The Impact of the Number of Neurons in the Hidden Layer on the Performance of MLP Neural Network: Application to the Fast Identification of Toxic Gases

Slimane Ouhmad, Abdellah Halimi

Abstract—In this work, neural networks methods MLP type were applied to a database from an array of six sensors for the detection of three toxic gases. The choice of the number of hidden layers and the weight values are influential on the convergence of the learning algorithm. We proposed, in this article, a mathematical formula to determine the optimal number of hidden layers and good weight values based on the method of back propagation of errors. The results of this modeling have improved discrimination of these gases and optimized the computation time. The model presented here has proven to be an effective application for the fast identification of toxic gases.

Keywords—Back-propagation, Computing time, Fast identification, MLP neural network, Number of neurons in the hidden layer.

I. INTRODUCTION

THE control of the air quality becomes an essential priority for researchers from different backgrounds, particularly when it comes to the pollutant gases emitted by the production units [1]-[10]. These gases have a direct and fatal impact on human health [11]-[13]. It is necessary to have a continuous monitoring system in order to measure air security in factories and units of production involving toxic gases. Recent developments of intelligent sensors have provided a rapid progression in the detection of toxic gases due to their broad spectrum of sensitivity [14]-[21]. The detection of these gases has resulted in a different set of descriptors (the electrical signals from the sensors) as demonstrated in the ANN models. These models have been frequently used for modeling air pollutant concentrations [22]. The ANNs in this sense can be successfully used for modeling complex nonlinear systems and for forecasting signals in a wide range of engineering applications [23]-[29].

The number of hidden layers, the neuron numbers in each layer and the activation functions are parameters which influence the performance of neural networks in different learning algorithms [30]-[34].F or this we proposed a mathematical formula to determine these optimal parameters of MLP model.

The main objective of this work is to apply our model to

Slimane Ouhmad is a PhD Candidate with the Department of Physics, Faculty of Sciences Moulay Ismail University, Bp 11201, Zitoune, Meknes Morocco (corresponding author e-mail: ouhmad _slimane @ yahoo.fr.)

Abdellah Halimi is a PhD Candidate with Department of Physics, Faculty of Sciences Moulay Ismail University, Bp11201, Zitoune, Meknes Morocco (e-mail: halimi_abdellah@ yahoo.fr).

identify three toxic gases namely: nitrogen dioxide (NO2), hydrogen sulfide (H2S) and their mixture in low concentration. To achieve this objective, we used a database obtained from a gas sensor array which consists of six chemical sensors of type TGS (called electronic noses) based on metal oxide [35], [36]. Each sensor emits an electrical signal characterized by four variables: the initial conductance (G_0), the dynamic slope of the conductance (dG_S/dt), the conductance curve (A) and the steady-state conductance (G_S) [5], [35]-[37].

In this study, we evaluated the performance of MLP with one hidden layer using sigmoid activation function through back propagation algorithm. First, we developed our model of such multilayer perceptron (MLP) to supervise the learning on-line for the optimization of the neuron numbers in one hidden layer and computation time. Second, we compared this type of learning responses to the challenges of the gases identification with the values of mean square error in minimum iterations [38]. Finally, we compared the results to some previous research.

The remaining of the paper is organized as follows: Section II deals with the methodology which refers to the materials and methods. Section III presents the results and the discussion. Finally, Section IV concludes the paper.

II. MATERIALS AND METHODS

A. Feature Extraction

The features obtained were extracted from the temporal responses of the sensor array for the detection of the toxic gases H2S, NO2 and their mixture (H2S-NO2) realized by [35], [36]. The sensor array comprised six TGS-xx (with xx = 800, 813, 822, 825, 832 and 2105), Taguchi Gas Sensor obtained from Figaro Engineering, Incorporation [37]. We are concerned with information obtained from this experiment. For this, we deal with first three basic representative features (G_0, dG/dt, Gs) from the response signal extracted with the ignorance of the fourth feature (A).

1) G_0: the initial conductance of a sensor calculated as the average value of its conductance during the first minute of a measurement.
2) Gs: the steady-state conductance calculated as the average value of its conductance during the latest minute of a measurement.

