GC-IMS, is well suited for clinical situations from the maintenance point of view because of its simplicity and robustness. GC-IMS and MCC-IMS are also used to increase the analytical dimensionality [49].

3.5 E-noses based instrumentation

Simple sensors and electronic noses (E-noses) are also components of analytical equipment used for online measurements. They are cheap, simple to use, and can monitor real-time data using algorithms for pattern recognition. Their fundamental shortcomings are a lack of selectivity, inability to detect VOCs, and the possibility of interferences affecting reproducibility, which reduces their dependability and robustness [50].

4. Material and methods

Convolutional Neural Network (CNN) is a popular deep learning technique used mostly for image recognition, object detection, and picture classification. The CNN network used real-time tools like Amazon product recommendations, Facebook face detection, and Google picture search. Convolutional, pooling, and fully connected are the three main layers of the CNN model [51, 52].

The LSTM network, a variant of the RNN, performs exceptionally well in managing sequential data, time series analysis, weather forecasting, picture captioning, and text synthesis tasks [38, 53]. The lengthier sequences network is recursive in the LSTM sequences using a gating mechanism that helps it to remember previous data [54]. The network receives processed input data from the input layers. The network uses memory cells and cell states to store historical data. The forget gate determines whether specific information is necessary or should be deleted [55].

The Bi-LSTM network is an expanded variant of the LSTM network and is a successful model for processing textual data. The ability to read encoder unit contents from hidden decoder unit layers is applied forward and backward [52].

4.1 Dataset description

Mendeley has provided the raw data for the dataset used in this investigation. (https://data.mendeley.com/datasets/h5pcn99zw4/5). The SP-3, MQ-3, TGS-822, MQ-138, MQ-137, TGS-813, TGS-800, and MQ-135 sensors have been employed. Data collection was conducted involving three distinct groups: individuals with COPD, smokers, and healthy individuals. Seventy-eight samples were collected from Pasto Nariño, Colombia. Among the 78 participants, there were 40 samples from individuals with COPD, 8 samples from smokers, 20 from the control group, and 10 from the air. Figure 1 illustrates the sampling protocol.

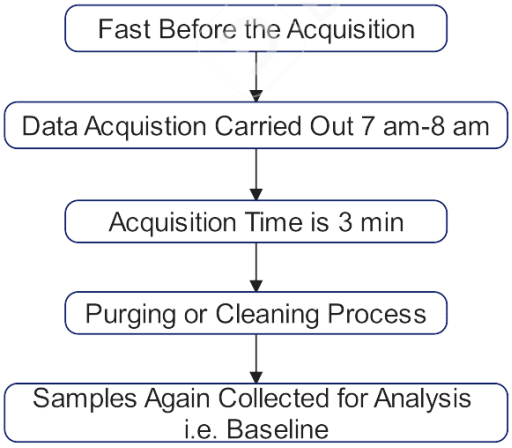

Figure 1. Sampling procedure.

The participant must fast before measurement acquisition to reduce inaccuracy caused by scents or consumed items that may interfere with the exhaled VOCs. The data was collected from 7 am to 8 am. Furthermore, the acquisition duration was approximately 3 minutes. The study's inclusion criteria include participants aged 18 and older, those diagnosed with COPD, and the signed informed consent form requirement. The patient must position their mouth on the disposable mouthpiece affixed to the device's inlet; using a nasal clip is unnecessary. The acquisition occurs when the operator activates the device by pressing the acquisition button, which emits a beep signalling the patient to commence exhalation. During exhale, the sensor array detects the introduced VOCs in the sensor chamber, and the data is collected and stored.

The participants were instructed to exhale continuously and comfortably for the volunteers. Upon completion of the acquisition, the device undergoes a purging or cleaning process, wherein synthetic air or oxygen is introduced under pressure to cleanse the chamber and restore the sensor to its original state [56].

4.2 Proposed methodology

The COPD Mendeley dataset, an open source used in this study to evaluate the model, distinguishes between COPD diseases, smoking, and healthy individuals. This implementation uses the input sequence, Bi-LSTM, fully connected, softmax, and classification output layers. Figure 2 shows the block diagram for the Bi-LSTM. Figure 3 shows the designed Bi-LSTM architecture.

Table 2 contains the parameter values and training parameters for the designed Bi-LSTM architecture.

Table 2 and figure 3 demonstrate the utilization of five Bi-LSTM layers. Based on the information acquired from these layers, the categorization of individuals into either COPD or a healthy state is provided as input to the fully connected layer. The classification process is then finalized using Softmax. This method utilizes parameter values in the suggested architecture, such as the number of neurons in the hidden layer, activation functions, etc. During the training phase, the Adam optimization progressively decreases the error in each iteration. Adam optimization is tailored explicitly to train deep neural networks.

5. Bi-LSTM model

1. Input sequence: Bi-LSTM accepts a token or vector input sequence as its input.
2. Forward LSTM equations: The input sequence X is processed from left to right by the forward LSTM. It saves a hidden state vector at every step that records the pertinent historical data. The previous hidden state is $h_f(t-1)$, and the previous cell state is $c_f(t-1)$ for the input at step t. Here are the forward LSTM equations:

a. Forget gate:

$$f_{ft} = \sigma(W_{ff} \cdot x_t + U_{ff} \cdot h_{f_{t-1}} + b_{ff}) \quad (1)$$

b. Input gate:

$$i_{ft} = \sigma(W_{if} \cdot x_t + U_{if} \cdot h_{f_{t-1}} + b_{if}) \quad (2)$$

c. Candidate cell state:

$$g_{ft} = \tan h(W_{cf} \cdot x_t + U_{cf} \cdot h_{f_{t-1}} + b_{cf}) \quad (3)$$

d. New cell state:

