

Lung Cancer Identification by an Electronic Nose based on an Array of MOS Sensors

Rossella Blatt, Andrea Bonarini, Elisa Calabró, Matteo Della Torre, Matteo Matteucci, Ugo Pastorino

Abstract— We present a method to recognize the presence of lung cancer in individuals by classifying the olfactory signal acquired through an electronic nose based on an array of MOS sensors. We analyzed the breath of 101 persons, of which 58 as control and 43 suffering from different types of lung cancer (primary and not) at different stages. In order to find the components able to discriminate between the two classes ‘healthy’ and ‘sick’ as best as possible and to reduce the dimensionality of the problem, we extracted the most significant features and projected them into a lower dimensional space, using Non Parametric Linear Discriminant Analysis. Finally, we used these features as input to several supervised pattern classification techniques, based on different k -nearest neighbors (k -NN) approaches (classic, modified and Fuzzy k -NN), linear and quadratic discriminant classifiers and on a feedforward artificial neural network (ANN). The observed results, all validated using cross-validation, have been satisfactory, achieving an accuracy of 92.6%, a sensitivity of 95.3% and a specificity of 90.5%. These results put the electronic nose as a valid implementation of lung cancer diagnostic technique, being able to obtain excellent results with a non invasive, small, low cost and very fast instrument.

I. INTRODUCTION

Nowadays the research on olfactory systems has become very lively, most of all because of the multitude of applications in which it has been successfully used [1]. The olfactory signal, as well as other signals perceived through human senses, transports much more information than human beings are able to perceive; an electronic nose is an instrument that allows to acquire this kind of signal. An electronic nose is composed by an array of electronic chemical sensors with partial specificity and an appropriate pattern recognition system able to recognize simple or complex odors [2]. In the medical field, clinicians have always considered odor as a fundamental information for the diagnosis, according to the fundamental principle of clinical chemistry, namely the fact that every pathology changes people chemical composition, modifying the concentration of some chemicals in the human body. For this reason an electronic nose could be used to automatically analyze substances produced and emitted from

Rossella Blatt, Andrea Bonarini, and Matteo Matteucci are with the Department of Electronics and Information, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy. E-mail: blatt@elet.polimi.it, bonarini@elet.polimi.it, matteucci@elet.polimi.it, phone: +39-02-2399-3482

Ugo Pastorino and Elisa Calabró are with the Toracic Surgery Department, Istituto Nazionale Tumori, Milano, Italy. E-mail: ugo.pastorino@istitutotumori.mi.it, elisa.calabro@istitutotumori.mi.it, phone: +39-02-23902770

Matteo Della Torre is with Automation & Inspection Systems SACMI Imola S.C., Via Provinciale Selice 17/a, 40026 Imola, (BO), Italy. E-mail: matteo.della.torre@sacmi.it, phone: +39-0542-607809

the human body finding, in a rapid and non invasive way, several diseases [3], [4].

This paper focuses on the diagnosis of lung cancer; it has been demonstrated that this type of cancer alters the percentage of some volatile organic compounds (VOCs) present in the human breath [5], [6], which may be considered as markers of this disease. Whenever the breath of a healthy person gives a different pattern than that of a sick one, this difference could be detected by the electronic nose and an appropriate pattern classification algorithm, without the need for expensive techniques to extract VOCs (e.g., gas chromatography). The main objective of this paper is to demonstrate that it is possible to recognize individuals affected by lung cancer, analyzing the olfactory signal of their breath by the use of an electronic nose with an appropriate classification algorithm.

II. FUNCTIONING OF THE ELECTRONIC NOSE

An electronic nose is an instrument able to detect and recognize odors, namely the volatile organic compounds present in an analyzed substance. It is composed of an array of non specific electronic devices able to convert a physical or chemical information into an electrical signal. Each sensor reacts in a different way to the analyzed substance providing multidimensional data that can be considered as an olfactory blueprint of the substance itself. The output of an electronic nose can be the detection of a specific substance, an estimate of the concentration of the odor or some particular characteristic of the odor that allows to associate it to a particular class/situation.

