International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 09, September-2016

Prediction of Sulpherdioxide Pollution Component in Delhi Metropolitan City using Multi Layer Perceptron

Monika Singh [1]Dr, S.P Mahapatra [1] Dr Sudhir Nigam[2] Mridu Sahu [1]
NIT RIAPUR [1]
LNCTS BHOPAL [2]

***Abstract* – With increase in industrialization and immense emission from vehicles due to huge transportation and much more, India is facing serious problems due to rising air pollution. Among all the pollutants released, SOx is one of the dangerous ingredients of polluted air. Apart from various health problems such as heart failure and breathing difficulties, Sox causes environmental hazards such as damage and destruction of water courses, vegetation, construction material and deterioration in soil quality. AQI is an indicator of air quality which depicts how polluted the breathing air is and forecasts how polluted it is to become. AQI helps government agencies to take preventive measures and steps to control the pollution level. Artificial neural network (ANN) is a computational model based on the structure and functions of the biological neuron and can be used for prediction after proper training is given to the network**

Keywords— Pollution, AQI, Artifical Neural Network

I. INTRODUCTION

India is a developing country and with this era of industrialization most cities are suffering from serious air quality problems.[14] Along with industrialization, rapid growth in population is also a major cause for increasing air pollution[15]Air pollutants, such as carbon monoxide (CO), sulfur dioxide (SO_2), nitrogen oxides (NOx), volatile organic compounds (VOCs), ozone (O_3), heavy metals, and respirable particulate matter (PM2.5 and PM10), differ in their chemical composition, reaction properties, emission, time of disintegration and ability to diffuse in long or short distances. [16]. Air pollution is caused from both human as well as natural actions. Naturally, air pollution may be caused due to forest fires, wind erosion, volcanic eruptions, evaporation of organic compounds and natural radioactivity. Pollution from these natural actions is not very often. Human activities that lead to air pollution are heavy transportation, fuel combustion in burning of fossil fuels like coal, wood, dry grass, and construction activity. Fuel combustion and burning of fossil fuels emits Oxides of Sulphur (mainly, SO_2). Natural sulfur compound emissions are SO4 aerosols produced in sea spray and H_2S from the decomposition of organic matter in swamp areas, bogs, and tidal flats [17]Motor vehicles emits high amount of Carbon Monoxide (CO) and Hydrocarbons (HC) and Nitrogen Oxides (NO). Activities involving construction, bad roads and burning of fossil fuels cause Dust (particulate matter) Pollution. Residential and Commercial activities also contribute to Air Pollution.

The main attention in this paper is SO_2 which is predicted using artificial neural network. The two main sources of SO_2 emission are fossil fuel combustion at power plants (73%) and other industrial facilities (20%). Smaller sources of SO_2 emissions include industrial processes such as extracting metal from ore, and the burning of fuels containing large amount of sulphur by locomotives, large-sized ships, and non-road equipment. Sulfur dioxide (SO_2), a colorless, reactive gas, pungent and irritating gas,.SO_2 is recognized as a pollutant because of its role in forming cold time smog [13]SO_2 can affect adverse effects on human health especially in children .it effects respiratory system. its other effects are asthma, bronco Cho constriction, wheezing ,chest tightness ,shortness in breath ,make worse condition for person in existing heart and lung disease, alter the lungs defense mechanism and cardiovascular disease ,involuntary counting reflex, eye irritation, nausea, vomiting, abdominal pain, sore throat and often pneumonia and constriction in nose, nerve ,throat and lungs,(frank ,1064:wolfet.all ,1975) more serious problem is faced by asthmatic patients, At very high levels, sulfur dioxide may cause wheezing chest tightness, and shortness of breath in people who do not have asthma also. It shows negative effects on plants and environmental also [13]. Sulphates particles cause acid rain .The water of acid rain makes water acidic and changes the makeup of water and soil and damages forests and crops and make water unsuitable for fishes. Continued exposure disturbs balance of ecosystem the gas is absorbed into the mesophyll of the leaves through the stomata. Toxicity is due largely to the reducing properties of the gas. [19].SO_2 causes damage to buildings and the arch logical monuments and paints etc.

