IOSR Journal of Environmental Science, Toxicology and Food Technology (IOSR-JESTFT)
e-ISSN: 2319-2402,p- ISSN: 2319-2399.Volume 12, Issue 11 Ver. I (November. 2018), PP 01-07
www.iosrjournals.org

Prediction of Air Pollutant Concentration Based On Artificial Intelligence Technique

Shameem Hasan
Department of Electrical and Electronics Engineering, Islamic University of Technology, Bangladesh

Abstract : *The aim of this paper is to predict the average concentration of particulate matter of a particular region using artificial neural network. Air pollution represents a long lasting concern for the public authorities, the scientists as well as the civil society because it has a long term impact on health and the environment. The dispersions or suspensions of the particles in solid and the liquid forms in atmosphere are termed as aerosols. The term particulate matter (PM) is used to define the suspended solid-phase matter in the atmosphere. It is the mixture of the different elements. Further pollutants like SO_2, CH_4, CO, NO, O_3 and NO_2 are largely found in the industrial waste. The evidences reveal that sulfate and organic matter are the two main contributing factors for annual PM 10 and PM 2.5 concentrations and its consequences are like health hazards and ecological imbalance. In this paper, the average concentration of pollutants like PM10 and PM2.5 in air have been predicted efficiently. The detailed analysishave been examined with the help of Artificial Neural Network(ANN).*
Keywords: *Artificial Neural network, Air pollution, Human health, Mean Square Error, Particulate Matter*

Date of Submission: 30-10-2018 Date of acceptance: 15-11-2018

I. Introduction

Air pollution is one of the most important environmental problems. Pollution is caused by both natural and man-made sources. Major man-made sources of air pollution include industries, burning oil, coal, gasoline and other fossil fuels. Nitrogen oxides (NOx) and Carbon monoxide (CO) are very important air polluting chemical compounds in industrial site [1,2]. Oxides of nitrogen can be formed by combustion or thermal fixation of atmospheric nitrogen in the air. Carbon monoxide results from the incomplete burning of "common" fuels such as natural or liquefied petroleum gas (LP-Gas), oil, wood, or coal [3-5].

In the environmental engineering literature, the terms aerosols and PM are often used interchangeably. The aerosols come from the wider range of anthropogenic and natural sources from the earth. Airborne PM is not a single pollutant, rather it is the mixture of many subclasses of pollutants [8]. The fate and transport of gas-phase components in atmosphere is closely linked to the aerosols in contrast to the PM which are either directly emitted by emission sources or are formed due to reaction between the gases. For example, reaction between ammonia and oxides of sulfur or nitrogen results in PM [9]. PM having aerodynamic diameter <10 µm (PM10) or <2.5 µm (PM2.5) is of major concern in public health issues. The fact that suspended particulate matter which enters the lungs is more dangerous to health than larger particulate up to 100 microns is well established. It is important to remember that the ratio of RSPM to SPM will be specific to an area and the measurement of the one should be able to infer the other if the ratio has been experimentally determined. Different study shows that there are mainly two objectives of atmospheric aerosols: firstly, the direct impact on the health as a result of near exposure on the surface of the earth and secondly, the role of aerosols in the physical processes and atmospheric chemicals and the way it is affecting the local and global climate. Moreover, the recent studies show that concentration of PM10 and PM2.5 airborne aerosols in urban areas is quite high compared to rural areas. PM2.5 is also known as fine particles. The long-term exposure to PM10 leads to inflammation in lungs. The dominant pollutant, particularly PM2.5 in haze pollution is epidemiologically associated with the risk of dangerous health effects on heart and lung diseases. The lungs and heart get affected due to the inhalation of air particles [11]. However, the wide spread sources of PM2.5 including industrial process, energy production from power stations, vehicular traffic, residential heating, transport, natural disasters, coupled with the complicated physical and chemical processes, make the PM2.5 forecasting a difficult task. Several types of approaches have been used. Whole world is going through environmental and climate change now-a-days. Madrid, capital of Spain experiences climate change that is quite evident. In the following sections, the statistical analysis, methodologies and results have been analyzed.