An electronic nose consists in three principal components:

- 1) Gas Acquisition System
- 2) Pre-processing and Dimensionality Reduction
- 3) Classification Algorithm

In particular, the *acquisition* of the olfactory signal is done through a sensor array that measures a given physical or chemical quantity; the recorded data are then sent to a *processing system* that reduces the impact of the noise and extracts the most significant features from the signal. Once the most representative characteristics are found, it is possible to proceed with the analysis of the data, that, in this particular case, consists of a *pattern recognition* algorithm.

Current types of electronic noses are different according to the sensor technology used and, therefore, according to the measured physical or chemical quantity. MOS sensors are characterized by high sensitivity (in the order of parts per billion ppb), low cost, high speed response and a relatively

Fig. 1. Example of a typical sensor response.

simple electronics. Considering that most of the VOCs markers of lung cancer are present in the diseased people's breath in very small quantities, varying from parts per million to parts per billion, we have chosen to use this kind of sensors rather than others. In particular, we used an array composed of six MOS sensors (developed by SACMI s.c.), that react to gases with a variation of resistance. The VOCs interact with a doped semiconducting material deposited between two metal contacts over a resistive heating element, which operates from 200 °C to 400 °C. As a VOC passes over the doped oxide material, the resistance between the two metal contacts changes in proportion to the concentration of the VOC. The registered signal corresponds to the change of resistance through time produced by the gas flow [7].

In Figure 1 it is possible to see a typical response of a MOS sensor. In particular, each measure consists of three main phases:

- 1) **Before:** during this time the instrument inhales the reference air, showing in its graph a relatively constant curve;
- 2) **During:** it is the period in which the electronic nose inhales the analyzed gas, producing a change of the sensors' resistance. It is the most important part of the measurement because it contains informations about how sensors react to the particular substance;
- 3) **After:** during this phase the instrument returns to the reference line.

III. DATA PROCESSING AND DIMENSIONALITY REDUCTION

The first computational phase, after the electronic nose has acquired the olfactory signal, is pre-processing; its purpose is to reduce the effect of humidity, to normalize the obtained signal and to manipulate the baseline. The latter transforms the sensor response w.r.t. its baseline (e.g., response to a reference analyte) for the purposes of contrast enhancement and drift compensation [1]. In this particular case, the drift effects were not so relevant and we noticed that the performance with or without calibration were quite the same, probably because the humidity correction partially solved also the drift problem. In order to reduce the humidity impact, we subtracted its correlation to the signal from the signal itself. It is important to point out that also the humidity is susceptible of the drift phenomenon and this is the reason why humidity correction indirectly affects calibration. Normalization is used to compensate for the scale difference between the sensors in the array; for this reason we forced all sensors to have zero mean and variance equal to 1.

After pre-processing we performed dimensionality reduction to extract only the most relevant information from the signal. We reached these objectives using three different techniques:

- 1) features extraction

- 2) features selection
- 3) features projection in a lower dimensional space

The first operation extracts those descriptors from the sensors' responses able to represent data characteristics in the most efficient way. We computed 10 features, based on the variation of resistance, the course of the curve, its derivative, integral and Fast Fourier Transform (FFT). Some of these features returned more than one value (like the FFT), for a total of 39 descriptors for each measurement. Considering that we used 6 sensors, each measure would be described by 234 descriptors. Therefore, the need for a feature selection.

Among all features it is necessary to find those that maximize the informative components and, thus, the accuracy of the classifier. For this reason we applied the non-parametric test of Mann-Whitney-Wilcoxon [8] with a significance level equal to $\alpha = 0.0001$ to select only discriminant descriptors. The choice of using a non-parametric test instead of a parametric one, is due to a previous analysis of the features distribution and a Lilliefors test. In order to evaluate the discriminative ability of the combination of more features, we performed an Analysis of Variance (ANOVA) [8] and several scatter plots.

