Kernel Approach on Detection of Ethanol Concentration Using ZnO Gas Sensor

Laxmi Shaw, Sangeeta Bagha, Ambarish G. Mohapatra, and Narayan Nayak *Member, IEEE*

***Abstract*—Ethanol is the major consumable biofuel which is highly inflammable. So the detection of ethanol at its different concentration is certainly required. In this paper a cost effective thick-film gas sensor based on ethanol detection has been described. The sensor exhibits excellent ethanol sensing characteristics at the temperatures between 175 °C to 300 °C and the characteristics of the sensor for different concentration of the ethanol gas has been successfully studied by performing a nonlinear form of principal component analysis (PCA) to cluster the different concentration of ethanol. The present method is to classify the 230 samples of each concentration levels of gas using kernel principal component analysis (KPCA). Ethanol gas is a volatile organic compound (VOC) which is little difficult to distinguish the different concentration level in their nominal method.In this regard, classifiers having the machine learning ability can be of great benefit by automatically including the newly presented patterns in the training dataset without affecting class integrity of the previously trained system. In the presented paper, the kernel principal component analysis (KPCA) is used in the clustering algorithm for ethanol concentration discrimination with the help of thick film gas sensor.**

***Index Terms*—Cluster computing, kernel, principal component analysis, feature space, nonlinear mapping, pattern recognition.**

I. Introduction

The thick film ZnO gas sensor has been successfully employed for detection and measure of different concentration of various gases viz., CO, CO_2, and ethanol etc [1]. This paper, clustering of different concentration level of ethanol gas has been considered, and in this context, the applicability of kernel principal component analysis for data clustering has been demonstrated.The proposed system has been used to detect and display different gas concentration in the LabView front panel. The sensor is a thick film based zinc oxide gas sensor. The sensor is designed to have an inner heating coil which is heated up to a temperature range of 75°C to 300°C.ZnO is one of the semiconductor materials that are sensitive to many gases of interest like hydrocarbons, hydrogen and other volatile organic compounds etc.

They exhibit high sensitivity, satisfactory stability and rapid response to even small concentrations of gases (p.p.m. level). In this paper we will concern ourselves with the sensitivity of ZnO thick film resistor for sensing ethanol gas at 100 p.p.m. of gas concentration at various temperatures from 75°C to 300°C in order to find the temperature for maximum sensitivity [2]. Also the sensitivity for different concentrations of ethanol (from 5 p.p.m. to 95 p.p.m.) was tested. The proposed system contains an array of sensing units for the detection and transmission of real-time data to the user side.

Manuscript received December 25, 2011; revised February 21, 2012.

Laxmi Shaw, Sangeeta Bagha, Ambarish G. Mohapatra and Narayan Nayak are with the Silicon Institute of Technology, Bhubaneswar, Odisha, India (e-mail: laxmishaw1983@ gmail.com; sangeetabagha@ gmail.com; ambarish.mohapatra@gmail.com; narayan.nayak@silicon.ac.in).

II. System Configuration

In the system we have used the following units to detect the Zinc-Oxide characteristics.

a. Zinc Oxide based thick-film sensor with its arrangement
b. Internal Heating Coil
c. Data Acquisition hardware and software
d. Digital temperature recorder
e. Power supply Unit
f. Power source for heating coil

Zinc-Oxide Sensor Preparation

By making use of change in resistivity on exposure to the relevant gas, ZnO has been traditionally employed as a gas sensor. ZnO thick films are used on an alumina substrate for sensing ethanol. The substrate was an alumina tube about 2cm long. Electrodes were wound over it using nichrome wires. Thick film preparation method was very simple and inexpensive. First, the commercial ZnO powder was reduced to a very fine powder by constant grinding for about an hour. A paste of it was then prepared by simply adding few drops of distilled water. Finally the thick film was made by applying the paste of the grinded commercial ZnO powder over an alumina tube (substrate) containing the electrodes using a fine paint brush. Applying a smooth coating such that no air gaps or cracks were present between the coating and the electrodes. The sample was then dried in the air for 24 hours and finally annealed at 300°C for 2 hours [1]. The coating that we got was very smooth and had good adherence with the alumina tube after annealing. Finally a coiled nichrome wire was inserted through the tube to act as a heater. The complete structure of the thick film over the alumina substrate is required for the data collection.