3) dG/dt: the dynamic slope of the conductance calculated between 2 and 30 to 50 minutes of a measurement.
4) A: the area conductance curve in a time interval between 2 minutes and d G/dt time of measurement. This area was calculated by the trapeze method.

These first three features were extracted from the response of each sensor. Given the fact that there were 6 sensors within the array, each measurement was described by 18 features.

B. Database

The consideration of database is composed of 180 samples distributed in three classes (NO_2, H_2S, and NO_2-H_2S (mixture)). The data were divided into two parts: 80 % are used for learning and 20% are used for testing.

The selection of samples used in learning and testing are done randomly and none randomly.

C. Artificial Neural Networks

The first concept of Artificial Neural Network was introduced in 1943 by [38], [39]. It is based on mathematical laws and structures. ANN is a set of processing elements (or nodes) loosely analogous to neurons in the brain. This model is fortified and developed by the introduction of the Back Propagation (BP) training algorithm feed forward in 1986 to overcome difficulties detected by [39]. In brief, ANNs architecture consists of neurons arranged in a layered topology containing an input layer, a hidden layers, and an output layer, which are all interconnected. Many theoretical and experimental works have shown that a single hidden layer is sufficient for ANNs to approximate any complex non-linear function [40]. For this reason, our paper focuses on ANN study of MLP type, this model is based on a feed-forward neural-network with a single hidden layer. The back propagation algorithm is used to train the network. Furthermore, the chosen activation functions are two sigmoid functions in the hidden and output layers.

D. Artificial Neural Network-Sensors Modeling

The compilation of the feature sequentially extracted from a system of multi-sensor electrical signals characterized by [36]. The procedure configures and trains features sequentially using our developed model, neural-network MLP of the on-line learning type based on the back propagation algorithm, which is executed in MATLAB. The neural network is inputted with the feature matrix of sensors variables that corresponds to the number of input neurons. However, the output layer neurons correspond to the number of the classes of toxic gases. The complexity of the MLP model can be varied by varying the number of neurons in hidden layer, the coefficient of activation functions, good weight values and rate learning. Therefore, we have developed a model of MLP to determine these optimal parameters. When the neural network is properly configured, training is started. Testing begins after the training has ended.

E. The Development of Multi-Layer Perceptron Model

The MLP neural network model is used in this work. It consists of matrix (6x3) input neurons, 6 sensors each one with three variables (G0, dG/dt, G_S), one hidden layer and output neurons (NO_2, H_2S and NO_2-H_2S (mixture)). These Classes encoded as follow: [1, -1, -1] = NO_2, [-1, -1, 1] = H_2S and [-1, 1, -1] = NO_2-H_2S (mixture). Also, the chosen activation functions are sigmoid $\varphi_1(x)$ and $\varphi_2(x)$ functions in the hidden and output layers, respectively.

The rest is explained in the following notations:

w_{ij}, u_{jk}: the weights connecting input with hidden layer and the weights connecting hidden layer with output respectively.

x_i : Input vectors.

y_j : Calculated output of the hidden layer.

t_k, z_k : Calculated output and desired output of neural network respectively.

η : Learning rate

N, P, K: Numbers of units in input, hidden and output layers respectively.

The outputs of the hidden layer are also weighted by the weights w_{ij} and then transmitted to the output layer according to:

$$c_j = \sum_{i=1}^{N} w_{ij} x_i + a_j \tag{1}$$

$$y_j = \varphi_1(c_j) \tag{2}$$

The outputs of the hidden layer are also weighted by the weight u_{jk} and then transmitted to the output layer according to the following equations:

$$o_k = \sum_{j=1}^{P} u_{jk} y_j + b_k \tag{3}$$

$$t_k = \varphi_2(o_k) \tag{4}$$

The objective function of the proposed model is the error MSE between the calculated output and the desired output:

$$\min(E_{mse}) = \min(\frac{1}{2n} \sum_{n=1}^{n} \sum_{k=1}^{K} (t_{k,n} - z_{k,n})^2) < 10^{-4} \tag{5}$$

with n: Number of samples, n=1 "real time Learning"

The last constraints guarantee two conditions:
1) The existence of only one hidden layer.
2) Time of iterations (of samples) $< (G_0/2)$ time, and excellent Identification Accuracy.

