CHEST

# Diagnosing Non-Small Cell Lung Cancer by Exhaled Breath Profiling Using an Electronic Nose

## A Multicenter Validation Study

*Sharina Kort, MD, PhD; Marjolein Brusse-Keizer, PhD; Hugo Schouwink, MD, PhD; Emanuel Citgez, MD; Frans H. de Jongh, PhD; Jan W. G. van Putten, MD, PhD; Ben van den Borne, MD, PhD; Elisabeth A. Kastelijn, MD, PhD; Daiana Stolz, MD, PhD; Milou Schuurbiers, MD; Michel M. van den Heuvel, MD, PhD; Wouter H. van Geffen, MD, PhD; and Job van der Palen, PhD*

**BACKGROUND:** Despite the potential of exhaled breath analysis of volatile organic compounds to diagnose lung cancer, clinical implementation has not been realized, partly due to the lack of validation studies.

**RESEARCH QUESTION:** This study addressed two questions. First, can we simultaneously train and validate a prediction model to distinguish patients with non-small cell lung cancer from non-lung cancer subjects based on exhaled breath patterns? Second, does addition of clinical variables to exhaled breath data improve the diagnosis of lung cancer?

**STUDY DESIGN AND METHODS:** In this multicenter study, subjects with non-small cell lung cancer and control subjects performed 5 min of tidal breathing through the aeoNose, a handheld electronic nose device. A training cohort was used for developing a prediction model based on breath data, and a blinded cohort was used for validation. Multivariable logistic regression analysis was performed, including breath data and clinical variables, in which the formula and cutoff value for the probability of lung cancer were applied to the validation data.

**RESULTS:** A total of 376 subjects formed the training set, and 199 subjects formed the validation set. The full training model (including exhaled breath data and clinical parameters from the training set) were combined in a multivariable logistic regression analysis, maintaining a cut off of 16% probability of lung cancer, resulting in a sensitivity of 95%, a specificity of 51%, and a negative predictive value of 94%; the area under the receiver-operating characteristic curve was 0.87. Performance of the prediction model on the validation cohort showed corresponding results with a sensitivity of 95%, a specificity of 49%, a negative predictive value of 94%, and an area under the receiver-operating characteristic curve of 0.86.

**INTERPRETATION:** Combining exhaled breath data and clinical variables in a multicenter, multi-device validation study can adequately distinguish patients with lung cancer from subjects without lung cancer in a noninvasive manner. This study paves the way to implement exhaled breath analysis in the daily practice of diagnosing lung cancer.

**CLINICAL TRIAL REGISTRATION:** The Netherlands Trial Register; No.: NL7025; URL: https://trialregister.nl/trial/7025

CHEST 2023; 163(3):697-706

**KEY WORDS:** electronic nose; exhaled breath; lung cancer; validation

FOR EDITORIAL COMMENT, SEE PAGE 479

**ABBREVIATIONS:** ANN = artificial neural network; AUC-ROC = area under the receiver-operating characteristic curve; LDCT = low-dose CT; NPV = negative predictive value; NSCLC = non-small cell lung cancer; RF = random forest; VOC = volatile organic compound

**AFFILIATIONS:** From the Department of Respiratory Medicine (S. K., H. S., E. C., and F. d. J.), Medisch Spectrum Twente Enschede; Medical School Twente (M. B.-K. and J. v. d. P.), Enschede; Department of Respiratory Medicine (J. W. G. v. P.), Martini Ziekenhuis, Groningen; Department of Respiratory Medicine (B. v. d. B.), Catharina

## Take-home Points

**Study Question:** Can exhaled breath patterns of patients with NSCLC and without NSCLC adequately be discriminated with an electronic nose in a multicenter, multi-device validation study?
**Results:** Exhaled breath data can adequately distinguish patients with lung cancer from subjects without lung cancer in a noninvasive manner in this multicenter, multi-device study that included 575 subjects. Adding clinical variables relevantly improved the diagnostic performance to diagnose lung cancer.
**Interpretation:** Validation of a prediction model, as performed in this study, is a pivotal step for clinical integration of exhaled breath analysis in the diagnostic path of lung cancer.

Lung cancer is the leading cause of cancer mortality worldwide.[1,2] Its high mortality rate is generally a consequence of advanced-stage disease at the time of initial diagnosis. Despite striking progress in treatment options in advanced-stage lung cancer, such as molecular-targeted therapies and immunotherapy, an essential step to reducing lung cancer mortality is early detection through noninvasive, point-of-care strategies.[3-6]

Exhaled breath contains a gas mixture of thousands of volatile organic compounds (VOCs) in low concentrations that reflect metabolic processes at the tissue level.[7,8] Exhaled breath analysis is based on shifts of this VOC composition due to biochemical changes in different (patho) physiological processes. This method has been investigated extensively in clinical research as a noninvasive tool to diagnose a variety of conditions.[9-11] Studies on pattern recognition for classification of VOC mixtures through nonspecific cross-reactive sensors mimicking human and animal olfaction (eg, electronic noses) as well as identifying individual VOCs by using separation methods (eg, gas chromatography mass spectrometry) have shown promising results in pilot studies for the diagnosis of lung cancer.[12-18]

In addition, studies based on imaging techniques have been shown to be effective for screening purposes in the diagnosis of lung cancer in high-risk asymptomatic subjects. Significant mortality reduction in high-risk subjects was observed in the National Lung Screening Trial (NLST) and the Dutch-Belgian Lung Cancer Screening Trial (NELSON).[19,20] However screening of high-risk subjects has not yet been implemented in Europe. Furthermore, determination of accurate screening criteria remains debatable because only subjects at the highest risk for lung cancer are targeted in current screening programs.

