



OPEN

Determination of lung cancer exhaled breath biomarkers using machine learning-a new analysis framework

Tlotlo Cassandra Setlhare¹, Atlang Gild Mpolokang¹, Emmanuel Flahaut² & George Chimowa¹✉

Exhaled breath samples of lung cancer patients (LC), tuberculosis (TB) patients and asymptomatic controls (C) were analyzed using gas chromatography-mass spectrometry (GC-MS). Ten volatile organic compounds (VOCs) were identified as possible biomarkers after confounders were statistically eliminated to enhance biomarker specificity. The diagnostic potential of these possible biomarkers was evaluated using multiple machine learning models and their performance for classifying patients and controls was compared. Partial least squares-discriminant analysis (PLS-DA) emerged as the best-performing model for separating lung cancer from controls, with a recall (sensitivity) of 82%, precision of 90%, accuracy of 80% and F1-score of 86%. To further validate this model, TB data was introduced as a confounding disease, and the model achieved precision, recall, accuracy and F1-score of 88% each, in distinguishing lung cancer from TB. These findings address the inter-disease variability and underscores the reliability of the reported VOCs as potential biomarkers of lung cancer. This study establishes a new framework integrating machine learning and confounder elimination for biomarker confirmation.

Lung cancer remains a serious global health challenge, representing one of the deadliest malignancies with increasing prevalence worldwide among all cancers¹. According to the World Health Organization, it claims over 1.8 million lives annually². Biologically, lung cancer is divided into non-small cell lung cancer (NSCLC), which accounts for about 85% of cases, and small cell lung cancer (SCLC), which comprises approximately 15%³. In developing countries, the incidence and mortality rates of lung cancer are constantly on the rise⁴. The insidious onset of lung cancer often results in late stage diagnosis, since initial symptoms such as chest pain, weight loss, coughing, and hemoptysis only emerge as the disease progresses⁵. Therefore, early diagnosis of lung cancer is crucial. Statistics reveal that the 5-year survival rate for early-stage lung cancer is 90% while it plummets to less than 5% for those diagnosed at a late stage⁶. Cancer diagnosis often requires many tests, some of which are invasive surgical procedures. Existing non-invasive methods often have limitations. For example, low-dose spiral CT scans, offers more sensitivity compared to conventional chest X-rays^{7,8}, however, despite its advancements, it yields high false positive rate and exposes patients to radiation, potentially accelerating cancer progression⁹. This underlines the need for a non-invasive, low-cost and rapid diagnostic tool that allows for early detection of lung cancer, which can give patients a fighting chance to overcome the disease.

In response to these challenges, special attention is drawn to volatile organic compounds (VOCs) in exhaled human breath because its easy and comfortable for patients as its non-invasive and can even be applied to patients in intensive care units^{10,11}. Several studies have underscored the viability of this approach^{12,13}. Alternative approaches for exhaled breath analysis such as proton transfer reaction mass spectrometry (PTR-MS)¹⁴, single-photon ionization mass spectrometry (SPI-MS)¹⁵ and selected-ion flow-tube mass spectrometry (SIFT-MS)¹⁶ have been explored. However, they suffer from poor sensitivity and specificity in VOC identification and quantification when compared to the gas chromatography-mass spectrometry (GC-MS)¹⁷. The GC-MS has advantages in the separation of the compounds based on their volatility in the chromatographic column which makes identification more accurate. Previous studies have been devoted to identifying lung cancer biomarkers using various GC-MS conditions^{12,18,19} in conjunction with analytical software tools such as chemstation²⁰, MZmine²¹, Xcalibur²² and R packages²³.

¹Department of Physics and Astronomy, Botswana International University of Science and Technology, Private Bag 16, Boseja Ward, Palapye, Botswana. ²CIRIMAT, Toulouse INP, CNR, Université de Toulouse, 118 Route de Narbonne, 31062 Toulouse Cedex 9, France. ✉email: chimowag@biust.ac.bw

Machine learning methods including random forest, principal component analysis and decision tree, just to name a few, have been utilised to classify lung cancer status based on VOC profiles. Despite all these approaches there is no standardized technique for diagnosing lung cancer via exhaled breath. This explains why there is no single unique biomarker associated with lung cancer, as the VOCs identified vary from one research to the other. The variability could be associated with sample treatment, clinical characteristics, analysis methods and varied external factors. Therefore, there is a need to introduce a method that uses relative VOC concentrations and retention times to identify lung cancer associated biomarkers and also eliminate all VOCs that are influenced by exogenous factors. Previous studies have not been rigorous in dealing with these exogenous factors and this work attempts to address this challenge.

This study focuses on profiling VOCs associated with lung cancer using a single quadrupole GC-MS²⁴ coupled with Openchrom²⁵, Automated Mass Spectral Deconvolution and Identification System (AMDIS)²⁶ and National Institute of Standards and Technology (NIST)²⁷ data library. The study uniquely identifies VOCs not previously linked to lung cancer, potentially offering an early diagnostic technique. By integrating machine learning models and statistical methods, we determine the significance of these VOCs, eliminating those influenced by external factors such as smoking and diet. Furthermore, different histological subtypes and stages of lung cancer can cause breath patterns to vary, adding to the disease's heterogeneity. This study sought to address these challenges. Earlier research work has demonstrated that even with a small dataset, robust statistical and machine learning methods can identify meaningful biomarker trends²⁸.

Results

Ninety-nine exhaled breath samples from 52 participants were analyzed using gas chromatography–mass spectrometry (GC-MS). Figure 1 shows a typical total ion chromatogram (TIC) representing the breath profile of a lung cancer patient. Key volatile organic compounds (VOCs) identified in this study are labeled and annotated based on their retention times and mass spectral matches using the NIST library. The chromatogram gives an overview of the VOC profile observed in a lung cancer patient and displays several distinct peaks corresponding to compounds that were either absent or present at very low concentrations in control samples. This highlights the chemical complexity of exhaled breath in affected individuals. Figure 1 visually summarizes the specific VOCs that may be associated with lung cancer and have potential utility as diagnostic biomarkers.

To validate the performance of the GC-MS instrument, calibration curves were established for o-cymene and hexadecane using external standards (see Supplementary Figures S1 and S2). Both curves showed excellent linearity ($R^2 = 0.998$ and 0.997), confirming that the instrument response was proportional to concentration (Supplementary Table S1). Hexadecane exhibited a significantly steeper slope (1.23×10^9 a.u./ppm) than o-cymene (8548.06 a.u./ppm), indicating higher detector sensitivity. Calculated LODs and LOQs were 4.89/14.83 ppm for o-cymene and 0.08/0.24 ppm for hexadecane, confirming trace-level detection capability. Replicate analysis yielded low RSDs (2.16% and 3.14%), supporting method precision within the accepted <5% range²⁹. VOCs without calibration standards were identified via NIST spectral matching (match and reverse match >80%).