Let define $R(t)$ the curve representing the resistance variation during the measurement and R_0 the value of the resistance at the beginning of the measurement (as indicated in Figure 1), we found as the most discriminative features between the two classes 'healthy' and 'sick':

- *Delta:* resistance change of sensors during measurement:

$$\delta = R_0 - \min(R(t)) \quad (1)$$

- *Classic:* the ratio between the reference line and the minimum value of resistance reached during the measurement:

$$C = R_0 / \min(R(t)) \quad (2)$$

- *Relative Integral:* calculated as:

$$I = \int R(t) / (t \cdot R_0) \quad (3)$$

- *Phase Integral:* the closed area determined by the plot of the state graph of the measurement [9]:

$$x = R, \quad y = dR/dt \quad (4)$$

- *Single Point:* the minimum value of resistance reached during the measurement.

$$S = \min(R(t)) \quad (5)$$

After feature selection we performed data projection: we considered Principal Component Analysis (PCA) [10] and Nonparametric Linear Discriminant Analysis (NPLDA) [11], that is based on nonparametric extensions of commonly used Fisher's linear discriminant analysis [10]. PCA transforms data in a linear way projecting features into the directions with maximum variance. It is important to notice that PCA does not consider category labels; this means that the discarded directions could be exactly the most suitable for

the classification purpose. This limit can be overcome by NPLDA, which looks for the projection able to maximize differences between different classes and minimize those intra-class. In particular, NPLDA removes the unimodal gaussian assumption by computing the between scatter-matrix S_b using local information and the k nearest neighbors rule; as a result of this, the matrix S_b is full-rank, allowing to extract more than $c-1$ features (where c is equal to the number of considered classes) and the projections are able to preserve the structure of the data more closely [11]. As evident from Figure 2, NPLDA is able to separate the projected features more clearly than PCA, which plot shows a more evident overlap of samples. This means that NPLDA is more suitable, for the problem considered, in terms of classification performance. Moreover, the plot and the obtained eigenvalues clearly indicated that only one principal component is needed.

IV. CLASSIFICATION

After the olfactory signal has been processed and after extracting the most significant features, it has been possible to perform classification. We considered three families of classifiers:

- 1) Nearest Neighbors Classifiers (k -NN);
- 2) Linear and Quadratic Discriminant Function based Classifiers (LD and QD);
- 3) Artificial Neural Network (ANN).

A. Nearest Neighbors

The basic idea of this simple and powerful algorithm is to assign a sample to the class of the k closest samples in the training set. This method is able to do a non linear classification starting from a small number of samples. The algorithm is based on a measure of the distance (in this case, the Euclidean one) between the normalized features, and it has been demonstrated [10], that the k -NN is formally a non parametric approximation of the Maximum A Posteriori MAP criterion. The asymptotic performance of this algorithm, is almost optimum: with an infinite number of samples and setting $k=1$, the minimum error is never higher than the double of the Bayesian error (that is the theoretical lower bound reachable) [12].

One of the most critical aspects of this method regards the choice of parameter k with a limited number of samples: if k is too large, then the problem is too much simplified and the local information loses its relevance. On the other hand, a too small k leads to a density estimation too sensitive to outliers. For this reason, in addition to the classic k -NN, we implemented two other versions of this technique: the Modified k -NN and the Fuzzy k -NN. In the former, k means the number of closest neighbors to look for (as in the classic approach), but belonging all to the same class. This dynamically modify the neighborhood according to the noise in the dataset. Fuzzy k -NN, a variation of the classic k -NN based on a Fuzzy logic approach [13], assigns a fuzzy class membership to each sample and provides an output in a fuzzy form. In particular, the membership value of unlabeled

sample x to i^{th} class is influenced by the inverse of the distances from neighbors and their class memberships:

$$\mu_i(x) = \frac{\sum_{j=1}^k \mu_{ij} (\|x - x_j\|)^{\frac{-2}{m-1}}}{\sum_{j=1}^k (\|x - x_j\|)^{\frac{-2}{m-1}}} \quad (6)$$

where μ_{ij} represents the membership of labeled sample x_j to the i^{th} class. This value can be crisp or it can be calculated according to a particular fuzzy rule: in this work we defined a fuzzy triangular membership function with maximum value at the average of the class and null outside the minimum and maximum values of it. In this way, the closer the sample j is to the average point of class i , the closer its membership value μ_{ij} will be to 1 and vice versa. The parameter m determines how heavily the distance is weighted when calculating each neighbor's contribution to the membership value [14]; we chose $m = 2$, but almost the same error rates have been obtained on these data over a wide range of values of m .