The aeoNose (the eNose Company) is a handheld electronic nose device featuring an array of three metal-oxide sensors that enables real-time breath analysis. The technology and breath sampling method have been described previously in detail.[21,22] Following exposure to VOCs, consecutive conductivity changes at the sensors are recorded, resulting in a digital exhaled breath profile consisting of conductivity values. Exhaled breath profiles of patients with lung cancer can then be distinguished from profiles of non-lung cancer subjects by using artificial intelligence techniques. Once a model has been developed for separating the groups, a new breath profile can be classified using this model. In previous studies, several malignant and nonmalignant conditions have been investigated using the aeoNose.[13,23-25]

We have previously reported the results of a proof-of-concept multicenter study performed with the aeoNose in which a prediction model, based on exhaled breath profiles, was developed using supervised machine learning techniques to discriminate subjects with and without non-small cell lung cancer (NSCLC) in a hospital setting.[13] An artificial neural network (ANN) trained with 290 subjects was able to classify breath samples with a sensitivity of 94%, a specificity of 33%, and an area under the receiver-operating characteristic curve (AUC-ROC) of 0.76. Resampling techniques, including leave-10%-out cross-validation and bootstrapping, were incorporated to reduce the risk of overfitting of the diagnostic model. Adding readily available clinical information (ie, sex, age, number of pack-years, smoking status, COPD status) to the exhaled breath data resulted in a relevant improvement in diagnosing patients with lung cancer.[26]

Ziekenhuis, Eindhoven; Department of Respiratory Medicine (L. K.), Sint Antonius Ziekenhuis, Utrecht, The Netherlands; Clinic for Pulmonary Medicine and Respiratory Cell Research (D. S.), Universitätspital Basel, Basel, Switzerland; Clinic for Respiratory Medicine (D. S.), Medical Center, University of Freiburg, Faculty of Medicine, Freiburg, Germany; Department of Respiratory Medicine (M. S. and M. v. d. H.), Radboud UMC, Nijmegen; Department of Respiratory Medicine (W. H. v. G), Medisch Centrum Leeuwarden, Leeuwarden; and the Universiteit Twente (M. B.-K., F. d. J., J. v. d. P.), Faculty of Behavioural Management and Social Sciences, Enschede, The Netherlands.

**CORRESPONDENCE TO:** Sharina Kort, MD, PhD; email: s.kort@mst.nl

**DOI:** https://doi.org/10.1016/j.chest.2022.09.042

To date, no single breath test has yet been approved for clinical practice to diagnose lung cancer. For this, validation studies are required, preferably involving multiple devices in multiple centers, where part of the data are used for developing a diagnostic model, and the remainder remain blinded to validate the model. Several studies on external validation of breath biomarkers in lung cancer have been performed; however, these studies were aimed at identification of specific VOCs rather than exhaled breath patterns.[14,27,28] Regarding pattern recognition techniques, Fens et al[29] and Bos et al[30] assessed validation of exhaled breath molecular patterns in pulmonary diseases other than lung cancer, based on previous created training sets, showing moderate to high accuracy.

The objective of this prospective multicenter study using multiple devices was to train and subsequently validate a prediction model to distinguish NSCLC patients from subjects initially suspected of lung cancer but considered negative and healthy control subjects, based on their exhaled breath patterns.

## Study Design and Methods

### Study Design and Participants

Participants suspected of having lung cancer were recruited from seven outpatient pulmonary departments between May 2018 and April 2020. The participating hospitals comprised Medisch Spectrum Twente Enschede, Radboud UMC Nijmegen, Medisch Centrum Leeuwarden, Martini Ziekenhuis Groningen, Catharina Ziekenhuis Eindhoven, Sint Antonius Ziekenhuis Utrecht (all in The Netherlands), and Universitätsspital Basel (Switzerland). Each center used one aeoNose device, except for the Basel site, which used two devices. A single aeoNose device needs, as a rule of thumb, a minimum number of 30 observations in the smallest group (in this case, positive measurements) to calibrate the device and hence form reliable conclusions considering the training data; thus, data from devices with an insufficient number of measurements were not used for further analyses.

Subjects suspected of having lung cancer, based on symptom reports or abnormal imaging, were divided into a group with confirmed NSCLC based on pathology and a group with a rejected diagnosis of lung cancer (control subjects) based on imaging and/or pathology. Types of lung cancer other than NSCLC were excluded. Additional healthy control subjects with a minimum age of 55 years were recruited through an alert at the hospitals' websites. In case of pathologically confirmed lung cancer, staging was established according to the eighth edition of the American Joint Committee on Cancer TNM staging system.[31] Patients suspected of lung cancer in whom pathology (ie, the gold standard) was not performed due to insufficient clinical performance were excluded from the analyses. Demographic data and data on some highly prevalent comorbidities (ie, COPD, diabetes mellitus, hypertension) were collected for all subjects. All participants were asked to complete a short questionnaire on recent smoking, eating, and alcohol intake, and were instructed to perform tidal breathing through the non-rebreathing aeoNose device for 5 min with their nose clipped.

The study protocol was approved by the institutional review board of Medisch Spectrum Twente and the board of directors of all participating institutions (e-Appendix 1). All eligible patients provided written informed consent.

The second-generation, CE-certified aeoNose device was used in this study. Because the training study was performed with the first-generation, CE-uncertified device, these previously collected data were deemed not compatible and therefore not used.[13] Instead, we decided to create a split-sample study design in which we enabled development and subsequent validation of new prediction models, which conform to the European Respiratory Society criteria for exhaled biomarkers.[32] Collected breath data were split into a training cohort for supervised learning and internal cross-validation, and a validation cohort, which was kept blinded, for model validation. A random subset of subjects was assigned to the validation cohort, based on the sample size calculation of this validation cohort. Also, an equal prevalence of patients with lung cancer in both sets was taken into consideration.