II. POLLUTION IN DELHI

A. Delhi has world's dirtiest air

Among world's top 20 polluted cities, 13 belong to India.[18] The PM counts of these cities are shown in figure 1. A study from World Health Organization (WHO) of 1,600 cities revealed that air pollution had worsened significantly and that Delhi which is spread over 1484 square kilometers officially has the world's dirtiest air, by having 153 mg m^{-3} as its PM2.5 count [6]. The following shows the PM2.5 count of top 20 polluted cities of the world: -

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 09, September-2016

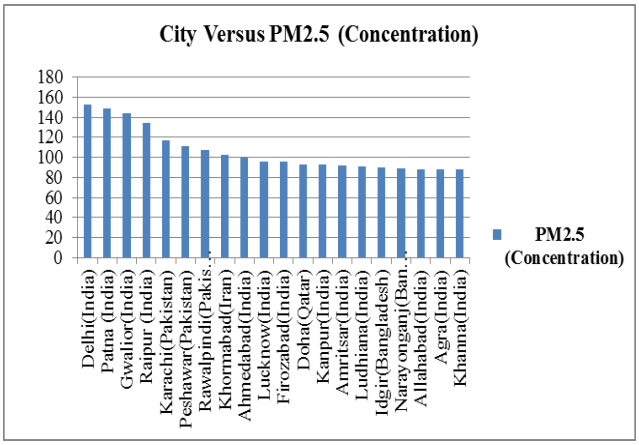

Fig. **[1]**PM counts of top 20 polluted cities of world

The capital of India faces heavy pollution due to large population, high vehicular exhaust, industrial emissions and burning crops in adjoining states. To reduce to level of pollution Delhi government has undertaken many policies for example odd even car driving, car pools are taking in existence, etc. Different programs have been adopted to spread awareness among people to reduce outdoor and indoor pollution. Old vehicles running on diesels are also in eye to reduce pollution by them. Many air quality management programs are also run by government not only in Delhi but by many other countries, in which air quality monitoring is also an important part [8] [9]. It helps in forecasting of pollutants which aides government in alarming people of adverse effects of specific pollutant and take necessary action to prevent them.

B. Reasons for pollution in Delhi

Based on air quality data of Department of Environment, Delhi, and air pollution in Delhi's National Capital Region (NCR) is complex mix of pollution from both human activities (vehicle emissions, construction, industry and residential fuel burning) as well as natural sources like sea salt and dust [1]. The heavy concentration of particulate matter (PM) is greatly affected by meteorological conditions in the winter; cool air causes inversions that stagnant the air which traps pollution close to the ground [3]. Air flowing from Afghanistan and Pakistan pick up emissions as they move over the densely urbanized regions of Punjab and Haryana where farmers burn the straw in their fields and pull this pollution into Delhi. Pre-monsoon dust storms also contribute to air pollution in this region. [6]

II.AIR QUALTIY INDEX(AQI)

Air quality in cities is a result of complex interaction between natural and anthropogenic environmental conditions [7]. The AQI is an index for reporting daily air quality. It tells you how clean or polluted your air is, and what necessary related health precautions should be taken.
Air quality index is a numeric quantity for measuring air quality with respect to its effect on human health [6]. EPA(Environmental Protection Agency)(1999) proposed a method for calculating AQI taking five pollutants under consideration, the five pollutants namely, particulate matter(PM), SO_2, NO_2, CO and O_3 [8]. EPA uses the AQI for five major air pollutants regulated by the Clean Air Act: ground-level ozone, particulate matter, carbon monoxide, sulfur dioxide, and nitrogen dioxide. For each of these pollutants, EPA has established national air quality standards to protect against harmful health effects. Air quality index AQI is a quantity that is measured from 0 to 500 [1]. The higher the AQI value, the greater the level of air pollution and the greater the health danger. For instance consider an AQI value of 50 represents good air quality and little potential to affect public health, while an AQI value over 300 represents hazardous air quality. An AQI value of 100 generally corresponds to the national air quality standard for the pollutant, which is the level EPA, has set to protect public health. So, AQI values below 100 are generally satisfactory. When AQI values are above 100, air quality is considered to be unhealthy for sensitive groups of people to a certain limit and then for everyone as AQI value gets higher [8].