$$c_{ft} = f_{ft} \cdot c_{f_{t-1}} + i_{ft} \cdot g_{ft} \quad (4)$$

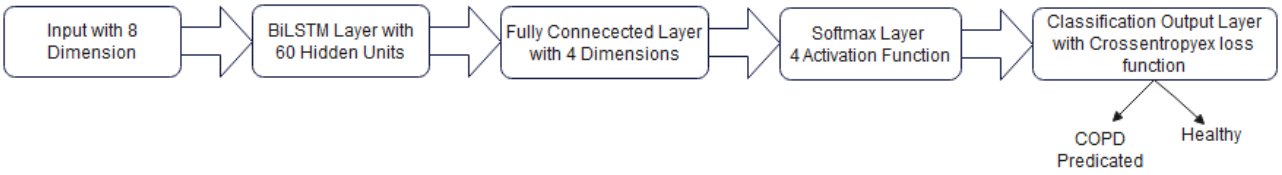

Figure 2. Proposed block diagram of the Bi-LSTM.

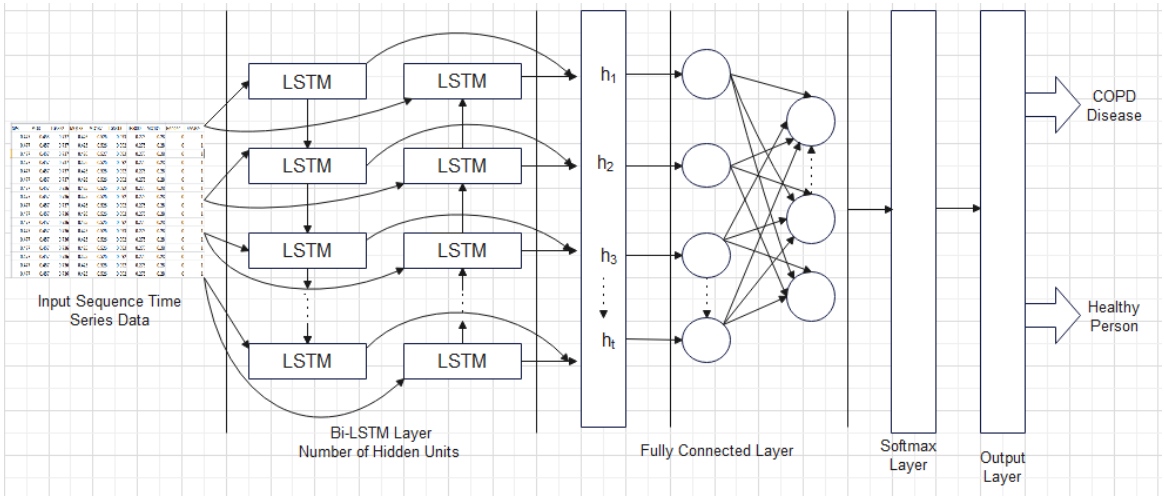

Figure 3. Shows the Bi-LSTM architecture.

Table 2. Hyperparameters of layers used for architecture and training options.

Bi-LSTM						
Number of hidden units		**State activation function**		**Gate activation function**		**Output Size**
100		tanh		sigmoid		8
Training Parameters						
Optimizer	Max. Epoch	Mini Batch Size	Input Size	Number of Hidden Units	Number of Classes	Loss Function
Adam	200	25	8	60	4	Crossentropyex

e. Output gate:

$$o_{ft} = \sigma(W_{of} \cdot x_t + U_{of} \cdot h_{f_{t-1}} + b_{of}) \quad (5)$$

f. Hidden state:

$$h_{ft} = o_{ft} \cdot \tan h(c_{ft}) \quad (6)$$

3. Backward LSTM equations: The input sequence X is processed by the backward LSTM in the opposite direction, going from right to left. A time step is always maintained, and a hidden state vector h_b keeps track of the pertinent information from the future. The previous hidden state is h_b (time step $t+1$), while the previous cell state is c_b (time step t). These are the backward LSTM equations:

a. Forget gate:

$$f_{bt} = \sigma(W_{fb} \cdot x_t + U_{fb} \cdot h_{b_{t+1}} + b_{fb}) \quad (7)$$

b. Input gate:

$$i_{bt} = \sigma(W_{ib} \cdot x_t + U_{ib} \cdot h_{b_{t+1}} + b_{ib}) \quad (8)$$

c. Candidate cell state:

$$g_{bt} = \tan h(W_{cb} \cdot x_t + U_{cb} \cdot h_{b_{t+1}} + b_{cb}) \quad (9)$$

d. New cell state:

$$c_{bt} = f_{bt} \cdot c_{b_{t+1}} i_{bt} \cdot g_{bt} \quad (10)$$

e. Output gate:

$$o_{bt} = \sigma(W_{ob} \cdot x_t + U_{ob} \cdot h_{b_{i+1}} + b_{ob}) \quad (11)$$

f. Hidden state:

$$h_{bt} = O_{bt} \cdot \tan h(C_{bt}) \quad (12)$$

4. Fully connected layer: It adds a bias vector after multiplying the input by a weight matrix. This implementation has four classes: COPD, healthy, smoking, and ambient air.

$$\text{Layers} = \text{fully} - \text{connected layer (numClasses)} \quad (13)$$

5. Softmax and classification layer: A softmax layer is used to apply a softmax function to the input. A softmax layer for classification issues typically follows the final, fully linked layer and a classification layer afterwards.
The corresponding algorithm for the training process of the BiLSTM model is described in algorithm 1.