B. Discriminant Functions Classifier

Classification based on discriminant functions represents a geometric approach in which the features space is divided in c decision regions each one corresponding to a particular class. The idea is to represent the classifier as a family of discriminant functions $g_i(x)$ with only one output that should minimize a certain cost function. We considered two types of discriminant functions: the linear (LD) and the quadratic one (QD). A classifier based on a linear discriminant function divides the features space by planes and it is therefore optimum when the problem is linearly separable. In any case, this technique is able to lead to good performances also when the problem is not linearly separable. We implemented the Minimum Distance to Means (MDM) approach, in which the representatives of each class have been calculated as the mean value of samples belonging to that class. This approach is very simple and lead to good generalization; the drawback is that it compresses all information in only one representative value. If the problem is not linearly separable, a quadratic discrimination function could be more suitable, as has also verified in this work.

C. Artificial Neural Network

Artificial Neural Networks (ANN) are non-linear statistical modeling tools that can be used to model complex relationships between inputs and outputs or to find patterns in data. It can be demonstrated that an ANN, given a sufficient number of sigmoidal neurons in the hidden levels, is able to approximate any non linear function on a compact set. Moreover ANNs asymptotically (with an infinite number of examples) approximate the a-posteriori probability as with the Bayesian classifiers [15].

One of the main drawbacks of this method regards the impossibility to decide a priori the best topology to use. This choice has therefore been made through an empirical approach. In particular, we chose to use a feedforward neural network with one hidden layer, in which inputs are the first

Fig. 2. The result of dimensionality reduction through PCA on the left and NPLDA on the right.

principal component obtained by NPLDA and the output is a single neuron assuming the value 1 if the presence of the disease is detected and 0 otherwise. All neurons have a sigmoidal function as activation function. The net has been trained using the Resilient Backpropagation algorithm, based on the gradient descent approach, in which only the sign of the derivative is used to determine the direction of the weights update [16]. This choice is due to the fact that this algorithm was able to offer the best compromise between the error on the validation and convergence. Finally, we set the number of neurons in the hidden layer equal to 3; this value has been obtained by training a set of networks with increasing number of hidden neurons and picking the smallest one with a good validation error. Since ANN's results depend on the values of the initialization, we trained the net 20 times and we choose the best configuration (according to the early stopping error) to evaluate the test set.

V. METHODOLOGY

The experiment has been developed within the Italian MILD (Multicentric Italian Lung Detection) project, promoted by the Istituto Nazionale Tumori, Italy. We analyzed the breath of 101 volunteers, of which 58 healthy and 43 suffering from different types of lung cancer. All cases were hospitalized at the Istituto Nazionale Tumori of Milan. Among them, 23 have a primary lung cancer, while 20 of them have different kinds of pulmonary metastasis. Control people have no pulmonary disease and have negative chest CT scan. The study has been approved from the Ethical Committee of the Institute and we asked everybody to sign an agreement for the participation to the study.

The breath acquisition has been made by inviting all volunteers to blow into a nalophan bag of approximately 400cm^3 . Considering that the breath exhaled directly from lung is contained only in the last part of exhalation, we decided to consider only this portion of the breath. We used a spirometer to evaluate each volunteer exhalation capacity and, at the end of the exhalation, we diverted the flow into the bag. Finally, the air contained in the bag has been input to the electronic nose and analyzed. From each bag we took two measures, obtaining a total of 202 measurements, of which 116 correspond to the breath of healthy people and 86 to diseased ones.

The performance of the classifiers has been evaluated through the obtained confusion matrices and performance indexes; being 'TruePositive' (TP) a sick sample classified as sick, 'TrueNegative' (TN) a healthy sample classified as healthy, 'FalsePositive' (FP) a healthy sample classified as sick and 'FalseNegative' (FN) a sick sample classified as healthy we used:

- **Accuracy** (Non Error Rate NER): the probability of

doing a generic correct classification;

$$NER = \frac{TP + TN}{TP + FP + TN + FN} \quad (7)$$

- **Sensitivity** (True Positive Rate TPR): the probability to classify a person as sick when this is true;

$$TPR = \frac{TP}{TP + FN} \quad (8)$$

- **Specificity** (True Negative Rate TNR): the probability of classifying a person as healthy when this is true;

$$TNR = \frac{TN}{TN + FP} \quad (9)$$

- **Precision w.r.t. diseased people** ($PREC_{POS}$): the probability that, having assigned a sample to the class of diseased people, it actually belongs to that class.

