Development of a polymer modified Quartz Tuning Fork (QTF) sensor array based Volatile Organic Compound (VOC) classifier

Suresh Panchal, Apoorva Phadke, Unnikrishnan Gopinathan and Suwarna Datar

Abstract—Polymer modified Quartz Tuning Fork sensor array with three sensors has been developed to classify 1,4-dimethoxy-2, 3-butanediol (BD), cyclohexanone (CH), ethanol, methanol and acetone with various concentrations. Linear Discriminant Analysis (LDA), Decision Tree (DT), and Random Forrest (RF) have been used and evaluated as the classifiers predicting the VOCs and their concentration. The classifiers were trained with three features: response time, recovery time, and shift in the resonant frequency of the sensors. The results have been compared and analyzed for the performance of each classifier for the present system. RF gives the best performance as a classifier with an accuracy of classification more than 95%. We present results of VOC classifier performance for the QTF sensor array.

Index Terms—Quartz Tuning Fork, electronic nose, Sensor Array, Volatile Organic Compound, feature selection, Machine Learning, Classifier, Linear Discriminant Analysis, Decision Tree, Random Forest

I. INTRODUCTION

The present-day automation systems make decisions by analysis of data from a sensor array [1], [2]. Machine learning algorithms have significantly altered the way in which sensor data is analyzed, especially the role of classification in traditional data analysis techniques [3]. Environmental, industrial emission [4]–[6], food degradation [7] and indoor air monitoring [5], [8] are problems that can be addressed by the detection of volatile organic compounds (VOC's). Another area where sensor arrays for VOC detection are useful is an emerging field of human breath analysis.

Several types of sensors support the recent development in the field of robotics and automated systems. These sensors may be optical, mechanical, and electrical. There may be several sets of sensors called an array of sensors detecting various parameters. These may perform the tasks such as object detection, computer vision, pressure and force measurement, or detection of gases in the surroundings. The detection of gases in the surrounding is a very complex task as several types

of gases are present. Several sensors in the array can respond to a particular gas, or there might be a gas to which no sensor responds. All sensors' collective responses may improve the selectivity in sensing certain gasses or any proportionality imbalance in the air. These types of systems are called electronic noses. E-nose devices simulate different stages of the human olfactory system, olfactory receptor, olfactory bulb, and olfactory cortex by the electronic sensor array, pre-processor, and pattern recognition respectively [9], [10]. While the e-nose, cannot replace the standard spectroscopic characterization techniques but it can support in classifying gases for a large number of samples with a portable device, where using the spectroscopic techniques are not feasible [11], [12]. The application of this technology spans many fields of application, from detection of explosives (Defence/Security application) to human health monitoring (Medical application). Human health monitoring is a repetitive task over a small period, therefore, using e-nose for breath analysis can be a good non-invasive alternative to traditional techniques [7], [13]. Human breath contains several VOC's which are signatures of various chemical activities in the respiratory or digestive system. A disease can alter the concentration of these VOCs or exhalation of certain VOCs specific to certain diseases called breath biomarkers for those diseases. Human breath analysis can be used as a non-invasive technique for identification of such diseases [7], [13]. A healthy person has a pattern of VOCs present in the breath [13], [14]. Once labeled data is collected, there is a need for some data classification or clustering techniques for prediction [15]. The electronic nose (e-nose) technology can

Suresh Panchal is with Department of Applied Physics, Defence Institute of Advanced Technology (DU), Pune-411025, INDIA.

Apoorva Phadke was with Department of Applied Physics, Defence Institute of Advanced Technology (DU), Pune-411025, INDIA.

Unnikrishnan Gopinathan is with Photonics Division, Instruments Research and Development Establishment, Raipur Road, Dehradun-248008, INDIA.

Suwarna Datar is with Department of Applied Physics, Defence Institute of Advanced Technology (DU), Pune-411025, INDIA. (e-mail: suwarna.datar@gmail.com).

Suresh Panchal and Apoorva Phadke contributed equally.

provide a sort of fingerprint of exhaled breath by detecting different VOCs through an array of sensors. The e-nose provides distinctive breath fingerprints for some diseases. [4], [10], [16]–[20]. Identification of such biomarkers can be done using spectroscopies such as gas chromatography-mass spectrometry (GC-MS) or proton transfer reaction mass spectrometry (PTR-MS) [16], [21], [22]. Once a breath bio-marker is identified a portable device will be enough to identify any abnormality in breath sample indicative of a particular disease. For this, a data needs to be generated from sensor array to detect several VOCs which include control as well as bio-markers. The machine learning algorithm can then be trained to detect bio-markers. Breath analysis, compared to blood and urine tests, is quite simple and samples may be collected easily at any point in required quantities.

