

# Recognition of Future Air Quality Index using Artificial Neural Network

Ruchi Raturi<sup>1</sup> Dr. J.R. Prasad<sup>2</sup>

<sup>1</sup>Student <sup>2</sup>Professor

<sup>1,2</sup>Department of Computer Engineering

<sup>1,2</sup>SCOE Vadgaon Pune, Maharashtra, India

**Abstract**— The environment is affected in terms of global climate change and adverse effects for plants and eco systems due to urbanization in recent years. Vehicles are a significant source of emissions into the atmosphere. The commonly occurred air pollutants are CO, NO, NO<sub>2</sub>, PM 10, O<sub>3</sub>, SO<sub>2</sub> and several organic compounds which gives dangerous effects on biological structure of human being like loss of human lives and spoil other living organisms. Hence there is need to monitor air pollution. Many decision support systems are designed for monitoring the data but they are largely limited. We are implementing a system with data mining method for prediction of next day or next month air pollution count with Multilayer perceptron (MLP) and Artificial Neural Network (ANN).

**Key words:** Air Quality Index, Artificial Neural Network (ANN)

## I. INTRODUCTION

Day to day increase a pollution so its required to predict pollution of the next date, next months, next year. Using some previous air related data.

Air pollution is escalating day by day caused by different human activities, and inception of chemicals, particulates, or biological materials into the environment that cause distress, or death of human lives, impairment of other livelihood and disturb the normal environment. In reality, air pollution is main environmental problem in urban cities. So it is necessary to forecast pollution and evade the problems caused due to.

We can say that Artificial Neural Network is tool for predicting and data evaluation for worth of Air. With the data mining prediction of pollution is complicated job. Various schemes are intended to support air pollution data storing, inventory management and generation of simple statistics. Some systems use decision support systems, but they are largely limited. They can answer simple queries like “What is the maximum limit of air pollution”, “which area have a maximum pollution” However, they cannot answer complex queries like “Predict next month air pollution count.”, “Given me tomorrows pollution details” this type of prediction techniques are used in this system.

The increase in global warming, many health disorders that are found in human body as well in animals because of global warming. Even the food, fruits and vegetables we eat is not safe, it also get affected by various air pollutants [2]. We come to know that all human beings, animals, plants and the environment surrounding them get affected due to air pollution can be avoided by predicting it before it causes harm to the people. It is more useful to the people who has less immune power i.e by predicting it before let the people know to go in that particular area of not.

## II. LITERATURE REVIEW

The comparison of two methods of feature selection i.e. the genetic algorithm, and linear method of stepwise fit is given in[3]. Depending on that the factors responsible to influence the prediction is explained. For last prediction the system uses neural networks.

The paper gives comprehensive review on existing air quality forecasting techniques through soft computing. Normally Artificial Neural Network is measured to be an excellent predictive tool [2].

The paper proposed MANET (Mobile Ad Hoc Network) routing algorithm[4] which has nearly 28 mobile nodes (Vehicles) provide a exposure region of 300 meters around the city. The method gives the constraint like data type, the speed of transmission, coverage of the system, coverage area size, and no. of vehicles to closely monitor the system. Authors have used principal module regression to envisage air quality index (AQI). The AQI is calculated for 2000- 2006. The K-means clustering and Logistic model[5] is implemented to forecast the carbon content emissions in China from 2014 to 2023. The indicators, which include urbanization rate, GDP, and the proportion of the second industry, the energy efficiency and the carbon emission intensity are selected.

The data mining techniques to forecast which day have maximum pollution of Boushehr city [6] seeing that an aspect of air pollution with the 53 years data indicating in between 1951 to 2003. They export decision rule from data by using the air pressure, humidity and dusty days of January, February and March of each year and we describe the relation between polluted days of upcoming month and this last three-month feature.

