International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 03 | Mar 2020 www.irjet.net p-ISSN: 2395-0072

Prediction of Air Pollutant Concentration using Deep Learning

Sangeetha A[1], Vijaya Subramaniam P R[2]

[1]PG Scholar, Environmental Engineering, Kumaraguru College of Technology, Coimbatore, India
[2]PG Scholar, Environmental Engineering, Kumaraguru College of Technology, Coimbatore, India

---

Abstract - *This paper focuses on applying on Artificial Neural Networks (ANNs) by using of Feed-Forward Back-Propagation to make performance predictions of $PM_{2.5}$ at Manali, Chennai which is an industrial area, in terms of NO_2, SO_2, CO and meteorological factors such as relative humidity and wind speed that data gathered for research over a three year period. The model studies have 1084 observed data, the uncertainties of inputs and output operation are trained using a designed network. Furthermore, ANN effectively analyzes and diagnoses the modelled performance of nonlinear data. The study provides the ANN model with the predictive performance of $PM_{2.5}$ with a correlation coefficient (R) between the variables of predicted and observed output that was 0.8.*

Key Words: Artificial Neural Network, Air Quality Prediction, Deep learning, Transfer functions.

1. INTRODUCTION

Mathematical models are used to describe how the dynamics of air pollutants were processed that includes formation, emission, transport and disappearance of air pollutants. For more complex more information is required by the model for their application to have sufficient certainty that the results will have technical or scientific value (Russell, 2000). These deterministic models require much information that is not always possible to obtain; the data available have not always resulted in successful outcomes upon application of the model (Roth, 1999), or the cost of obtaining reliable data can be prohibitive (Louis, 2000). The methods that require less information generally make use of statistical techniques such as regression or other data-fitting methods using numerical techniques. These methods establish the respective relationships between the various physicochemical parameters and variables that are taken based on daily measured historical data. Flexibility, efficiency and accuracy of modelling and prediction is achieved greater in Artificial Neural Network than other models. ANN has different features similar to those of the brain. So they are,

i. Capable of learning from experience.
ii. Generalize from previous data to new data.
iii. Abstract the essential element from inputs containing irrelevant information.

They use adaptive learning; it performs tasks based on training experience. ANN can generate their own distribution of the weights of the links through learning and it does not need any algorithm to solve a problem. Because of these features, ANN has low computational requirements and their construction is less complex.

The pollutant of interest in this study is ambient air pollution, that is caused due to the vehicles, industries etc. It is the main component in creating the type of air pollution known as smog. According to the Central Pollution Control Board (CPCB) and the newspaper articles, the metropolitan zone of Chennai, TamilNadu is in fourth place in India in exceeding the ambient air pollution standards given by Central Pollution Control Board. Ambient air pollution has $PM_{2.5}$, NO_x, CO and SO_2 as the major pollutants with harmful effects on human health, causing respiratory problems and ailments such as headaches, irritation of eyes and affecting vegetation, metals and construction materials.

The meteorological factors are ambient temperatures, relative humidity, wind speed, wind direction, and gaseous pollutants are particulate matter 2.5 & 10, Sulphur dioxide (SO_2), carbon monoxide (CO) and nitrogen dioxide (NO_2). Ambient air quality was strongly influenced by meteorological factors through the various mechanisms in the atmosphere, both directly and indirectly. Wind speed and wind direction are responsible for certain processes of particle emission in the atmosphere, that is it results in re-suspension of particles and diffusion, as well as the dispersion of particles. Increasing wind speed results in decreased pollutant concentration, it dilutes the pollutant. Through scavenging process particulates in the atmosphere are removed and it is able to dissolve other gaseous pollutants. Heavy rainfall results in better air quality. Ambient temperature gives additional support in air quality assessment, air quality modeling and forecasting model.