The Global Cancer Observatory (GCO)'s report Globocan 2020 shows 19,292,789 new cases of cancer were found in 2020, in which 2,261,419 were breast cancer. Nearly 11.7 % of total cancer cases, which is the largest number as compared to other cancer cases. There are 684,996 deaths reported due to breast cancer which is a serious concern. Early-stage diagnosis of breast cancer reduces morbidity rate and hence mortality rate. While, VOCs like, ethanol, methanol and acetone are present in the breath in lower concentrations, there are several VOC's reported as breath biomarker for the breast cancer [23]–[27] 1,4-dimethoxy-2, 3-butanediol (BD), and cyclohexanone (CH) appear as one of the prominent breath bio-marker for breast cancer patients [28]–[30]. If detected at early states, the technique can be a useful as an non-invasive method for further screening. In the present work we have focused on detection of these two bio-markers along with the controls ethanol, methanol and acetone. In the work we have tried to use multiple sensor data and used different machine learning algorithms to discriminate the presence of lower concentrations of 5 VOCs which include 3 control VOCs and 2 breath bio-markers for breast cancer patients.

Quartz Tuning Forks (QTFs) are electro-mechanical oscillators that have been used in wristwatches and other high precision timing circuits for decades. The piezoelectric material quartz allows the QTF to be excited electrically with the oscillator circuit and its resonant frequency to be read out in the form of an electrical signal. In recent years, QTFs have been used as sensors for the detection of several VOCs. [3], [31]–[33]

In the present work, a sensor array system has been developed such that all the sensors were simultaneously exposed to the VOC, and then the data was collected. A three-element sensor array comprising of polymer wires/films modified QTFs were used to detect and label an unknown VOC. Three classifiers - LDA, DT and RF- were used to label, detect and classify five different VOCs namely 1,4-dimethoxy-2, 3-butanediol (BD), and cyclohexanone (CH), ethanol, methanol and acetone. Each of the three QTF sensors measured three parameters -response time, recovery time and shift in the resonant frequency-that formed nine features used for detection and classification of unknown VOC. A study has been on each classifier's performance and the performance of sensor array in detection and classification of 5 VOCs and their

concentrations. [34]

II. EXPERIMENTAL

A. Polymer modification of QTF

99.5% pure Poly-methyl-methacrylate (PMMA), Polystyrene (PS) and 99.0% pure aniline were used to modify the QTF. 2wt%, of PMMA and 5wt%, 10wt % of PS pellets were dissolved in 5ml of aniline by keeping it overnight at room temperature to form the respective polymer solutions of required uniform viscosity. Figure 1a represents the step-wise procedure to modify QTF with polymer films and wire. The QTF and polymer system was allowed to dry naturally at room temperature for a day. The shift in a resonant frequency of modified QTF was then measured and ensured that the QTF resonant frequency increased after modification. Sensor 1 and sensor 2 were modified with 2wt%, of PMMA and 5wt% of PS film respectively by the method shown in Figure 1a and sensor 3 was modified with 10wt% of PS by as shown in Figure 1b.

(a) Schematic illustrating the step-wise method for QTF modification using polymer film with scanning electron microscope image showing polymer film on top of QTF

(b) Schematic illustrating the step-wise method for QTF modification using polymer wire with optical and scanning electron microscope image of the modified QTF

Fig. 1: Schematic illustrating the step-wise method for QTF modification by two methods. The corresponding microscopic images are also shown with the schematic.

B. QTF based gas sensing

In our previous works, we have elaborately described the working of the polymer modified QTFs. [3] The working of the

system is discussed in short in the present work context. The QTF which oscillates at a resonant frequency of 32.768kHz (which is sealed) is opened and exposed to air. The resonant frequency of QTF is given by Eqn. 1.

$$f_0 = \frac{1}{2\pi} \sqrt{\frac{k}{m}} \quad (1)$$

where:

$$\begin{aligned} f_0 &= \text{Resonant frequency of a QTF,} \\ k &= \text{Spring constant of a QTF,} \\ m &= \text{Mass of a QTF.} \end{aligned}$$

The opened QTF can be modified by polymer which can interact with the analyte changing its resonant frequency indicative of the presence of analyte. The interaction of the polymer with the analyte depends on several factors pertaining to its solubility in the polymer and thus in turn leads to change in mechanical properties of polymer which affect the resonant frequency of the QTF. Therefore, the way polymer is deposited on the QTF and the size and shape of the polymer also affects the response of the QTF towards particular analyte.