Fig. 1: The Map of Madrid, Spain

II. Statistical Analysis

In this research work, many statistical approaches such as Bayesian regularization (BR), Levenberg–Marquardt algorithm (LM), and scaled conjugate gradient (SCG) have been used for the evaluation and accuracy of the performance and results. Bayesian classification is the technique to construct the classifiers. Classifiers are nothing but the models that assign the class labels to the problem instance. Levenberg–Marquardt algorithm is also known as DLS that is damped least squares used for solving generic curve fitting problems by finding the local minimum [10]. Scaled conjugate gradient is feed-forward and supervised algorithm for neural networks. Feed forward here means that in connections there is no loop between the units. The general equations corresponding to each are mentioned below:

Bayesian regularization:

$$x = \arg_{b\in\{1,....B\}} \max p(C_b)\prod_{i=1} np(y_i|C_b) \tag{1}$$

Levenberg–Marquardt algorithm:

$$H(\beta) = \sum_{j=1} m[x_j - f(y_j, \beta)]^2 \tag{2}$$

Scaled conjugate:

$$S_k = \frac{\dot{E}(W_k + \sigma_k P_k) - \dot{E}(W_k)}{2} \tag{3}$$

Bayesian regularization, Levenberg–Marquardt algorithm and scaled conjugate are the various algorithmic parameters and functions used in the neural networks. BR can eliminate or reduce the need for lengthy cross-validations and it is more robust than the standard back-propagation methods, whereas to solve nonlinear least squares problems, the LM technique is considered to be the standard one as it shows lower performance in terms of predictive ability. On the other hand, SCG needs O(n) of memory where n represents the number of weights in the network although it uses second order of information from neural networks [9,10]. Among these three, the BR is considered to be the optimal one as it develops the nonlinear relationships and it has more predictive abilities. To get better and refined results, the data was tested through ten hidden layers, and on observing the results, it can be seen that the BR shows least mean square error (TABLE 1) for both PM10 and PM2.5. In further sections, a brief introduction of neural networks is cited and the results in regard to BR have been shown and explained in the sections after it.

Table 1: No. of Hidden layers, Mean Square Error (MSE) and Iterations for PM10

Serial No.	No. of hidden layers	Mean square error	No. of iterations
1	10	8.3135	720
2	15	10.4159	940
3	20	13.8069	961
4	25	8.187	476
5	30	11.2145	405

Table 2: No. of Hidden layers, Mean Square Error (MSE) and Iterations for PM2.5

Serial No.	No. of hidden layers	Mean square error	No. of iterations
1	10	12.2039	92
2	15	11.9942	48
3	20	11.6796	106
4	25	9.8073	134
5	30	11.4545	473

2.1 Neural Networks

It is one of the concepts which has been inspired by the functionality of the human brain and its performance in identification of phenomena. Neurons (a single neuron shown in Fig. 1) are placed in different layers in multilayer neuron network. Input layer being the first layer receiving information and till its capability with other neurons, it transfers the information in the form of input signals to the other next layers. Neuron weight is the communication ability of each neuron with other neurons. The number of neurons in each layer depends on the weight of neuron and the previous layers' neurons. In addition to the input layer, the neural network also consists of the hidden layers and the output layers. Some of the advantages of using artificial neural networks(ANN) are its immaculate an on-point accuracy on a wide variety range of problems, less requirement of formal statistical training, offering various multiple training algorithms and having the implicit ability of detecting nonlinear complex relationships between independent and dependent variables. In this, neuron is the main processor and adding neurons to hidden layers will reduce calculation error but will be more time consuming for calculations. In the next section, the methodology used in the paper and results have been discussed.

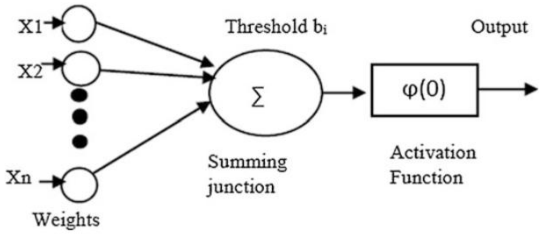

Fig. 2: Single neuron

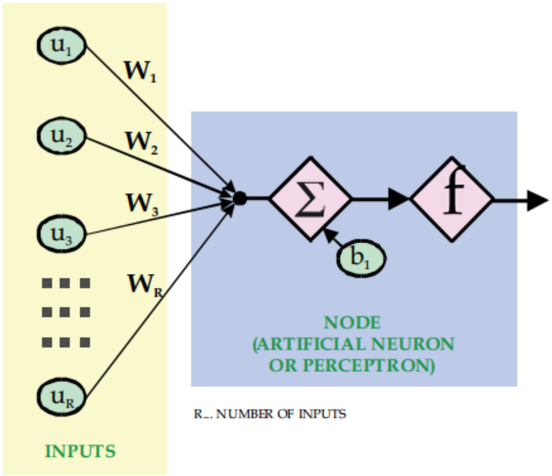

Fig. 3: Node (Artificial neuron or perceptron)