The neural architecture optimization problem can be formulated

$$\left\| (\varphi_2(\sum_{j=1}^{P} \varphi_1(\sum_{i=1}^{N} w_{ij} x_i + a_j) u_{jk} + b_k)) - z_k)^2 \right\| \leq 10^{-4} \tag{6}$$

with

$$\varphi_i(x) = \frac{1 - e^{-\gamma_i x}}{1 + e^{-\gamma_i x}}, i = 1:2 \qquad (7)$$

and $w, u, a, b \in [-1, 1]$, and $0 \le \eta \le 1$.

The determination of these key parameters (w, u, a, b, and coefficients of activation functions) present the solution of this problem.

For this reason, we used the gradient descent in order to minimize error functions. The rest algorithm comes after to adjust the weight connection iteratively [40].

Adaptation of weights ujk:

$$u_{jk}(t+1) = u_{jk}(t) + \eta(t_k - z_k)\varphi_2'(o_k)y_j \qquad (8)$$

Adaptation of weights wij:

$$w_{ij}(t+1) = w_{ij}(t) + \eta(\sum_{k=1}^{K}(t_k - z_k)\varphi_2'(o_k)u_{jk})\varphi_1'(c_j))(x_i) \qquad (9)$$

F. Other Performance Measures

The generalization performance of the classification model was accessed using the standard statistical measure accuracy.

$$Accuracy = \frac{\text{Number of correctly classified data}}{\text{total number of data}} \times 100 \qquad (10)$$

III. RESULTS AND DISCUSSION

In the previous method, we attempted better key parameters of the neural model which includes the architecture of the network. To achieve this objective, we fixed only one hidden layer instead of other layers. In addition to that, we varied the neuron numbers, rate learning and coefficients of activation functions ($\gamma1 \ll 1$ and $0.6 < \gamma2 < 1$). This simulation is based on certain performance criteria in terms of statistical data (MSEs, Discrimination and identification Accuracy) and time computation account. The latter must be less than the initial conductance of a sensor time.

Fig. 1 shows the training time influenced by the number of nodes in one hidden layer.

Fig. 2 contains the average errors of 150 errors to each structure started from 2 till N of neurons in one hidden layer. For this, we notice that the optimal number of nodes is n (matrix includes 16 nodes) in normal on-line training taking into consideration the stopping criterion G0/6 time (=10 seconds). With the possibility of achieving fast training by relying on stopping criterion 3 seconds (using ordinary laptop), only with 2x2 nodes, instead of 10 seconds, which gave similar performance Fig. 2. In addition, we observed that the average error increases when we augment the neuron number in a fixed minimal iteration. However, it decreases if we augment the iteration number in this case.

Figs. 3-5 explain the minimum MSEs during learning and testing in different subsets randomly and non-randomly.
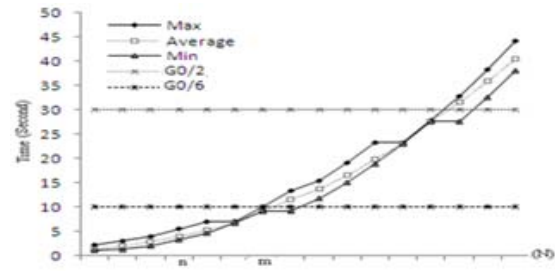

Fig. 1 Influence of number of neurons in one hidden layer over training time
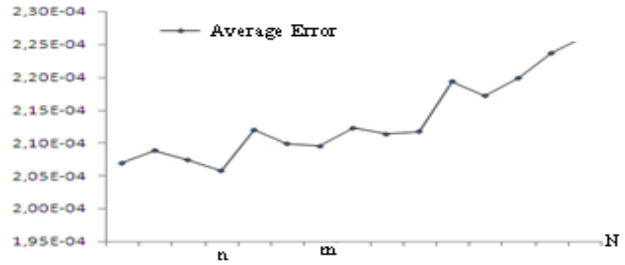