1.1 Study Area

Chennai is located on TamilNadu Coromandel coast of Bay of Bengal, with around 7,088,000 inhabitants of various nationalities. It is located in the south eastern part of India and north eastern part of Tamil Nadu. It has a land area of 426 km^2, is one of the most densely populated metropolises in India, and consists of river, the Kortalaiyar, travels through the northern part of the city and drains into the Bay of Bengal, at Ennore, Adyar and Cooum rivers and the Buckingham Canal, about 4 km inland, runs parallel to the coast and also the Otteri Nullah, an east–west stream, runs through north Chennai and meets the Buckingham Canal at Basin Bridge. Dry Summer tropical wet and dry climate and the hottest part of the year is from late May to early June. And humid with

occasional rainfall, the coolest part of the year occurs in the month of January. The recorded annual average rainfall in Chennai is about 140cm (55 in). The primary pollutants are carbon monoxide and sulfur dioxide, nitrogen dioxide and particulate matter emissions from vehicles and industries.

Manali is the industrial area which has largest petrochemical complexes in India and it is spread over an area of about 2000 hectares. It is located in the suburb of Chennai city about 20 Km north. It extends its border at Thiruvottiyur on the East, Chennai city by South, Kossathaliyar river in North and west by villages of Manjambakkam, Mathur and Madhavaram. The petrochemical complex is connected by Ennore High road and by NH-5A of Chennai to Kolkata. The Manali town is located at the west nearer to this complex. The population as per 2011 census in Manali was 35,248. Types of industries located in the Manali industrial area were Highly polluting industries about 12, Red category industries about 11, Orange and Green category industries about 5. The CEPI score established by MOEF and CPCB for Manali was 76.32. The monitoring station was set up in Manali by the TamilNadu pollution control board.

2. METHODS AND MATERIALS

2.1 Data Sets

The data used in this research are relative humidity, wind speed, and daily twenty four hour average concentration of CO, NO_2, SO_2 and $PM_{2.5}$ in Manali, Chennai for three years period from 2017 to 2019. All of this data was provided on the Central Pollution Control Board website. The data has to be divided for feeding in ANN network for learning, training and testing it helps in the verification of efficiency and correctness of the model developed.

Table -1: Correlation coefficient of different air pollutants

Variables	$PM_{2.5}$	SO_2	NO_2	CO	WS	RH
$PM_{2.5}$	1.00	0.12	0.04	-0.05	-0.02	-0.04
SO_2	0.12	1.00	0.02	0.09	-0.12	0.02
NO_2	0.04	0.02	1.00	0.04	0.07	-0.11
CO	-0.05	0.09	0.04	1.00	0.03	-0.11
WS	-0.02	-0.12	0.07	0.03	1.00	-0.47
RH	-0.04	0.02	-0.11	-0.11	-0.47	1.00

Table -2: Statistics of collected values from 01 January 2017 to 31 December 2019

Variable	Unit	Range	Mean	St. dev.
$PM_{2.5}$	$\mu g/m^3$	[0.08, 4.43]	1.20	0.55
SO_2	$\mu g/m^3$	[31.27, 99.23]	75.45	12.94
NO_2	$\mu g/m^3$	[0.15, 4.7]	1.01	0.36
CO	mg/m^3	[0.76, 141.56]	13.71	13.68
WS	m/s	[1.76, 167.07]	22.60	12.88
RH	%	[3.79, 346.21]	61.02	36.36

The above table-1 shows the correlation coefficients of all the air pollutants and the table-2 gives the parameters to be used in prediction, its units, minimum and maximum range of values, their mean and standard deviation.

2.2 Software

MATLAB Neural Network Toolbox (The MathWorks Inc. USA) is used for development of the air quality prediction model because it is more flexible and easier to apply. This software has a Neural Network Toolbox which has a broad variety of parameters for developing the neural network.

2.3 ANN Model for Air Quality Prediction

Feed forward back propagation network was used in this research. The input layer contains following the input variables CO, NO_2, SO_2 and $PM_{2.5}$ and two support variables which are relative humidity, wind speed. Therefore, there were six neurons in the input layer. The number of hidden layers and neurons in each hidden layer are the important parameters that result in better prediction in the model. Optimization of ANN performance is done with the help of two hidden layers and different values of neurons. The output layer consists of the target of the prediction model.