The neural network has 8 input layer, several hidden and one output layer. The input parameters are the concentration of pollutant in the previous moment, the autoregression component and the meteorological variables for which the statistical analysis has shown strongest influence (Fig. 4). The number of autoregression components m is determined by the correlation coefficients between the predicted value and its previous values persistence of the predicted variable. The goal is forecasting one step ahead. The number of neurons in the first and second layers is determined by the criterion of the minimum squared error(MSE).

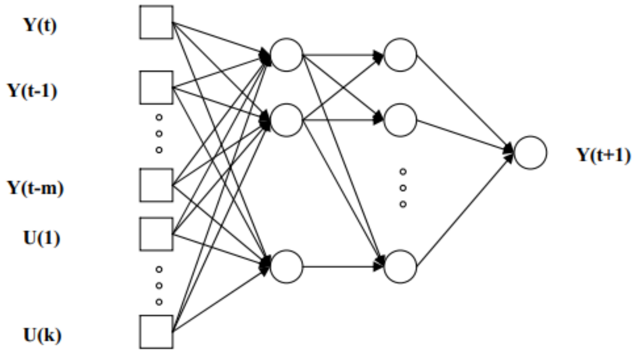

Fig. 4: Architecture of Neural Network

III. Data and Methodology

The data set used in this paper is the recent one of Madrid, capital of Spain. The values of CO, SO_2, NO, NO_2, NO_x, O_3, CH_4, PM 10 and PM 2.5 for the one and a half years (January,2017- June,2018) have been used. The unit of values are in µg/m^3. There are three portions of the data that act as training set, validation set and testing set in neural networks having 70%, 15%, and 15% weightage respectively. These values can be changed but these particular values give better results. For this paper, 400 sample data have been used collected from internet sources.

The nftool (neural fitting tool) of MATLAB has been used in our proposed work to determine the performance and the results. In this tool, the number of input data to layer and the no. of hidden layers have to be defined. In this paper 10, 15, 20, 25 and 30 hidden layers have been used one by one to get the output and their mean square errors have been compared to get the performance measures. Since layer 25 gives most refined results with least value of error, the data has been trained till 25 layers.