We have observed previously that there are some analytes which increase the elastic modulus of the polymer and therefore, effectively increase the resonant frequency of the QTF, on the other hand some analytes, swell the polymer making it softer and therefore, decrease the resonance frequency of the QTF. The relative change in resonant frequency of QTF sensor as a function of change in elastic modulus of polymer is given by Eqn. 2 [35]

$$\frac{\Delta f_0}{f_0} \cong \frac{A}{2Lk_{fork}} \Delta E \quad (2)$$

where:

$$\begin{aligned} \Delta E &= \text{change in elastic modulus of the polymer,} \\ L, A &= \text{Length and area of cross-section of polymer wire,} \\ \Delta f_0/f_0 &= \text{Shift in resonant frequency of QTF} \\ k_{fork} &= \text{Spring constant of QTF.} \end{aligned}$$

The detailed work on understanding this phenomenon has been reported elsewhere. An oscillator circuit is used to convert the change in resonant frequency to a measured electrical signal. [36], [37]. The three sensors were exposed to various VOCs simultaneously in a chamber and change in the resonant frequency was recorded along with the response time of the sensor. Later the sensors were exposed to ambient conditions for recovery. Several such repeat cycles were recorded for these sensors and the data for change in resonant frequency, response and recovery time was monitored. One such frequency response is shown in Figure 2 to describe the collection of three features of the classifier, namely, shift in frequency, response and recovery time.

Figure 3 shows the schematic of the QTF sensor array comprising of three modified QTF. The output from the QTF is collected using three channels of the data acquisition system (National Instrument's PXIe 1082 chassis housing a 5162 High-Speed Digitizer card).

In the present work classification of analytes concentrations from 0.5ppm to 20ppm as mentioned in table I has been done.

Fig. 2: Feature collection from raw data response, response time and recovery time.

Fig. 3: Schematic of the QTF sensor array based VOC classifier where array of sensors situated inside the fume chamber. The output response of each of the sensor is given to data acquisition system away from the sensor.

TABLE I: Analyte and their concentration used for experiments

Analyte	Concentration (ppm)
(R,R)-(+)-1,4-Dimethoxy- 2,3-butanediol(BD)	0.5, 1
Cyclohexanone(CH)	10, 20
Ethanol	10, 20
Methanol	10, 20
Acetone	10, 20

Analyte molecule concentration present in the head space of bottles was calculated using the following Eqn. 3.

$$\text{Concentration}(ppm) = \frac{\text{Vapour Pressure (mmHg)}}{760(mmHg)} 10^6 \quad (3)$$

To take a lower concentration of analytes in the fume chamber as maintained in the table, dilution was done in two steps. First, the concentration was reduced using Tedlar bags (TB) and then it was further reduced to a particular concentration in the fume chamber using the following Eqn.4.

$$C_c(ppm) = \frac{C_{tb}(ppm)V_a(ml)}{V_c(ml)} \quad (4)$$

Where C_c is the concentration of the analyte in fume chamber, C_{tb} is the concentration of analyte inside the TB

(500 ppm for this set of experiments), V_a is the volume of analyte being injected using a syringe (1 to 6 mL), and V_c is the volume of the fume chamber (50mL for this set of experiments).

Each experiment cycle consisted of following steps: (i) Start data acquisition (ii) Wait for 60sec (iii) Close fume chamber (iv) Wait for 60sec (v) Inject analyte in chamber (vi) Wait for 60sec (vi) Open it for 60 sec. These cycles were repeated six times in a single experiment. All QTFs were exposed simultaneously to a particular analyte and their response was used to train the data for classification. The shift in a frequency was considered for further analysis instead of frequency before and after exposing the analyte.

C. Classification of QTF sensor measurements

Nine features, three features from each of three sensors were extracted from QTF sensor array measurements for identification of VOCs and its concentration. The classification involves two phases: Training and Testing as shown in Figure 4.

Fig. 4: Schematic showing the training and testing phase used in the VOC classifier

III. RESULTS AND DISCUSSION

Three classifiers - LDA (linear discriminant analysis), DT and RF- were used for classification of sensor data. The data was divided into 80% as training examples and 20% as testing examples. An unknown VOC has to be classified into one of the 10 classes using 9 features measured by the sensor array. The description of the features and classes are given in table II and III.