### Statistical Analysis

Clinical characteristics are reported as means with SDs in case of a normal distribution, or as medians with interquartile ranges. Nominal variables are reported as numbers with corresponding percentages. To assess differences between the groups, *t* tests, *U* tests, or $\chi^2$ tests were applied, as appropriate.

Analysis of exhaled breath data was executed by Aethena, a proprietary software package, incorporating data pre-processing, data compression, machine learning algorithms for classification (eg, ANN, Support Vector Machine, Random Forest [RF], XGBoost, logistic regression), internal validation techniques (leave-10%-out cross-validation and bootstrapping), and model selection. Analyses yielded values between −1 and 1 per subject, indicating the degree to which the subject was classified as having lung cancer (maximum value, 1) or not having lung cancer (minimum value, −1). Details on the software package Aethena have been published previously.[22]

We selected and trained five different models (each using a different classifier: ANN, Logistic Regression, RF, RF Extreme, and XGBoost), with each showing proper discriminative performance. Because the various classifying techniques could interpret the data differently, we envisioned that averaging results over these five models would increase classification robustness. A cutoff value for the probability of lung cancer was determined for the training set to obtain a high sensitivity and negative predictive value (NPV), together with an acceptable number of false-positive cases, as deemed relevant for clinical practice. ROC curves were composed and AUCs were calculated with 95% CIs.

Subsequently, clinical variables (ie, sex, age, number of pack-years, COPD, diabetes, hypertension, BMI, the absolute value obtained from the aeoNose [between −1 and 1]) were entered in a multivariable logistic regression analysis. Nonsignificant variables were eliminated according to the backward method until the fit of the model decreased significantly, based on the −2 log likelihood. Age and sex were included regardless of their significance. A cutoff value for the probability of lung cancer based on this multivariable model was again chosen to obtain a high sensitivity and NPV together with an acceptable number of false-positive cases.

The diagnostic performance of this final logistic regression model, based on the training data, was validated on the blinded data set,

where the β-coefficients were fixed. The same cutoff value, chosen for the training data to determine the presence of lung cancer, was applied to the logistic regression analysis in the validation set. Results are expressed as sensitivity, specificity, predictive values, and AUC-ROC.

A calibration plot was constructed to show how well the predicted probability of lung cancer matches the observed probability of lung cancer.

Stratification for variables to evaluate possible influences on exhaled breath outcomes was performed in explorative analyses for sex, age, presence of COPD, lung cancer stage, and type of histology. Early stage lung cancer was classified as either stage I or II, and late-stage lung cancer was classified as stage III or IV. To compare the final prediction model including breath data and clinical variables vs a nodule calculator, we calculated Spearman rho and the diagnostic performance of both models expressed in terms of sensitivity, specificity, and predictive values.

### Sample Size

Taking into consideration a sensitivity of 95%, as acceptable in clinical practice, estimated with a precision of 5%, with a prevalence of lung cancer in our population of 40%, and an expected specificity of 50%, we would need 183 subjects in our validation cohort. This would lead to a negative predictive value of 93% (95% CI, 0.85-0.98).

SPSS version 24.0 (IBM SPSS Statistics, IBM Corporation) was used for analysis. All statistical tests were two-sided with a significance level at .05.

## Results

A total of 575 subjects were enrolled in the analyses (Fig 1). Approximately two-thirds formed the training set (376 subjects [160 patients with lung cancer, 51 suspected but negative, and 165 healthy control subjects]), and the remaining one-third comprised the validation set (199 subjects [79 patients with lung cancer, 32 suspected but negative, and 88 healthy control subjects]). Subject characteristics are described in Table 1. Data were obtained using five aeoNose devices.

The training model, exclusively based on breath data from the aeoNose, showed, at a cutoff value of –0.36, an AUC-ROC of 0.83 (95% CI, 0.79-0.87), a sensitivity of 91%, a specificity of 54%, and an NPV of 89%. The diagnostic performance of the aeoNose, maintaining the same cutoff value in the validation set, reached an AUC-ROC of 0.79 (95% CI, 0.72-0.85), with a sensitivity of 88%, a specificity of 52%, and an NPV of 87%, which conforms to the training model.

Due to the multicollinearity of smoking status and number of pack-years, we chose to include number of pack-years in these analyses because this parameter contained the most detailed information. The multivariable analysis based on solely clinical data from the training set, including sex, age, and number of pack-years, yielded an AUC-ROC of 0.67 (95% CI, 0.61-0.72); the validation set yielded an AUC-ROC of 0.75 (95% CI, 0.68-0.82).

