Prediction of Air Quality in Urban Area,

Chennai

**Abstract**:-

The quality of air in ALANDUR, Chennai is polluted by Particulate Matter (PM2.5) over the years. Reports prove that particulates affect the health of humans and environment. Development of accurate forecasting models to find PM2.5 concentration in air helps to take control measures, early warning and relieve measures. In this study, the performance of non-linear model (Feed Forward Back Propagation using LEARNGD function) with meteorological data and gaseous pollutants as input parameters from the year 2015–2019 at ALANDUR with different surrounding activities of urban area. In this paper, the prediction of PM2.5 in the study area is mainly focused to find the effects of harmful emissions. To predict PM2.5, an artificial neural network (ANN) prediction model is developed. The data obtained from the monitoring station on the ALANDUR. Bus depot ALANDUR area in Chennai is given as input variable. The prediction model is validated and evaluated by statistical calculations, and then it was found that it performed well in the prediction of PM2.5. The performance of the developed model was evaluated by Mean Square Error (MSE) and value of R2. The best prediction performance was observed in the model for Purelin transfer function with R2 value of 0.96 and MSE of 0.094 and for Tansig transfer function with R2 value of 0.97 and MSE of 0.103 from the framed networks.

Keywords— Air quality prediction, Artificial neural network, PM2.5 Prediction, MATLAB. I.

# INTRODUCTION

Technological advancements lead to the emissions of air pollutants over the decades. Major concerns in industrial cities which experience air pollution, can be harmful not only for the environment but also for human health. Due to this urban resident are more likely to live in less polluted neighborhoods to avoid the health impact of air pollution. Atmospheric pollution can be classified into three types based on the sources mobile, stationery and area sources.

Mobile sources are due to the motor vehicles, airplanes, locomotives and other engines and equipment that are able to move to different locations. Stationary sources include foundries, fossil fuel burning, food processing plants, power plants, refineries and other industrial sources. Area sources is caused by certain local actions. Air pollution can be caused due to the pollutants which are emitted directly from a source or which are not directly emitted as such. It can result in the degradation of ambient air quality in the industrial cities. Also daily exposure of people to air pollution results in diseases like asthma, wheezing, and bronchitis. Air pollutants such as Sulphur dioxide (SO2), nitrogen oxide (NOx), nitric oxide (NO), nitrogen dioxide (NO2), carbon monoxide (CO), Ozone (O3), respirable suspended particulates (RSPs) are some of the major airborne pollutants which exerts impact on physical and biological environment.

Air quality monitoring data are used to check the concentration with the ambient air quality standards provided by the government. The purpose of prediction is to develop effective emission control strategies and also helps to find the contribution of each source causing pollution. There are two types of prediction methods, deterministic and stochastic. In this work, deterministic method is used for the prediction. This methods works on the basis of physical and chemical transportation process of pollutants with the influences of meteorological variables, by mathematical models. Artificial neural networks help to

forecast the pollutants in complicated non-linear functions. The accuracy of prediction by artificial neural networks is higher than other methods. The learning process of ANN is similar to animal brain and it can process nonlinear and complex data. It can learn and identify correlated patterns for input data sets to corresponding target values. After training, ANN is used to predict the output of new independent input data. In this research, feed-forward back propagation neural network model is used for prediction of air quality where data collected for the last five years is prediction. This research is done due to the lack of awareness about the real time air quality status among the society. The prediction model by ANN is done by MATLAB software. The objective is to collect the PM2.5 and meteorological data that play a major role in ambient air pollution and to predict the concentration of PM2.5 by ANN.



# Study Area

Chennai, the capital of Tamil Nadu in India is located on the Coromandel Coast off the Bay of Bengal. It is the economic and educational centre of south India. Chennai lies on the south–eastern coast of India.

# Data set

The first and foremost step in modelling is to collect and group the relevant data, both past data and data from air quality monitoring. The data is collected from the website of Central Pollution Control Board. It is very important that the required data and the factors that cause pollution are collected. The daily 24-hour average data for five years (2015-2019) is collected for the following parameters; wind speed, relative humidity, wind direction, temperature, Sulphur dioxide, oxides of nitrogen, PM2.5. The five year mean of above parameters.

# Modelling

The second step to the modelling process is the implementation of the modelling software, this research uses Artificial Neural Network to determine the input and output of the model. Preprocessing of data, removing errors in data and dividing for training, validation and evaluation has to be done to get better results. After this data is ready to be implemented in ANN. In this research the methodology implies the Feed Forward Backpropagation neural network with three layers (input, hidden, and output). The network has input, target and output files. The input layer has the pollutant and meteorological data which is multiplied by coefficient of weights that is obtained by training process. And this meteorological data should have influence on output data.

# Evaluation

The next step is to evaluate the model, with the response of training or validation process. The earlier prepared set of the input and output data are used to evaluate the model and the data response is compared with modelled and measured. The evaluation and validation is similar in their process, the difference lies in the number of numerical quality measures used.

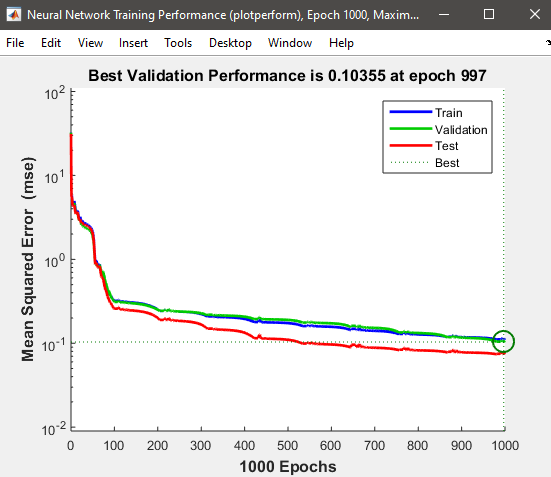
In general,

* Mean Squared Error (MSE),
* Root Mean Squared Error (RMSE),
* Mean Absolute Error (MAE),
* Mean Squared Relative Error (MSRE),
* Coefficient of Determination (R2),
* Index of Agreement,
* Percentage to BIAS (PBIAS),
* Root Mean Squared Error to Standard Deviation (RSR) are recommended.

|  |  |
| --- | --- |
| **Variables** | **2015 to 2019** |
| PM2.5 (μg/m3) | 44.63 |
| SO2 (μg/m3) | 14.46 |
| NOx (μg/m3) | 10.25 |
| Temperature (˚C) | 29.04 |
| Relative humidity (%) | 66.51 |
| Wind speed (m/s) | 0.96 |
| Wind direction (degree) | 194.23 |

1. **Performance of Tansig Function**

The following figure 5 appears during the training process. This graph shows the performance of network versus the number of epochs. During training the performance of the network starts from a large value at first and the weights are altered to have minimum epoch value in the function. In the graph, the black dashed line represents the best performance validation of the network. The green line represents the validation training set, when it intersects with the black line the training process stops. The network performance function is shown in the figure 5. The best performance is achieved by the model using tansig transfer function with the minimum MSE of 0.103.



1. **Performance of Purelin Transfer Function**

The following figure 7 shows that the best performance is achieved by the model using purelin transfer function with the minimum MSE of 0.094.