## III. PROPOSED SYSTEM

In this paper, a multilayer perceptron algorithm and Time series Analysis is proposed for pollution prediction to get better accuracy for prediction. The data is fetched from <https://data.gov.in/> website. Parameters used are "SO<sub>2</sub>", "NO<sub>2</sub>", "RSPM (Respirable Suspended Particulate Matter) [1]. Below figure shows the architecture of given scheme.

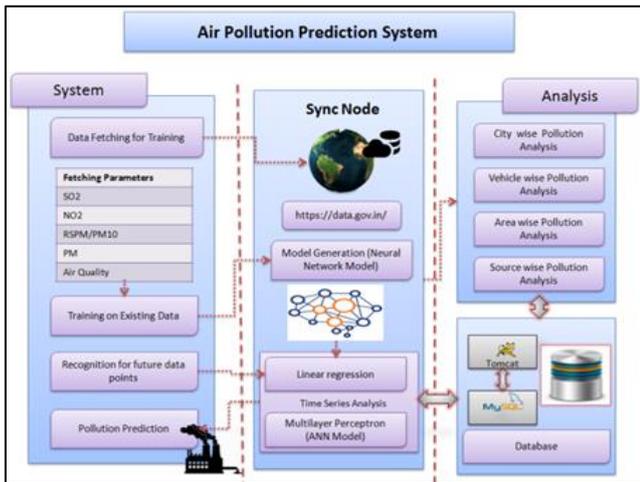


Fig. 1: Architecture Diagram

Existing data parameters along with their air quality parameter will be provided in training set. Output of the training set will be AIR Quality index. Data has to be provided in time series in the form of (SO<sub>2</sub>, NO<sub>2</sub>, RSPM, PM, air quality day wise) train the data getting from above website by using multilayer perceptron (MLP) is an artificial neural network (ANN) model that plots set of input data into a set of suitable outputs. A MLP contains several layers such that every layer fully associated to the subsequent one. Apart from the input nodes, each node is element through which processing can be done of a nonlinear activation function. The flow of the projected scheme is as given in below figure. The cause and effect of parameter on climate change is as follows:

#### A. SO<sub>2</sub>

It generally formed by burning of remnant fuel. It can cause acid rain and smog (smoke + fog) which create visibility problem [14].

#### B. NO<sub>2</sub>

It is produced during burning of fuel under high temperature. Responsible to increase temperature and it also creates smog [14].

#### C. O<sub>3</sub>

Is formed when a chemical reaction of explosive crude compounds is happen and no occurs in sunlight.

#### D. RSPM

A Respirable Suspended Particulate Matter is a complex pollutant and consists of variety of components. As fine particles in the atmosphere are increased as it cause health issues like heart disease or lung cancer [14].

#### E. CO<sub>2</sub>

It is one of the main pollutants of causing global warming. Because of greenhouse effect, the heat on the earth is increased and seasonal change is occurred because of the increase of CO<sub>2</sub>[14].

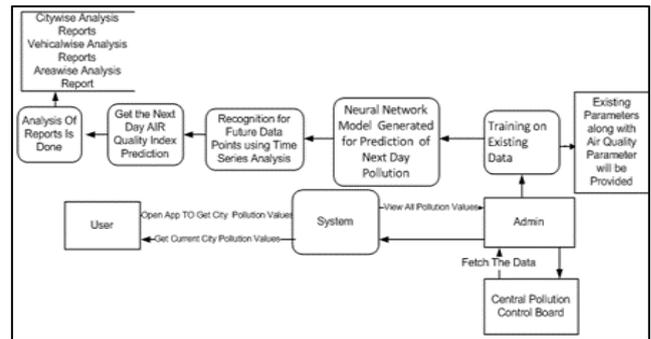


Fig. 2: Flow Diagram

## IV. ALGORITHM USED

### A. Linear Regression

Linear regression is an arithmetical technique that gives the study of relationships among two continuous variables  $y$  and one or more descriptive variables indicated by  $X$ .