III. Experimental Procedure

Gas sensing characteristics of ZnO thick-film sensor were measured with a home built apparatus consisting of a testing glass chamber, sensor holder, RTD, heating coil arrangement as a heater, temperature controller, National Instruments data acquisition software, digital temperature detector, required

circuitry and a Host controller. In the temperature range of 75°C to 300°C, the ZnO sensor was successfully tested for ethanol vapor. Upto 100 p.p.m. of ethanol was injected by micropipette into the testing chamber. The working temperature of the sensor was adjusted by changing the voltage across the heater side [1]. The working circuit shown in the Fig. 1 is used to monitor the output voltage across the load resistor, corresponding voltage across the sensor and the resistance of the gas sensor in dry air as well as in test gas to measure the concentration level. The Zinc-Oxide gas sensor is placed inside a glass envelop with its inlet valve for the gas tesing and concentraion measurement.Fig. 2. Shows the characteristics of ZnO sensor at different concentration in Matlab plot.

Fig. 1. Zinc Oxide Gas Sensor with heating coil arrangement [1]

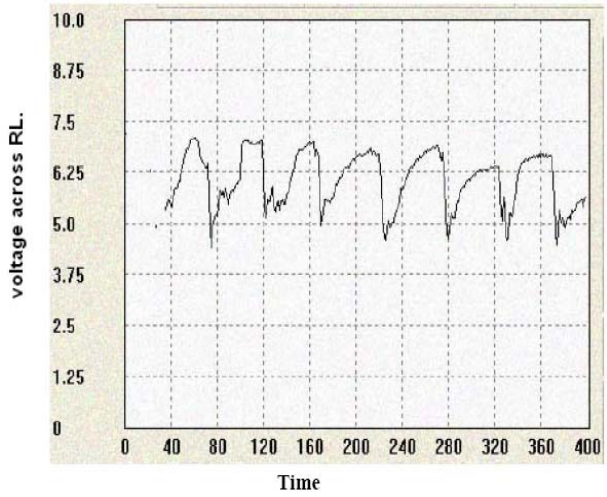

Fig. 2. Gas sensitivity of the sensor at different concentration of Ethanol

From Table I V_{air} and V_{gas} are the voltage obtained across the load resistor R_L. The gas sensitivity, S, was determined as $[(R_{air} – R_{gas})/R_{air}]*100$ .Where R_{air} is the resistance of the gas sensor in the dry air while R_{gas} is that of the sensor in the test gas. The sensor resistance can be found out from the simple formula: R_S (sensor in air or gas) = $R_L*(V_S- V_L)/ V_L$, where V_S is the source voltage (here 10 volts) and V_L is the load voltage which we are measuring directly as V_{air} and V_{gas}.As soon as ethanol gas was introduced into the sensing chamber, the electrical resistance (R_{gas}) of the thick film sensor started to decrease as compared to what was in dry air (R_{air}). It is clear that the resistance of the sensor was in the 160 KΩ to 17 MΩ range in the dry air and was in the 60 KΩ to 15 MΩ range for ethanol at 75-300°C.The sensor responses can also be affected by the slight change of temperature. In Table I, we have shown the sensitivity measured at different temperatures. It should be noticed that at 250°C the sensor shows highest sensitivity which is around 75%.Above this temperature the sensitivity starts decreasing.