6. Results and discussion

This study uses the Bi-LSTM model to detect the exhaled breath of a patient with COPD. The data collected is processed using the Bi-LSTM in deep learning and artificial intelligence. Bi-LSTM models are commonly employed to classify sequential data and acquire knowledge about underlying processes. These networks identify enduring connections between the time phases of the data. The collected data is categorized into four groups: ambient air, smoking, control, and COPD patients. The sensor subsequently partitions the data. In figure 4 (a)-(d), the time steps are represented by a horizontal line, while vertical lines show the sensor's analogue responses for each breath sample.
The training portion of the first sequence, i.e., Xtrain, and the testing portion of the second sequence, i.e., Ytrain, comprises the training and testing data set. The declaration of the lengths of the sequences for each observation is prioritized initially. The variable Xtrain quantifies the total number of observations. Depending on the quantity of observations, the sequence is extracted, and the length of the series is specified. Each iteration of the loop results in either an increase or decrease in the sequence length. The Bi-LSTM network architecture comprises five layers and necessitates an input layer with a size of 8. Figure 5 shows the training and testing of the COPD data, and figure 6 shows the 5x1 Bi-LSTM layer.
The final element of the sequence will be output by the Bi-LSTM layer, which has 60 hidden units. The neural network architecture incorporates a fully connected layer whose dimensions are determined by the ailment being considered. It is followed by a softmax and classification layers responsible for assigning classes to the input data. The training choices are configured such that the solver 'Adam' and the gradient threshold are set to 1. The upper limit for the number of epochs is set at 200, while the mini-batch size is fixed at 25. The central processing unit (CPU) is considered a more appropriate choice for training because the mini-batches used in the training process consist of a limited number of concise sequences. The execution environment known as the CPU has been chosen. Once the test set is loaded, the sequences are categorized and sorted into the Xtest and Ytest groups.
There are 23 values selected for Testing. The test values

Algorithm 1. The training process of BiLSTM.

Input: Exhaled breath gases: Time series data $\{X1, X2, \ldots, XT\}$

Output: Classification of COPD disease and healthy persons

1. Load datasets and downsample by 10 to reduce training time
2. For $i = 1$ to size, do
   2.1 Obtain the time series X_T with (1);
3. End for 4. Divide datasets into the number of samples taken XTrain = cell ([78,1]);

// COPD = 40, Healthy = 20, Smoker = 8, Air = 10, Train = 1

5. Train BiLSTM network
6. Get sequence lengths for each observation
7. For each epoch, do
8. For each batch, do
   8.1 The equations for the forward pass of the Bidirectional LSTM model are as follows: 1 through 6
   8.2 The equations for the backward pass of the Bidirectional LSTM model are as follows: From 7 to 12
   8.3 Generate ht and ct via the Bi-Lstm model using the following equation.

$$\overrightarrow{h_t} = \overrightarrow{LSTM}(h_{t-1}, x_t, c_{t-1}), t\varepsilon[1, T]$$
$$\overleftarrow{h_t} = \overleftarrow{LSTM}(h_{t+1}, x_t, c_{t+1}), t\varepsilon[T, 1]$$

   With the input $\overline{x_t}$ and h_{t-1}

9. End for
10. End for
11. Generate fully connected layer and softmax layer
12. Return the prediction matrix

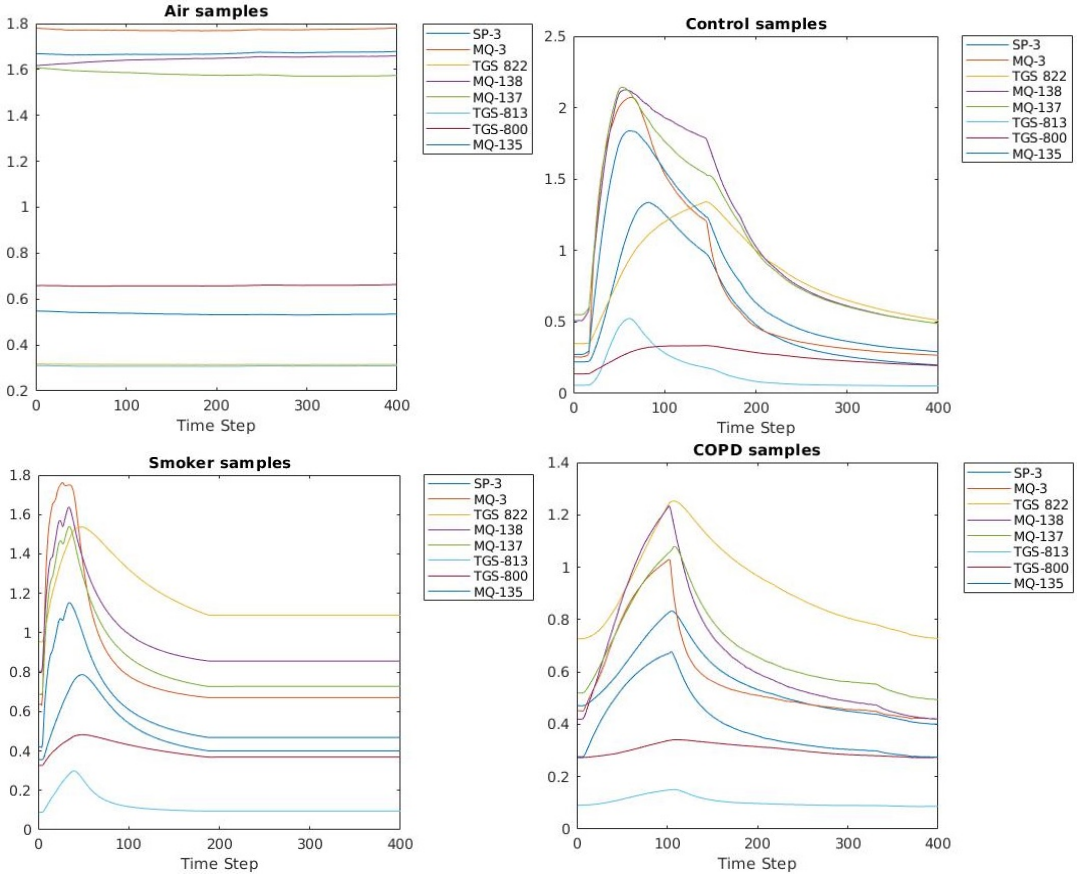

Figure 4. (a) Ambient air samples, (b) Control samples, (c) Smokers samples, (d) COPD Samples