TABLE II: Shows the feature set used for training and test of the VOC classifiers

Sr.No.	Feature description	Nomenclature
1	Response time for QTF 1 (2% wt. PMMA films)	2-dt
2	Recovery time for QTF 1 (2% wt. PMMA films)	2-dr
3	Shift in resonant frequency for QTF 1 (2% wt. PMMA films)	2-df
4	Response time for QTF 2 (5% wt. PS films)	5-dt
5	Recovery time for QTF 2 (5% wt. PS films)	5-dr
6	Shift in resonant frequency for QTF 2 (5% wt. PS films)	5-df
7	Response time for QTF 3 (10% wt. PS wire)	10-dt
8	Recovery time for QTF 3 (10% wt. PS wire)	10-dr
9	Shift in resonant frequency for QTF 3 (10% wt. PS wire)	10-df

TABLE III: The description of the 10 classes used for training and testing phase

Sr.No.	Class description	Class label (Nomenclature)
1	(R,R)-(+)-1,4-Dimethoxy- 2,3-butanediol (Conc. 0.1 ppm)	BD0.5
2	(R,R)-(+)-1,4-Dimethoxy- 2,3-butanediol (Conc. 1 ppm)	BD1
3	Cyclohexanone (Conc. 10 ppm)	CH10
4	Cyclohexanone (Conc. 20 ppm)	CH20
5	Ethanol (Conc. 10 ppm)	E10
6	Ethanol (Conc. 20 ppm)	E20
7	Methanol (Conc. 10 ppm)	M10
8	Methanol (Conc. 20 ppm)	M20
9	Acetone (Conc. 10 ppm)	AC10
10	Acetone (Conc. 20 ppm)	AC20

Figure 5 shows the average response of the sensor array towards the 10 classes mentioned in terms of change in frequency, response time and recovery time in the form of box plot. One can see that while there is some variation in all the parameters of sensor array towards different classes, there is no visible pattern which can be identified looking at the data. Furthermore, all sensors do not respond to all the analytes. For example, BD is detected only by 2 % wt PMMA film, whereas, 10 wt % wire is selective towards ethanol and acetone. Classification using various algorithm can identify these patterns and use them to get accuracy in classifying different VOCs and their concentrations.

The performance evaluation was carried out using the metric predication accuracy as given in Eqn.5. Figure 6 shows the overall performance of all the classifiers for training and testing data.

$$\text{Prediction Accuracy}(\%) = \frac{n}{N} \times 100\% \quad (5)$$

where:

n = No Of Times Model Predicted Correctly,
 N = Total No Of Predictions.

For a c-class classification problem, LDA classifier projects a d-dimensional data from a d-dimensional space to a (d-1) dimensional space in such a way that the metric Fischer ratio is a maximum [38]. Fisher ratio is defined as the ratio of Between Class Scatter to Within Class Scatter. For the present classification problem, the dimensionality of the data d is 9 (size of feature vector) and the number of classes c is 10. The predication accuracy was 88.37% for training and 81.81% for the test data. Since here the features are transformed from higher dimensional feature space to lower-dimensional feature space, while projecting data points in lower-dimensional space it reduces part of the information which describes the data point position. Therefore, the Fisher ratio is maximized which makes LDA computationally less expensive than DT and RF however, reduces the prediction accuracy. The same can be seen in the classification of our sensor data when LDA is used.

DT is a tree based supervised learning algorithm used for classification. A decision is made at each node to split into two or more sub-nodes based on the value of an attribute.

Fig. 5: Box plot of response of individual sensors with sensing attributes (a) shift in frequency , (b) response time, and (c) recovery time

Fig. 6: Chart showing Prediction accuracy of the three techniques discussed earlier where blue bars shows the prediction accuracy of an algorithms (classifiers) for the data used in training while orange bars for the unseen data to the algorithms (classifiers).

The attribute is chosen from a set of attributes that would result in more homogeneous or pure sub-nodes. For this the prediction accuracy for training improved to 100% but test accuracy remained at 81.81%. This could be due to the fact that many a times DT tends to over-fit the data. The DT obtained is shown in Figure 9. The DT many a times fails while generalizing for unknown objects and therefore suffers from a high variance problem. Therefore, although it improves accuracy it suffers from over-fitting. In DT it must prioritize

the attribute-based test even if there exists some other tests which reduces disorderness in the data by similar values with a negligible difference. This drawback of DT is addressed by generating many trees with randomly selected features which is done is RF. The Law of Large Numbers addresses the problem of high variance hence the RF becomes more effective. In RF prediction accuracy for training was also 100% but test accuracy improved to 95.45%. RF gives an overall good performance for both training and testing compared to LDA and DT [39]–[41]. Figure 7 and 8 shows confusion matrix for DT and RF.