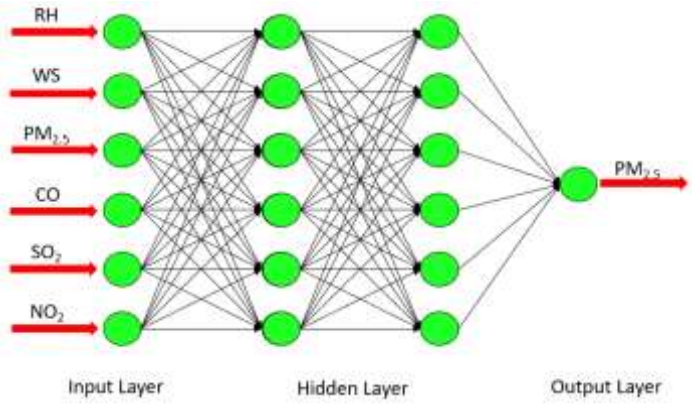

Fig -1: ANN architecture for air quality prediction

Here, $PM_{2.5}$ is used as an output variable. Hyperbolic tangent sigmoid function and Linear transfer function were used as the transfer function. The database was divided into three sections, training the networks uses 70% of data, validation set uses 15% of data and the testing employed with remaining 15% of data. The network performance was measured using one of the statistical methods Mean Square Error (MSE). The architecture of ANN is shown in fig-1.

2.4 Feed Forward Neural Network Based on Back Propagation (FFANN – BP)

Artificial neural networks use approximate functions that require large numbers of inputs that work on the basis of biological neural networks. Artificial neural network structure consists of layers and interconnected nodes. The simplified mathematical model, Feed forward artificial

neural networks are commonly used information processing units, and it is able to perform well in capturing complex interactions within the given input parameters with satisfactory performance.

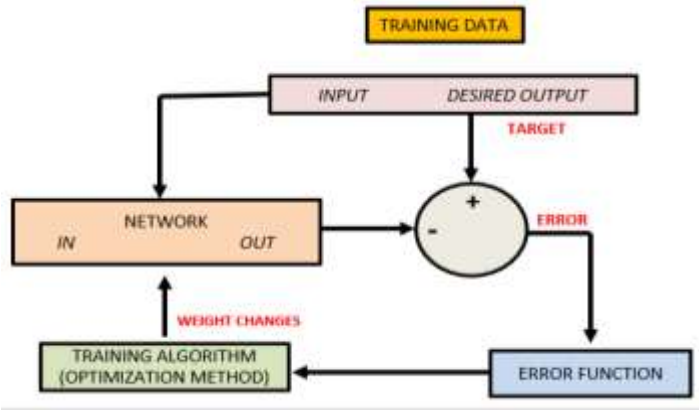

Fig -2: Schematic representation of the ANN model

Feed forward back propagation is the most widely used supervised learning method in this method and the teacher is required to get the desired output for any given input. This system is made of interconnected neurons that pass messages between each other. It will assign the numeric weights that can be changed based on experience. Traditional approaches cannot capture the influential factors of air pollution concentration, but feed forward back propagation networks are able to capture and provide relationships between the pollutant concentration and their influential factors.

Feed forward back propagation networks consist of an input layer, hidden layer and the output layer. After the input, hidden and the targets are fed, the network starts to learn the system, this process is called the learning process. The learning process is stopped when the error of the neural network is reduced to the desired minimum. Followed by a learning process, training process is carried out. In Feed forward back propagation first the data stream is propagated in the forward direction and then the error signal from the data stream is backward - propagated

$$x_{jq}^{l} = \Sigma_i w_{jiq}^{l} y_{iq}^{l-1} \quad (1)$$

Where, w_{jiq}^{l} is the weight which connects the i^{th} neuron in the l-1^{th} layer and the j^{th} neuron in the l layer, $y_{iq}^{l-1} = f\left(x_{jq}^{l}\right) - \theta_{jq}^{l}$ is the response of the j^{th} neuron in the l^{th} layer, and f is the activation function, θ_{jq}^{l} is the bias of the neuron. Fig-2 gives the schematic representation of the ANN model.