The descriptive variable is also labeled as simple linear regression. In linear regression, the relationships are represented with linear predictor functions whose factors are calculated from the data. Similar models are called linear models.

The least square regression line for the set of  $n$  data points is given by

$$y = ax + b$$

Where  $a$  and  $b$  are given by

### B. Multilayer Perceptron (ANN)

A Multilayer Perceptron (MLP) is a forwards network model of the artificial neural network that has the map sets input data on the set of proper outputs. An MLP contains various layers of nodes in a directed graph, with each layer completely linked to the next one. Accept the input nodes; each node has a nonlinear establishment function with a neuron. MLP uses supervise learning technique that is called Back Propagation for training the network. MLP is an adaptation in the ordinary linear perceptron and can differentiate data that are nonlinearly distinguishable.

#### 1) Activation Function

If all neurons have a linear activation function in multilevel perceptron, then it is a linear function that plan the inputs and the output of particular neuron, so that linear algebra of some layers can be decreased into standard two-layer input-output models.

$$y(v_i) = \tanh(v_i) \text{ and } y(v_i) = (1 + e^{-v_i})^{-1},$$

In which the former function is a hyperbolic tangent which ranges from -1 to 1, and the latter, the logistic function, is similar in shape, but ranges from 0 to 1.

#### 2) Layers

It includes three or more layers of nonlinear-active nodes and thus considers a deep neural network.

#### 3) Learning through Back Propagation

Comparing the expected results, the number of errors in the output, along with the process of converting the weight of each part of the data, changes in the perceptron. It is a model to manage learning, and it approved out through back propagation.

We show a fault in output node  $j$  in the  $n$ th data point by

$$e_j(n) = d_j(n) - y_j(n)$$

Where  $d$  the target is value and  $y$  is the value formed by the perceptron.

After that we get better weight of the nodes on the basis of those corrections, which reduces the error in whole output, given by Using the gradient descent, we find our

$$\mathcal{E}(n) = \frac{1}{2} \sum_j e_j^2(n).$$

Alteration in every weight, where ( $Y_i$ ) is the outcome of

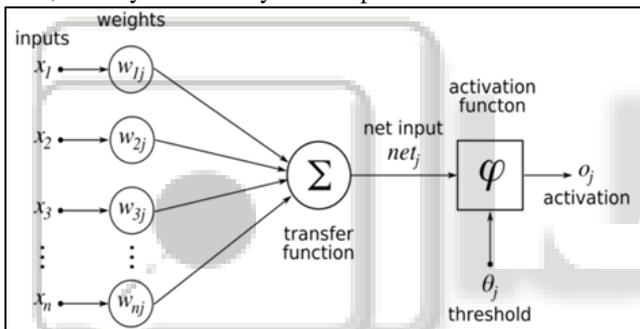
$$\Delta w_{ji}(n) = -\eta \frac{\partial \mathcal{E}(n)}{\partial v_j(n)} y_i(n)$$

The previous neuron and is the learning rate.

The calculated derivative derived from the induced local field, which changes itself. The derivative of the activation function given above is.

$$-\frac{\partial \mathcal{E}(n)}{\partial v_j(n)} = e_j(n) \phi'(v_j(n))$$

Conservatively the input layer is layer 0, and if we are thinking about an  $N$  layer network, you can say that the  $N$  layers of weights and  $N$  non-input layers of processing units. Thus, two layers Multilayer Perceptron takes the form.



By using MLP (ANN) model for recognition for future data points. Depending upon the trained model, new values will be provided to get the subsequent day AIR quality prediction. Also you can predict next month next year wise air quality prediction using this MLP model. And generate model for future use.

The inspection can be completed on the data such as City wise and Vehicle wise to know which vehicle is most affected.