Fig. 7: Confusion matrix for DT

Fig. 8: Confusion matrix for RF

The overall performance of a classifier can be evaluated by calculating Precision, Recall and F1-score for an individual class using Eqn. 6, 7 and 8 respectively. Where TP are true positive and FP, false positive predictions in the confusion matrix.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (6)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (7)$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

These metrics can be calculated globally or for individual class. A perfect score of 1 for all the metrics gives 100% accuracy for individual classes. Tables IV, V and VI show the class-wise performance of each classifier. From Tables IV one can see that CH10 and CH20 are classified well with LDA, but M20 and E10 have lower precision, indicating high false positive cases. On the other hand E20 and Ac10 show more false negative cases reflected in recall. Similarly, one can see that while DT performed better in training, its performance measures during testing are comparable to LDA (see table V). All scores for RF for all classes except M20 are high as can be seen from Tables VI, indicating very high global performance which can be improved by adding another sensor in the array specifically to address methanol concentration identification.

TABLE IV: Performance measure of LDA on test data

	TP	TN	FP	FN	Precision	Recall	F1-score
CH10	4	18	0	0	1	1	1
CH20	2	20	0	0	1	1	1
M10	0	21	0	1	-	0	0
M20	1	20	1	0	0.5	1	0.66
E10	3	17	2	0	0.6	1	0.75
E20	3	18	0	1	1	0.75	0.85
Ac10	1	19	0	2	1	0.33	0.5
Ac20	0	21	1	0	0	-	0
BD0.5	4	18	0	0	1	1	1
BD1	0	22	0	0	-	-	-
Total	18	194	4	4			
Accuracy							0.81

TABLE V: Performance measure of DT on test data

	TP	TN	FP	FN	Precision	Recall	F1-score
CH10	4	18	0	0	1	1	1
CH20	2	20	0	0	1	1	1
M10	0	22	0	0	-	-	-
M20	0	20	2	0	0	-	0
E10	5	17	0	0	1	1	1
E20	1	18	2	1	0.33	0.5	0.4
Ac10	1	21	0	0	1	1	1
Ac20	1	18	0	3	1	0.25	0.4
BD0.5	4	18	0	0	1	1	1
BD1	0	22	0	0	-	-	-
Total	18	194	4	4			
Accuracy							0.81

IV. CONCLUSION

In present work, we have developed polymer modified QTF sensor array of three sensors with polymer wires/films for detection of 5 VOCs and their concentrations. We used three classifiers- LDA, DT and RF. The features used for classification were – the response time, recovery time and shift in the resonant frequency of the three sensors in the array. All these sensors give their individual response at the same time for each of the injected analyte in the chamber. The performance of RF in classifying 10 classes which include different concentrations

TABLE VI: Performance measure of RF on test data

	TP	TN	FP	FN	Precision	Recall	F1-score
CH10	4	18	0	0	1	1	1
CH20	2	20	0	0	1	1	1
M10	0	22	0	0	-	-	-
M20	1	20	1	0	0.5	1	0.66
E10	5	17	0	0	1	1	1
E20	3	18	0	1	1	0.75	0.85
Ac10	1	21	0	0	1	1	1
Ac20	1	21	0	0	1	1	1
BD0.5	4	18	0	0	1	1	1
BD1	0	22	0	0	-	-	-
Total	21	197	1	1			
Accuracy							0.95

of 1,4-dimethoxy-2, 3-butanediol (BD), and Cyclohexanol (CH), ethanol, methanol and acetone is quite high with the accuracy of more than 95 %. Based on the confusion matrix of DT and RF one can see that while testing accuracy of RF is better, accuracy of DT to detect CH and BD is also good which are the bio-markers to be detected. RF could not detect ethanol accurately which can be improved with more training. The performance measures of all classifiers also indicate fairly good performance of RF for all classes except methanol and ethanol. A better selective sensor towards ethanol or methanol added to the array can resolve this problem along with more data collection. The present work demonstrates the potential of QTF sensor in an array to detect multiple analytes in e-nose applications.