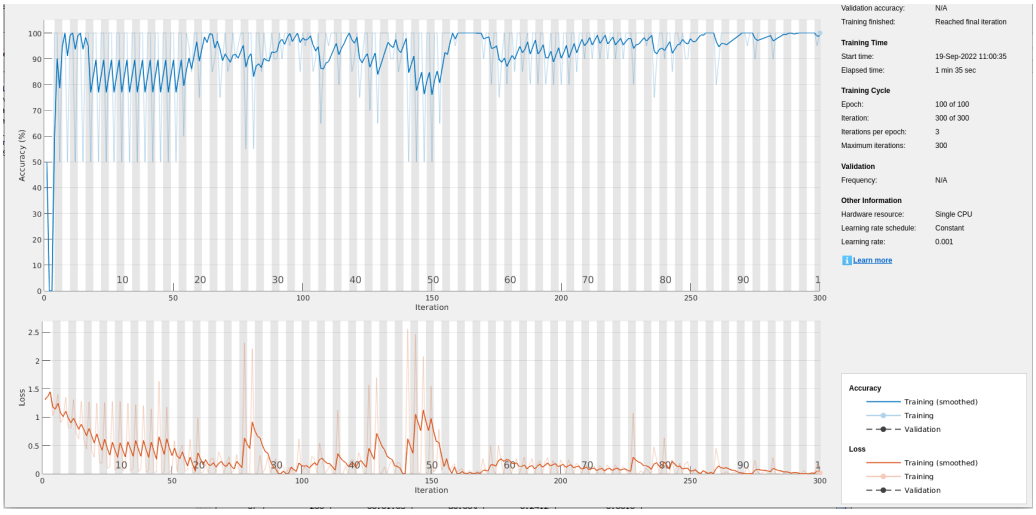

Figure 5. Training and testing of the COPD data.

are set to the output of the randperm (78, 23), which produces a row vector of 23 distinct numbers drawn at random from a range of 1:78. The eight sensors are SP-3, MQ-3, TGS-822, MQ-138, MQ-137, TGS-813, TGS-800, and MQ-135. *X*test is a sequence cell of an array dimension of varied length, and *Y*test is a categorical vector of labels corresponding to each of the eight sensors.

The expiratory flow rate, breath-holding duration, and anatomical dead space may considerably influence "breath prints." The aspects above may impact the classification accuracy of electronic noses in distinguishing diseases from health; by modifying collection-related factors, the discrimination capability may be enhanced [57]. Specific investigations identified notable disparities in the volatile chemical patterns of exhaled breath between healthy individuals and lung transplant recipients, demonstrating a significant correlation with the "breath print" in lung transplant patients. Patients with transplanted lungs can be distinguished from healthy individuals by the profile of exhaled breath volatile organic compounds (VOCs). Post-lung transplantation treatment must be considered while utilizing an e-nose since medicine can significantly affect breath prints [58].

This study has analyzed and displayed the classification results using smokers, healthy individuals, COPD patients, and ambient air data. Various performance indicators are computed to assess performance, including accuracy, recall, specificity, sensitivity, precision, F1 score, and area under curve (AUC). Table 3 displays the performance metrics of

```
layers =

  5×1 Layer array with layers:

     1   ''   Sequence Input          Sequence input with 8 dimensions
     2   ''   BiLSTM                 BiLSTM with 60 hidden units
     3   ''   Fully Connected         4 fully connected layer
     4   ''   Softmax                softmax
     5   ''   Classification Output   crossentropyex
```

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Mini-batch Loss	Base Learning Rate
1	1	00:00:02	40.00%	1.2625	0.0010
17	50	00:00:30	100.00%	0.1662	0.0010
34	100	00:00:42	100.00%	0.0419	0.0010
50	150	00:00:53	72.00%	0.9603	0.0010
67	200	00:01:05	100.00%	0.0413	0.0010
84	250	00:01:19	100.00%	0.0144	0.0010
100	300	00:01:32	100.00%	0.0826	0.0010
117	350	00:01:45	100.00%	0.0307	0.0010
134	400	00:01:57	100.00%	0.0069	0.0010
150	450	00:02:09	100.00%	0.0102	0.0010
167	500	00:02:21	100.00%	0.0213	0.0010
184	550	00:02:33	100.00%	0.0036	0.0010
200	600	00:02:45	100.00%	0.0039	0.0010

Figure 6. 5x1 Bi-LSTM layer.

Table 3. Performance metrics of the proposed architecture.

Accuracy	Recall	Precision	F1 Score	Sensitivity	Specificity	AUC
99%	1	1	1	99%	99%	99%

the suggested architecture, while Figure 7 exhibits the ROC curves.

This work's scientific uniqueness is in selecting a MOS sensor capable of providing precise results in predicting COPD using a deep learning algorithm. The E-nose system is economical, highly portable, and suitable for deployment in distant locations. Most prior studies on respiratory disorders employing electronic noses utilized ten or more sensors. The utilized data yielded favourable outcomes with a minimum of eight gas sensors. The Bi-LSTM time series deep learning method, as presented in Table 3, exhibits an accuracy, sensitivity, and specificity of 99%, along with a recall, precision, and F1-score of 1. This indicates that control data and patients may be easily distinguished. This study compared COPD patients to healthy controls; nevertheless, the sample size was limited.