2.5 Model Development

In the Data Manager window of MATLAB Toolbox there is a Neural Network that lets the user to import, create, and then export neural networks and data are created. Neural Network Training window was illustrated as follows:

- Network inputs: CO, NO_2, SO_2 $PM_{2.5}$, relative humidity and wind speed.
- Network target: $PM_{2.5}$
- Network type: Feed-Forward Back-Propagation.
- Function of training: TRAINGDX
- Adjustment learning function: LEARNGD
- Function of performance: MSE.
- Number of hidden layers: 2.

3. RESULT AND DISCUSSION

The aim of the experiment is to examine the performance of the model used in ANN for the prediction of $PM_{2.5}$ concentrations. ANNs model that provides the maximum Correlation coefficient (R^2) and minimum Mean Square Error (MSE) or Mean Absolute Error (MAE), is said to be a better predicted ANN model.

Feed-forward back propagation neural networks have been used in this study. The transfer functions tansig and purelin were used for the neurons in the hidden layer and output layer respectively. Using gradient-descent with momentum back-propagation the weights and bias were adjusted in the training phase. The network performance was validated by selecting Mean Square Error. The parameters and their values were shown in table-3.

Table -3: Training results of neural network

S.No.	Transfer function	Momentum constant	Learning rate	MSE	R^2
1	Purelin	0.6	0.1	0.06	0.99
2	Tansig	0.6	0.1	1.04	0.976

The network performance versus the number of epochs in the training phase were represented in the following figure. In training, the weights are adjusted to minimize the large values at the first epoch. Black dashed line in the performance graph represents the best validation performance of the network. The training process cutoff when the green line which represents the validation training set (network performance) intersects with the black line. The network performance function is shown in fig-3, fig-4.

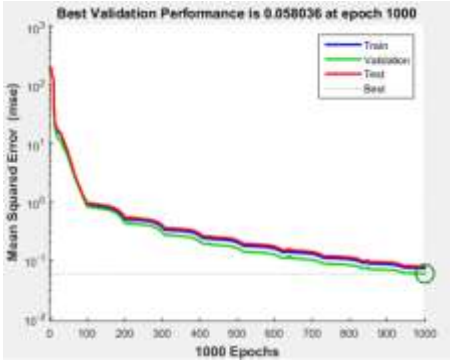

Fig -3: Performance of Purelin function during training

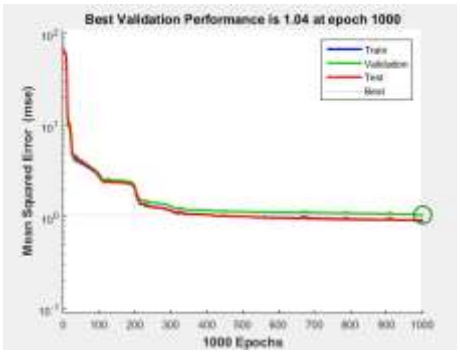

Fig -4: Performance of Tansig function during training

After the completion of training, testing and validation process, regression analysis was performed to compare the correlation between the actual and predicted results based on the value of correlation coefficient, R. The value of R which is equal to 1 represents the perfect fit between training data and the produced results. Fig-5, fig-6 shows the regression analysis plots of the network structure. In a regression plot, the solid line indicates the perfect fit which shows the perfect correlation between the predicted and targets. The best fit produced by the algorithm is indicated in the dashed line.

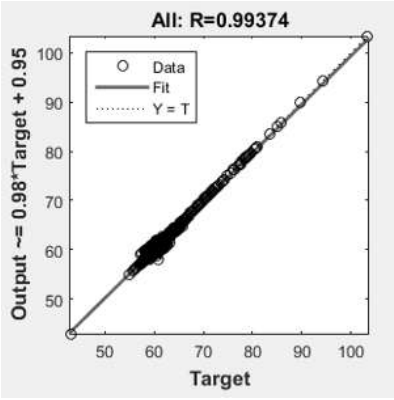

Fig -5: Regression plot for Purelin function

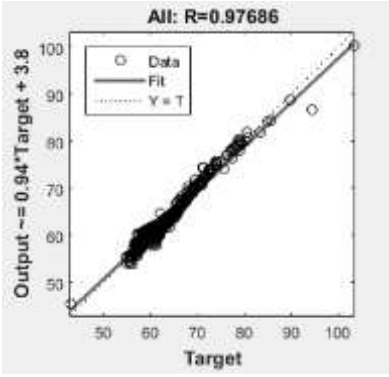

Fig -6: Regression plot for Tansig function

The best models for the air quality prediction is the one which has the lowest values. Here, the MSE is 0.06 and coefficient of determination, R^2 is 0.99. The best model shows a good agreement between predicted and measured data based on the value of correlation coefficient, R.