Exhaled breath data and clinical parameters from the training set were combined in a multivariable logistic regression analysis, maintaining a cutoff of 16% probability of lung cancer, resulting in a sensitivity of 95%, a specificity of 51%, and an NPV of 94%, which was based on clinical relevance (Tables 2-3). This corresponded to an AUC-ROC of 0.87 (95% CI, 0.83-0.90). When applying the identical multivariable logistic regression model on the validation set, maintaining the selected cutoff probability of 16%, we observed a

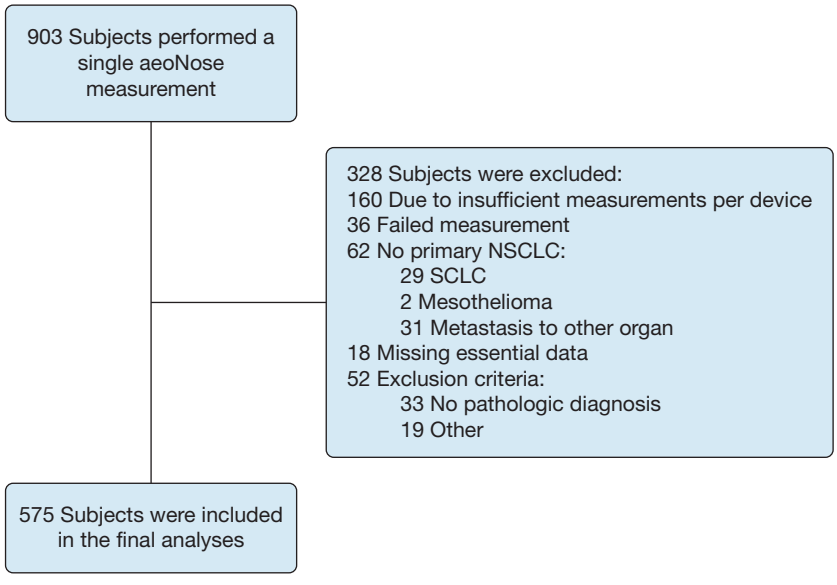


Figure 1 – *Flowchart study cohort. NSCLC = non-small cell lung cancer; SCLC = small cell lung cancer.*

**TABLE 1** ] Clinical Characteristics of All Enrolled Subjects

| Characteristic | Training Set (n = 376) | | | Validation Set (n = 199) | | |
|---|---|---|---|---|---|---|
| | Lung Cancer (n = 160) | Control Subjects (n = 216) | *P* Value | Lung Cancer (n = 79) | Control Subjects (n = 120) | *P* Value |
| Age, mean ± SD, y | 68.4 ± 8.6 | 64.6 ± 8.2 | < .001 | 69.0 ± 7.9 | 63.4 ± 9.4 | < .001 |
| Male | 97 (60.6) | 131 (60.6) | .996 | 49 (62.0) | 59 (49.2) | .075 |
| Smoking status | | | < .001 | | | .001 |
| Current smoker | 48 (30.0) | 41 (19.0) | | 30 (38.0) | 27 (22.5) | |
| Ex-smoker | 103 (64.4) | 130 (60.2) | | 45 (57.0) | 65 (54.2) | |
| Never smoker | 9 (5.6) | 45 (20.8) | | 4 (5.1) | 28 (23.3) | |
| Pack-years[a] | | | < .001 | | | < .001 |
| 0 | 8 (5.1) | 45 (20.8) | | 2 (2.6) | 28 (23.3) | |
| 1-20 | 37 (23.4) | 56 (25.9) | | 14 (18.4) | 38 (31.7) | |
| 21-40 | 52 (32.9) | 55 (25.5) | | 28 (36.8) | 19 (15.8) | |
| > 40 | 61 (38.6) | 60 (27.8) | | 32 (42.1) | 35 (29.2) | |
| COPD | 71 (44.4) | 94 (43.5) | .869 | 37 (46.8) | 52 (43.3) | .627 |
| Hypertension[b] | 66 (41.3) | 74 (34.3) | .166 | 27 (34.6) | 38 (31.9) | .695 |
| Diabetes[b] | 15 (9.4) | 22 (10.2) | .794 | 11 (13.9) | 10 (8.4) | .217 |
| BMI, mean ± SD, kg/m$^2$ | 26.4 ± 4.4 | 25.8 ± 4.7 | .210 | 26.2 ± 5.0 | 25.7 ± 4.4 | .402 |
| Type of NSCLC | | | | | | |
| Adenocarcinoma | 101 (63.1) | | | 39 (50.0) | … | … |
| Squamous cell carcinoma | 43 (26.9) | | | 32 (41.0) | … | … |
| Large cell carcinoma | 6 (3.8) | | | 4 (5.1) | … | … |
| NOS | 10 (6.3) | | | 3 (3.8) | … | … |
| Stage[c] | | | | | | |
| I | 54 (33.8) | | | 21 (26.6) | … | … |
| II | 23 (14.4) | | | 15 (19.0) | … | … |
| III | 38 (23.8) | | | 19 (24.1) | … | … |
| IV | 45 (28.2) | | | 24 (30.4) | … | … |
| Hospital | | | | | | |
| MST | 66 (41.3) | 69 (31.9) | | 30 (38.0) | 30 (25.0) | … |
| Radboud UMC | 31 (19.4) | 29 (13.4) | | 20 (25.3) | 12 (10.0) | … |
| MCL Leeuwarden | 29 (18.1) | 34 (15.7) | | 17 (21.5) | 33 (27.5) | … |
| US Basel | 34 (21.3) | 84 (38.9) | | 12 (15.2) | 45 (37.5) | … |

Data are presented as No. (%) unless otherwise indicated. MCL = Medisch Centrum Leeuwarden; MST = Medisch Spectrum Twente; NOS = not otherwise specified; NSCLC = non-small cell lung cancer.
[a]Five missing subjects.
[b]One missing subject.
[c]According to the eighth edition of the American Joint Committee on Cancer TNM staging system.

sensitivity of 95%, a specificity of 49%, a positive predictive value of 54%, and an NPV of 94%, with a corresponding AUC-ROC of 0.86 (95% CI, 0.81-0.91) (Fig 2, Table 3). In case of this cutoff probability of 16%, 63 of the 196 subjects (32%) were classified as "no lung cancer" (Table 4). Corresponding performance of breath data only, with an equal cutoff probability of lung cancer in the training and validation set, is also displayed in Table 3.