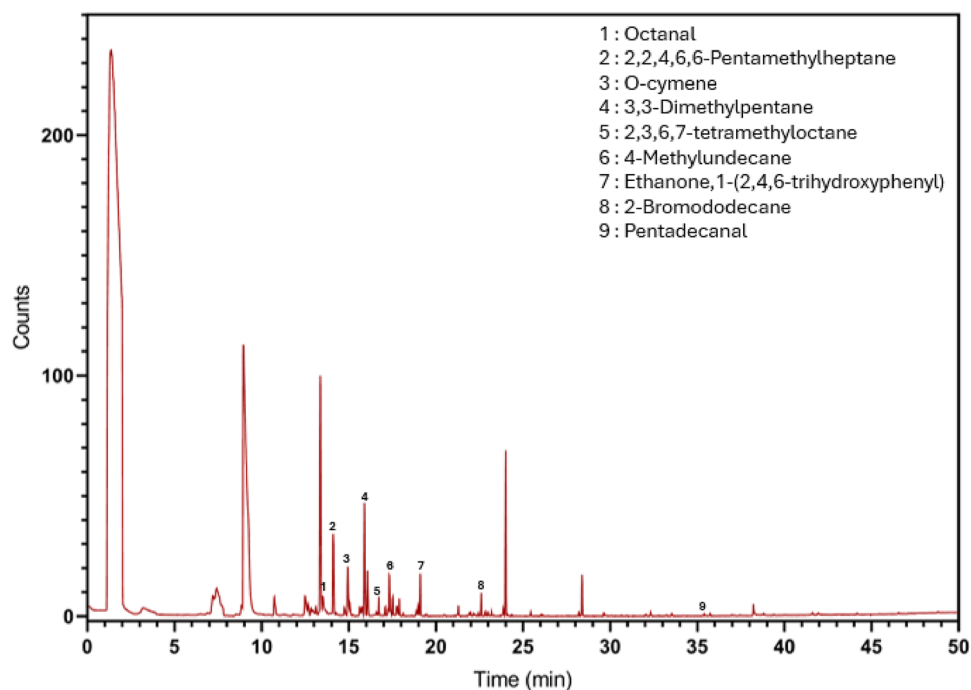


Fig. 1. A typical total ion chromatogram (TIC) of a lung cancer patient's breath profile showing multiple peaks at different retention times. Key VOCs identified in the study are labeled and annotated based on their retention times and mass spectral matches using the NIST library.

Internal standards were not added to patient samples to avoid altering the integrity of the breath samples, a critical concern in diagnostic studies.

In the initial statistical analysis, the criterion for identifying significant VOCs was to ensure that each compound's mass spectrum matched a reference spectrum in the NIST library with a match factor of 80% or higher. The peak areas of the 20 identified VOCs were used as input parameters for statistical analysis. Since most variables did not follow a normal distribution (as confirmed by the Shapiro-Wilk test), a non-parametric Mann-Whitney U test was applied to determine significant differences between groups. All 20 VOCs showed significant differences ($p < 0.05$) between lung cancer patients and controls, showing their potential for lung cancer diagnosis as shown in Table 1. Most importantly, these compounds were found to have elevated levels in lung cancer patients compared to controls, hinting at their association with the disease. Any compound that demonstrated statistical significance ($p < 0.05$) in relation to factors such as smoking history, gender, lung cancer stage, histology and has exogenous association was excluded as a potential biomarker to ensure that the biomarkers identified are not influenced by these variables. The results in Table 1 show that males had considerably higher levels of *hexanedioic acid, bis(2-ethylhexyl) ester* and dodecane than females ($p < 0.05$). While the biological significance of this observation remains uncertain, previous studies suggest that sex-specific differences in lipid metabolism and oxidative stress pathways may influence VOC profiles³⁰. However, confirming any such relationship would require larger, controlled studies designed to specifically investigate gender-based metabolic variations in lung cancer. In our study population, a higher incidence of lung cancer was observed among males compared to females (Fig. 5a), which may reflect both biological and lifestyle-related influences. These findings underscore the importance of controlling for gender-related factors in breath biomarker studies and highlight the need for targeted research into sex-specific metabolic processes that may contribute to disease risk.

Phenyl acetate showed significantly elevated levels in smokers than in non-smokers ($p < 0.05$) and researchers have indicated that smoking can also affect the VOCs in exhaled breath³¹. Additionally, other VOCs such as 1,2-dichlorobenzene, 3-methylheptyl acetate, 2-Phenoxyethanol, 1,2-benzenedicarboxylic acid bis(2-methylpropyl) ester, and decanal were also influenced by both smoking and gender^{32–34}. Chen et al. observed higher levels of decanal in smokers than in non-smokers³⁵. Moreover, 3-methylheptyl acetate, an ester compound, shares structural similarity with esters commonly used as flavorants in tobacco products, suggesting that its presence in breath samples may be related to smoking behavior³⁶. Given that a significant proportion of the male participants in the study were smokers, the gender-based differences observed in the levels of the above mentioned VOCs could likely be attributed to smoking behavior. This suggests that the influence of these VOCs may be more strongly associated with lifestyle factors rather than with lung cancer-specific mechanisms and hence can be eliminated as possible biomarkers.

In our study, the distribution of lung cancer stages depicted in Fig. 5b demonstrates a prevalence of stage III diagnoses in NSCLC patients and stage IV in SCLC patients, aligning with established literature that underscores the typically slower progression of NSCLC compared to the more aggressive progression seen in SCLC³⁷. This observation reflects the differing biological behaviors of these two types of lung cancer and the information is

VOC name	Lung cancer vs Control (p-value)	Histology (NSCLC vs SCLC) (p-value)	Stages (III vs IV) (p-value)	Smoking (p-value)	Gender (p-value)
1,2 Benzenedicarboxylic acid, bis(2- methylpropyl)ester *	0.007	n.s	n.s	0.034	0.045
1,2-Dichlorobenzene	0.07	n.s	n.s	0.025	0.034
2,2,4,6,6-Pentamethylheptane	0.002	n.s	n.s	n.s	n.s
2,3,6,7-Tetramethyloctane	0.006	n.s	n.s	n.s	n.s
2-Bromododecane	0.013	n.s	n.s	n.s	n.s
2-Phenoxyethanol	0.044	n.s	n.s	0.017	0.026
2,5,9-Trimethyldecane	0.07	n.s	n.s	n.s	n.s
3,3-Dimethylpentane	0.003	n.s	n.s	n.s	n.s
3-Methylheptyl acetate	0.024	n.s	n.s	0.040	0.037
4-Methylundecane	0.013	n.s	n.s	n.s	n.s
Decanal	0.014	n.s	n.s	0.048	0.030
Diethyl Phthalate *	0.03	n.s	0.020	n.s	n.s
Dodecane	0.009	n.s	n.s	n.s	0.033
Ethanone 1-(2,4,6-trihydroxyphenyl)	0.003	n.s	n.s	n.s	n.s
Hexadecane	0.013	n.s	0.003	n.s	n.s
Hexanedioic acid, bis(2-ethylhexyl)ester *	0.002	n.s	n.s	n.s	0.023
O-cymene	0.007	n.s	n.s	n.s	n.s
Octanal	0.005	n.s	n.s	n.s	n.s
Pentadecanal	0.027	n.s	n.s	n.s	n.s
Phenyl acetate	0.015	n.s	n.s	0.050	n.s

Table 1. List of VOCs used to study intergroup differences and confounder elimination. Compounds marked with an asterisk (*) are common plasticizers.

crucial as it provides a context for understanding the variations in volatile organic compounds (VOCs) observed across these two major histological groups of lung cancer. Interestingly, several studies suggest that VOC profiles differ across lung cancer subtypes, Corradi et al. showed that ethylbenzene was higher in adenocarcinoma than squamous cell carcinoma³⁸. However, other studies did not find a significant impact of histology on exhaled VOCs^{39–41}, which is consistent with our study that did not identify histology-specific VOCs as we can observe from Table 1 that all the identified compound were not statistically significant (n.s). This suggests that the VOCs identified are not sensitive to a particular lung cancer type. A few studies also compared the lung cancer stages, but there was no specific difference^{40,42,43} noted in the composition of the VOCs for this comparison. However, in our study stage-dependent variations were observed, with hexadecane and *diethyl phthalate* showing elevated level in stage III compared to stage IV. This stage-based variation may reflect metabolic changes associated with tumor progression and highlights the potential of these VOCs for disease monitoring rather than early diagnosis. However, more studies still need to be done to confirm this.