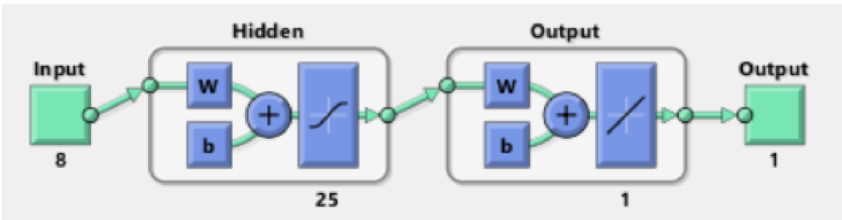

Fig. 5: Neural network working for PM10 and PM2.5

Fig. 6: Performance graph for PM10 showing mean square error

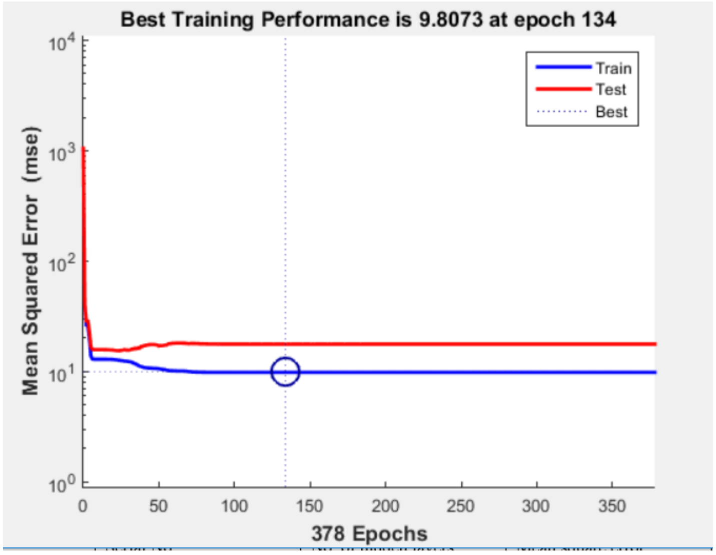

Fig. 7: Performance graph for PM10 showing mean square error

The neural network is given with the inputs such as the values of CO, SO_2, NO, NO_2, NO_x, O_3, CH_4, PM 10, PM2.5 the SPM, the hidden layers, and the output having three desired values such as hazardous (1), moderate (0.5), and good (0). These are levels of the PM10 in which the range of its values 200 and above is considered to be hazardous level, range from 100 to 200 is moderate, and 0–100 is good. These levels further suggest how much harmful is the PM10 for the human life and the ecosystem explained in the next section. Table 1 shows the number of hidden layers with their mean square.

TABLE 1 and Fig. 6 show that the validation, test data sets and the performance of training with respect to epochs of PM10 level prediction in the atmosphere. Also TABLE 2 and Fig. 7 show that the validation, test data sets and the performance of training with respect to epochs of PM2.5 level prediction in the atmosphere. As it is clearly seen the best performance measure is at Serial No. 4 that is with the 25 hidden layers and with minimum mean square value that is 8.187 for PM10 and 9.8073 for PM2.5. Further Fig. 8 and Fig. 10 show the regression curve between the target data and the output of levels of the pollutant level prediction in the atmosphere and the error histogram of pollutant levels in the atmosphere for PM10 respectively. Again, Fig. 9 and Fig. 11 show the regression curve between the target data and the output of levels of the pollutant level prediction in the atmosphere and the error histogram of pollutant levels in the atmosphere for PM2.5 respectively.