REFERENCES

- [1] Y. Do and J. Kim, "Infrared range sensor array for 3d sensing in robotic applications," *International Journal of Advanced Robotic Systems*, vol. 10, no. 4, p. 193, 2013.
- [2] H. Park, S. Baek, and S. Lee, "Ir sensor array for a mobile robot," in *Proceedings, 2005 IEEE/ASME International Conference on Advanced Intelligent Mechatronics*. IEEE, 2005, pp. 928–933.
- [3] A. Sampson, S. Panchal, A. Phadke, A. Kashyap, J. Suman, G. Unnikrishnan, and S. Datar, "Quartz tuning fork based sensor for detection of volatile organic compounds: towards breath analysis," *Materials Research Express*, vol. 5, no. 4, p. 045407, 2018.
- [4] R. A. Incalzi, G. Pennazza, S. Scarlata, M. Santonico, M. Petriaggi, D. Chiurco, C. Pedone, Arnaldo, and D'Amico, "Reproducibility and respiratory function correlates of exhaled breath fingerprint in chronic obstructive pulmonary disease," *PLoS One*, vol. 7, no. 10, p. e45396, 2012.
- [5] R. G. Derwent, "Sources, distributions, and fates of vocs in the atmosphere," *Issues in Environmental Science and Technology*, vol. 4, pp. 1–16, 1995.
- [6] A. Lay-Ekuakille and A. Trotta, "Predicting voc concentration measurements: Cognitive approach for sensor networks," *IEEE Sensors Journal*, vol. 11, no. 11, pp. 3023–3030, 2011.
- [7] B. de Lacy Costello, A. Amann, H. Al-Kateb, C. Flynn, W. Filipiak, T. Khalid, D. Osborne, and N. M. Ratcliffe, "A review of the volatiles from the healthy human body," *Journal of breath research*, vol. 8, no. 1, p. 014001, 2014.
- [8] D. Panagiotaras, D. Nikolopoulos, E. Petraki, S. Kottou, D. Koulougliotis, P. Yannakopoulos, and S. Kaplanis, "Comprehensive experience for indoor air quality assessment: a review on the determination of volatile organic compounds (vocs)," *Journal of Physical Chemistry & Biophysics*, vol. 4, no. 5, p. 1, 2014.
- [9] A. P. Turner and N. Magan, "Electronic noses and disease diagnostics," *Nature Reviews Microbiology*, vol. 2, no. 2, pp. 161–166, 2004.
- [10] E. R. Thaler and C. W. Hanson, "Medical applications of electronic nose technology," *Expert review of medical devices*, vol. 2, no. 5, pp. 559–566, 2005.
- [11] F. Röck, N. Barsan, and U. Weimar, "Electronic nose: current status and future trends," *Chemical reviews*, vol. 108, no. 2, pp. 705–725, 2008.

Fig. 9: Decision tree classifying each of the analyte CH10, CH20, M10, M20, E10, E20, Ac10, Ac20, BD0.5, and BD1 as described in table III at leaf node using features 2-dt, 2-dr, 2-df, 5-dt, 5-dr, 5-df, 10-dt, 10-dr, 10-df described in table II