TABLE I: SENSITIVITY OF THE GAS SENSOR AT DIFFERENT TEMPERATURE RANGES [1]

Temp (^{0}C)	V_{air} (Volt)	R_{air} (Ohm)	V_{gas} (Volt)	R_{gas} (Ohm)	Sensitivity (S in %)
75	0.187	16.79E03	0.23	13.59E03	19.05
100	1.062	2.693E03	1.35	2.050E03	23.86
125	1,917	1.349E03	2.39	1.024E03	24.06
150	2.540	939	3.95	490.12	47.80
175	2.50	960	4.2	441.90	53.90
200	3.19	683.13	5.40	275.59	59.66
225	3.75	533.33	6.15	200.32	62.43
250	4.03	474.04	7.09	131.33	72.29
275	5.54	257.61	7.97	81.50	68.36
300	6.63	162.6	8.43	59.596	63.16

Data Collection

The response for different concentrations of ethanol was tested with the help of ZnO sensor. The thick film was tested for various concentrations of ethanol gas varying from 5 p.p.m. to 95 p.p.m. It was found that with the increase in concentration of ethanol the resistance continuously decreased. Thus the corresponding sensitivity also increased. The response time for the sensor was also calculated. The rise time defined as the time required for the conductance to reach 90% of the equilibrium value after the gas is injected [3]. The fall time is the time needed by the sensor to acquire 10% of the above the original value in air after the gas is removed. In Fig.2 we display the variation in voltage of the sensor at the different concentration ranging from 5 to 95 ppm at 250°C at different time intervals.

IV. SOFTWARE IMPLEMENTATION

The method of working is based on simple measurement procedures. The sensor is connected to a resistive voltage divider circuit as the output of the sensor is a resistance change with respect to the input gas concentration. Data acquisition hardware and software are used for signal processing of the sensor data. The gas concentration was calculated using National Instruments Data acquisition software (LabView-2009) which is shown in Fig.3. [4], [5].

The transfer characteristics of the sensor were studied at various ppm labels and are shown in Fig. 4. The method of working is based on simple measurement procedures. The signal from the sensor is connected to a resistive voltage divider circuit as the output of the sensor is a resistance

change with respect to the input gas concentration. The output of the voltage divider circuit is then fed to NI DAQ card for the processing using Lab View software. From the sensitivity curve the temperature for maximum sensitivity is found out.

Fig. 3. Experimental setup with heating coil at 250°C [1]

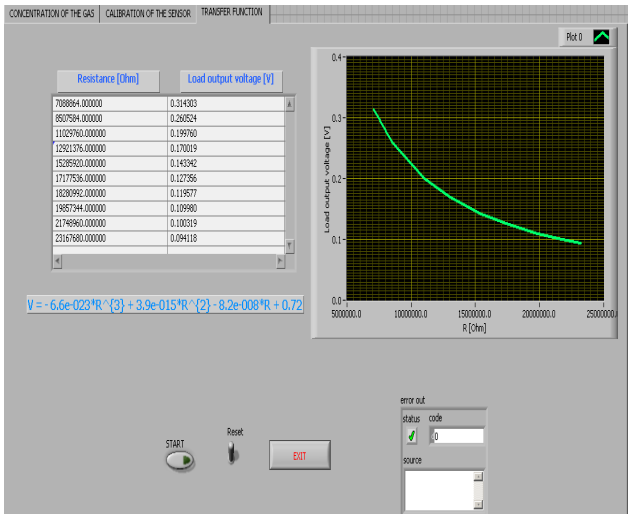

Fig. 4. Transfer function of the gas sensor

V. Data Analysis

A. Preprocessing Techniques

Data collection is the initial step for data investigation. The ZnO gas sensor generates electrical signals when subjected to the odour molecules and the signals are converted to numeric data using analog-to-digital converters. These data are then used for computation and finally for classification. This procedure often causes the difficulty in the classification problem as the original signal may be distorted due to the characteristics and limitations of the sensor [6]. In the context of ethanol gas sensor, the signals, usually considered, are the maximum of the sensor response, and these are the raw or non-normalized data. Prior to classification using a suitable pattern recognition algorithm, these raw data are normalized. An appropriate normalization technique may improve the pattern classification system in an gas sensor system, but there are no general guidelines to determine the appropriate normalization techniques [7].