Fig. 2 Influence of the number of neurons in one hidden layer on average error
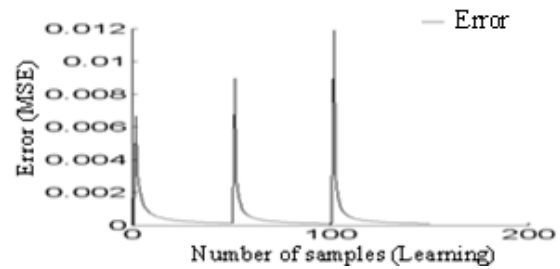

Fig. 3 Evolution of the minimums MSEs during learning with 100 iterations (nonrandomly)
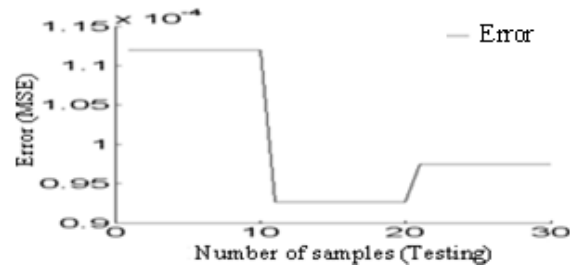

Fig. 4 Evolution of the minimums MSEs during testing (nonrandomly)

Figs. 7 and 8 are added to verify the power of our model in terms of criterion MSEs when it augments the iteration number; this shows how our model is useful if the time of computation is ignored.

Figs. 4 and 7 indicate more aspects than error-testing. These encompass the number of gas classes and the number of samples in each group. In addition, this figure shows MSEs=7×10^{-06} which refers to the mixture of gases.
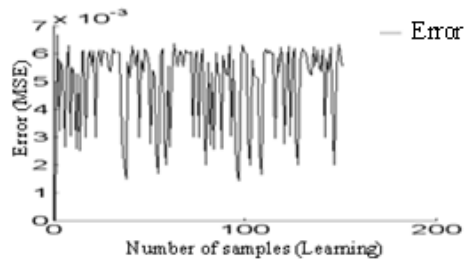
Fig. 5 Evolution of the minimums MSEs during learning with 100 iterations (randomly)
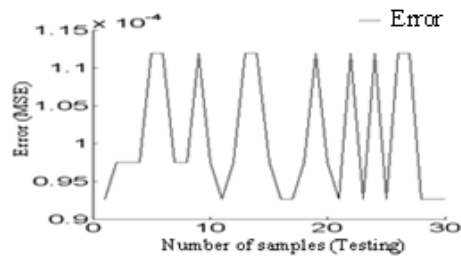

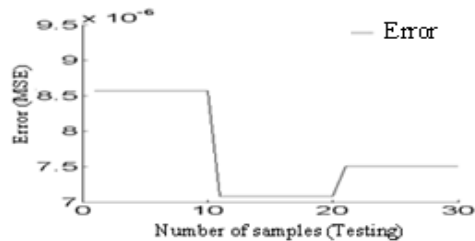Fig. 6 Evolution of the minimums MSEs during testing (randomly)

Fig. 7 Evolution of the minimums MSEs during testing with 1300 iterations (nonrandomly)
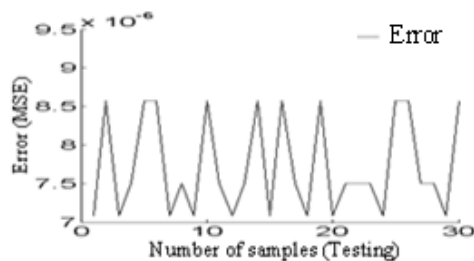

Fig. 8 Evolution of the minimums MSEs during testing with 1300 iterations (randomly)