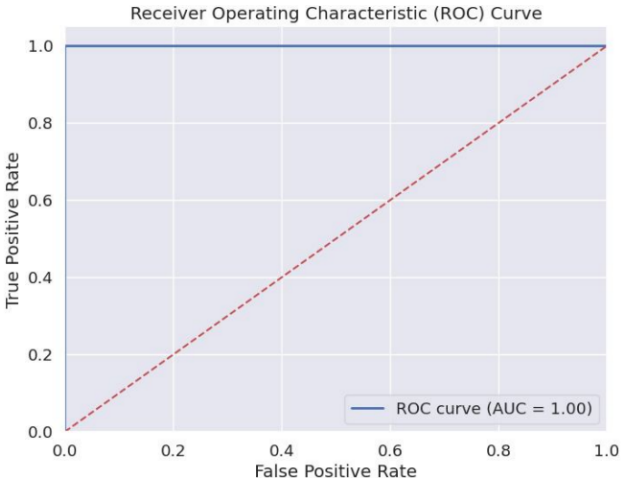

Figure 7. Receiver operating characteristic curve.

7. Conclusion

This study recommends the Bi-LSTM algorithm for differentiating between people with COPD and healthy people. The outcomes demonstrate that the system can distinguish between people with COPD, ambient air, controls, and smokers; the classification achieved is also 99%. In an exhaled breath of COPD patients, acetone, ethanol, benzene, hexane, carbon monoxide, ammonia, toluene, alcohol, and hydrogen gases are more sensitive than the control group. Alcohol gas is more sensitive in smokers than in COPD and control group. This technology benefits hospitals and health facilities, where it is envisioned as a diagnostic aid. The system is used for early disease diagnosis but is limited only to COPD disease, so more diseases are included in the future for real-time disease diagnosis.

Authors contributions
Authors have contributed equally in preparing and writing the manuscript.

Availability of data and materials
The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflict of interests
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] S. Swaminathan, K. Qirko, T. Smith, E. Corcoran, N. G. Wysham, G. Bazaz, G. Kappel, and A. N. Gerber. "**A machine learning approach to triaging patients with chronic obstructive pulmonary disease**.". 12(11):e0188532, 2017.
DOI: https://doi.org/10.1371/journal.pone.0188532.

[2] D. M. Mannino and A. S. Buist. "**Global burden of COPD: Risk factors, prevalence, and future trends**.". *Lancet*, 370(9589):765–73, 2007.
DOI: https://doi.org/10.1016/S0140-6736(07)61380-4.

[3] C. F. Vogelmeier et al. "**Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global strategy for the diagnosis, management, and prevention of COPD 2017 report. GOLD Executive Summary**.". 2017. URL https://doi.org/10.1164/rccm.201701-0218PP.

[4] B. R. Celli and W. MacNee. "**American Thoracic Society/European Respiratory Society Task Force, Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper**.". *Eur Respir J.*, 23(6):932–46, 2004.
DOI: https://doi.org/10.1183/09031936.04.00014304.

[5] J. L. López-Campos, W. Tan, and J. B. Soriano. "**Global Burden of COPD**.". *Respirology*, 21(1):14–23, 2016.
DOI: https://doi.org/10.1111/resp.12660.

[6] M. Rodríguez-Aguilar, L. D. de Leon-Martínez, P. Gorocica-Rosete, R. Perez Padilla, I. Thirion-Romero, O. Ornelas-Rebolledo, and R. F. Ramirez. "**Identification of breath prints for the COPD detection associated with smoking and household air pollution by E-nose**.". *Respiratory Medicine*, 163:105901, 2020.
DOI: https://doi.org/10.1016/j.rmed.2020.105901.

[7] V. A. Binson, M. Subramoniam, and L. Mathew. "**Non-invasive detection of COPD and Lung Cancer through breath analysis using MOS sensor array-based e-nose**.". *Expert Rev Mol Diagn*, 21(11):1223–1233, 2021.
DOI: https://doi.org/10.1080/14737159.2021.1971079.

[8] J. Y. Choi and C. K. Rhee. "**Diagnosis and treatment of early Chronic Obstructive Lung Disease (COPD)**.". *J Clin Med.*, 9(11):3426, 2020.
DOI: https://doi.org/10.3390/jcm9113426.

[9] M. Basanta, B. Ibrahim, R. Dockry, D. Douce, M. Morris, D. Singh, A. Woodcock, and S. J. Fowler. "**Exhaled volatile organic compounds for phenotyping chronic obstructive pulmonary disease: a cross-sectional study**.". *Respiratory Research*, 13(72), 2012.
DOI: https://doi.org/10.1186/1465-9921-13-72.

[10] A. Bikov, Z. Lazar, K. Schandl, B. M. Antus, G. Losonczy, and I. Horvath. "**Exercise changes volatiles in exhaled breath assessed by an electronic nose**.". *Acta Physiol Hung*, 98(3):321–8, 2011. DOI: https://doi.org/10.1556/aphysiol.98.2011.3.9.

[11] A. Bikov, J. Pako, D. Kovacs, L. Tamasi, Z. Lazar, J. Rigo, G. Losonczy, and I. Horvath. "**Exhaled breath volatile alterations in pregnancy assessed with electronic nose**.". *Biomarkers*, 16(6):476–84, 2011. DOI: https://doi.org/10.3109/1354750X.2011.598562.

[12] V. A. Binson, Sania T., Prinu Chacko P., P. Anish T.and Pratap, and Nithin John P. "**Detection of early lung cancer cases in patients with COPD using e-nose technology: A promising non-invasive approach**.". *IEEE International Conference on Recent Advances in Systems Science and Engineering (RASSE)*, 2023. DOI: https://doi.org/10.1109/RASSE60029.2023.10363510.