Table [1]: AQI of EPA for SO_2

Air-quality level	AQI	SO_2
Good	0 – 50	0 – 90
Moderate	51 – 100	91 – 383
Unhealthy for sensitive groups	101 – 150	384 – 596
Unhealthy	151 – 200	597 – 809
Very unhealthy	201 – 300	810 – 1607
Hazardous	301 – 500	1608 - 2973

Maximum Operator Function method was used for calculating the AQI. Suggested by Tiwari and Ali(1987) and followed by Kuashik et. all.[20] The AQI value for each pollutant is calculated by the formula given. This method has been used by USEPA and also by CPCB (Central Pollution Control Board) for AQI estimation. CPCB uses exceedance factor in this formula where a factor of 100 as multiple is not used. According to this method,

$$AQI = \frac{pollutant * 100}{pollutantstandardconcentra} \qquad (1)$$

There is a direct need for proper action and control strategy for the effects of air quality. This scheme predicts the next day's AQI level and helps in giving warning as per the situations. Air quality warning systems is needed to notify that the ambient air quality standard may exceed the given standard limit and proper actions should be taken for the same in time.

Urban air quality management and proper information and alarming is necessary to aware the people after prediction of pollutant concentration of forthcoming days and for providing proper control management and air quality warning systems which are required in order to get accurate advance notice that the ambient air concentration levels might exceed the air quality guideline or the limit value. These alerts are required by the health care

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 09, September-2016

department as well as by traffic and environmental management system to minimize the adverse effect of particular pollutant [1]. These warning systems should be easily understandable, reliable and accepted by the society .Several deterministic models exist to evaluate and predict the pollutant dispersion in urban areas and the Gaussian dispersion models are generally used in most of the air pollution studies for predicting the concentration of air pollutants [18.]

III.DATA USED FOR EXPERIMENT

CPCB has an automatic monitoring station in ITO interaction in New Delhi. At this station respirable suspended particles, carbon monoxide, ozone, sulphur di-oxide, nitrogen dioxide and suspended particulate matter are being monitored and information is weekly updated. The data used for this experiment is collected from CPCB's website.

Table [2]: Level of Health Concern

Index Values	Levels of Health Concern	Health Concern Statements
0-50	Good	None
51 - 100*	Moderate	None
101 – 150	Unhealthy for Sensitive Groups	People with asthma should consider limiting Outdoor exertion.
151 – 200	Unhealthy	Children, asthmatics ,and people with heart or lung disease should limit outdoor exertion
201 – 300	Very Unhealthy	Children, asthmatics, and people with heart or lung disease should avoid outdoor exertion; everyone else should limit outdoor exertion
301 - 500	Hazardous	Children, asthmatics, and people with heart or lung disease should remain indoors; everyone else should avoid outdoor exertion

IV.PREDICTION OF SO_2 THROUGH MLP

A. *ARTIFICIAL NEURAL NETWORK (ANN)*

The neural network models are based on the working of human brain and artificial model of the brain is known as artificial neural network. An artificial neural network (ANN) is a computational model based on the structure and functions of the biological neural networks [2]. Information that flows through the neural network affects the structure of the ANN because a neural network changes or learns in a sense based on that input and output.ANN is considered nonlinear statically data modeling tools where the complex relationships between input and outputs are modeled or patterns are found. ANN has many advantages and is used in various fields. It can actually learn by observing data.

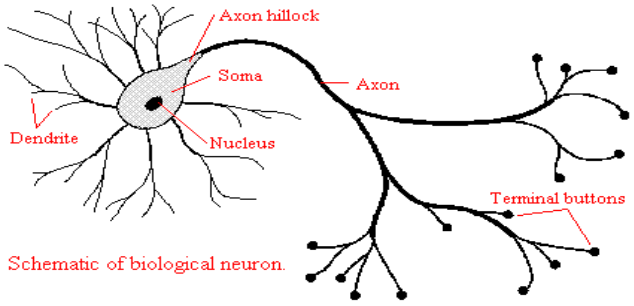

Fig. [2] .Biological neurons

http://vv.carleton.ca/~neil/neural/neuron1.gif[12]

B. *MULTI-LAYER PERCEPTRON (MLP)*

Multi-layer perceptron is a feed forward artificial neural network model that maps sets of input data on to a set of appropriate outputs [2]. A MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one

MLP is a special class of neural networks which consists of an input layer, an output layer and one or more hidden layers of computation modes. A simple MLP with one hidden layer is shown in figure 2. The MLPs are used with supervised learning which may be defined as the process of providing the network with a series of sample inputs and comparing the outputs with the expected responses. Until the network provides expected response its training continues. Back propagation (BP) method is used for training MLPs. BP is a multilayer forward network using extend gradient-descent based delta learning rule. Back propagation provides a computationally efficient method for changing the weights in a feed forward network, with differentiable activation function units, to learn a training set of input-output examples.