Fig. 9 (b) shows an excellent classification and identification accuracy 100%, it is visualized in Fig. 10. They include these classes. These results reflect the power of our model despite of the complex conditions which are: the concentration 1ppm, small data base and the stopping criteria 10 seconds, with training errors in 50 iterations as in Fig. 9 (a). We are concerned with different iterations to reflect the performance of our model in many situations shown.
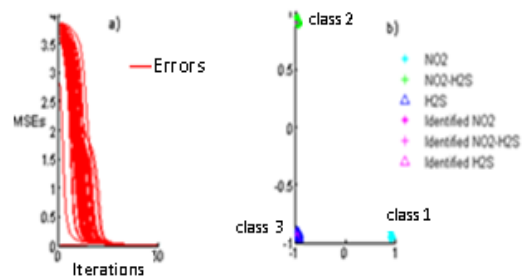

Fig. 9 Classification and identification of three toxic gazes over number of iterations 50
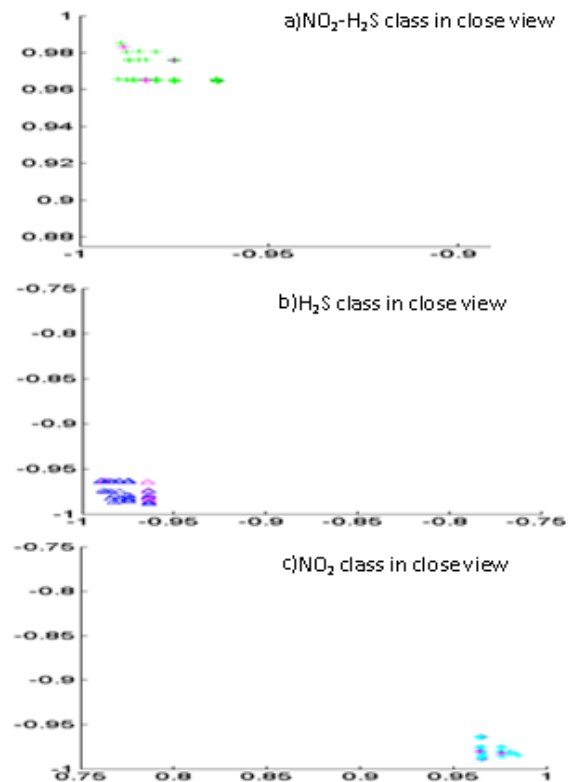

Fig. 10 Visualization of three toxic gases classified and identified

Compared to previous research that contains great architecture (hidden layers), learning batch (off-line) using the toolbox without taking into consideration computing time [35], [36], [41]. Also the discrimination of different concentration; however, our model is concerned only with one ppm (complex concentration) in real time by using complex learning on-line (real time) with fast identification in term of computation time plus high accuracy.

IV. CONCLUSION

In this research, we present an artificial neural network (ANN) model based on the multi-layer Perceptron (MLP) for classifying and identifying the three toxic gases in low concentration (one part per million). In order to prove the power of our model, it is applied to give useful results in the

accurate identification and classification with the possibility to make the training faster based only on 3 seconds instead of 10 seconds.

In addition, it can be an excellent application in engineering to monitor the identification of gases in real time when there is a danger that cannot be detected by human smell sense. It is a practical concept to install the computer models in a high industrial area for air security.

REFERENCES

[1]   M. W. Gardner, S.R. Dorling, "Neural network modeling and prediction of hourly NO_X and NO_2 concentrations in urban air in London," *Atmospheric Environment*, 33, 5, pp. 709–719, 1999.

[2]   Kukkonen,et al, "Extensive evaluation of neural network models for the prediction of NO2 and PM10 concentrations compared with a deterministic modeling system and measurements in central Helsinki," *Atmospheric Environment*,37,pp. 4539–4550, 2003.

[3]   D. Jiang, Y. Zhang, X. Hu, Y. Zeng, J. Tan, D. Shao, "in developing an ANN model for air pollution index forecast," *Atmospheric Environment*, 38,pp. 7055–7064, 2004.

[4]   U. Brunelli, V. Piazza, L. Pignato, F. Sorbello, S. Vitabile, "Two-day ahead prediction of daily maximum concentrations of SO2, O3, PM10, NO2, CO in the urban area of Palermo," *Atmospheric Environment*, 41,pp. 2967–2995, 2007.

[5]   K.D. Karatzas, S. Kaltsatos, "Air pollution modelling with the aid of computational intelligence methods," *in Thessaloniki*. Greece, *Simulation Modelling Practice and Theory*, pp. 1310–1319, doi. 10.1016/ j.simpat.2007.09.005.