4. CONCLUSIONS

In this work a successful execution of prediction of $PM_{2.5}$ by using artificial neural networks (ANNs) including Multi-layer perception (MLP) with feed forward with back propagation algorithm has been accomplished. The analysis of the daily data set depended on the influences of NO_2, SO_2, CO, WS and RH as an input and $PM_{2.5}$ as output. ANN was modeled using the actual observed daily data for the period 2017–2019. TRAINGDX algorithm was applied to the data, to get better predicted outcomes which prove the used algorithm's excellence.

The result indicates the higher value of correlation coefficient (R^2) between the variables of predicted and measured output getting up to 0.99 and Mean Square Value (MSE) value of best validation performance is 0.06.

Hence, this model development provided the satisfactory for accuracy and capability and is suitable for the prediction models. The outcome, depending on the high value accuracy of ANN in prediction, the neural network modeling might successfully simulate and predict the particulate matter ($PM_{2.5}$) emission.

REFERENCES

[1] Afshin Khoshand, Mahshid Shahbazi et al., "Prediction of Ground-Level Air Pollution using Artificial Neural Network in Tehran," Anthropogenic Pollution Journal, vol.1(1), 2001, pp.61-67.

[2] Ana Russo, Pedro G. Lind et al., "Neural Network Forecast of Daily Pollution Concentration using Optimal Meteorological Data at Synoptic and Local Scales," Atmospheric Pollution Research, 2015, Vol.6, pp.540-549.

[3] Ignacio Garcia, Jose G. Rodriguez et al., "Artificial Neural Network Models for Prediction of Ozone Concentrations in Guadalajara, Mexico," Air Quality Models and application, 2011, pp.35-52.

[4] Jiangshe Zhang, Weifu Ding, "Prediction of Air Pollutants Concentration Based on an Extreme Learning Machine: The Case of Hong Kong," Environmental Research and Public health, vol.14 (114), 2017.

[5] Maitha H. Al Shamisi, Ali H. Assi et al., "Using MATLAB to Develop Artificial Neural Network Models for Predicting Global Solar Radiation in Al Ain City - UAE," Intech open science.

[6] Mohammed Dorofki, Ahmed H. Elshafie et al., "Comparison of Artificial Neural Network Transfer Functions Abilities to Simulate Extreme Runoff Data," vol.33, 2012, pp.39-44.

[7] Prachi, Kumar Nishant et al., "Artificial Neural Network Applications in Air Quality Monitoring and management," International Journal for Environmental Rehabilitation and Conservation, vol.2(1), 2011, pp.30-64.

[8] Suhasini V. Kottur, Dr.S.S. Mantha, "An Integrated Model using Artificial Neural Network (ANN) and Kriging for Forecasting Air Pollutants using Meteorological Data," International Journal of Advanced Research in Computer and Communication Engineering, vol.4(1), 2015, pp.146-152.

[9] Suraya Ghazali, Lokman Hakim Ismail, "Air Quality Prediction using Artificial Neural Network".

BIOGRAPHIES

"**A. Sangeetha** doing my Masters in Environmental Engineering at Kumaraguru College of Technology, Coimbatore. I have completed my Bachelor's Degree in Civil Engineering in the year 2018 at Karpagam College of Engineering, Coimbatore, with CGPA of 9.12. Student member in Indian Geotechnical Society 2015 -2018. Student member in Indian Concrete Institute 2017-2018. Participated in the National and International conferences".

"**P. R. Vijaya Subramaniam** doing my Masters in Environmental Engineering at Kumaraguru College of Technology, Coimbatore. I have completed my Bachelor's Degree in Civil Engineering in the year 2018 at National College of Engineering, Tirunelveli, with CGPA of 7.97, Participated in National and International conferences".