The MLP shown in figure 2 consists of 3 layers: 1 input, 1 hidden and 1 output layer containing n, p and m number of neurons respectively. This network also consists of 2 biased neurons whose weights are always 1. Few weights are shown in the figure and the rest are named in similar manner.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 09, September-2016

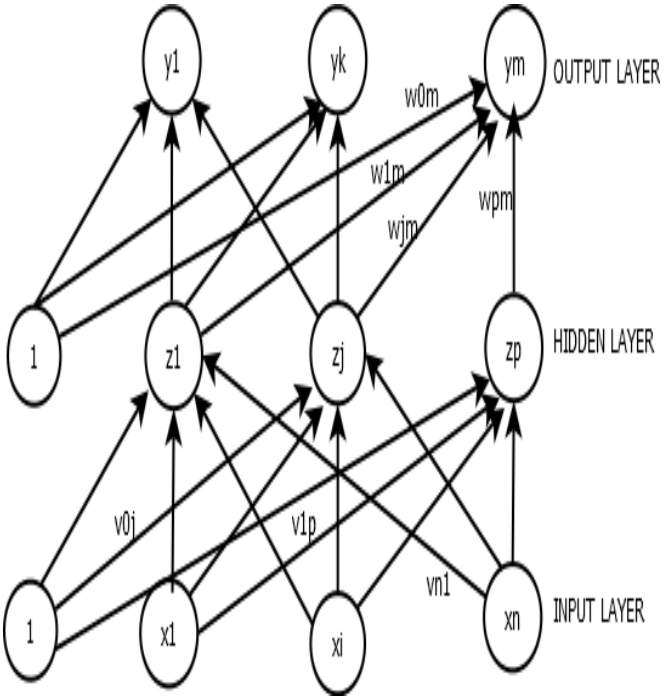

Fig. [3]: MLP with One Hidden Layer

C. *PREDICTING SO_2 USING MLP*

This paper describes the development of an Artificial Neural Network (ANN) model for SO_2 prediction in Delhi. It is a multilayer perceptron (MLP) network, with meteorological forecasting data as the main input, to output the next day SO_2 value. The training of MLP is based on BP error correction, which uses gradient descent optimization for error reduction. The training in BP can be done in two ways, either sequential training or batch training. In sequential training a given input is propagated forward, the error is determined and back propagated and the weights are updated. In batch training the weights are updated only after the entire set of training network has been presented to the network. Thus the weight update is only performed after each iteration. In this experiment sequential training is given to the dataset.

The dataset has been accessed through Central Pollution Control Board website. The daily concentration for 365 days of six major pollutants namely; CO, NO_x, So_x, $PM_{2.5}$, ozone and dust has been provided.
The sample dataset is shown in table 2. The table consists of counts of CO, So_x, NO_x, $PM_{2.5}$, Ozone, and dust. Class label of first 10 days of the dataset taken for the experiment. NA (not applicable) is specified to the cell for which data is not available. At the time of calculation this NA value is replaced by the average of readings of one day prior to that day and one day posterior to that day.

Table [3]: 10 Days Readings of Pollutants

Co	SOx	NOx	Pm2.5	Ozone	Dust	Class
3	1	4	4	1	3	1
3	1	4	5	1	3	1
2	1	3	5	1	2	1
2	1	3	4	1	3	1
2	1	4	5	1	3	1
2	1	3	4	1	3	1
2	1	3	7	1	6	1
6	1	7	NA	1	7	1
6	1	7	7	1	7	1
4	1	6	6	1	4	1

Table [4]: Summary Function over Entire Dataset

Pollutant	Minimum	Maximum	Mean	Median
CO	1.000	6.000	2.836	3.000
NO_x	1.000	7.000	3.542	3.000
$PM_{2.5}$	0.000	7.000	3.493	3.000
OZONE	1.000	3.000	1.159	1.000
DUST	1.000	7.000	2.718	2.000