[6]   J. Galloway, et al, "Nitrogen cycles: past, present and future," *Biogeochemistry*, 70, pp.153–226, 2004.

[7]   H.J.S. Fernando, et al, "Forecasting PM10 in metropolitan areas: Efficacy of neural networks," *Environmental Pollution*, 163, pp.62-67, 2012.

[8]   J. Fenger, "Air pollution in the last 50 years—from local to global," *Atmos Environ*, pp. 13–22, 2009.

[9]   D. Gacquer, et al, "Comparative study of supervised classification algorithms for the detection of atmospheric pollution," *Engineering Applications of Artificial Intelligence*, 24, pp. 1070–1083, 2011.

[10] A.B. Chelani, M.Z. Hasan, "Forecasting nitrogen dioxide concentration in ambient air using artificial neural-networks," *International Journal of Environmental Studies*, 58, 4,pp. 487–499, 2001.

[11] K.M. Venables. et al, "Thunderstorm-related asthma - the epidemic of 24/25 June 1994," *Clinical and Experimental Allergy*, 27, pp.725-736, 1997.

[12] N. Lokeshwari, G. Srinikethan, V. S. Hegde, "Management of Air Pollutants from Point Sources," *International Journal of Environmental Earth Science and Engineering*, Vol.7, pp.12, 2013.

[13] J. Burlakovs, M. Klavins, R. Ernsteins, A. Ruskulis, "Contamination in Industrial Areas and Environmental Management *in Latvia*," *World Academy of Science. Engineering and Technology*, Vol.76, pp.04-23, 2013.

[14] R. Lalauze, C. Pijolat, "A new approach to selective detection of gas by an SnO2 solid-state sensor," *Sensors and Actuators*, 5, pp. 55–63, 1984.

[15] P. Romppainen, V. Lantto, S. Leppavuori, "Effect of water vapour on the CO response behavior of tin dioxide sensors in constant temperature and temperature-pulsed modes of operation," *Sensors and Actuators, B. Chemicall*, pp. 73–78, 1990.

[16] P. van Geloven, M. Honore, J. Roggen, "The influence of relative humidity on the response of tin oxide gas sensors to carbon monoxide," *Sensors and Actuators, B. Chemical*, 4,pp.185–188, 1991.

[17] J. Watson, K. Ihokura, G.S.V. Coles, " The tin dioxide gas sensor," Measurement Science and Technology, 4,pp. 711–719, 1993.

[18] H. Yamaura, T. Jinkawa, J. Tamaki, K. Moriya, N. Miura, N. Yamazoe,"Indium oxide-based gas sensor for selective detection of CO," *Sensors and Actuators, B: Chemical*, 35–56,pp.325–332, 1996.

[19] H. Debeda, P. Massok, C. Lucati, F. Menil, J.L. Aucouturier, "Methane sensing:from sensitive thick films to a reliable selective device," *Measurement Science and Technology*,vol.8, pp.99–110, 1997.

[20] J. Judith Vijaya, L. John Kennedy, G. Sekaran, K.S. Nagaraja, "Utilization of Sr(II)-added calcium aluminate for the detection of volatile organic compounds," *Industrial and Engineering Chemistry Research*, vol. 46.19,pp. 6251–6258, 2007.

[21] B.W. Licznerski, K. Nitsch, H. Teterycz, T. Soba´nski, K. Wi´sniewski, "Characterization of electrical parameters for multilayer SnO2gas sensors," *Sensors and Actuators B: Chemical*. vol. 103, 1–2, pp. 69–75, 2004.

[22] A. Gulbag, F. Temurtas, "A study on quantitative classification of binary gas mixture using neural networks and adaptive neuro -fuzzy inference systems," *Sensors and Actuators.B. Chemical*, vol. 115, pp. 252–262, 2006.

[23] U. Brunelli, V. Piazza, L. Pignato, F. Sorbello, S. Vitabile, "days ahead prediction of daily maximum concentrations of SO2, O3, PM10, NO2, CO in th eurban area of Palermo.Italy,". *Atmospheric Environment*, vol.41,pp.2967–2995, 2007.