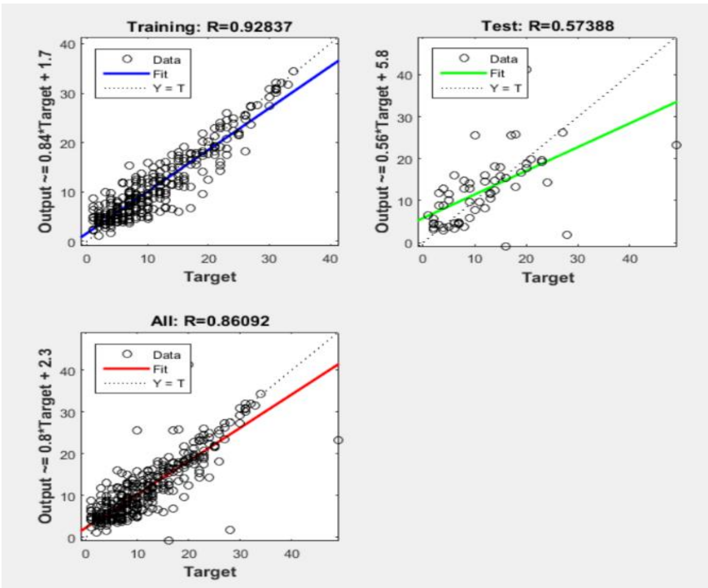

Fig. 8: Regression curve for PM10 levels in atmosphere

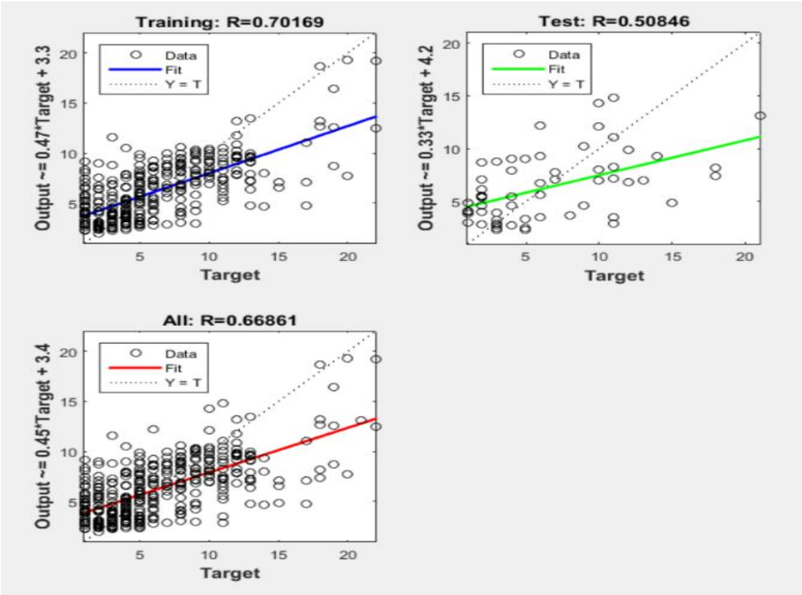

Fig. 9: Regression curve for PM2.5 levels in atmosphere

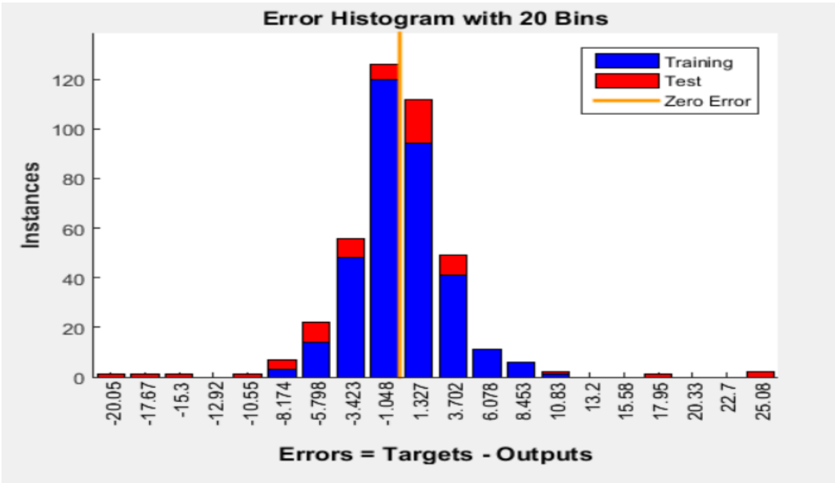

Fig. 10: Error histogram for PM10 levels in atmosphere

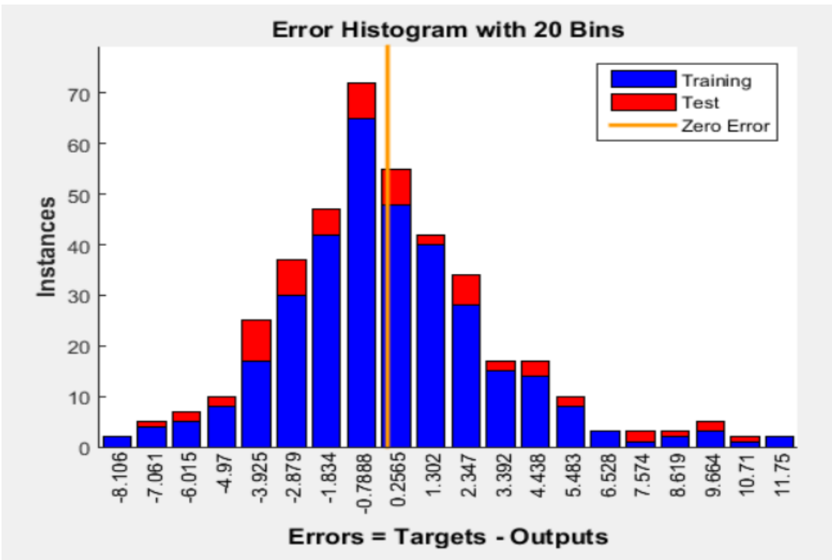

Fig. 11: Error histogram for PM2.5 levels in atmosphere

IV. Impact on Human Health and Ecosystem

The recent rate of increase in particulate matter in the atmosphere is a great matter of concern as it is largely affecting the human health. PM particles are things like organic dust, airborne bacteria, construction dust, and coal particles from power plants. The deposition of PM on the vegetation surface is mainly having three sources that is dry deposition, wet deposition, and the occult deposition. On the other hand, the particles of size between 2.5 µm –10 µm are removed by the upper track of respiratory system of humans but the particles of size less than 2.5 µm (PM 2.5) get deposited on the bronchi walls in the bronchi system or bronchi tree. Further analysis shows that the employees working in municipal solid-waste area were observed with increased rate of symptoms of different respiratory diseases. The experiments were conducted by exposing rats to different PM levels at different times and the reports reveal the increase in size of the lungs' and inflammatory changes [8]. Similarly, in plants, the most exposed part are the leaves which consistently absorb the polluted environment and the dust particles [12]. The rise in levels of CH_4, SO_2, O_3, NO and NO_2 has largely affected the regional weather changes. The global problem like greenhouse gas effect is also the result of air pollution. Depending on the deposition of the particles, there are differing phytotoxic responses due to the exposure to the airborne particulate matter. It has also led to the heavy acidifying deposition of the sulfates and nitrates on the plant surfaces. The further consequences are reduced growth in plants, less yield and decrease in the reproduction of plants.