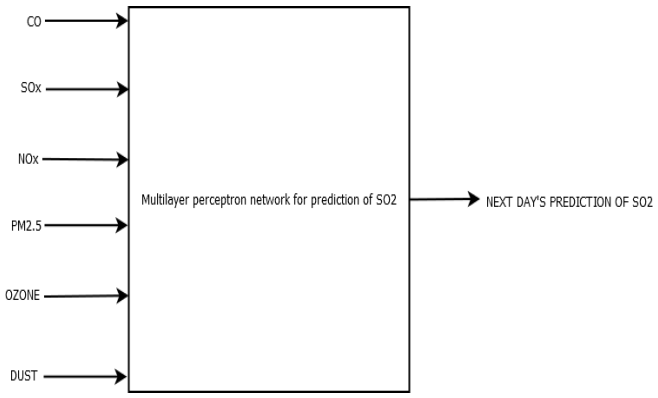

Fig. [4]: Multilayer perceptron network for SO prediction

Figure 3,4 and 5 shows the trend of CO, NOx and SOx for 365 days as per the dataset provided

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 09, September-2016

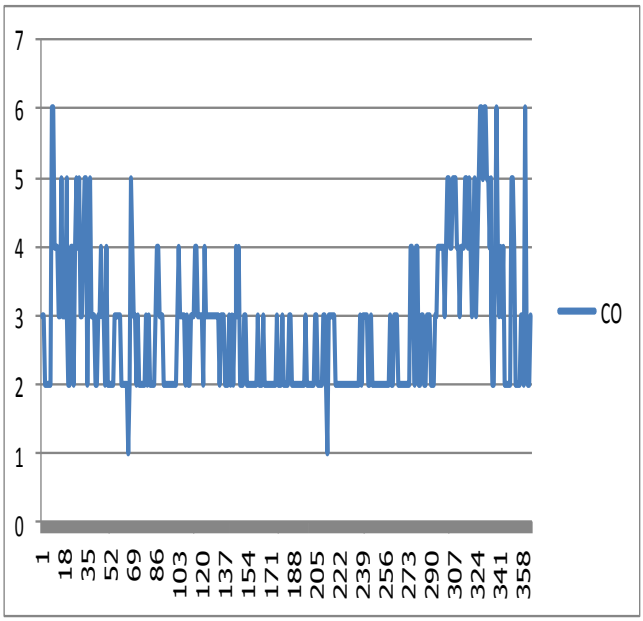

Fig. [5]: Daily Concentration of CO for 365 days

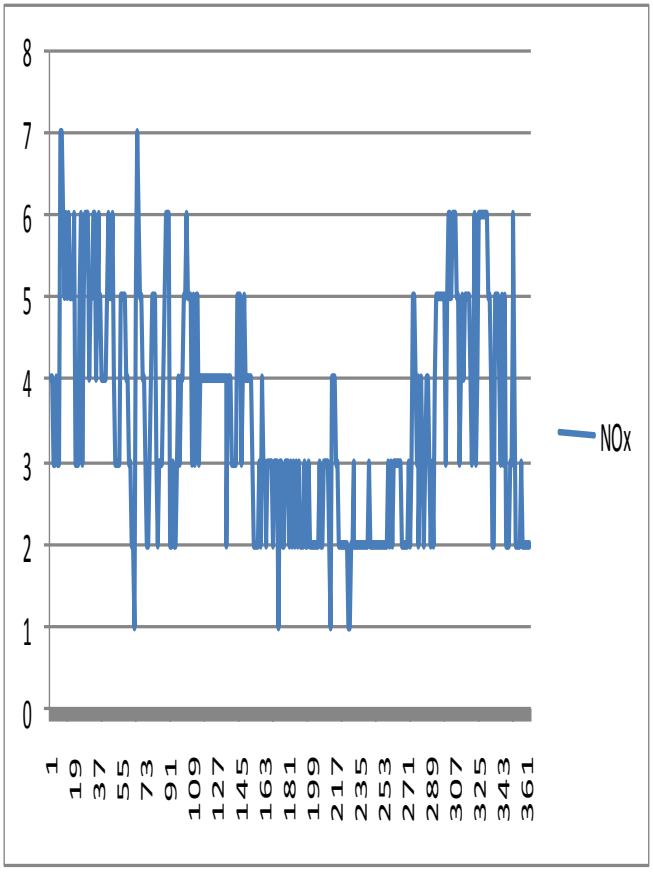

Fig.[6]: Daily Concentration of NOx for 365 days

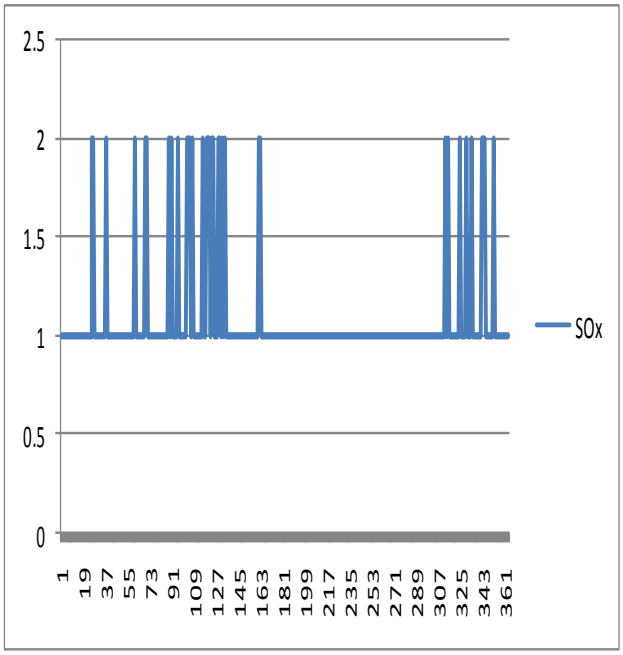

Fig..[7]: Daily Concentration of SOx for 365 Days

The SO_2 level for 366th day is predicted using this dataset. Among 365 readings, training is given to the data of days 1-300 and testing for the same is done on data from 301-320. The neural network presenting the above experiment is shown in figure 6.

The entire experiment is performed using open source software R. The summary function of R is used to give the minimum, maximum, mean and median of each pollutant mentioned in the dataset.

Table [5]: Statistical Parameters for Pollutant

Component	Minimum	Maximum	Mean	Median
CO	1.000	6.000	2.836	3.000
NOx	1.000	7.000	3.542	3.000
PM2.5	0.000	7.000	3.493	3.000
OZONE	1.000	3.000	1.159	1.000
DUST	1.000	7.000	2.718	2.000

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 09, September-2016

After performing the summary function, training is provided to dataset for values from 1-300 and testing is done for 301-320 values. The Sox value for days 301-320 is predicted and accuracy and kappa value is calculated for the same.

VI. RESULTS AND DISCUSSION

Table [6]: Example Pollution Data Set

S. NO	CO	SOX	NOX	PM2.5	OZONE	DUST	CLASS LABEL
1	3	1	4	4	1	3	1
2	3	1	4	5	1	4	1
3	2	1	3	5	1	2	1
4	2	1	3	4	1	3	1
5	2	1	4	5	1	3	1
6	2	1	3	4	1	3	1

The result generated from proposed methodology is shown in Table [7], this says predicted and actual value of Sox using multilayer perceptron, and the prediction accuracy is 90% obtained. The classification task also completed using function based classifier. This classifier performing well as compare to tree based or other classifier present in machine learning environment. Classification accuracy 90.411 obtained and the measures like True Positive (TP), True Negative (TP),False Positive (FP), and False Negative (FP) shown in Table [9] and Table [10].