A calibration plot with the predicted probability of lung cancer in deciles of the validation cohort is shown in e-Figure 1. The figure shows good concordance between the predicted probability of lung cancer in each decile, and the observed prevalence of lung cancer in the same decile.

Explorative subgroup analyses show equal performance of the aeoNose in early- and late-stage lung cancer, in

**TABLE 2** ] Results of the Multivariable Logistic Regression Analysis for Diagnosing Lung Cancer

| Variable | Multivariable Analysis: OR (95% CI) | Regression Coefficient (β) |
|---|---|---|
| | | −4.949 (intercept) |
| Sex (ref female) | 0.68 (0.38-1.19) | −0.393 |
| Age | 1.05 (1.02-1.09) | 0.049 |
| Pack-years | | |
| 0 | Ref | |
| 1-20 | 5.19 (1.91-14.1) | 1.647 |
| 21-40 | 8.11 (3.02-21.76) | 2.092 |
| > 40 | 8.69 (3.22-23.50) | 2.162 |
| Classification value aeoNose | 27.9 (14.0-55.5) | 3.328 |

Ref = reference.

both sexes, different age groups, and different types of histology (e-Tables 1-12). In stage I and II lung cancer, sensitivity and NPV were 94% and 97%, respectively; in stage III and IV lung cancer, sensitivity and NPV were 84% and 90%. Furthermore, we compared our full model with the Mayo Clinic nodule calculator that included the following clinical variables: age, smoking history, previous malignancy, nodule size, upper lobe, spiculation, and PET scan result (e-Tables 13-14).[33,34] This analysis included 55 patients who met the inclusion criteria for this nodule calculator. A fair correlation was observed between the predicted value of lung cancer being present with a Spearman rho of 0.31 ($P = .021$). When using our predefined cutoff probability of lung cancer of 16% for both the full model, including breath data and clinical variables, and for the Mayo Clinic nodule calculator, the sensitivity was 94.6% and 83.8% for the study model and the Mayo Clinic model. Similarly, specificity was 22.2%, and 61.1%, NPV was 66.7% and 64.7%, and positive predictive value was 71.4% and 81.6%.

## Discussion

In the current study, we trained and subsequently validated exhaled breath data to distinguish between patients with NSCLC and clinically relevant control subjects in a multicenter setting using multiple devices. The findings show that patients with NSCLC can successfully be discriminated from subjects without NSCLC by using exhaled breath patterns based on a training set concentrating on a high NPV to exclude the diagnosis of lung cancer in a noninvasive manner. Discrimination between both groups improves significantly when readily available clinical variables (ie, age, sex, number of pack-years) are added to the prediction model. Classifying new subjects, not used for training of the aeoNose, exhibited excellent performance. The multivariable analysis based on solely clinical data showed worse discrimination compared with the full model, underlining the important additional value of exhaled breath data.

Our previously performed training study indicated that exhaled breath patterns differ between patients with lung cancer and subjects without lung cancer.[13] The current study provided the necessary essential step in which a prediction model based on a training set was validated on "blind" subjects in a multicenter and multinational setting, using multiple devices.

To our knowledge, this is the first NSCLC study to validate blinded exhaled breath profiles based on pattern recognition techniques in a multicenter split-sample design, including readily available clinical variables, while using multiple electronic nose devices. Previously, Machado et al[35] performed a similar study in which they used a split-sample design to validate a prediction model to distinguish patients with lung cancer from control subjects. However, although they reported promising results, the study was performed in a single-center setting and had a very small study population (14 individuals with bronchogenic carcinoma in both the training phase and validation phase, respectively). In addition, Mazzone et al[16] performed a split-sample study design using pattern recognition techniques based on exhaled breath to distinguish patients with lung cancer from control subjects. This concerned a two-center study with the application of only one electronic nose device, and diagnostic performance in the validation set could be considered moderate. A recent study of Long et al[27] showed interesting results in an external validation study of exhaled breath biomarkers

**TABLE 3** ] Diagnostic Accuracy of Exhaled Breath Analysis in the Training and Validation Set

| Prediction Model | Cutoff Probability[a] | Sensitivity | Specificity | PPV | NPV | AUC-ROC (95% CI) |
|---|---|---|---|---|---|---|
| Training clinical data only[b] | 20% | 47.5 | 76.4 | 59.5 | 66.5 | 0.67 (0.61-0.72) |
| Validation clinical data only[a,b] | 20% | 53.9 | 77.5 | 60.3 | 72.7 | 0.75 (0.68-0.82) |
| Training breath data only[b] | 20% | 93.0 | 54.2 | 59.8 | 91.4 | 0.83 (0.79-0.87) |
| Validation breath data only[a,b] | 20% | 88.2 | 48.3 | 51.9 | 86.6 | 0.79 (0.72-0.85) |
| Training clinical parameters + breath data[b] | 16% | 94.9 | 50.5 | 58.4 | 93.2 | 0.87 (0.83-0.90) |
| Validation clinical parameters + breath data[a,b] | 16% | 94.7 | 49.2 | 54.1 | 93.7 | 0.86 (0.81-0.91) |

AUC-ROC = area under the receiver-operating characteristic curve; NPV = negative predictive value; PPV = positive predictive value.
[a]Corresponding cutoff values and fixed β-coefficients based on logistic regression analyses in the training set.
[b]All analyses are performed in subjects without missing data (training data, n = 374; validation data, n = 196).

to diagnose lung cancer. Although the investigators used the gas chromatography mass spectrometry technique, with several Tedlar bags and one gas chromatography mass spectrometry station, to identify molecules in exhaled breath, they also focused on the possible origin of breath biomarkers by explaining specific metabolic processes in lung cancer pathogenesis. However, this strict study protocol cannot easily be implemented in daily clinical practice and, contrary to the aeoNose, it does not offer a point-of-care solution.