An average of 819 compounds per sample were identified using OpenChrom software. Contaminants associated with the tedlar bag, such as N,N-dimethylacetamide and phenol^{44,45} were excluded and a total of 2265 features were extracted. This huge amount of data requires dimensionality reduction software to interpret and visualize. We first applied Principal component analysis (PCA) as an unsupervised technique to reduce the complexity of the data while retaining the majority of its variance. This was followed by Partial least discriminant analysis (PLS-DA), a supervised method, to visualize the class separation between; lung cancer (blue), TB (green) and control (red), and a prediction model was established as shown in Fig. 2. PLS components, namely PLS1 and PLS2, primarily capture the variations based on the characteristic peaks of the VOCs as identified by their retention times and peak area(s). PLS1 captures the largest variance that is involved in class separation. A significant spread can be observed in this axis indicating that these components are highly effective in distinguishing between the three classes, PLS2 captures additional variance orthogonal to PLS1 and further contributes to the class separation. The green cluster (TB) shows to be well separate from the blue cluster (lung cancer) along PLS1, which suggests that this component captures features in the breath profile that differentiate lung cancer from TB. The tight clustering of the green points indicates consistency in the TB patients and the spread of the blue points indicated that there is great variability in lung cancer samples which could be due to the different types of lung cancer types (NSCLC and SCLC) and different stages. The red cluster (control) is distinct from both green and blue along PLS2, which indicates that this component captures variance related to the separation of the healthy controls from the diseased patients.

There is some observed overlap between blue points (lung cancer) and red points (controls), the observed overlap can be partly attributed to the presence of shared VOCs with a subset of controls. It is noted that these controls fall within the same age range typically associated with early signs of lung cancer⁴⁶. Given this demographic similarity and the asymptomatic nature of the controls, it is conceivable that they might be more susceptible to developing lung cancer as they age, however, this needs further investigations. As observed in

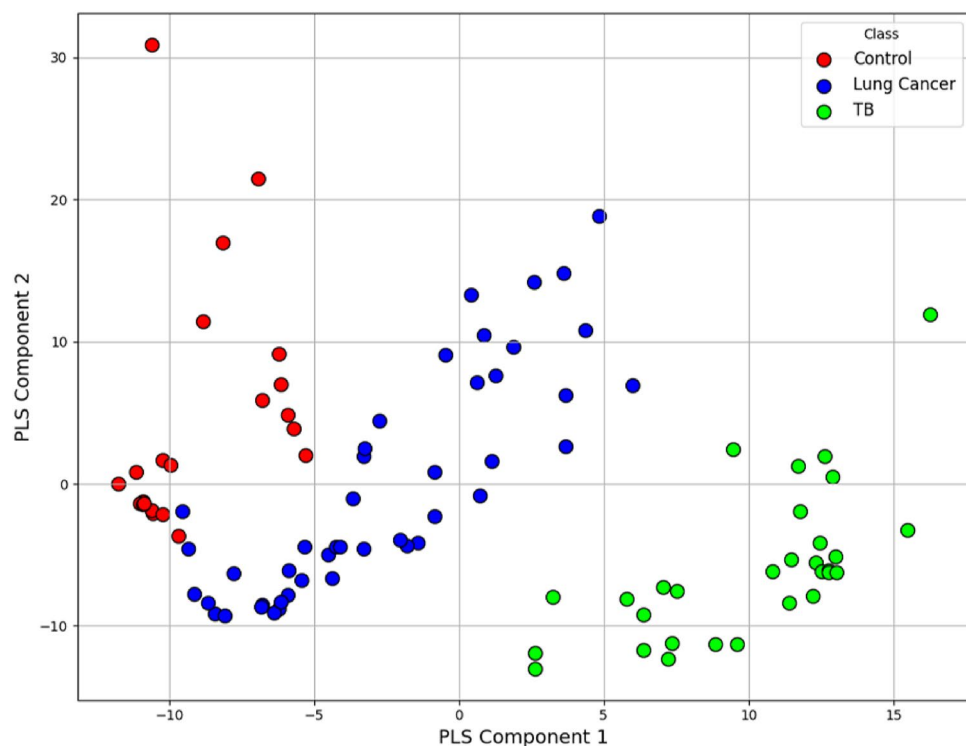


Fig. 2. Score plot from the PLS-DA analysis of lung cancer patients, TB patients and controls. The lung cancer/TB/control status was used as the supervisory variable ($R^2Y = 0.87$ and $Q^2Y = 0.70$).

Table 2, certain compounds identified in lung cancer were also detected in a small subset of control samples. According to the R^2Y value of 0.87 it means that 87% of the variability is explained by the model which indicates that the chosen VOCs are effective in distinguishing between lung cancer patients, TB patients, and controls based on their breath samples. When it is applied to new, unseen data, the model can be able to predict 70% of the variability in the disease condition, as indicated by its Q^2Y of 0.70. This strongly shows that the model is robust and has possible applicability in real-world situations. In addition to fitting the current data well, a high Q^2Y score in relation to R^2Y indicates that the model can also generalize well to new samples. After excluding VOCs affected by gender, smoking history, exogenous factors, lung cancer stages and histology, a refined list of 10 potential VOCs unique to lung cancer was established as shown in Table 2. All compounds exhibited large Cliff's delta effect sizes ($\delta \geq 0.474$), reinforcing their discriminatory power. Notably, 2,2,4,6,6-Pentamethylheptane and 2,3,6,7-Tetramethyloctane showed the highest discrimination, with $\delta = 0.60$ and 0.63 respectively. These VOCs remain strong candidates for breathalyzer development.

In Table 2, three VOCs, 3,3-Dimethylpentane, Octanal and 2,2,4,6,6-Pentamethylheptane have been previously reported as candidate exhaled breath biomarkers for lung cancer. However, to the best of our knowledge, this is the first time that the following VOCs are being reported in association with lung cancer: 2-Bromododecane, Pentadecanal, 2,3,6,7-tetramethyloctane, 4-Methylundecane, Ethanone,1-(2,4,6-trihydroxyphenyl), 2,5,9-Trimethyldecane and o-cymene. This VOCs were all showing a high Variable Importance in Projection (VIP) score (>1) as shown in Table 2, which underscores their contribution to the class separation between lung cancer and controls. The inclusion of both previously reported and new VOCs confirms the reliability of the new identified biomarkers, shown in this study. To evaluate the discriminative power of these VOCs, machine learning classification models such as PLS-DA, Support vector machine (SVM) and K-nearest neighbor(KNN) were performed using Python. Figure 3a shows the bar graph of the prediction metrics (Recall (sensitivity), precision, accuracy and F1-score) of these models for classifying lung cancer patients and controls. PLS-DA outperformed the other models, showing its superior prediction accuracy with a sensitivity of 82%, precision of 90%, accuracy of 80% and F1-score of 86%.

Because of its capability to manage multicollinearity and focus on maximizing the separation between predefined classes, PLS-DA performs better than SVM and KNN. PLS-DA combines dimensionality reduction and class discrimination, unlike SVM, which focuses on identifying the best hyperplane for classification, or KNN, which uses proximity-based classification. PLS-DA uses the variation in the data which is most relevant for class separation, and this is very useful when analyzing high-dimensional datasets such as breath samples with large variability, whereby the number of VOC features is usually more than the number of samples. Furthermore, because of its ability to generate VIP scores it allows for the identification of the most significant compounds that are causing the class separation. In light of these strengths, PLS-DA was further tested by adding TB data as a confounding disease to validate the VOCs that were observed. Since PLS-DA had been successful in differentiating between lung cancer and controls, the goal was to see if it could also distinguish between TB and lung cancer. It is pleasing to report that the model was again very successful in differentiating LC from TB as shown by Figure 2, with a recall, F1-score, precision and accuracy of 88%. This proves that the VOCs identified are potential lung cancer biomarkers and rule out false positives that might be associated with similar metabolic pathways between TB and lung cancer.