[13] Binson V. A., P. Mathew, S. Thomas, and L. Mathew. "**Detection of lung cancer and stages via breath analysis using a self-made electronic nose device**.". *Expert Rev Mol Diagn*, 24(4):341–353, 2024. DOI: https://doi.org/10.1080/14737159.2024.2316755.

[14] V. A. Binson, M. Subramoniam, and L. Mathew. "**Prediction of lung cancer with a sensor array based e-nose system using machine learning methods**.". *Microsyst Technol*, 30:1421–1434, 2024. DOI: https://doi.org/10.1007/s00542-024-05656-5.

[15] V. A. Binson, S. Thomas, G. K. Ragesh, and A. Kumar. "**Non-invasive diagnosis of COPD with E-nose using XGBoost algorithm**.". *2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS)*, 2021. DOI: https://doi.org/10.1109/ACCESS51619.2021.9563303.

[16] V. A. Binson and M. Subramoniam. "**Design and development of an e-nose system for the diagnosis of pulmonary diseases**.". *Acta Bioeng Biomech.*, 23(1):35–44, 2021. DOI: https://doi.org/10.37190/ABB-01737-2020-03.

[17] C. A. Cuastumal and C. M. Duran. "**Wireless electronic smell system for the deletion of diseases through the exhaled breath**.". *Chemical Engineering Transactions*, 68:391–396, 2018. DOI: https://doi.org/10.3303/CET1868066.

[18] M. Basanta, R. M. Jarvis, Y. Xu, G. Blackburn, R. Tal-Singer, A. Woodcock, D. Singh, R. Goodacre, C. L. Paul Thomas, and S. J. Fowler. "**Non-invasive metabolomics analysis of breath using differential mobility spectrometry in patients with chronic obstructive pulmonary disease and healthy smokers**.". *Analyst*, 135(2):315–20, 2010. DOI: https://doi.org/10.1039/B916374C.

[19] N. Fens, A. H. Zwinderman, M. P. van der Schee, S. B. de Nijs, E. Dijkers, A. C. Roldaan, D. Cheung, E. H. Bel, and P. J. Sterk. "**Exhaled breath profiling enables discrimination of chronic obstructive pulmonary disease and asthma**.". *American Journal Of Respiratory And Critical Care Medicine*, 180, 2009. DOI: https://doi.org/10.1164/rccm.200906-0939OC.

[20] A. Gaida, O. Holz, C. Nell, S. Schuchardt, B. Lavae-Mokhtari, L. Kruse, U. Boas, J. Langejuergen, M. Allers, and S. Zimmermann. "**A dual center study to compare breath volatile organic compounds from smokers and non-smokers with and without COPD**.". *Journal of Breath Research*, 10(5):1315, 2023. DOI: https://doi.org/10.1088/1752-7155/10/2/026006.

[21] M. Rodríguez-Aguilar, L. Díaz de León-Martínez, P. Gorocica-Rosete, R. Pérez Padilla, I. Thirión-Romero, O. Ornelas-Rebolledo, and R. Flores-Ramírez. "**Identification of breath prints for the COPD detection associated with smoking and household air pollution by electronic nose**.". *Respir Med*, 163:105901, 2020. DOI: https://doi.org/10.1016/j.rmed.2020.105901.

[22] J. J. M. H. van Bragt, P. Brinkman, R. de Vries, S. J. H. Vijverberg, E. J. M. Weersink, E. G. Haarman, F. H. C. de Jongh, S. Kester, A. Lucas, et al. "**Identification of recent exacerbations in COPD patients by electronic nose**.". *ERJ Open Res*, (6), 2020. DOI: https://doi.org/10.1183/23120541.00307-2020.

[23] P. van Velzen, P. Brinkman, H. H. Knobel, J. W. K. van den Berg, R. E. Jonkers, R. J. Loijmans, J. M. Prins, and P. J. Sterk. "**Exhaled breath profiles before, during and after exacerbation of COPD: A prospective follow-up study**.". *COPD*, 16(5-6):330–337, 2019. DOI: https://doi.org/10.1080/15412555.2019.1669550.

[24] M. T. Gaugg, Y. Nussbaumer-Ochsner, L. Bregy, A. Engler, N. Stebler, T. Gaisl, T. Bruderer, N. Nowak, P. Sinues, R. Zenobi, and M. Kohler. "**Real-time breath analysis reveals specific metabolic signatures of COPD exacerbation**.". *Chest*, 156(2):269–276, 2019. DOI: https://doi.org/10.1016/j.chest.2018.12.023.

[25] C. O. Phillips, Y. Syed, N. M. Parthaláin, R. Zwiggelaar, T. C. Claypole, and K. E. Lewis. "**Machine learning methods on exhaled volatile organic compounds for distinguishing COPD patients from healthy controls**.". *J Breath Res.*, 6(3):036003, 2012. DOI: https://doi.org/10.1088/1752-7155/6/3/036003.

[26] A.-C. Hauschild, J. I. Baumbach, and J. Baumbach. "**Integrated statistical learning of metabolic ion mobility spectrometry profiles for pulmonary disease identification**.". *Genet Mol Res.*, 11(3): 2733–44, 2012. DOI: https://doi.org/10.4238/2012.July.10.17.

[27] S. Scarlata, P. Finamore, M. Meszaros, S. Dragonieri, and A. Bikov. "**The Role of Electronic Noses in Phenotyping Patients with Chronic Obstructive Pulmonary Disease**.". *Biosensors (Basel)*, 10 (11):171, 2020. DOI: https://doi.org/10.3390/bios10110171.

[28] I. Horváth et al. "**A European Respiratory Society technical standard: exhaled biomarkers in lung disease**.". *Eur Respir J.*, 49 (4):1600965, 2017. DOI: https://doi.org/10.1183/13993003.00965-2016.