Table [7]: Predicted and Actual Sox Value

S.No	Predicted Data of SO_x	ActualData of SO_x
1	1	1
2	1	1
3	1	1
4	1	1
5	1	1
6	1	1
7	1	1
8	1	1
9	1	1
10	1	1
11	1	1
12	1	1
13	1	1
14	1	2
15	1	1
16	1	2
17	1	1
18	1	1
19	1	1
20	1	1

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	PRC Area	Class
0.947	0.857	0.912	0.947	0.929	0.750	0.963	1
0.143	0.053	0.222	0.143	0.174	0.750	0.232	2
0.870	0.780	0.846	0.870	0.857	o.750	0.893	

Table 8: Accuracy Measures for SO_2

	A	B	Class Label
5	12	7	A=1
12		2	B=2

Table 9: Confusion Matrix for SO_2

A	B	Class
TP	FN	A=1
FP	TN	B=2

Table 10: Accuracy Measures

VI. CONCLUSION

Present work is first study of prediction of SO_2 in environmental Delhi metropolitan data set. The presented data set is CPCB data set. SO2 concentration is a main pollutant among several pollutant components. This study is based on multilayer perceptron, because the nature of data in non linear so many linear algorithms have not given better prediction than multilayer perceptron. Classifier accuracy also measured and article found function based classifier is well performer for present work. In future this work will apply to other pollutant and find out better classifier for the same

ACKNOWLEDGMENT

R has few packages for creating neural network models. Some of them are neuralnet, nnet, RSNNS etc. In this experiment nnet package has been used which was created by Brian Ripley. This package helps to develop and validate multilayer perceptron. The nnet package helps in providing training, followed by validation and then satisfactory prediction.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 09, September-2016

REFERENCES

[1] Mohan, M., &Kandya, A. (2007). An analysis of the annual and seasonal trends of air quality index of Delhi. Environmental monitoring and assessment, 131(1-3), 267-277.