### C. River Waves Algorithm

It is a part of ANN (Artificial Neural network) algorithm it used to give the performance of organic structure formulated as “neurons”. It can be used for the forecasting the numerical values from the given data. The data given by the River Waves algorithm is fiscal data. River waves include a large number of connected processing units that work together to process information. They also generate meaningful results from it. It contains the following 3 layers:

- Input layer

The action of the input units correspond to the rough information that can feed into the network.

- Hidden layer

To determine the action of every unknown unit. There might be one or additional hidden layers.

- Output layer

The performance of the output units based on the action of the hidden units and the weights among the hidden and output units.

The algorithm with auto correction which displays some fundamental ideas.

For easy understanding purpose, we'll consider a flow in the ocean: "river waves". The major round gives the variation of number of points (N) in the window of time prediction. A head period is find out by N in each distinction from "river waves". The number of unknown layers is  $2*N+1$  automatic. As figure of merit will be used the

$$R = \sum |Fobs| - |Fcalc| / \sum |Fobs|,$$

So called R-factor through full time series data. For every N will be variated the training sets: the elements will be detached successively to attain the least of R-factor and to get most positive relationship among the under fitting and over fitting. The middle value through committee is the result of this simple automatic flow.

## V. EXPERIMENTAL RESULTS

JavaScript and HTML 5 designed for developing client-side applications which are open source, cross-platform from which user can see the pollution values. JavaScript is runtime environment. The systems GUI was designed in JSP. Core Technologies used were JSP, J2EE etc. The development was done in the Eclipse IDE and for DB we used MY SQL GUI browser. The user interface was integrated with the modules which made the use web services. Data flow originated from the database built in MySQL. The database basically used for user storing user details like usernames, password, locations, and pollution data.

For analysis and testing purpose data is collected from www.data.gov.in. In which data collection for different cities from various location can done.

The major air pollutants SO2, NO2 and PM10 are taken into consideration. The below figure shows the daily pollution by three pollutants. The x axis gives pollution prediction occurred at the different time in a day .The y axis gives the pollution values in mg/m3.The prediction is done on the data gathered for last one month.

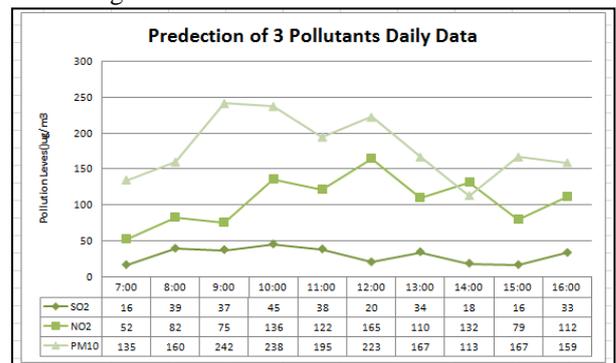


Fig. 3: Daily Data of Pollution

By using linear regression approach we can increase the accuracy up to 80% in prediction of pollution as compared to other algorithms as SVM classifier and random forest.

Algorithm Comparison				
Pollutants	Algorithms	Tested Records	Accuracy	% Accuracy
SO2	SVM Classifier	200	122	0.61
	Random Forest	200	120	0.6
	Linear Regression (Our Approach)	200	160	0.8
NO2	SVM Classifier	200	124	0.62
	Random Forest	200	120	0.6
	Linear Regression (Our Approach)	200	160	0.8
O3	SVM Classifier	200	124	0.62
	Random Forest	200	120	0.6
	Linear Regression (Our Approach)	200	160	0.8