To further assess the classification performance of each model, receiver operating characteristic (ROC) curve analysis was performed for each model. The ROC curve for the PLS-DA model, shown in Fig. 3b, demonstrated excellent discriminative ability with an area under the curve (AUC) of 0.96. This AUC confirms the model's ability in distinguishing between lung cancer patients and controls, indicating a very low false positive rate

VOC name	Retention time(minutes) ± 0.5	Molecular mass [base peak] (m/z)	Cas. No	VIP score	Cliffs Delta (δ)	Detection Frequency	
						Lung Cancer (%)	Control (%)
2,2,4,6,6-Pentamethylheptane	14.11	170.0 [57]	13475-82-6	1.80	0.60	35.0	0
2,3,6,7-Tetramethyloctane	16.77	170.0 [43]	52670-34-5	1.51	0.63	50.0	13.6
2-Bromododecane	22.32	248.0 [57]	13187-99-0	1.16	0.51	30.0	4.5
2,5,9-Trimethyldecane	15.10	184.0 [57]	62108-22-9	1.19	0.50	25.0	0
3,3-Dimethylpentane	15.89	100.0 [43]	562-49-2	1.26	0.50	35.0	9.1
4-Methylundecane	17.31	170.0 [43]	2980-69-0	1.08	0.51	30.0	4.5
Ethanone,1-(2,4,6-trihydroxyphenyl)	18.97	168.0 [153]	480-66-0	1.70	0.55	30.0	0
O-cymene	15.01	134.0 [119]	527-84-4	1.38	0.59	40.0	9.1
Octanal	13.61	128.0 [43]	124-13-0	1.17	0.57	35.0	4.5
Pentadecanal	35.57	226.0 [82]	2765-11-9	1.58	0.54	55.0	18.2

Table 2. VOCs selected as possible biomarkers of lung cancer. The new potential Biomarkers reported for the first time are highlighted in bold. The “molecular mass [base peak] (m/z)” column presents the calculated molecular weight alongside the most abundant fragment ion observed in the mass spectrum (base peak). Retention times are given with an approximate variability of ± 0.5 , reflecting minor shifts due to chromatographic conditions and the percentage values represent the detection frequency of each compound within the lung cancer and control groups.

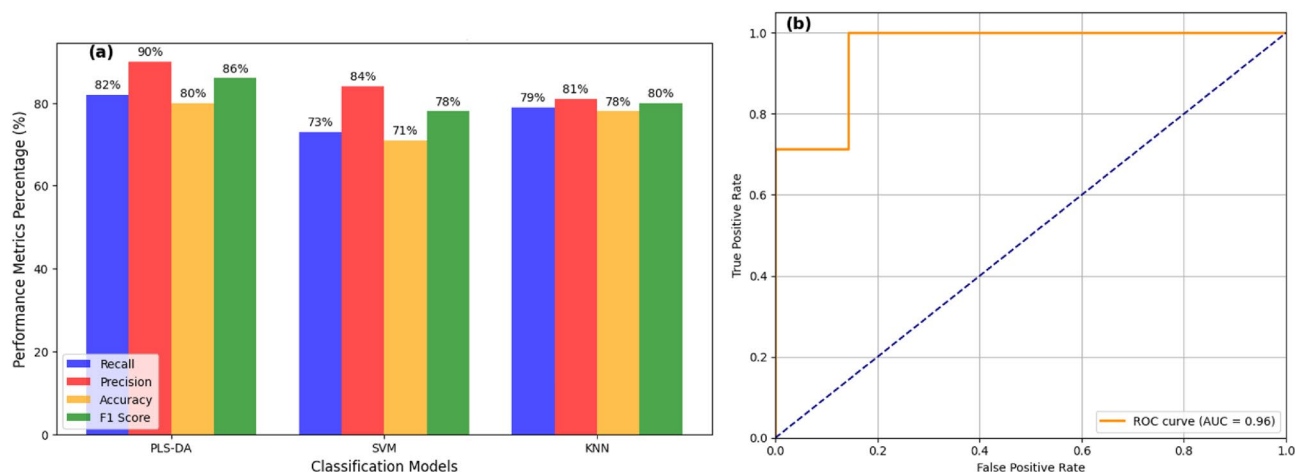


Fig. 3. (a) The performance metrics of the prediction models for classification of lung cancer patients and control and (b) PLS-DA Receiver operating characteristics (ROC) curve for classification of lung cancer patients and controls.

and a high true positive rate. The PLS-DA AUC in comparison to the SVM and KNN AUC of 0.89 and 0.85, respectively (Supplementary Figure S3) underscores the effectiveness of our VOC identification approach and highlight its potential for clinical applications. This research not only advances the field of VOC identification but also lays a strong foundation in developing a lung cancer breathalyzer. This device could transform lung cancer diagnostics, offering a rapid, non-invasive, and cost-effective solution at the point of care, significantly enhancing early detection and, consequently, survival rates.

Discussion

Previous research has shown that VOCs due to oxidative stress are likely not to be disease-specific, as similar pathways are activated in various diseases^{47,48}. However, the more cancer-specific processes, such as gene mutations, altered protein expression, and the Warburg effect, are likely to contribute to the abnormal VOC profiles observed in lung cancer patients⁴⁹. To ensure that the identified VOCs were reliable and specific to lung cancer, we evaluated confounding factors such as smoking, gender, lung cancer stage, and histology. Smoking, a major risk factor for lung cancer, influenced VOCs such as phenyl acetate, and benzene derivatives, consistent with prior studies⁵⁰. VOCs influenced by both gender and smoking (e.g. 1,2-dichlorobenzene, 3-methylheptyl acetate, 2-Phenoxyethanol, 1,2-benzenedicarboxylic acid bis(2-methylpropyl) ester and decanal) highlight the intertwined effects of these variables, reinforcing the importance of controlling for such factors in biomarker identification. Although these compounds were reported in the results section for transparency, all VOCs significantly influenced by exogenous sources (e.g., plasticizers), lung cancer stage, histology, smoking, or gender were excluded from the final biomarker panel (Table 2). This filtering step ensured that only biologically relevant and disease-specific compounds were considered in the interpretation of lung cancer biomarkers. The inclusion of previously reported VOCs, such as 3,3-Dimethylpentane⁵¹, octanal^{38,52–54}, 2,2,4,6,6-Pentamethylheptane^{55,56}, alongside novel candidates; 2-Bromododecane, Pentadecanal, Ethanone,1-(2,4,6-trihydroxyphenyl), 2,5,9-Trimethyldecane, o-cymene, 2,3,6,7-tetramethyloctane (previously detected in lung cancer cell cultures^{57,58}) and 4-Methylundecane (also been identified in lung cancer cell culture⁵⁸) underscores the credibility of our findings. The identification of o-cymene as a potential biomarker is particularly intriguing, given its structural similarity to p-cymene, which has been previously associated with lung cancer⁵⁹ as shown in Fig. 4b. This similarity is attributed to the positional difference of the 1-methylethyl group, suggesting a common source for these VOCs.

Furthermore, the presence of shared VOCs such as o-cymene, which shows varying concentrations across different disease groups (lung cancer, tuberculosis, and controls), as shown in Fig. 4a, was confirmed through external calibration. Despite its higher LOD (4.89 ppm) o-cymene was frequently detected in patient samples, TB group exhibited a much higher median concentration (79.13 ppm) compared to the lung cancer (17.62 ppm) and control (5.54 ppm) group, highlighting the complexity of diagnosing diseases using VOC profiles. These results suggest that o-cymene is a VOC of potential diagnostic relevance, albeit not specific to lung cancer. Notably the results indicated that TB patients exhibit higher concentrations of o-cymene compared to lung cancer patients and healthy controls. This highlights that the need for a comprehensive VOC profile, rather than individual compounds, is essential for differentiating between diseases affecting the same organ. Higher concentrations of o-cymene observed in TB and lung cancer samples compared to controls suggest a potential association with biological processes such as oxidative stress and inflammation. However, the direct role of o-cymene in these pathways remains speculative and warrants further investigation to establish a clear connection and understand its mechanistic significance in these disease states. By considering a broad spectrum of VOCs and their interactions with biological processes such as inflammation, researchers can better understand the underlying mechanisms of disease and enhance the performance metrics of diagnostic tools.