[24] M. Kolehmainen, H. Martikainen, J. Ruuskanen, "Neural networks and periodic components used in air quality forecasting," *Atmospheric Environment*, vol. 35, pp. 815–825, 2001.

[25] J. Yi, V.R. Prybutok, "A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area," *Environmental Pollution*, vol. 92, pp. 349–357, 1996.

[26] E. I. Vlahogianni, "Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach," *Transportation Research Part C*, vol. 1, pp. 211–234, 2005.

[27] A. Eswaradass, X.-H. Sun, M. Wu, "A neural network based predictive mechanism for available bandwidth", *proceeding of 19th, IEEE International Conference on Parallel and Distributed Proceeding Symposium,Workshop.10*,vol.11,p.228,2005.. doi>10.1109/IPDPS.p.234.2005.

[28] H. Yousefia, "Utilizing neural networks to reduce packet loss in self-similar tele traffic," *proceeding of IEEE International. Conference on Communications*, vol.3.pp. 1942–1946, 2003.

[29] M. Badura, et *al*, "Statistical assessment of quantification methods used in gas sensor system", *Sensors and Actuators: B. Chemical*, vol.188, pp.815– 823, November 2011.

[30] P. Chandra and Y. Singh, "An activation function adapting training algorithm for sigmoidal feedforward networks," *Neurocomputing*, vol.61, pp. 429–437, 2004.

[31] W. Duch and N. Jankowski, "Survey of neural transfer functions", *Neural Comput Appl*, vol. 2, pp.163–212, 1999.

[32] W. Duch and N. Jankowski, "Transfer functions: Hidden possibilities for better neural networks," *In 9th European symposium on artificial neural networks*, pp. 81–94, 2001.

[33] Y. Singh and P. Chandra, "A class +1 sigmoidal activation functions for FFANNs," *J Econ Dynamic Control*, vol. 28, pp.183–187, 2003.

[34] B.W. Licznerski, C K. Nitsch, H. Teterycz, T. Sobanski, K. Wisniewski, "haracterisation of electrical parameters for multilayer SnO2gas sensors", *Sensors and Actuators. B. Chemical*,vol.103, No. 1–2, pp. 69–75, 2004.

[35] A. Lfakir, "Identification and Quantification of a complex gaseous atmosphere with an intelligent multi-sensor system. Application to Detection of mixtures compounds H_2S, NO_2, SO_2 wet atmosphere variable", Thesis of Metz University, 2006. "Identification et Quantification d'une atmosphère gazeuse complexe à l'aide d'un système multi-capteurs intelligent. Application La détection de mélanges composés de H2S, NO2, SO2 en atmosphère humide variable", thèse d'Université Metz, 2006.

[36] O. Helli, "Gas multi -sensor for the design of an electronic nose for air pollution monitoring", Thesis of Metz University, 2003. "Multi-capteurs de gaz pour la conception d'un nez électronique de surveillance de la pollution atmosphérique", thèse d'Université Metz, 2003.

[37] T. Seiyama, N. Yamazoe, S. Yamauchi,*Chemical Sensor Technology*, Vol. 1-4, Elsevier Science Ltd, 1988.

[38] Rumelhart, D.E. Hinton, G.E. Williams, "Learning internal representations by error propagation," In Rumelhart, D.E. Mc Cleland, J.L. (Eds.), *Parallel Distributed Processing. MIT Press*, Cambridge, pp. 318–362, 1986.

[39] W.S. McCulloch, W.H. Pitts, "A logical calculus of the ideas immanent in nervous activity", *Bulletin of Mathematical Biophysics*, vol 5, pp. 115–133, 1943.

[40] Cybenko, "Approximation by superposition of a sigmoidal function," *Math. Cont. Sig. Syst 2*, pp. 303–314.

[41] S. Ouhmad, H. Roukhe, A, Roukhe "Real time identification of toxic gases based on artificial neural networks," *International Journal of Computational Engineering Research,*vol. 04, issue. 4, pp. 67-72, 2014.