[29] Christopher P., N. Mac Parthaláin, Y. Syed, D. Deganello, T. Claypole, and K. Lewis. "**Short-term infra-subject variation in exhaled volatile organic compounds (VOCs) in COPD patients and healthy controls and its effect on disease classification**.". *Metabolites*, 4(2):300–318, 2014. DOI: https://doi.org/10.3390/metabo4020300.

[30] P. Martinez-Lozano Sinues, L. Meier, C. Berchtold, M. Ivanov, N. Sievi, G. Camen, M. Kohler, and R. Zenobi. "**Breath Analysis in real-time by mass spectrometry in chronic obstructive pulmonary disease**.". *Respiration*, 87(4):301–10, 2014. DOI: https://doi.org/10.1159/000357785.

[31] R. de Vries, N. Farzan, T. Fabius, F. H. C. De Jongh, P. M. C. Jak, E. G. Haarman, E. Snoey, J. C. C. M. Iít Veen, Y. W. F. Dagelet, A.-H. Maitland-Van Der Zee, A. Lucas, M. M. V. Den Heuvel, M. Wolf-Lansdorf, M. Muller, P. Baas, and P. J. Sterk. "**Prospective detection of early lung cancer in patients with COPD in regular care by electronic nose analysis of exhaled breath**.". *Chest.*, pages 00754–7, 2023. DOI: https://doi.org/10.1016/j.chest.2023.04.050.

[32] L. Bregy, Y. Nussbaumer-Ochsner, P. Martinez-Lozano Sinues, D. Garcia-Gómez, Y. Suter, T. Gaisl, N. Stebler, M. T. Gaugg, M. Kohler, and R. Zenobi. "**Real-time mass spectrometric identification of metabolites characteristic of chronic obstructive pulmonary disease in exhaled breath**.". *Clinical Mass Spectrometry*, 7:29–35, 2018. DOI: https://doi.org/10.1016/j.clinms.2018.02.003.

[33] J. J. Jare no Esteban, M. Ángeles Muñoz Lucas, Óscar Gómez-Martín, S. Utrilla-Trigo, C. Gutiérrez-Ortega, A. Aguilar-Ros, L. Collado-Yurrita, and L. M. Callol-Sánchez. "**Study of 5 volatile organic compounds in exhaled breath in chronic obstructive pulmonary disease**.". *Arch Bronconeumol*, 53(5):251–256, 2017. DOI: https://doi.org/10.1016/j.arbres.2016.09.003.

[34] V. Besa, H. Teschler, I. Kurth, A. Maqbul Khan, P. Zarogoulidis, J. Ingo Baumbach, U. Sommerwerck, L. Freitag, and K. Darwiche. "**Exhaled volatile organic compounds discriminate patients with chronic obstructive pulmonary disease from healthy subjects**.". *Int J Chron Obstruct Pulmon Dis.*, 10:399–406, 2015.
DOI: https://doi.org/10.2147/COPD.S76212.

[35] A. Pizzini, W. Filipiak, J. Wille, C. Ager, H. Wiesenhofer, R. Kubinec, J. Blaško, C. Tschurtschenthaler, C. A. Mayhew, G. Weiss, and R. Bellmann-Weiler. "**Analysis of volatile organic compounds in the breath of patients with stable or acute exacerbation of chronic obstructive pulmonary disease**.". *J Breath Res.*, 12(3): 036002, 2023.
DOI: https://doi.org/10.1088/1752-7163/aaa4c5.

[36] A. D. M. Hattesohl, R. A. Jörres, H. Dressel, S. Schmid, C. Vogelmeier, T. Greulich, S. Noeske, R. Bals, and A. R. Koczulla. "**Discrimination between COPD patients with and without alpha 1-antitrypsin deficiency using an electronic nose**.". *Respirology*, 16(8):1258–64, 2011.
DOI: https://doi.org/10.1111/j.1440-1843.2011.02047.x.

[37] J. J. B. N. Van Berkel, J. W. Dallinga, G. M. Möller, R. W. L. Godschalk, E. J. Moonen, E. F. M. Wouters, and F. J. Van Schooten. "**A profile of volatile organic compounds in breath discriminates COPD patients from controls**.". *Respiratory Medicine*, 104(4): 557–563, 2010.
DOI: https://doi.org/10.1016/j.rmed.2009.10.018.

[38] N. Fens, A. C. Roldaan, M. P. van der Schee, R. J. Boksem, A. H. Zwinderman, E. H. Bel, and P. J. Sterk. "**External validation of exhaled breath profiling using an electronic nose in the discrimination of asthma with fixed airways obstruction and chronic obstructive pulmonary disease**.". *Clin Exp Allergy*, 41(10):1371–8, 2011.
DOI: https://doi.org/10.1111/j.1365-2222.2011.03800.x.

[39] D. Rodríguez-Aguilar, S. Ramirez Garcia, C. Ilizaliturri-Hernández, A. Gómez-Gómez, E. Van-Brussel, F. Díaz-Barriga, S. Medellín-Garibay, and R. Flores-Ramírez. "**Ultrafast GC coupled to E-nose to identify volatile biomarkers in exhaled breath from COPD: A pilot study**.". *Biomed Chromatogr*, 33(12):e4684, 2019.
DOI: https://doi.org/10.1002/bmc.4684.

[40] V. A. Binson, M. Subramoniam, and L. Mathew. "**Detection of COPD and lung cancer with electronic nose using ensemble learning methods**.". *Clin Chim Acta.*, 523:231–238, 2021.
DOI: https://doi.org/10.1016/j.cca.2021.10.005.

[41] R. A. P. Rosdiana. "**Design of E-nose for the detection of Tuberculosis (TB) and chronic obstructive pulmonary disease (COPD) using the method of artificial neural network (ANN)**.". *2nd International Conference on Science, Technology, and Modern Society (ICSTMS)*, 2020.
DOI: https://doi.org/10.2991/assehr.k.210909.092.