B. Principal Component Analysis

The directions in which the data with the most variation, can be found out in principal component analysis. By using PCA data may be expressed and presented in such a way as to highlight their similarities and differences. Since patterns in data can be difficult to observe in data of high dimension. In the vector space, PCA identifies the major directions, and the corresponding strengths, of variation in the data [7]. PCA achieves this by computing the eigenvectors and eigenvalues of the covariance matrix of the dataset. Keeping only a few eigenvectors corresponding to the largest eigenvalues, PCA can be also used as a tool to reduce the dimensions of the dataset while retaining the major variation of the data. Dimension of a huge data set can be trimmed down by using principal component analysis which is considered as one of the most prevalent and useful statistical method [7]–[9]. This method transforms the original data in to new dimensions. The new variables are formed by taking linear combinations of the original variables of the form:

$$\begin{aligned} P_1 &= b_1^{'}K = b_{11}K_1 + b_{12}K_2 + \ldots + b_{1m}K_m \\ P_2 &= b_2^{'}K = b_{21}K_1 + b_{21}K_2 + \ldots + b_{2m}K_m \\ &\ldots\ldots \\ P_n &= b_n^{'}K = b_{n1}K_1 + b_{n1}K_n + \ldots + b_{nm}K_m \end{aligned} \quad (1)$$

In matrix style, we can write, $P = B.K$, where $b_{11}, b_{12}, b_{13}, \ldots b_{nn}$ are known as the loading parameters. The new axes are attuned such that they are orthogonal to one another with utmost expand of information.

$$\begin{aligned} Var(P_i) &= b_i^{'}\sum b_i, i = 1,2,3.....n \\ Cov(P_i, P_j) &= b_i^{'}\sum b_j, i = 1,2,3.....n \end{aligned} \quad (2)$$

K_1 is the first principal component holding the prime variance. As the direct computation of matrix B is not achievable. So, in feature transformation, the first step is to ascertain the covariance matrix *U* which can be expressed as

$$U_{m\times q} = \frac{1}{m-1}\left[\sum_{i=1}^{m}(K_i - \overline{K})'.(K_i - \overline{K})\right] \quad (3)$$

where $\overline{K} = (\frac{1}{m})\sum_{i=1}^{m} X_i$ (4)

The next step is to determine the eigen values for the covariance matrix 'U'[10], [11]. Eventually, a linear transformation is defined by n eigen vectors match up to q eigen values from a m-dimensional space to q-dimensional space (q<m). Principal axes are also referred to as eigen vectors $E_1, E_2, E_3, \ldots, E_m$ correspond to eigen values $\lambda_1 + \lambda_2 + \lambda_3 + \ldots + \lambda_q$ .Generally; the first few principal components hold most of the information. Analysis of variances proportion represents the total number of principal components that should be retained from the dataset.

C. Kernel Principal Component Analysis

Kernel principal component analysis (KPCA) is an excellent statistical learning technique that is used to cluster the different groups of data samples. Due to its flexibility and good performance it is widely applied to various learning scenarios. Based on principal component analysis (PCA), this method provides a technique for nonlinear feature extraction in the sample data. The nonlinearity is introduced by first mapping the data from the original input space into a higher dimensional feature space F using a nonlinear map $\Phi : R^N \rightarrow F$,where R is the set of real numbers and N is the dimension of the original input space and linear PCA is then performed in F using the mapped samples $\Phi(\mathbf{X}_k)$, $\mathbf{X}_k$ being the sample data[12]. In general, linear operations of PCA are done in a reproducing kernel Hilbert space with a non-linear mapping. The basic idea of PCA and KPCA is shown in Fig. 5 and Fig. 6 respectively. In kernel PCA, an arbitrary function Φ is chosen, which is non-trivial [13], [14]. Generally the dimension of arbitrary function Φ is very high. Since we usually try to avoid working in the Φ -space, which is known as 'feature space'. By using the function Φ we can create the N-by-N kernel $\mathbf{K} = \mathrm{k(x,y)} = (\Phi(\mathrm{x}), \Phi(\mathrm{y}))$ matrix which represents the inner product of feature space. According to the idea of KPCA algorithm, any sample can be converted to high dimension by using the equation