$$PREC_{POS} = \frac{TP}{TP + FP} \quad (10)$$

- **Precision w.r.t. healthy people** ($PREC_{NEG}$): the probability that, having assigned a sample to the class of healthy people, it actually belongs to that class.

$$PREC_{NEG} = \frac{TN}{TN + FN} \quad (11)$$

To obtain indexes able to describe in a reliable way the performances of the algorithms, it is necessary to evaluate these parameters on new and unknown data, validating the obtained results. Considering the not so big dimension of population and that for every person we had two samples, we opted for a modified Leave-One-Out approach; each test set is composed by the pair of measurements corresponding to the same person, instead of a single measure as would be in the normal Leave-One-Out method. Doing this way, we avoided that one of these two measures could belong to the training set, while using the other in the test set.

In order to deeply understand the relevance of the obtained performance indexes, we calculated the corresponding confidence intervals, which lower and upper bounds are defined as:

$$\bar{X} - t_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu_x \leq \bar{X} + t_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \quad (12)$$

where \bar{X} is the registered index value, n is the number of the degrees of freedom, σ is the standard deviation and $t_{\frac{\alpha}{2}}$ is the quantile of the t-student distribution corresponding to the degrees of freedom-1 of the problem.

VI. RESULTS

All implemented algorithms have demonstrated a good ability to discriminate the two classes 'healthy' and 'sick'. Performance indexes are reported in Table I, where we considered the first principal component obtained from NPLDA. The first consideration regards the similarity of Modified and Classic k -NN: results are strongly comparable, but a

TABLE I

PERFORMANCE INDEXES AND CORRESPONDING CONFIDENCE INTERVALS (CI=95%) OBTAINED FROM CONSIDERED ALGORITHMS. FEATURES HAVE BEEN PREVIOUSLY PROJECTED BY NPLDA AND ONLY THE FIRST PRINCIPAL COMPONENT HAS BEEN KEPT FOR CLASSIFICATION. FOR k -NN TECHNIQUES, WE CONSIDERED $k=1,3,5,9,101$; FOR CLASSIC AND MODIFIED k -NN WE SHOW THE BEST ACHIEVED RESULTS (WHEN $k=9$). ON THE CONTRARY, FUZZY k -NN LED TO THE SAME RESULTS INDEPENDENTLY FROM k 'S VALUES.

Classifier	NER	TPR	TNR	PREC _{POS}	PREC _{NEG}
Classic 9-NN	90.1%	89.5%	90.5%	87.5%	92.1%
Confidence Interval	[85.7-94.5]	[85.3-93.8]	[86.0-95.0]	[81.6-93.4]	[86.8-97.4]
Modified 9-NN	91.1%	91.9%	90.5%	87.8%	93.7%
Confidence Interval	[86.8-95.4]	[87.9-95.9]	[86.0-95.0]	[81.9-93.7]	[89.1-98.4]
Fuzzy k -NN	92.6%	95.3%	90.5%	88.2%	96.3%
Confidence Interval	[88.5-96.7]	[91.8-98.9]	[86.0-95.0]	[82.3-94.1]	[93.2-99.4]
LD	89.6%	96.5%	84.5%	82.2%	97.0%
Confidence Interval	[85.0-94.2]	[93.7-99.3]	[79.1-89.9]	[75.2-89.1]	[93.9-100]
QD	92.6%	95.3%	90.5%	88.2%	96.3%
Confidence Interval	[88.5-96.7]	[91.8-98.9]	[86.0-95.0]	[82.3-94.1]	[93.2-99.4]
ANN	91.6%	91.9%	91.3793%	88.8%	93.8%
Confidence Interval	[87.4-95.8]	[87.9-95.9]	[87.0-95.8]	[84.1-93.4]	[88.2-99.4]

TABLE II

CONFUSION MATRIX OBTAINED FROM FUZZY k -NN AND QUADRATIC DISCRIMINANT FUNCTIONS ALGORITHMS.