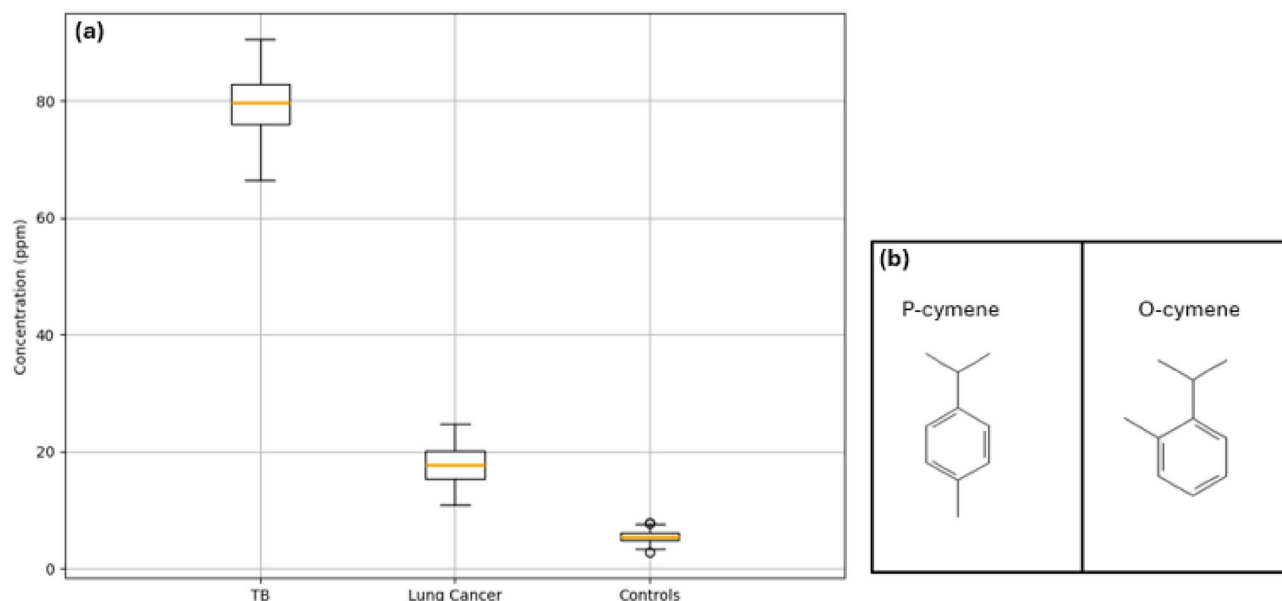


Fig. 4. (a) Box plot showing the concentration of o-cymene (in parts per million, ppm) across different diagnostic groups (lung cancer, tuberculosis, and controls), determined using external standard calibration. (b) Chemical structures of p-cymene and o-cymene presented side-by-side to highlight structural similarities.

The consistent detection of the VOCs in patients unlike in controls likely reflects underlying biological heterogeneity rather than technical variability. This is supported by the strong performance of the PLS-DA model, which achieved sensitivity, precision, accuracy and F1-score of 88% in distinguishing lung cancer from TB, a disease with overlapping oxidative stress signatures. The ROC curve for PLS-DA (AUC = 0.96) further confirms its excellent diagnostic capability (Fig. 3b), outperforming both SVM and KNN models, results are shown in Supplementary Figure S3. The findings of these compounds shows that breath VOC analysis may diagnose lung cancer and provide functional insights into understanding its metabolism.

Conclusion

Exhaled breath analysis presents a promising, non-invasive approach for disease diagnosis. However, the composition of exhaled VOCs is influenced by various factors unrelated to the disease, such as smoking and gender, which must be carefully considered to avoid false-positive or misleading results. This study identified 10 statistically significant VOCs that distinguish lung cancer patients from healthy controls and tuberculosis patients. The integration of advanced statistical analysis and machine learning models (new framework) ensured the reliability of these biomarkers. The superior performance of PLS-DA, with high sensitivity, precision, accuracy and F1-score in distinguishing lung cancer from controls and TB, demonstrates its potential as a robust diagnostic tool. Importantly, the ability to differentiate TB from lung cancer addresses a critical challenge in pulmonary disease diagnostics, as both conditions exhibit overlapping inflammatory and metabolic processes. These findings support the development of a point-of-care breathalyzer for early lung cancer detection. VOC profiling using machine learning is proposed as a new analysis framework. Advanced breath analysis could transform non-invasive cancer screening, boosting early diagnosis and patient outcomes.

Methods

Study population and clinical characteristics

Fifty two (52) participants were enrolled and voluntarily gave their informed consent, with legal guardians consenting where applicable. The cohort included 22 controls, 20 Lung cancer (14-NSCLC and 6-SCLC) patients and 10 active tuberculosis (TB) diagnosed patients. The number of TB patients included in the study was decided to reflect a balance between achieving sufficient statistical significance and accommodating the constraints of patient availability. Over three consecutive days (and in rare cases, two consecutive days) in order to minimize batch effects, 99 breath samples were collected and examined. Lung cancer patients were medically confirmed using histopathology and CT imaging while tuberculosis patients were confirmed via positive sputum culture tests and/or the GeneXpert Mycobacterium tuberculosis/Rifampin (MTB/RIF) assay. Controls were selected based on absence of any known medical or family history of lung cancer, TB and any known lung diseases. Figure 5a shows that all SCLC patients were males, indicating a significantly skewed gender distribution. Figure 5b describes the staging of the disease among participants, based on the Tumor, Lymph nodes, Metastasis (TNM) staging system which utilizes imaging and histological data obtained at diagnosis. It reveals that most NSCLC patients were diagnosed at stage III, whereas SCLC patients were predominantly diagnosed at stage IV.

From Table 3, the majority of SCLC patients are former smokers, whereas NSCLC patients include both former smokers and non-smokers. Participants were between the ages of 17 and 84. The mean age of

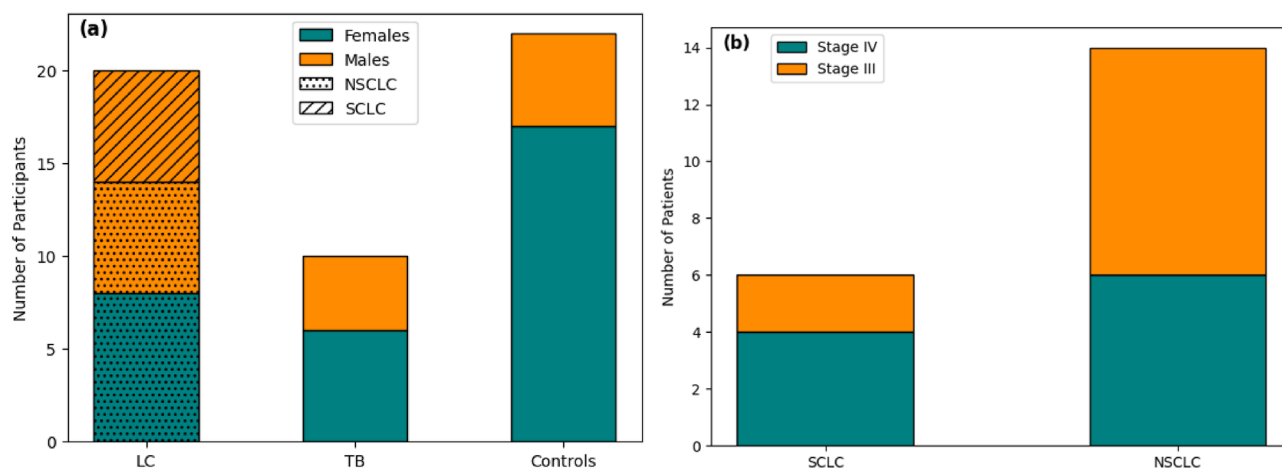


Fig. 5. Bar graph showing distribution of study groups by (a) gender of the participants. (b) lung cancer stage distribution.