$$Y_k = \left(\sum_{i=1}^{M} \alpha_i^k K(X_i, X) \right)_{j=1}^{k}$$

where M is the number of samples, α_i^k is the i^{th} value of k^{th} eigenvector of kernel matrix K, Y_k is the k^{th} (q…….M) value of sample after transforming, I is the i^{th} original sample, X is the original sample which is to be transformed, q is the sequence number of the first non zero eigen values in an ascending order. The dot product of the sample in the feature space is defined as $K(X_i,X_j)= \Phi(X_i)^T \Phi(X_j)$, where $X_i,X_j(i,j=1,2….M)$ are random samples of data sets. We use polynomial kernel function, which is defined as $K(X_i,X_j) = (\Phi(X_i)\cdot \Phi(X_j)+1)^d$. After getting the value of K, centralization of data is required to perform an effective principal component analysis, we 'centralize' K to become K′. $K'=K-1_NK-K1_N+1_NK1_N$.where $\mathbf{1_N}$ denotes a N×N matrix having each value 1. We use K′ to perform Kernel PCA[15]-[17].

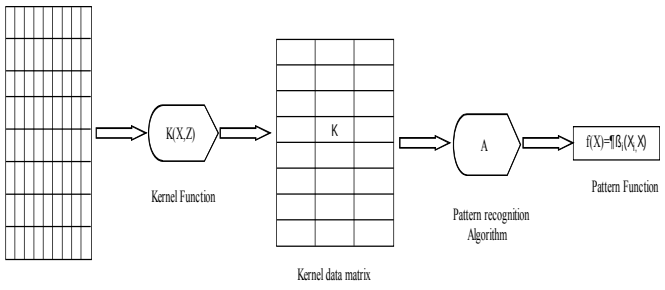

Fig. 5. A sketch of the modularity inherent in kernel based algorithm; the data is transformed into a kernel matrix, by using a kernel function, then the pattern analysis algorithm uses this information to find the suitable relations, which are all written in the form of kernel functions.

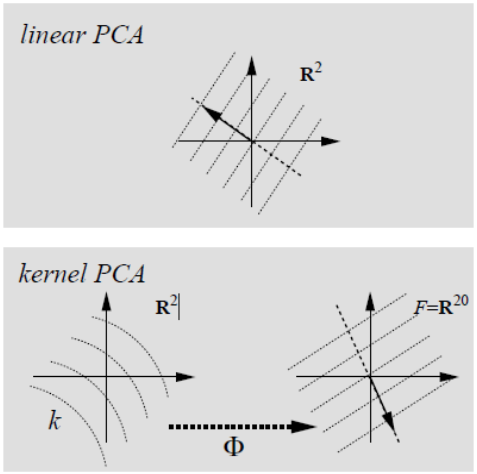

Fig.6. The basic idea of PCA and kernel PCA

VI. Result and Discussion

Experimentations with thick film ZnO gas sensor have been performed on 230 numbers of ethanol data samples and sensor output signatures are logged in the computer. Total data size becomes 230×10.i.e.230 samples and each sample having 10 dimensions, as we have 10 different concentrations in our gas sensor system. PCA & KPCA techniques have been applied on the 230 data samples .The PCA and KPCA plots are shown in Fig. 7 and 8, respectively.