V. Conclusion

In this paper, the predictive analysis of the pollutant levels at Madrid in Spain has been done through artificial neural networks. The different parameters of the neural network have been used to get the accurate results. The highest concentration of $SO_{2,}$ PM 2.5 and PM10 are found in industrial areas and that of NO_2 is found in commercial areas. So the topic of concern is it is slowly deteriorating the human health and the ecosystem. In this paper, analysis has been presented using Artificial Neural Network to predict the levels of the various pollutants of a place. Since the airborne PM is characterized by the diverse effects on climate, human health, ecosystem etc., further studies can be done in a manner like analyzing the levels of air pollutants causing pollution and the consequences in different areas. Epidemiological studies on large pollutants are unable to identify any threshold concentration below which ambient PM has no effects on health. This paper briefly points out the affects and therefore it is high time that steps should be taken to mitigate the problem before it becomes a serious threat to the living species in this world. In this paper the concentration of PM 10 and PM 2.5 of a particular place have been predicted but if sufficient data is available, this artificial intelligence technique can be applied to predict particulate matter concentration of other places as well.

References

[1]. Ma, J., Xu, F., Huang, K., Huang, R., Improvement on the linear and nonlinear auto-regressive model for predicting the NOx emission of diesel engine, Journal of Neuro-computing, 207(26), 2016, 150-164

[2]. Wang, J., Zheng, K., Singh, R., Lou, H., Hao, J., Wang, B., Cheng, F., Numerical simulation and cold experimental research of a low-NOx combustion technology for pulverized low-volatile coal, Applied Thermal Engineering, 114(5), 2017, pp 498-510

[3]. Das M., Çevik Y., Erel, Ö., Çorbacioglu SK., Ischemia-modified albumin levels in the prediction of acute critical neurological findings in carbon monoxide poisoning, The Kaohsiung Journal of Medical Sciences, 32(4), 2016, 201-206.

[4]. Ma, J., Ma, F., Zhou, Z., Liu, Y., Theoretical prediction on CO insertion reactions through the anionic complex [ClMg(η2-O2C]- as a catalyst, Journal of Atmospheric Pollution Research, 7(3), 2016, 412-418.

[5]. Seshadri, K., Bai, X., Rate-ratio asymptotic analysis of the influence of addition of carbon monoxide on the structure and mechanisms of extinction of no premixed methane flames with comparison to experiments, Journal of Combustion and Flame, 175(1), 2017, 107-117.

[6]. Li, S., Zhou, H., Hou, L., Ren, Z., An analytic model for the effects of nitrogen dilution and premixing characteristics On NOx formation in turbulent premixed hydrogen flames, International Journal of Hydrogen Energy, 2017

[7]. Rao, R., Honnery, D., A study of the relationship between NOx and the ion current in a direct-injection diesel engine, Journal of Combustion and Flame, 176(1), 2017, 309-317.

[8]. P. Raktim, K. Ki-Hyun, H. Yoon-Jung, and J. Eui-Chan. "The pollution status of atmospheric carbonyls in a highly industrialized area". Journal of Hazardous Materials, 153(3), 2008, 1122–1135

[9]. Gianola, D.; Okut, H.; Weigel, K.A.; Rosa, G.J.M. Rosa, Predicting complex quantitative traits with Bayesian neural networks: A case study with Jersey cows and wheat. BMC Genetics, 12(87), 2011, 1–37.

[10]. M.I.A. Lourakis. A brief description of the Levenberg-Marquardt algorithm implemented by levmar, Technical Report, Institute of Computer Science, Foundation for Research and Technology - Hellas, 2005.

[11]. M. Younger, H.R. Morrow-Almeida, S. M. Vindigni, and A. L. Dannenberg. "The Built Environment, Climate Change, and Health: Opportunities for Co-Benets". American Journal of Preventive Medicine, 35(5), 2008, 517–526.

Shameem Hasan "Prediction Of Air Pollutant Concentration Based On Artificial Intelligence Technique " IOSR Journal of Environmental Science, Toxicology and Food Technology (IOSR-JESTFT) 12.11 (2018): 01-07.