The reported AUC-ROC of 0.86 in the current study as obtained by the multivariable validation model provides very good accuracy but is lower than some of the reported accuracies by other studies using pattern recognition techniques in exhaled breath analysis to diagnose lung cancer.[12,18,36-38] Possible explanations for these discrepancies are incomparable study designs and control groups, a single-center vs multicenter setting, small data sets with the inherent risk of overfitting of models, the use of different sensor technologies, use of a

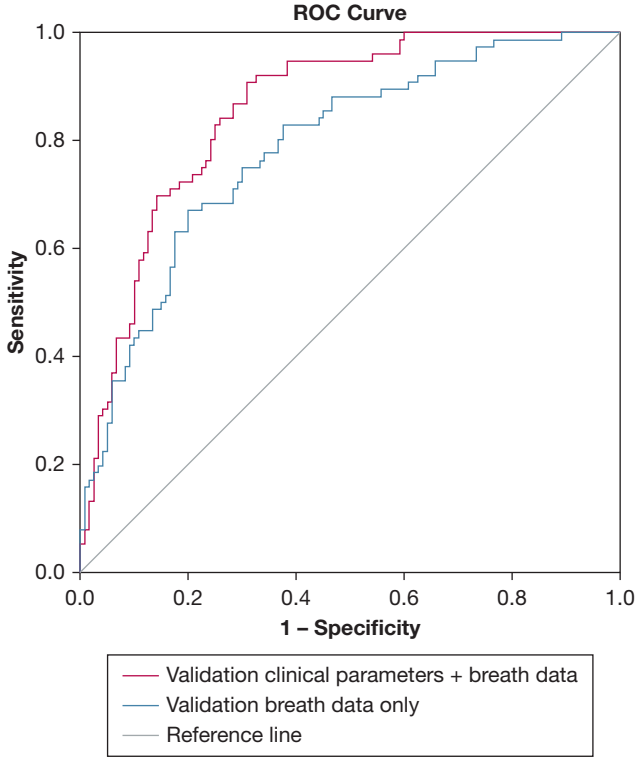


Figure 2 – *Combined ROC curve of the validation models based on either exhaled breath data only and a prediction model including clinical parameters. ROC = receiver-operating characteristic.*

**TABLE 4** ] Final Multivariable Prediction Model (2 × 2) Including Clinical Parameters + Breath Data, and a Cutoff Probability of 16%

| Outcome Prediction Model | Lung Cancer (Gold Standard) | No Lung Cancer (Gold Standard) | Total |
|---|---|---|---|
| Lung cancer (final model) | 72 | 61 | 133 |
| No lung cancer (final model) | 4 | 59 | 63 |
| Total | 76 | 120 | 196 |

single device, and reporting results based on training data that are not validated.

Validation of a prediction model, as performed in this study, is a pivotal step for clinical integration of exhaled breath analysis in the diagnostic path of lung cancer. To assess the feasibility and acceptability of the electronic nose in clinical practice, we envision using the aeoNose in parallel with current practice in a hospital setting. Although based on an exploratory analysis, the validated model seems to be able to distinguish early-stage lung cancer from non-lung cancer with relatively high accuracy. This finding supports the potential of the model to have a possible role in screening purposes. This would have to be investigated in the target population, and, in all likelihood, a different cutoff value needs to be chosen. In case of doubt (eg, based on CT scans), and a low probability of lung cancer, based on the validated model, a wait and see strategy could be used.

In addition, when comparing our full model vs a frequently used nodule calculator (ie, the Mayo Clinic nodule calculator), we found a higher sensitivity at the cost of lower specificity, and comparable NPV. However, it must be noted that this subanalysis was performed with relatively few subjects suspected of having lung cancer. Furthermore, the adjusted Mayo Clinic nodule calculator incorporates the PET result, which logically improves the diagnostic accuracy.

Exhaled breath analysis could have promise as an element in an integral lung cancer screening program, most likely combined with other noninvasive tests such as low-dose CT (LDCT) scan screening. However, the aeoNose should then be trained on a sample of subjects with an increased probability of lung cancer such as individuals who smoke heavily. Despite the fair number of early-stage lung cancer cases in the current cohort, we did not specifically analyze pulmonary nodules, which have inherently been the focus of LDCT scan screening. Future studies should focus on solitary pulmonary nodules and assess whether exhaled breath analysis can fulfill a substantial role in lung cancer screening, possibly serving a synergistic role combined with LDCT scan, and can guide risk assessment prior to LDCT scan screening as a preselection tool or following LDCT scan screening to determine surveillance intervals.[6] However, in such a setting, new prediction models must be built with data based on current screening criteria. In addition to assessment of lung cancer risk, exhaled breath analysis could serve as a prognostic biomarker to predict response on therapies and possible recurrence risk.[39,40]

A notable strength of the current study is the addition of clinical variables to the prediction model. This easily available information has previously been shown to be informative, including development of clinical prediction scores in lung cancer screening based on imaging.[33,41] Our results show significant improvement of the prediction model with the addition of clinical variables, which was confirmed in the validation cohort. Another strength is the excellent match between training and validation results. This is not straightforward, as artificial intelligence techniques are usually applied with far larger data sets.