	NSCLC	SCLC	Controls	TB
Age(mean)	65	63	48	40
Current smoker	0	0	3	–
Former smoker	6	5	3	–
Never smoked	8	1	16	–

Table 3. Age and smoking history of participants.

NSCLC patients was 65, SCLC patients was 63, that of TB patients was 40 and for controls it was 48. Careful consideration was given in choosing the age range of the controls, to make sure they provided a healthy baseline for comparison and reflected the larger demographics of the patients. The age range of 27 to 77 years for controls was purposefully selected to represent a broad spectrum of the adult population and closely correspond with the age distribution of lung cancer incidence, which generally rise with age. Lung cancer is mostly prevalent in older adults, as seen by the mean age of 65 years for NSCLC patients and 63 years for SCLC patients, which represents typical age ranges for diagnosis. In order to identify any early indicators of lung cancer that might exist prior to the usual age-related increases in incidence, we included younger persons in the control group, beginning at age 27. The relevance and applicability of our findings across various age groups were further improved by extending the age range of controls to 77 years, which permitted direct comparison across a similar lifespan. The same was done for TB, with an age range from 17 to 79 years. By using an age matched technique, it is possible to make sure that variation in VOCs found in breath samples are more likely caused by the illnesses and not by aging-related physiological changes. Furthermore, the broad age range offers reliable information for examining age as a possible confounding variable, enabling more in-depth comprehension and stratification in our research. To ensure a broad sample group, participants included outpatients from different parts of the country as well as inpatients from Princess Marina Hospital and Nyangabgwe Referral Hospital in Gaborone and Francistown, Botswana, respectively. This diversity is important because it includes a broad spectrum of lifestyle characteristics and environmental exposures linked to lung cancer, from big cities to rural locations. Additionally, the geographical diversity of the patients aids in capturing the variation in environmental exposures and genetic backgrounds that impacts the onset and course of the diseases. Breath samples were collected early in the morning before any meal or medication intake, to minimize confounding factors such as medication and diet.

Declaration

All sample collection, handling, and data analysis procedures adhered to guidelines approved by the Ministry of Health and Wellness in Botswana and the ethics committees of the following institutions in Botswana, Human Resources development Council (HRDC), Princess Marina Hospital and Nyangabgwe Referral Hospital. The experimental protocols followed international standards for research involving human participants namely the Declaration of Helsinki.

Sample collection and analytical procedure

A 0.5 liter Tedlar bag (Keika Ventures, USA) with a mouth piece, was used to collect breath samples. The participants were instructed to exhale into the bag until it was almost full while slightly pinching their nostrils to prevent re-inhalation. To minimize compound degradation, the filled tedlar bags were immediately stored in a cold environment and transported to the laboratory for analysis on the same day.

A single quadrupole GC/MS instrument (Agilent 7890B GC coupled with 5977A MSD) was used to analyse the samples. To maintain the chromatographic baseline integrity and make sure no residual chemicals from previous runs impacted the results, a blank run was conducted before each sample analysis. A blank run involved injecting an empty SPME fiber under the same analytical conditions to confirm the absence of carryover from previous runs. Using a solid-phase micro extraction (SPME) fiber (75 μ m Carboxen/Polydimethylsiloxane CAR/PDMS), volatile organic compounds were preconcentrated. The fiber was inserted into the tedlar bag and exposed to the breath sample for a few minutes and then desorbed for 2 minutes at 250°C in the hot GC injector, no chemical derivatization was performed. An Agilent J & W DB-5 capillary column (30 m $\times$ 0.25 mm i.d., 0.25 μ m film thickness) was used to separate the compounds. For a total of 50 minutes of analysis time, the temperature was held at 35°C for 5 minutes before being ramped to 250°C at a rate of 5°C/min and held for an additional 2 minutes. The carrier gas (helium) flow was maintained at a flow rate of 1.2 mL/min in splitless mode. The electron impact ionization was set at 70.0 eV, the quadrupole at 150°C, and the mass spectrometer (MS) ion source at 230°C, with selective ion monitoring (SIM). With a signal threshold of 100, mass spectra were obtained in full scan mode, encompassing the *m/z* range of 25–350, at a scan rate of 1.2 scans/second. The Automated Mass Spectral Deconvolution and Identification System (AMDIS, v2.73) software (<https://chemdata.nist.gov/dokuwiki/doku.php?id=chemdata:amdms>) and OpenChrom Lablicate Edition 1.5.0 (<https://www.openchrom.net/>) were used for chromatographic data acquisition, peak detection, and deconvolution. The NIST 11 MS library (v2.3) was used for compound identification, with minimum match and reverse match threshold of 80/100.

Instrumental quality control (QC) was ensured by injecting authentic standards of *o*-cymene and hexadecane (Sigma Aldrich) at regular intervals across the analytical sequence. These standards were used to monitor retention time stability, peak area reproducibility, and detector performance throughout the study. Authentic standards of *o*-cymene and hexadecane (Sigma Aldrich) were used to validate compound identity using retention time and mass spectral matching. Calibration curves constructed from these standards (see Supplementary Figures S1 and S2) showed excellent linearity ($R^2 = 0.99$) and were used to calculate the detector's sensitivity, limit of detection (LOD), limit of quantification (LOQ) and reproducibility (%RSD) as reported in the Supplementary Table S1. To minimize analytical variability, all samples were analyzed using the same instrument, operator and analytical protocol. Sample injection order was randomized to reduce run-order bias and a single SPME fiber type (CAR/PDMS) was used throughout the study to maintain consistent extraction consistency. For VOCs without available standards, identification was based on mass spectral matching (match factor and reverse match factor above 80%), supported by the instrument's up-to-date calibration and validated retention time performance.

Each breath sample was analyzed once due to the long analysis time (1 hour/sample) and the need to minimize degradation during storage and transport. To assess biological reproducibility, breath samples were collected from the same participants over three consecutive days, and each sample was analyzed separately. This is another additional quality control measure.

Statistical and machine learning application

All GC-MS data were collected in Total Ion Current (TIC) mode and processed using 1-norm normalization in OpenChrom. This method scales each sample such that the sum of absolute VOC intensities equals one, enabling fair comparisons across samples with differing total VOC abundances. A visualization of signal trends before and after normalization is provided in Supplementary Figure S4. These preprocessing steps, including mean-centering and auto-scaling, were essential to reduce variability and prepare the dataset for further analysis. The Mann-Whitney U test was used to ensure robust and reliable statistical analysis. A VOC with a *p*-value less than 0.05 was considered statistically significant. To control for potential confounding factors, subgroup comparisons were conducted to assess the influence of variables such as smoking status, gender, lung cancer stage and histology type. Specifically, the peak area distributions were compared between smokers and nonsmokers, males and females, stage III and stage IV, and NSCLC and SCLC, using the Mann-Whitney U test. VOCs that showed significant differences ($p < 0.05$) in either comparison were identified as potentially influenced by these variables and were excluded from the final biomarker panel to ensure disease specificity.