CONFUSION MATRIX		TRUE LABELS	
		Positive	Negative
ESTIMATED LABELS	Positive	82	11
	Negative	4	105

slight improvement is shown by Modified k -NN. Moreover Modified k -NN is able to achieve same performance as Classic k -NN with a lower k value. Another relevant consideration regards the robustness of Fuzzy k -NN to k changes: we considered different values of k ($k=1,3,5,9,101$), but the algorithm demonstrated to be robust to these changes, keeping its results invariant. In diagnostic field, sensitivity is more important than specificity because it is more relevant to recognize correctly a sick person instead of a healthy one; in the same way, precision on negative samples is more important than precision on positive ones, because it is worse to classify a person as healthy when he or she is actually sick, than the opposite. Considering larger importance of sensitivity and precision w.r.t. healthy samples, we can affirm that Fuzzy k -NN and Quadratic classifier are the algorithms able to achieve best results for the problem considered. The confusion matrix obtained by these algorithms is shown in Table II, where elements along the principal diagonal represent respectively the TruePositive (TP) and the TrueNegative (TN) values, while those off-diagonal are respectively the FalsePositive (FP) and the FalseNegative (FN) values.

Performing a Student's t-test between all pair of classifiers, no relevant differences emerged; this means that implemented classifiers' results are comparable for the considered problem.

TABLE III

COMPARISON OF LUNG CANCER DIAGNOSIS PERFORMANCE REACHED WITH THE ELECTRONIC NOSE PRESENTED IN THIS WORK AND CURRENT DIAGNOSIS TECHNIQUES (DATA FROM [17]).

Indexes	CAT	PET	E-Nose
Accuracy (NER)	Nd	Nd	92.6%
Confidence Interval			[88.5-96.7]
Sensitivity (TPR)	75%	91%	95.3%
Confidence Interval	[60-90]	[81-100]	[91.8-98.9]
Specificity (TNR)	66%	86%	90.5%
Confidence Interval	[55-77]	[78- 94]	[86.0-95.0]
PREC _{POS}	Nd	Nd	88.2%
Confidence Interval			[82.3-94.1]
PREC _{NEG}	Nd	Nd	96.3%
Confidence Interval			[93.2-99.4]

VII. CONCLUSION AND FURTHER DIRECTION OF RESEARCH

The use of an electronic nose as lung cancer diagnostic tool is reasonable if it gives some advantage compared to current lung cancer diagnostic techniques, namely Computed Axial Tomography (CAT) and Positron Emission Tomography (PET). Not only this is verified in terms of performances, as illustrated in Table III, but also because the electronic nose, unlike the classical approaches, is a low cost, robust, small (and thus eventually portable), very fast and, above all, non invasive instrument.

In literature there are three other main research works regarding lung cancer diagnosis by an electronic nose [18], [19], [20]. Accuracy indexes obtained from these works were respectively equal to 90.32%, 88.16% and 80%. Moreover, in [19] and [20], no cross-validation techniques has been applied to obtain such results; this means that results have been obtained from one realization and, therefore, they are not necessarily representative of the real generalization capability of the classifier.

Our work could be extended in two parallel directions: the first one regards the improvement of sensors technology with the development of longer-life and stable sensors. Moreover, the development of hybrid systems is desirable, in order to obtain both selective and sensitive sensors. The second direction regards the improvement of classification techniques in which we put in evidence the importance of evaluating other classification algorithms (as support vector machines, Bayesian approaches or other topologies of ANN), as well as the use of different algorithms to select the best subgroup of features for each classifier, instead of using ranking techniques based on statistical tests. It could be also very interesting to train the ANN in presence of noise, since it has been demonstrated that ANNs can compensate humidity, drift and temperature variation phenomenons [21] that affect olfactory signals.

According to the scientific literature, there are no studies on the variation of VOCs in the breath before and after the surgery: it may be interesting to evaluate the resolution of the disease due to surgery. An ambitious research prospective regards the individuation of risk factors connected to lung cancer (as smoke or food). Involving a larger population and partitioning it according to different disease stages, it would be possible to study the possibility of early diagnosis, that is the most important prospective of research that this work should follow.