The aeoNose device features the possibility of performing real-time analysis of breath data without the necessity of breath sample storage; in addition, it incorporates a washout period of 2 min in which the lungs are fully cleared of dead space ventilation, and analysis is solely performed on VOCs originating from metabolic processes in peripheral tissues. Other strong points worth mentioning are the multicenter and multinational design. Multiple devices were used for gathering training data, leading to a prediction model capable of classifying blinded samples, also collected with multiple devices.

The study follows the recommendations of the TRIPOD statement (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis).[42] Unfortunately in the current case, due to the use of the second-generation, CE-certified aeoNose device, we could not use previously collected data. Given the long time frame of collecting the necessary data, we decided to use a split-sample study design in which we simultaneously

trained and validated a prediction model. We intended to use breath data from all eight aeoNoses (seven hospitals) to create a training cohort for supervised learning and cross-validation, and a validation cohort, which was kept blinded, for validation. However, it turned out that in some of the participating hospitals, the amount of breath data, due to limited positive and negative lung cancer diagnoses, was not sufficient for adequate data analysis and could therefore result in a somewhat diminished accuracy.

## Interpretation

Combining exhaled breath data and clinical parameters in a multicenter, multi-device validation study can adequately distinguish patients with lung cancer from subjects without lung cancer in a noninvasive manner. This study paves the way to implement exhaled breath analysis in the daily practice of diagnosing lung cancer.


## Funding/Support

This work was supported by the eNose Company.


## Financial/Nonfinancial Disclosures

The authors have reported to *CHEST* the following: S. K. reports an unrestricted research grant paid to her institution by The eNose Company during the conduct of the study. None declared (M. B.-K., H. S., E. C., F. H. d. J., J. W. G. v. P., B. v. d. B., E. A. K., D. S., M. S., M. M. v. d. H., W. H. v. G., J. v. d. P.).


## Acknowledgments

**Author contributions:** S. K. is the guarantor of the content of the manuscript, data, and analysis, and was responsible for study conception and design, data analysis and interpretation, and manuscript writing. S. K., M. B.-K., M. M. v. d. H., W. H. v. G., and J. v. d. P. were involved in the conception and design of the study; S. K., H. S., E. C., J. W. G. v. P., B. v. d. B., E. A. K., D. S., M. S, M. M. v. d. H., and W. H. v. G. acquired the data; S. K., M. B.-K., and J. v. d. P. analyzed and interpreted the data; and S. K., M. B.-K., H. S., E. C., F. H. d. J., J. W. G. v. P., B. v. d. B., E. A. K., D. S., M. S., M. M. v. d. H., W. H. v. G., and J. W. G. v. P. wrote the article or were substantially involved in its revision before submission.

**Role of sponsors:** The funder of the study, the eNose Company, had no role in the study design, data collection, data interpretation, or writing of the report. However, they performed part of the data analysis, although for model validation they had explicitly no access to blinded data and the classification of these study subjects. S. K. and J. v. d. P. had access to raw data. S. K., the corresponding author, had full access to all the data in the study and the final responsibility for the decision to submit for publication.

**Additional information:** The e-Appendix, e-Figure, and e-Tables are available online under "Supplementary Data."


## References


1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin*. 2020;70(1):7-30.
2. Dagenais GR, Leong DP, Rangarajan S, et al. Variations in common diseases, hospital admissions, and deaths in middle-aged adults in 21 countries from five continents (PURE): a prospective cohort study. *Lancet*. 2020;395(10226):785-794.
3. Broodman I, Lindemans J, van Sten J, Bischoff R, Luider T. Serum protein markers for the early detection of lung cancer: a focus on autoantibodies. *J Proteome Res*. 2017;16(1):3-13.
4. Jiang R, Dong X, Zhu W, et al. Combining PET/CT with serum tumor markers to improve the evaluation of histological type of suspicious lung cancers. *PLoS One*. 2017;12(9):e0184338.
5. Rolfo C, Russo A. Liquid biopsy for early stage lung cancer moves ever closer. *Nat Rev Clin Oncol*. 2020;17(9):523-524.
6. Seijo LM, Peled N, Ajona D, et al. Biomarkers in lung cancer screening: achievements, promises, and challenges. *J Thorac Oncol*. 2019;14(3):343-357.
7. Mansurova M, Ebert BE, Blank LM, Ibanez AJ. A breath of information: the volatilome. *Curr Genet*. 2018;64(4):959-964.
8. Boots AW, van Berkel JJ, Dallinga JW, Smolinska A, Wouters EF, van Schooten FJ. The versatile use of exhaled volatile organic compounds in human health and disease. *J Breath Res*. 2012;6(2):027108.
9. Nakhleh MK, Amal H, Jeries R, et al. Diagnosis and classification of 17 diseases from 1404 subjects via pattern analysis of exhaled molecules. *ACS Nano*. 2017;11(1):112-125.
10. Wilson AD. Advances in electronic-nose technologies for the detection of volatile biomarker metabolites in the human breath. *Metabolites*. 2015;5(1):140-163.
11. Dragonieri S, Annema JT, Schot R, et al. An electronic nose in the discrimination of patients with non-small cell lung cancer and COPD. *Lung Cancer*. 2009;64(2):166-170.
12. Gasparri R, Santonico M, Valentini C, et al. Volatile signature for the early diagnosis of lung cancer. *J Breath Res*. 2016;10(1):016007.
13. Kort S, Tiggeloven MM, Brusse-Keizer M, et al. Multi-centre prospective study on diagnosing subtypes of lung cancer by exhaled-breath analysis. *Lung Cancer*. 2018;125:223-229.
14. Phillips M, Altorki N, Austin JH, et al. Prediction of lung cancer using volatile biomarkers in breath. *Cancer Biomark*. 2007;3(2):95-109.
15. van de Goor R, van Hooren M, Dingemans AM, Kremer B, Kross K. Training and validating a portable electronic nose for lung cancer screening. *J Thorac Oncol*. 2018;13(5):676-681.
16. Mazzone PJ, Wang XF, Lim S, et al. Progress in the development of volatile exhaled breath signatures of lung cancer. *Ann Am Thorac Soc*. 2015;12(5):752-757.
17. Rocco G, Pennazza G, Santonico M, et al. Breathprinting and early diagnosis of lung cancer. *J Thorac Oncol*. 2018;13(7):883-894.
18. Tirzite M, Bukovskis M, Strazda G, Jurka N, Taivans I. Detection of lung cancer with electronic nose and logistic regression analysis. *J Breath Res*. 2018;13(1):016006.
19. National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365(5):395-409.
20. de Koning HJ, van der Aalst CM, de Jong PA, et al. Reduced lung-cancer mortality with volume CT screening in a randomized trial. *N Engl J Med*. 2020;382(6):503-513.
21. Bruins MGJ, Gerritsen JW, Van de Sande W, Van Belkum A, Bos A. Enabling a transferable calibration model for metal-oxide type electronic noses. *Sensors Actuators B Chemical*. 2013;188:1187-1195.