To ensure biological relevance beyond statistical significance, Cliff's delta (δ) was calculated to quantify the strength of differences between lung cancer and control groups. Compounds with $\delta \geq 0.474$ were considered to have a large effect size and should be considered as potential lung cancer biomarkers. Principal component analysis (PCA) was applied as an unsupervised method to reduce data dimensionality. This helped to simplify the high-dimensional dataset while retaining most of the informative variance in the data. Partial least-squares discriminant analysis (PLS-DA) was then used to analyze differences in VOCs or variables. PLS-DA variables were deemed significant if their Variable Importance in Projection (VIP) score exceeded 1 and the model performance was assessed using R^2 (to evaluate the fit of the model) and Q^2 (to assess the predictability of the model). We also used Support vector machine (SVM) and K-nearest neighbor (KNN) to establish the best prediction model for lung cancer with all the identified VOCs as shown in Figure 6. To validate the model, we randomly split data into a training set (80%) for model derivation and a test set (20%) for model evaluation. Recall (sensitivity), precision, accuracy and F1-score were calculated in the test set to assess classification performance. Receiver Operating Characteristics (ROC) curves were also generated for each model to evaluate the model's overall diagnostic accuracy. To avoid overfitting, three-fold cross-validation was used to validate the models. All statistical analyses were conducted using Jamovi v2.3.28 (<https://www.jamovi.org/>) and Python v3.12.4 in Jupyter Notebook v7.2.1, (<https://jupyter.org/>). The top performing model was subsequently applied to the TB dataset to further validate the identified biomarkers. The schematic workflow in Fig. 6 summarizes the data analysis framework.

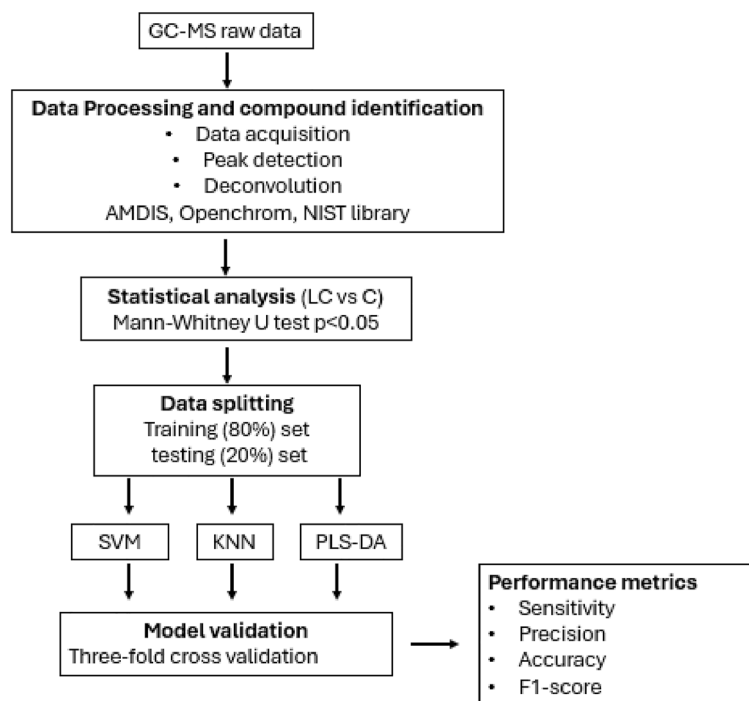


Fig. 6. Workflow diagram for lung cancer VOC analysis.

Data availability

The datasets generated and/or analysed during the current study are not publicly available due to the need to protect the privacy of participants and is still also being used for future work which has not yet been published but are available from the corresponding author on reasonable request.

Received: 2 April 2025; Accepted: 9 July 2025

Published online: 18 July 2025

References

1. Frick, C. et al. Quantitative estimates of preventable and treatable deaths from 36 cancers worldwide: A population-based study. *Lancet Glob. Health* **11**, e1700–e1712 (2023).
2. Li, Y., Wu, X., Yang, P., Jiang, G. & Luo, Y. Machine learning for lung cancer diagnosis, treatment, and prognosis. *Genom. Proteom. Bioinform.* **20**, 850–866 (2022).
3. Edition, S., Edge, S., Byrd, D. et al. AJCC Cancer Staging Manual (2017).
4. Brunetti, A. et al. A machine learning and radiomics approach in lung cancer for predicting histological subtype. *Appl. Sci.* **12**, 5829 (2022).
5. Wang, Y. & Chen, E. Interventional bronchoscopic treatment of lung cancer. *Laparosc. Endosc. Robot. Surg.* **5**, 52–56 (2022).
6. Dhaware, B. U. & Pise, A. C. Lung cancer detection using Bayesian classifier and fcm segmentation. In *2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICADOT)*. 170–174 (IEEE, 2016).
7. Bevilacqua, V. et al. A performance comparison between shallow and deeper neural networks supervised classification of tomosynthesis breast lesions images. *Cognit. Syst. Res.* **53**, 3–19 (2019).
8. Rahouma, K. H., Mabrouk, S. M. & Aouf, M. Lung cancer diagnosis based on Chan-Vese active contour and polynomial neural network. *Proc. Comput. Sci.* **194**, 22–31 (2021).
9. Dandil, E. A computer-aided pipeline for automatic lung cancer classification on computed tomography scans. *J. Healthc. Eng.* **2018**, 9409267 (2018).
10. Schubert, S. J. K., Miekisch, W. & Geiger, K. Exhaled breath markers in acute respiratory distress syndrome. In *Disease Markers in Exhaled Breath*. 394–411 (2002).
11. Schubert, J., Miekisch, W. & Nöldge-Schomburg, G. VOC breath markers in critically ill patients: Potentials and limitations. In *Breath Analysis for Clinical Diagnosis and Therapeutic Monitoring* 267–292 (2005).
12. Gashimova, E., Temerdashev, A., Perunov, D., Porkhanov, V. & Polyakov, I. Diagnosis of lung cancer through exhaled breath: A comprehensive study. *Mol. Diagn. Ther.* 1–14 (2024).
13. Li, X. et al. Volatile organic compounds in exhaled breath: A promising approach for accurate differentiation of lung adenocarcinoma and squamous cell carcinoma. *J. Breath Res.* **18**, 046007 (2024).
14. Jung, Y. J. et al. Advanced diagnostic technology of volatile organic compounds real time analysis analysis from exhaled breath of gastric cancer patients using proton-transfer-reaction time-of-flight mass spectrometry. *Front. Oncol.* **11**, 560591 (2021).
15. Hong, Y. et al. Exhaled breath analysis using on-line preconcentration mass spectrometry for gastric cancer diagnosis. *J. Mass Spectrom.* **56**, e4588 (2021).
16. Tsou, P.-H. et al. Exploring volatile organic compounds in breath for high-accuracy prediction of lung cancer. *Cancers* **13**, 1431 (2021).
17. Scheepers, M. H. et al. Detection of differentiated thyroid carcinoma in exhaled breath with an electronic nose. *J. Breath Res.* **16**, 036008 (2022).