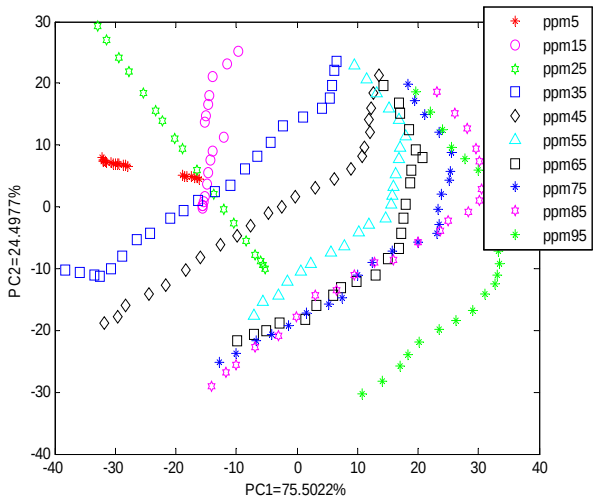

Fig.7. PCA plot for different concentration of ethanol

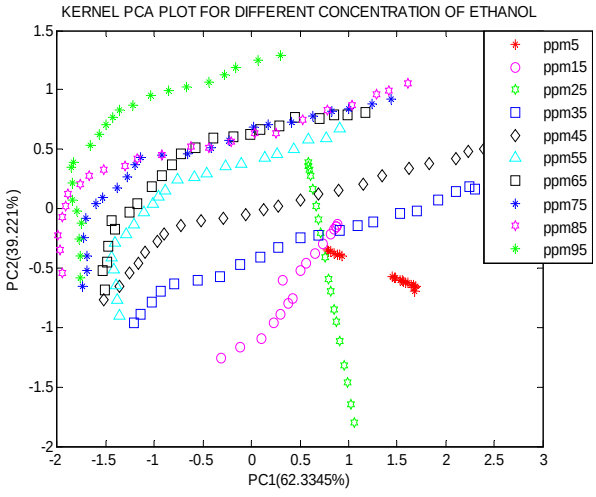

Fig.8. KPCA plot for different concentration of ethanol

It has been observed that some classes are very close to each other in PCA plot for different concentration of ethanol, but in KPCA more distinct plots have been achieved. Confirmations on results have been made in quantitative manner by calculating the separation index. In case of KPCA it gives very higher value (separation index value 83.0039) than the linear PCA (separation index value 51.7377).

VII. Conclusion

In this paper we have presented a novel approach to identify the presence of different concentration of ethanol gas which is useful in the detection and warning of DUI (Driving under the influence) using kernel method of pattern analysis technique i.e. kernel PCA. Main objective of our study was to get more improved clustering plots of thick film zinc oxide gas sensor obtained data with different concentration label. Compared to linear PCA, KPCA extracts more useful features for better clustering. From the mentioned technique, ethanol quality and concentration discrimination is achieved quite successfully.

Acknowledgment

The authors are deeply indebted to the Silicon Institute of technology, where research works were performed during the study. The work has been suggested and assisted by Prof A.K Tripathy, Director (Research) Silicon Institute of Technology, and Orissa. The authors also thank to Dr. R. N. Pal, retired Professor IIT Kharagpur, India and his encouragement and support with the evaluation.

References

[1] A.G Mohapatra, B.N. Rath, "Pattern Recognition Based Discrimination of Ethanol Concentration Using ZnO Thick-Film Gas Sensor for the Detection and Warning of Drunken Driving", *International Journal of Engineering Science and Technology* (*IJEST*), vol 3, no.6, pp. 4604 – 4613, June 2011.

[2] Liewhiran C, Phanichphant S, "Fabrication of various thickness of flame-made nano zinc oxide thick film and its response to ethanol", In proc. 2nd IEEE International conference on Nanoelectronics, 2008. INEC 2008.

[3] Bipan Tudu, Nabarun Bhattacharyya, Bikram Kow, and Rajib Bandyopadhyay, "Comparison of Multivariate Normalization Techniques as Applied to Electronic Nose Based Pattern Classification for Black Tea", *in Proc. IEEE 3rd International Conference on Sensing Technology*,Tainan,Nov.30- Dec.3, pp. 254-258, 2008.

[4] *Handbook of Machine Olfaction,* Wiley-VCH Verlag GmbH and Co,W.Gardner,H.T.Nagle, 2003.

[5] *Electronic noses: Principles and Applications*,Oxford University Press, J. W. Gardner, P. N. Barlett, 1999

[6] N. Bhattacharyya, R. Bandyopadhyay, M. Bhuyan, B. Tudu, D. Ghosh, and A. Jana, "Electronic nose for black tea classification and correlation of measurements with "Tea Taster" marks," *IEEE Trans. Inst. Meas,* vol. 57, pp. 1313-1321, Jul. 2008.