- [12] A. Staerz, F. Roeck, U. Weimar, and N. Barsan, "Electronic nose: current status and future trends," *Surface and Interface Science: Volume 9: Applications of Surface Science I*, vol. 9, pp. 335–379, 2020.
- [13] K. Schmidt and I. Podmore, "Current challenges in volatile organic compounds analysis as potential biomarkers of cancer," *Journal of biomarkers*, vol. 2015, 2015.
- [14] T. Ligor, M. Ligor, A. Amann, C. Ager, M. Bachler, A. Dzien, and B. Buszewski, "The analysis of healthy volunteers' exhaled breath by the use of solid-phase microextraction and gc-ms," *Journal of breath research*, vol. 2, no. 4, p. 046006, 2008.
- [15] K. L. Diehl, M. A. Ivy, S. Rabidou, S. M. Petry, G. Müller, and E. V. Anslyn, "Differential sensing for the regio- and stereoselective identification and quantitation of glycerides," *Proceedings of the National Academy of Sciences*, vol. 112, no. 30, pp. E3977–E3986, 2015.
- [16] P. Montuschi, M. Santonico, C. Mondino, G. Pennazza, G. Mantini, E. Martinelli, R. Capuano, G. Ciabattini, R. Paolesse, C. Di Natale, P. J. Barnes, and A. D'Amico, "Diagnostic performance of an electronic nose, fractional exhaled nitric oxide, and lung function testing in asthma," *Chest*, vol. 137, no. 4, pp. 790–796, 2010.
- [17] C. Di Natale, A. Macagnano, E. Martinelli, R. Paolesse, G. D'Arcangelo, C. Roscioni, A. Finazzi-Agro, and A. D'Amico, "Lung cancer identification by the analysis of breath by means of an array of non-selective gas sensors," *Biosensors and Bioelectronics*, vol. 18, no. 10, pp. 1209–1218, 2003.
- [18] J. K. Robinson, M. J. Bollinger, and J. W. Birks, "Luminol/h2o2 chemiluminescence detector for the analysis of nitric oxide in exhaled breath," *Analytical chemistry*, vol. 71, no. 22, pp. 5131–5136, 1999.
- [19] A. Nair, D. Menzies, and B. Lipworth, "P059 portable exhaled nitric oxide measurement: Comparison with the gold-standard technique," *Thorax*, vol. 61, 2006.
- [20] Y. Y. Broza and H. Haick, "Nanomaterial-based sensors for detection of disease by volatile organic compounds," *Nanomedicine*, vol. 8, no. 5, pp. 785–806, 2013.
- [21] M. Hakim, Y. Y. Broza, O. Barash, N. Peled, M. Phillips, A. Amann, and H. Haick, "Volatile organic compounds of lung cancer and possible biochemical pathways," *Chemical reviews*, vol. 112, no. 11, pp. 5949–5966, 2012.
- [22] P. Fuchs, C. Loeseken, J. K. Schubert, and W. Miekisch, "Breath gas aldehydes as biomarkers of lung cancer," *International Journal of Cancer*, vol. 126, no. 11, pp. 2663–2670, 2010.
- [23] J. Li, Y. Peng, and Y. Duan, "Diagnosis of breast cancer based on breath analysis: An emerging method," *Critical reviews in oncology/hematology*, vol. 87, no. 1, pp. 28–40, 2013.
- [24] M. Phillips, R. N. Cataneo, B. A. Ditkoff, P. Fisher, J. Greenberg, R. Gunawardena, C. S. Kwon, O. Tietje, and C. Wong, "Prediction of breast cancer using volatile biomarkers in the breath," *Breast cancer research and treatment*, vol. 99, no. 1, pp. 19–21, 2006.
- [25] M. Phillips, R. N. Cataneo, C. Saunders, P. Hope, P. Schmitt, and J. Wai, "Volatile biomarkers in the breath of women with breast cancer," *Journal of breath research*, vol. 4, no. 2, p. 026003, 2010.