18. Jia, Z., Thavasi, V., Venkatesan, T. & Lee, P. Breath analysis for lung cancer early detection—a clinical study. *Metabolites* **13**, 1197 (2023).
19. Zou, Y. et al. Exhaled metabolic markers and relevant dysregulated pathways of lung cancer: A pilot study. *Ann. Med.* **54**, 790–802 (2022).
20. Bajtarevic, A. et al. Noninvasive detection of lung cancer by analysis of exhaled breath. *BMC Cancer* **9**, 1–16 (2009).
21. Zang, X., Zhang, J., Jiao, P., Xue, X. & Lv, Z. Non-small cell lung cancer detection and subtyping by UPLC-HRMS-based tissue metabolomics. *J. Proteome Res.* **21**, 2011–2022 (2022).
22. Koureas, M. et al. Comparison of targeted and untargeted approaches in breath analysis for the discrimination of lung cancer from benign pulmonary diseases and healthy persons. *Molecules* **26**, 2609 (2021).
23. Team, R. C. R a language and environment for statistical computing. R Foundation for Statistical Computing (2020).
24. Xue, J. et al. Single quadrupole multiple fragment ion monitoring quantitative mass spectrometry. *Anal. Chem.* **93**, 10879–10889 (2021).
25. Wenig, P. & Odermatt, J. Openchrom: A cross-platform open source software for the mass spectrometric analysis of chromatographic data. *BMC Bioinform.* **11**, 1–9 (2010).
26. D'Arcy, P. & Mallard, W. G. Amdis—User guide. US Department of Commerce, Technology Administration, National Institute of Standards and Technology (NIST) [Google Scholar] (2004).
27. Levin, I. NIST inorganic crystal structure database, NIST Standard Reference Database Number 3 (National Institute of Standards and Technology, 2018) (2023).
28. Kuhn, M., Johnson, K. et al. *Applied Predictive Modeling*. Vol. 26 (Springer, 2013).
29. Guideline, I. H. T. et al. Validation of analytical procedures: Text and methodology. Q2 (R1) **1**, 05 (2005).
30. May, L., Shows, K., Nana-Sinkam, P., Li, H. & Landry, J. W. Sex differences in lung cancer. *Cancers* **15**, 3111 (2023).
31. Gordon, S. M., Wallace, L. A., Brinkman, M. C., Callahan, P. J. & Kenny, D. V. Volatile organic compounds as breath biomarkers for active and passive smoking. *Environ. Health Perspect.* **110**, 689–698 (2002).
32. Blanchet, L. et al. Factors that influence the volatile organic compound content in human breath. *J. Breath Res.* **11**, 016013 (2017).
33. Wu, G., Gong, S., He, Y. & Liu, D. Smoking is associated with elevated blood level of volatile organic compounds: A population-based analysis of nhanes 2017–2018. *Arch. Public Health* **81**, 55 (2023).
34. Keith, R. J. et al. Characterization of volatile organic compound metabolites in cigarette smokers, electronic nicotine device users, dual users, and nonusers of tobacco. *Nicotine Tobacco Res.* **22**, 264–272 (2020).
35. Chen, X. et al. Association of smoking with metabolic volatile organic compounds in exhaled breath. *Int. J. Mol. Sci.* **18**, 2235 (2017).
36. Fetterman, J. L. et al. Flavorings in tobacco products induce endothelial cell dysfunction. *Arterioscler. Thrombosis Vasc. Biol.* **38**, 1607–1615 (2018).
37. Ganti, A. K., Klein, A. B., Cotarla, I., Seal, B. & Chou, E. Update of incidence, prevalence, survival, and initial treatment in patients with non-small cell lung cancer in the us. *JAMA Oncol.* **7**, 1824–1832 (2021).
38. Corradi, M. et al. Exhaled breath analysis in suspected cases of non-small-cell lung cancer: A cross-sectional study. *J. Breath Res.* **9**, 027101 (2015).
39. Phillips, M. et al. Detection of lung cancer with volatile markers in the breath. *Chest* **123**, 2115–2123 (2003).
40. Zou, Y. et al. Optimization of volatile markers of lung cancer to exclude interferences of non-malignant disease. *Cancer Biomark.* **14**, 371–379 (2014).
41. Phillips, M. et al. Prediction of lung cancer using volatile biomarkers in breath. *Cancer Biomark.* **3**, 95–109 (2007).
42. Wang, Y. et al. The analysis of volatile organic compounds biomarkers for lung cancer in exhaled breath, tissues and cell lines. *Cancer Biomark.* **11**, 129–137 (2012).
43. Wang, M. et al. Confounding effect of benign pulmonary diseases in selecting volatile organic compounds as markers of lung cancer. *J. Breath Res.* **12**, 046013 (2018).
44. Beauchamp, J., Herbig, J., Gutmann, R. & Hansel, A. On the use of Tedlar® bags for breath-gas sampling and analysis. *J. Breath Res.* **2**, 046001 (2008).
45. Gashimova, E. et al. Investigation of different approaches for exhaled breath and tumor tissue analyses to identify lung cancer biomarkers. *Heliyon* **6** (2020).
46. Key statistics for lung cancer. howpublished <http://www.cancer.org/cancer/types/lung-cancer/about/key-statistics.html>. Accessed 16 Aug 2023 (2023).
47. Denisov, I. G., Makris, T. M., Sligar, S. G. & Schlichting, I. Structure and chemistry of cytochrome p450. *Chem. Rev.* **105**, 2253–2278 (2005).
48. Fruehauf, J. P. & Meyskens, F. L. Jr. Reactive oxygen species: A breath of life or death?. *Clin. Cancer Res.* **13**, 789–794 (2007).
49. Vousden, K. H. & Ryan, K. M. p53 and metabolism. *Nat. Rev. Cancer* **9**, 691–700 (2009).
50. Fowles, J. & Dybing, E. Application of toxicological risk assessment principles to the chemical constituents of cigarette smoke. *Tobacco Control* **12**, 424–430 (2003).
51. Kischkel, S. et al. Breath biomarkers for lung cancer detection and assessment of smoking related effects—Confounding variables, influence of normalization and statistical algorithms. *Clin. Chim. Acta* **411**, 1637–1644 (2010).
52. Poli, D. et al. Determination of aldehydes in exhaled breath of patients with lung cancer by means of on-fiber-derivatization SPME-GC/MS. *J. Chromatogr. B* **878**, 2643–2651 (2010).
53. Fuchs, P., Loeseken, C., Schubert, J. K. & Miekisch, W. Breath gas aldehydes as biomarkers of lung cancer. *Int. J. Cancer* **126**, 2663–2670 (2010).
54. Li, J. et al. Investigation of potential breath biomarkers for the early diagnosis of breast cancer using gas chromatography-mass spectrometry. *Clin. Chim. Acta* **436**, 59–67 (2014).
55. Phillips, M. et al. Volatile organic compounds in breath as markers of lung cancer: A cross-sectional study. *Lancet* **353**, 1930–1933 (1999).
56. Wang, C. et al. Exhaled volatile organic compounds as lung cancer biomarkers during one-lung ventilation. *Sci. Rep.* **4**, 7312 (2014).
57. Chu, Y. et al. Variable VOCs in plastic culture flasks and their potential impact on cell volatile biomarkers. *Anal. Bioanal. Chem.* **412**, 5397–5408 (2020).
58. Schmidt, K. *In Vitro Analysis of Volatile Organic Compounds in Search of Potential Biomarkers of Lung Cancer* (University of Salford (United Kingdom), 2016).
59. Peng, G. et al. Detection of lung, breast, colorectal, and prostate cancers from exhaled breath using a single array of nanosensors. *Br. J. Cancer* **103**, 542–551 (2010).

Acknowledgements

The authors extend their gratitude to all the participants who generously took part in this study, including the lung cancer patients and control subjects. Special thanks are also due to the Ministry of Health and Wellness in Botswana, as well as the staff and facilities of Nyangabgwe and Princess Marina Hospitals, for their invaluable support throughout this research. This work was funded by the European Union (Grant no. DCI-

PANAF/2020/420-028) under the African Research Initiative for Scientific Excellence (ARISE) pilot programme. ARISE is implemented by the African Academy of Sciences with the support of the European Commission and the African Union Commission. The authors bear sole responsibility for the content of this publication, which does not necessarily reflect the views or positions of the European Union, the African Academy of Sciences, or the African Union Commission.

Author contributions

G.C. Is the principal investigator, he conceived the experiment(s) and supervised the whole project, T.C. S. and A.G. M. Conducted the experiment(s) and analysed the results, with T.C.S. being the lead author, E.F. analysed the results. All authors reviewed the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-11365-4>.

Correspondence and requests for materials should be addressed to G.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025