22. Kort S, Brusse-Keizer M, Gerritsen JW, van der Palen J. Data analysis of electronic nose technology in lung cancer: generating prediction models by means of Aethena. *J Breath Res*. 2017;11(2):026006.

23. Peters Y, Schrauwen RWM, Tan AC, Bogers SK, de Jong B, Siersema PD. Detection of Barrett's oesophagus through exhaled breath using an electronic nose device. *Gut*. 2020;69(7):1169-1172.

24. van de Goor R, Hardy JCA, van Hooren MRA, Kremer B, Kross KW. Detecting recurrent head and neck cancer using electronic nose technology: a feasibility study. *Head Neck*. 2019;41(9):2983-2990.

25. Uslu HI, Dolle AR, Dullemen HM, Aktas H, Kolkman JJ, Venneman NG. Pancreatic ductal adenocarcinoma and chronic pancreatitis may be diagnosed by exhaled-breath profiles: a multicenter pilot study. *Clin Exp Gastroenterol*. 2019;12:385-390.

26. Kort S, Brusse-Keizer M, Gerritsen JW, et al. Improving lung cancer diagnosis by combining exhaled-breath data and clinical parameters. *ERJ Open Res*. 2020;6(1):00221-2019.

27. Long Y, Wang C, Wang T, et al. High performance exhaled breath biomarkers for diagnosis of lung cancer and potential biomarkers for classification of lung cancer. *J Breath Res*. 2021;15(1):016017.

28. Phillips M, Bauer TL, Cataneo RN, et al. Blinded validation of breath biomarkers of lung cancer, a potential ancillary to chest CT screening. *PLoS One*. 2015;10(12):e0142484.

29. Fens N, Roldaan AC, van der Schee MP, et al. 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*. 2011;41(10):1371-1378.

30. Bos LD, Schultz MJ, Sterk PJ. Exhaled breath profiling for diagnosing acute respiratory distress syndrome. *BMC Pulm Med*. 2014;14:72.

31. Goldstraw P, Chansky K, Crowley J, et al. The IASLC Lung Cancer Staging Project: proposals for revision of the TNM stage groupings in the forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. *J Thorac Oncol*. 2016;11(1):39-51.

32. Horvath I, Barnes PJ, Loukides S, et al. A European Respiratory Society technical standard: exhaled biomarkers in lung disease. *Eur Respir J*. 2017;49(4).

33. Herder GJ, van Tinteren H, Golding RP, et al. Clinical prediction model to characterize pulmonary nodules: validation and added value of 18F-fluorodeoxyglucose positron emission tomography. *Chest*. 2005;128(4):2490-2496.

34. Swensen SJ, Silverstein MD, Ilstrup DM, Schleck CD, Edell ES. The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules. *Arch Intern Med*. 1997;157(8):849-855.

35. Machado RF, Laskowski D, Deffenderfer O, et al. Detection of lung cancer by sensor array analyses of exhaled breath. *Am J Respir Crit Care Med*. 2005;171(11):1286-1291.

36. Li W, Liu H, Xie D, He Z, Pi X. Lung cancer screening based on type—different sensor arrays. *Sci Rep*. 2017;7(1):1969.

37. Mazzone PJ, Wang XF, Xu Y, et al. Exhaled breath analysis with a colorimetric sensor array for the identification and characterization of lung cancer. *J Thorac Oncol*. 2012;7(1):137-142.

38. D'Amico A, Pennazza G, Santonico M, et al. An investigation on electronic nose diagnosis of lung cancer. *Lung Cancer*. 2010;68(2):170-176.

39. de Vries R, Muller M, van der Noort V, et al. Prediction of response to anti-PD-1 therapy in patients with non-small-cell lung cancer by electronic nose analysis of exhaled breath. *Ann Oncol*. 2019;30(10):1660-1666.

40. Buma AIG, Muller M, de Vries R, et al. eNose analysis for early immunotherapy response monitoring in non-small cell lung cancer. *Lung Cancer*. 2021;160:36-43.

41. Baldwin DR, Callister ME, Guideline Development Group. The British Thoracic Society guidelines on the investigation and management of pulmonary nodules. *Thorax*. 2015;70(8):794-798.

42. Patzer RE, Kaji AH, Fong Y. TRIPOD reporting guidelines for diagnostic and prognostic studies. *JAMA Surg*. 2021;156(7):675-676.