[42] A. Mohan, B. Ranjan Patnaik, R. Vadala, N. Bhatraju, D. Rai, S. Mittal, et al. "**Discriminating chronic respiratory diseases from lung cancer using exhaled breath signatures by E-nose: A pilot study**.". *LUNG CANCER*, 164(4):A4146–A4147, 2023.
DOI: https://doi.org/10.1016/j.chest.2023.07.2700.

[43] V. A. Binson, M. Subramoniam, Y. Sunny, and L. Mathew. "**Prediction of Pulmonary Diseases With Electronic Nose Using SVM and XGBoost**.". *IEEE Sensors Journal*, 21(18):20886–20895, 2021.
DOI: https://doi.org/10.1109/JSEN.2021.3100390.

[44] V. A. Binson, M. Subramoniam, and L. Mathew. "**Discrimination of COPD and lung cancer from controls through breath analysis using a self-developed e-nose**.". *J Breath Res.*, 15(4), 2021.
DOI: https://doi.org/10.1088/1752-7163/ac1326.

[45] F. Monedeiro, R. B. Dos Reis, F. M. Peria, C. T. Gomes Sares, and B. S. De Martinis. "**Investigation of sweat VOC profiles in assessment of cancer biomarkers using HS-GC-MS**.". *J Breath Res.*, 14(2):026009, 2020.
DOI: https://doi.org/10.1088/1752-7163/ab5b3c.

[46] M. Malásková, B. Henderson, P. D. Chellayah, V. Ruzsanyi, P. Mochalski, S. M. Cristescu, and C. A. Mayhew. "**Proton transfer reaction time-of-flight mass spectrometric measurements of volatile compounds contained in peppermint oil capsules of relevance to real-time pharmacokinetic breath studies**.". *Journal of Breath Research*, 13(4), 2019.
DOI: https://doi.org/10.1088/1752-7163/ab26e2.

[47] P. Španěl and D. Smith. "**Quantification of volatile metabolites in exhaled breath by selected ion flow tube mass spectrometry, SIFT-MS**.". *Clinical Mass Spectrometry*, 16:18–24, 2020.
DOI: https://doi.org/10.1016/j.clinms.2020.02.001.

[48] I. Andreea Ratiu, T. Ligor, V. Bocos-Bintintan, C. A. Mayhew, and B. Buszewski. "**Review Volatile Organic Compounds in Exhaled Breath as Fingerprints of Lung Cancer, Asthma and COPD**.". *J Clin Med.*, 10(1):32, 2020.
DOI: https://doi.org/10.3390/jcm10010032.

[49] C. Szegedy, A. Toshev, and D. Erhan. "**Deep neural networks for object detection**.". *Proc. Adv. Neural Inf. Process. Syst.*, 26:1–9, 2013. URL https://proceedings.neurips.cc/paper_files/paper/2013/file/f7cade80b7cc92b991cf4d2806d6bd78-Paper.pdf.

[50] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders. "**Selective search for object recognition**.". *Int. J. Comput. Vis.*, 104(2):154–171, 2013. URL http://www.huppelen.nl/publications/selectiveSearchDraft.pdf.

[51] G. Dharmale, D. Patil, S. Shekapure, and A. Chougule. "**AI-Based Medicine Intake Tracker**.". *Nature-Inspired Methods for Smart Healthcare Systems and Medical Data*, 2023.
DOI: https://doi.org/10.1007/978-3-031-45952-8_2.

[52] T. Vignesh, K. K. Thyagharajan, R. B. Jeyavathana, and K. V. Kanimozhi. "**Land use and land cover classification using recurrent neural networks with shared layered architecture**.". *Proc. Int. Conf. Comput. Commun. Information. (ICCCI)*, pages 1–6, 2021.
URL 10.1109/ICCCI50826.2021.9402638.

[53] D. A. Meshram and D. D. Patil. "**Predicting Sour or Sweet: Exploring Advance DL Methods for Odor Perception Based on Molecular Properties**.". *International Journal of Intelligent Systems and Applications in Engineering*, 2023. URL https://ijisae.org/index.php/IJISAE/article/view/3433.

[54] G. Hinton and L. van der Maaten. "**Visualizing data using t-SNE**.". 9(86):2579–2605, 2008. URL https://jmlr.org/papers/v9/vandermaaten08a.html.

[55] R. Bhuvaneswari and S. G. Vaidyanathan. "**Classification and grading of diabetic retinopathy images using mixture of ensemble classifiers**.". *J. Intell. Fuzzy Syst.*, 41(6):7407–7419, 2021.
DOI: https://doi.org/10.3233/JIFS-211364.

[56] C. M. Durán Acevedo, C. A. Cuastumal Vasquez, and J. K. Carrillo Gómez. "**Electronic nose dataset for COPD detection from smokers and healthy people through exhaled breath analysis**.". *Data in Brief*, 35:106767, 2021.
DOI: https://doi.org/10.1016/j.dib.2021.106767.

[57] A. Bikov, M. Hernadi, B. Zita Korosi, L. Kunos, G. Zsamboki, Z. Sutto, A. D. Tarnoki, D. L. Tarnoki, G. Losonczy, and I. Horvath. "**Expiratory flow rate, breath hold and anatomic dead space influence electronic nose ability to detect lung cancer**.". *BMC Pulm Med.*, 16(14):202, 2014.
DOI: https://doi.org/10.1186/1471-2466-14-202.

[58] D. Kovacs, A. Bikov, G. Losonczy, G. Murakozy, and I. Horvath. "**Follow-up of lung transplant recipients using an electronic nose**.". *J Breath Res.*, 7(1):017117, 2013.
DOI: https://doi.org/10.1088/1752-7155/7/1/017117. Epub.