REFERENCES

- [1] Osuna R.G., Nagle H.T., Shiffman S.S., The how and why of electronic nose, *IEEE Spectrum*, pp. 22-34, September 1998
- [2] Gardner J.W., Bartlett P.N., *Electronic noses. Principles and applications*, Oxford University Press, USA, 1999
- [3] Gardner J.W., Shin H.W., Hines E.L., An electronic nose system to diagnose illness, *Sensors and Actuators B: Chemical*, vol. 70, no 1, pp. 19-24, 2000
- [4] Kodogiannis V.S., Chountas P., Pavlou A., Petrounias I., Chowdrey H.S., Temponi C., *Electronic nose: clinical diagnosis based on soft computing methodologies*, *Intelligent Systems, 2002-Proceedings-2002 First International IEEE Symposium*, vol. 1, pp 254- 259, 2002
- [5] Gordon S.M., Szidon J.P., Krotoszynski B.K., Gibbons R.D., O'Neill H.J., Volatile organic compounds in exhaled air from patients with lung cancer, *Clinical chemistry*, vol. 31, no 8, pp. 1278-1282, 1985
- [6] Phillips M.D., Cataneo R.N., Cummin A.R.C., Gagliardi A.J., Gleeson K., Greenberg J., Maxfield R.A., Rom W.N., Detection of lung cancer with volatile markers in the breath, *Chest*, vol. 123, no 6, pp. 2115-2123, 2003
- [7] Pardo M., Sberveglieri G., *Electronic olfactory systems based on metal oxide semiconductor sensor arrays*, In *Material Research Society Bulletin*, vol. 29, no 10, October 2004
- [8] Lyman Ott R., Longnecker M.T., *Introduction to statistical methods and data analysis*, Duxbury Press, 5th edition, 2001
- [9] E. Martinelli, C. Falconi, A. D'Amico, C. Di Natale, Feature Extraction of chemical sensors in phase space, *Sensors and Actuators B: Chemical*, (95)1, pp. 132-139, Elsevier Science, 2003
- [10] Fukunaga K., *Introduction to statistical pattern recognition*, Academic Press, Second Edition, San Diego, 1990
- [11] Fukunaga K., Mantock J.M., Nonparametric discriminant analysis, *IEEE Transactions on pattern analysis and machine intelligence*, vol. PAMI 5, no 6, pp. 671-678, 1983
- [12] Duda R.O., Hart P.E., Stork D.G., *Pattern classification*, Wiley-Interscience, 2nd edition, October 2000
- [13] Zadeh L., Fuzzy Sets, *Information and control*, vol. 8, pp. 338-353, 1965
- [14] Keller J.M., Gray M.R., Givens J.A., A fuzzy k -Nearest neighbor algorithm, *IEEE Transactions on systems, man and cybernetics*, vol. 15, no 4, pp. 580-585, July-August 1985
- [15] Bishop C.M., *Neural networks for pattern recognition*, Oxford University Press, 1996
- [16] Demuth H., Beale M., *Neural network toolbox for use with MATLAB*, The MathWorks, 1998
- [17] Pieterman R.M., Van Putten J.W.G., Meuzelaar J.J., Mooyaart E.L., Vaalburg W., Koter G.H., Fidler V., Pruijm J., Groen H.J.M., Preoperative staging of non-small-cell lung cancer with positron-emission tomography, *The New England journal of medicine*, vol. 343, no 4, pp. 254-261, 2000
- [18] Di Natale C., Macagnano A., Martinelli E., Paolesse R., D'Arcangelo G., Roscioni C., Finazzi-Agro A., D'Amico A., 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
- [19] Machado R.F., Laskowski D., Deffenderfer O., Burch T., Zheng S., Mazzone P.J., Mekhail T., Jennings C., Stoller J.K., Pyle J., Duncan J., Dweik R.A., Erzurum S., Detection of Lung Cancer by Sensor Array Analyses of Exhaled Breath, *American Journal of Respiratory and Critical Care Medicine*, vol. 171, pp. 1286-1291, 2005
- [20] Chen X., Cao M., Li Y., Hu W., Wang P., Ying K., Pan H., A study of an electronic nose for detection of lung cancer based on a virtual SAW gas sensors array and imaging recognition method, *Measurement science & technology*, vol. 16, no 8, pp. 1535-1546 , 2005
- [21] Brezmes J., Canyellas N., Llobet E., Vilanova X., Application of artificial neural networks to the design and implementation of electronic olfactory system, 5th seminar on Neural Network Applications in Electrical Engineering, NEUREL-2000