[2] Kurt, A., Gulbagci, B., Karaca, F., &Alagha, O. (2008). An online air pollution forecasting system using neural networks. Environment International, 34(5), 592-598.

[3] Cogliani, E. (2001). Air pollution forecast in cities by an air pollution index highly correlated with meteorological variables. Atmospheric Environment,35 (16), 2871-2877.

[4] Jiang, D., Zhang, Y., Hu, X., Zeng, Y., Tan, J., & Shao, D. (2004). Progress in developing an ANN model for air pollution index forecast. *Atmospheric Environment*, *38*(40), 7055-7064.

[5] Cacciola, M., Pellicanò, D., Megali, G., Lay-Ekuakille, A., Versaci, M., &Morabito, F. C. (2013, June). Aspects about air pollution prediction on urban environment. In 4th Imeko TC19 symposium on environmental instrumentation and measurements protecting environment, climate changes and pollution control (pp. 15-20).

[6] Rizwan, S. A., Nongkynrih, B., & Gupta, S. K. (2013). Air pollution in Delhi: its magnitude and effects on health. Indian Journal of Community Medicine,38(1), 4.

[7] Mayer, H. (1999). Air pollution in cities. Atmospheric environment, 33(24), 4029-4037.

[8] Murena, F. (2004). Measuring air quality over large urban areas: development and application of an air pollution index at the urban area of Naples. Atmospheric Environment, 38(36), 6195-6202.

[9] Cheng, W. L., Kuo, Y. C., Lin, P. L., Chang, K. H., Chen, Y. S., Lin, T. M., & Huang, R. (2004). Revised air quality index derived from an entropy function. Atmospheric Environment, 38(3), 383-391.

[10] Cheng, W. L., Chen, Y. S., Zhang, J., Lyons, T. J., Pai, J. L., & Chang, S. H. (2007). Comparison of the revised air quality index with the PSI and AQI indices. Science of the Total Environment, 382(2), 191-198.

[11] Kang, D., Mathur, R., &Rao, S. T. (2010). Real-time bias-adjusted O3 and PM 2.5 air quality index forecasts and their performance evaluations over the continental United States. Atmospheric Environment, 44(18), 2203-2212.

[12] http://vv.carleton.ca/~neil/neural/neuron1.gif

[13] Ayodele, J. T., &Abubakar, F. (2010). Sulphur dioxide as indoor pollutant in Kano Municipality Kano-Nigeria. Journal of environmental chemistry and ecotoxicology, 2(1), 9-13.

[14] McMichael, A. J. (2000). The urban environment and health in a world of increasing globalization: issues for developing countries. *Bulletin of the World Health Organization*, *78*(9), 1117-1126.

[15] Dursun, S., Kunt, F., &Taylan, O. (2015). Modellingsulphur dioxide levels of Konya city using artificial intelligent related to ozone, nitrogen dioxide and meteorological factors. *International Journal of Environmental Science and Technology*, *12*(12), 3915-3928.

[16] Kampa, M., &Castanas, E. (2008). Human health effects of air pollution.*Environmental pollution*, *151*(2), 362-367.

[17] Robinson, E., & Robbins, R. C. (1970). Gaseous sulfur pollutants from urban and natural sources. *Journal of the Air Pollution Control Association*, *20*(4), 233-235.

[18] Goyal, P. (2003). Present scenario of air quality in Delhi: a case study of CNG implementation.*Atmosphericenvironment*, *37*(38), 5423-5431.

[19] Thomas, M. D. (1961). Effects of air pollution on plants. Air pollution, 239

[20] Arman, S. C., Kumar, N., Singh, R., Kisku, G. C., Khan, A. H., Kidwai, M. M., ... & Jain, G. (2010). Assessment of urban air pollution and it's probable health impact. Journal of Environmental Biology, 31(6), 913-920