- [26] M. Phillips, R. N. Cataneo, C. Lebauer, M. Mundada, and C. Saunders, "Breath mass ion biomarkers of breast cancer," *Journal of breath research*, vol. 11, no. 1, p. 016004, 2017.
- [27] G. Shuster, Z. Gallimidi, A. H. Reiss, E. Dovgolevsky, S. Billan, R. Abdah-Bortnyak, A. Kuten, A. Engel, A. Shiban, U. Tisch *et al.*, "Classification of breast cancer precursors through exhaled breath," *Breast cancer research and treatment*, vol. 126, no. 3, pp. 791–796, 2011.
- [28] L. D. de León-Martínez, M. Rodríguez-Aguilar, P. Gorocica-Rosete, C. A. Domínguez-Reyes, V. Martínez-Bustos, J. A. Tenorio-Torres, O. Ornelas-Rebolledo, J. A. Cruz-Ramos, B. Balderas-Segura, and R. Flores-Ramírez, "Identification of profiles of volatile organic compounds in exhaled breath by means of an electronic nose as a proposal for a screening method for breast cancer: a case-control study," *Journal of Breath Research*, vol. 14, no. 4, p. 046009, 2020.
- [29] W. Miekisch, J. Herbig, and J. K. Schubert, "Data interpretation in breath biomarker research: pitfalls and directions," *Journal of breath research*, vol. 6, no. 3, p. 036007, 2012.
- [30] C. Wang, B. Sun, L. Guo, X. Wang, C. Ke, S. Liu, W. Zhao, S. Luo, Z. Guo, Y. Zhang *et al.*, "Volatile organic metabolites identify patients with breast cancer, cyclomastopathy and mammary gland fibroma," *Scientific reports*, vol. 4, no. 1, pp. 1–6, 2014.
- [31] F. Tsow, E. Forzani, A. Rai, R. Wang, R. Tsui, S. Mastroianni, C. Knobbe, A. J. Gandolfi, and N. Tao, "A wearable and wireless sensor system for real-time monitoring of toxic environmental volatile organic compounds," *IEEE Sensors Journal*, vol. 9, no. 12, pp. 1734–1740, 2009.
- [32] I. Negi, F. Tsow, K. Tanwar, L. Zhang, R. A. Iglesias, C. Chen, A. Rai, E. S. Forzani, and N. Tao, "Novel monitor paradigm for real-time exposure assessment," *Journal of exposure science & environmental epidemiology*, vol. 21, no. 4, pp. 419–426, 2011.
- [33] R. A. Iglesias, F. Tsow, R. Wang, E. S. Forzani, and N. Tao, "Hybrid separation and detection device for analysis of benzene, toluene, ethylbenzene, and xylenes in complex samples," *Analytical chemistry*, vol. 81, no. 21, pp. 8930–8935, 2009.
- [34] R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms," in *Proceedings of the 23rd international conference on Machine learning*, 2006, pp. 161–168.
- [35] M. Ren, E. S. Forzani, and N. Tao, "Chemical sensor based on microfabricated wristwatch tuning forks," *Analytical chemistry*, vol. 77, no. 9, pp. 2700–2707, 2005.
- [36] S. A. Sampson, S. V. Panchal, K. S. Date, and S. S. Datar, "Qtf based methanol sensors," in *2015 2nd International Symposium on Physics and Technology of Sensors (ISPTS)*. IEEE, 2015, pp. 303–306.
- [37] S. A. Sampson, K. S. Date, S. V. Panchal, A. Ambrale, and S. S. Datar, "Investigation of qtf based gas sensors," *Sensors and Actuators B: Chemical*, vol. 216, pp. 586–594, 2015.
- [38] R. O. Duda, P. E. Hart *et al.*, *Pattern classification*. John Wiley & Sons, 2006.
- [39] L. Guo, Y. Ma, B. Cukic, and H. Singh, "Robust prediction of fault-proneness by random forests," in *15th international symposium on software reliability engineering*. IEEE, 2004, pp. 417–428.
- [40] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [41] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE transactions on pattern analysis and machine intelligence*, vol. 20, no. 8, pp. 832–844, 1998.

Suresh Panchal received the MSc Physics from the Department of Physics, University of Pune, Pune in 2014 and joined on the project 'DRDO-DIAT project on nanomaterial's. In 2017, he registered for PhD in the Department of Applied Physics, Defence institute of Advanced Technology, Pune under Dr. Unnikrishnan Gopinathan and Dr. Suvarna Datar's guidance. Currently, he is working in the field of Computational Imaging and Sensing as Ph.D. research topic. He is applying machine learning techniques to sensor array

data analytics. Recently he addressed problems in image acquisition using Scanning Electron Microscope and Mid-Wave Infrared Camera by Deep Learning. His current research interest is in the field of Sensing, Imaging, Microscopy (Electron Microscopy, Atomic Force Microscopy, and Optical Microscopy), MW-IR Photonics, and Data Analytics.

Apoorva Phadke is a recent graduate in Statistical Science (Data Science track) from Leiden University, Netherlands. Before that, in 2017–2018, she worked as a research assistant at DIAT in pattern recognition and clustering of sensor data. In her master's thesis, she worked with the data team of Red Cross Netherlands, to use and implement machine learning for text classification. It included web scraping of the newspaper articles on disasters, comparison of statistical and machine learning methods to classify the articles. She also worked on data extraction from the articles using Natural Language Processing.

Stuttgart, Stuttgart, Germany.

Unnikrishnan Gopinathan received the Ph.D degree in physics from Indian Institute of Technology Delhi, New Delhi, India, in 2002. He is currently a Scientist with the Instrument Research and Development Establishment, Dehradun, India. His research interest is broadly in the area of Coherent optical signal processing. Dr. Gopinathan is a member of Optical Society of America. He was an Alexander Von Humboldt Fellow during the year 2007–2008 with the Institut für Technische Optik, Universität

Suvarna Datar received her PhD from the University of Pune, Pune, Maharashtra, India in Scanning Probe Microscopy and Spectroscopy Techniques. She is presently an Assistant Professor in Department of Applied Physics in Defence Institute of Advanced Technology (DIAT) a Deemed University. Her research interests are probe microscopy, graphene and graphene composites for EMI shielding, MEMS based sensors, and breathomics.