[7] M. Padro, G. Niederjaufner, G. Benussi, G. Faglia, G. Sberveglieri, M. Holmberg, and I. Lundstrom, "Data preprocessing enhances the classification of different brands of Espresso coffee with an electronic nose", *Sens. Actuators B,Chem.*, vol. 69, pp. 397-403, 2007.

[8] A. Legin, A. Rudnitskaya, L. Lvova, Y. Vlasov, C. Di Natale, A. D'Amico,"Evaluation of Italian wine by the electronic tongue: recognition, quantitative analysis and correlation with human sensory perception", *Anal. Chim. Acta 484* , pp. 33–44. 2003.

[9] H. Goto, Y. Hasegawa, and M. Tanaka, "Efficient Scheduling Focusing on the Duality of MPL Representatives," *Proc. IEEE Symp. Computational Intelligence in Scheduling (SCIS 07), IEEE Press,* pp. 57-64,.2007.

[10] K. Toko, Electronic tongue, Biosens. *Bioelectron*. Vol. 6,pp. 701–709. 1998.

[11] X.-w. Chen, T. Huang, "Facial expression recognition: a clustering based approach, Pattern Recognition", Lett. 24, pp. 1295–1302,2003.

[12] K.-R. Muller, S. Mika, et al., "An introduction to kernel-based learning algorithms",*IEEE Trans. Neural Networks 12* , pp. 181–20, 2001.

[13] K. Mizutani, S. Miyamoto, "Possibilistic approach to kernel-based fuzzy c-means clustering with entropy regularization, Modeling Decisions for Artificial Intelligence", *Second International Conference, MDAI 2005*, Tsukuba, Japan, pp. 144–155, 2005.

[14] G. Baudat, F. Anouar, "Generalized discriminant analysis using a kernel approach", *Neural Comput*. Vol 10, pp .2385–2404,2000.

[15] *Neural Networks for Optimization and Signal Processing,* 1st ed. Chichester, U.K.: Wiley, A. Cichocki and R. Unbehaven, 1993, ch. 2, pp. 45-47.

[16] *Kernel methods for Pattern analysis,* 1st ed. Cambridge University Press, J. S.Taylo rand N. Cristianini, New York.: , 2004, ch. 2, pp. 146-151.

[17] P. C. Jurs, G. A. Bakken, and H. E. McClelland, "Computational methods for the analysis of chemical sensor array data from volatile analytes", *Chemical Rev*., Vol. 100, pp. 2649- 2678, 2000.

Laxmi Shaw has born in Rourkela, Odisha and she received the M.Tech degree in Instrumentation and Electronics Engineering from Jadavpur University, Kolkata, India, in 2010.She is currently Assistant Professor with the Department of Applied Electronics and Instrumentation Engineering, Silicon Institute of Technology, Odisha, India. Her main research interests include pattern recognition, Computational Intelligence; signal processing, biomedical signal processing, machine learning and data mining and sensors.

Sageeta Bagha has born in Bhubaneswar, Odisha and she received the M.Tech degree from Karunya University, Tamilnadu, India, in 2011.She is currently Assistant Professor with the Department of Applied Electronics and Instrumentation Engineering, Silicon Institute of Technology, Odisha, India. Her main research interests include biomedical signal processing, Biomedical Instrumentation and biosensors.

Ambarish G. Mohapatra received the M.Tech. degree in Sensor System Technology from Vellore Institute of Technology, VIT University, Vellore, India, in 2008 and Bachelor degree in Electronics and Communication Engineering (ECE) from National Institute of Science and Technology (NIST), Berhampur, Inida, in 2004.He is working as Assistant Professor in Silicon Institute of Technology, Bhubaneswar, India. He has number of national and international papers in the field of sensing and control, where his interests are sensors and transducers, MEMS, wireless sensor network.

Narayan Nayak has received M.Tech degree from Biju Patnaik University of technology, Rourkela, India, in 2007.He is currently Associate Professor with the Department of Applied Electronics and Instrumentation Engineering, Silicon Institute of Technology, Odisha, India. His main research interests include Industrial automation and control, Digital signal processing, sensors and signals.