Fig. 4: Algorithm Comparison

Figure 4 shows the comparison of the proposed method with existing one

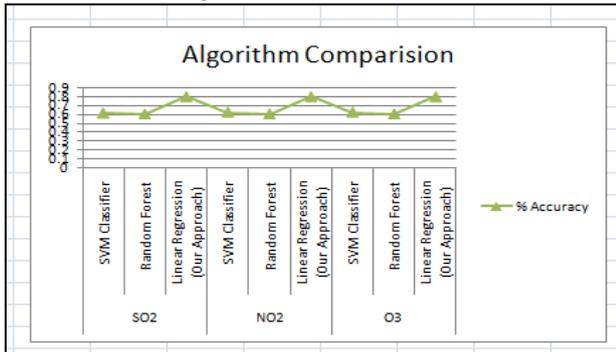


Fig. 5: Plot of Algorithm Comparison

The graph of algorithm comparison with greater accuracy using our approach is given. ANN are good modeling method which can produce good results for prediction. So it helps in analyzing pollution details and predicts future pollution.

## VI. CONCLUSION

A multilayer perceptron (MLP) feed forward artificial neural network model is proposed in this system to increase the accuracy in the prediction. By Using MLP (ANN) model for recognition for future data points can be done Based on the trained model and new values will be provided to get the next day, next month, next year AIR quality index prediction.

The parameters like SO2, NO2, PM, Ozone, Air Quality can take into the consideration.

Finally end result shows that the artificial neural network model is better than previously used algorithm.

In future other prediction algorithm can be used for greater accuracy in prediction of pollution.

## REFERENCES

- [1] "FORECASTING THE URBAN AIR QUALITY USING VARIOUS STATISTICAL TECHNIQUES", Anurag Kandya and Manju Mohan, The seventh International Conference on Urban Climate, 29 June - 3 July 2009, Yokohama, Japan
- [2] "A survey on Air Quality forecasting Techniques" Niharika, Venkatadri M, Padma S.Rao, International Journal of Computer Science and Information Technologies, Vol. 5 (1), 2014, 103-107.
- [3] "Data mining methods for prediction of air pollution", Krzysztof SIWEK, Stanislaw OSOWSKI.
- [4] Kumar, P. Goyal, Forecasting of air quality in Delhi using principal component regression technique.,

- Atmospheric Pollution Research, vol.2 Issue 4, October 2011, pp. 436444.
- [5] Lijun Ma, Kangqing Lin, Meijiao Guan, Meiyuan Lin, The prediction of carbon emission in all provinces of China with the K-means cluster based Logistic model, Humanities and Social Sciences Foundation of Ministry of Education of China with No. 14YJC630096.
- [6] Ebrahim Sahafizadeh, Esmail Ahmadi Prediction of Air Pollution of Boushehr City Using Data Mining, 2009 Second International Conference on Environmental and Computer Science.
- [7] R. Sindhvani, Assessment of gaseous and respirable suspended Particulate matter (PM10) emission estimates over megacity Delhi: past trends and future scenario (2000-2020), Center for atmospheric sciences, November 2012, pp. 123-140.
- [8] D. Mishra, P. Goyal, Development of artificial intelligence based NO2 forecasting models at Taj Mahal, Agra centre for atmospheric sciences, Atmospheric Pollution Research, Apr.2015, pp. 99-106.
- [9] Bureau of energy statistics of the national bureau of statistics. China's energy statistics yearbook 2014.
- [10] Kumar, P. Goyal, Forecasting of air quality in Delhi using principal component regression technique., Atmospheric Pollution Research, vol. 2 Issue 4, October 2011, pp. 436444.
- [11] Khunarak, C., Lutz, M., Kerdcharoen, T. Wi-Fi Electronic Nose For Indoor Air Monitoring IEEE Conference, pp. 1-4, (May 2012).
- [12] Zhu, Q., Peng, X., 2012. The impacts of population change on carbon emissions in China during 1978-2008. Environ. Impact Assess. Rev. 36, 1-8.
- [13] Intergovernmental Panel on Climate Change, 2014. The IPCC Fifth Assessment Report. <http://www.ipcc.ch/tabs-2>.
- [14] "Weather Forecasting using Incremental K-means Clustering", Prof. N.K.NAGWANI, SANJAY CHAKRABORTY, LOPAMUDRA DEY.