DATA ANALYSIS & PREDICTION OF AIR QUALITY PARAMETERS IN BANGALORE CITY

Kavyashree N V[1], Rekha H.B.[2]

[1]PG Student, Deparatment of Civil Engineering, UVCE, Bangalore University, Bengaluru, Karnataka, India.
[2]Associate Professor, Deparatment of Civil Engineering, UVCE, Bangalore University, Bengaluru, Karnataka, India.

---***---

Abstract: Air quality has advanced as a major factor affecting the quality of living in urban areas, specifically in densely populated and industrialized areas. Air pollution control is required in order to avoid deteriorating of air quality in the long run. At the same time, short-term forecasting of air quality is crucial in order to take precautionary and clear action during incidents of atmospheric air pollution. During the previous few years,, the usage of artificial neural networks (ANNs) has improved in many areas of engineering. Artificial neural network have been useful to many environmental engineering problems and have established some degree of success. The aim of this research is to create a neural network-based air quality prediction model. The Multilayer Perceptron is used to create a prediction method in this analysis. Several variables, including Lead, NH_3, PM_{10}, $PM_{2.5,}$ NO_2, $S0_2$ temperature, relative humidity, rainfall and wind velocity are considered in this study. The performance of the developed model was evaluated through a measure of Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Relative Error (RE). Data obtained from best prediction model are used for Forecasting Air Quality Parameters for Future and assessed through value of R^2. Air Quality Index is calculated for the Forecasted data and verified which category they fall into. Suitable remedial measures are suggested.

Key words: Artificial Neural network, Prediction, Neurons, errors.

1. Introduction

The atmosphere of the Earth is made up of air. The air we breathe is made up of a number of gases and dust particles. The clear gas used in vehicles and industries is polluting all of the air that living beings breathe at an unprecedented pace. One of the most significant natural resources for the creation and survival of all life on this planet is air. Plants and animals alike depend on air for their survival. As a result, all living organisms require good quality air that is free of harmful gases in order to survive. It is the ever-increasing population. Long-term and short-term health effects may be caused by air pollution. It has been discovered that air pollution has a greater impact on the elderly and young children. (Gaganjot Kaur Kang et,al 2018).

Air pollution is characterised as the presence of toxic chemicals or compounds in the air (including those of biological origin) at levels that are hazardous to human health. In a wider context, air pollution refers to the presence of pollutants or substances in the air that are not normally present and that degrade the quality of the air or have a negative impact on human health(Gaganjot Kaur Kang et,al 2018).

With the continued development and growth of the urban population, a number of environmental concerns, such as deforestation, the release of hazardous materials, solid waste disposal, air pollution, and others, have gained much more attention than ever before. The issue of urban air pollution has grown to the point that timely information about changes in pollution levels is needed. Because of the health effects caused by these airborne pollutants in urban areas during pollution episodes, air quality forecasting is now one of the main topics of air pollution research. As a result, it is critical to establish successful forecasting models for major air pollutants in urban areas. Air quality monitoring and forecasting tools are used to take preventative steps against air pollution, such as mitigating the effects of an expected air pollution peak on the local population and environment. In this research, a single Artificial Neural Network (ANN) is used to forecast future air quality (Enzo Grossi et,al 2009).

ANNs are artificial adaptive mechanisms that are influenced by the human brain's working processes. They are structures that can alter their internal structure in response to a function target. The nodes, also known as processing elements (PE), and the links form the foundation of the ANN. Each node has its own input, which it uses to receive communications from other nodes and/or the world, as well as an output, which it uses to communicate with other nodes or the environment. Finally, every node has a function that converts its own global input into output. (Enzo Grossi et,al 2009).

Moustris K.P et al (2012) research was to look at the one-day forecast of the daily maximum surface ozone concentration in the Greater Athens Area by using predictive models based on multiple regression analysis and artificial neural networks. According to the results of the study, the coefficient of determination (MLR: 0.653, ANN: 0.666) and index of agreement (MLR: 0.887, ANN: 0.892) between the observed and predicted ozone concentrations for the year 2005 are statistically significant (P <0.01Finally, the capacity of ANN models to forecast presents a small precedent as compared to MLR models. Furthermore, based on the availability of those data over space and time, the built ANN model's predictive ability may be enhanced by the use of many other parameters such as nitrogen oxides concentration, solar radiation strength, and sunlight length.

Suraya Ghazali1 et al (2012): Artificial neural networks (ANNs) have become increasingly common in many fields of engineering in recent years. Artificial neural networks have been used to solve a number of environmental engineering problems with varying degrees of performance. The aim of this research is to create a neural network-based air quality prediction model. A prediction method based on feed-forward neural networks is developed in this research. This research takes into account a number of variables, including sulphur dioxide (SO_2), carbon monoxide (CO), nitrogen dioxide (NO_2), nitric oxide (NO), temperature, relative humidity, and air velocity. A measure of Mean Square Error (MSE) and the value of R2 were used to evaluate the established model's accuracy. The model with network structure 7-20-4, R^2 value 0.57, and MSE 0.062 had the best prediction performance among the built networks.

Objectives of the study include Collection of data of air pollutant parameters for the study area in Bangalore. Collection of data from India Meteorological Department like wind velocity, Rainfall, Relative Humidity, Temperature, Analysis of Data and achieving best prediction model, Forecasting the air quality data for future years and Calculation of Air Quality Index (AQI).

AQI CALCULATION

The air quality index (AQI) is a number that is used to report the air quality on any given day: it essentially tells you how clean the air is. It monitors the amount of particles and chemicals in the air that are harmful to people's health (and ignores those that do not). Extreme pollution can have serious health consequences in places. These side effects can vary from minor irritations to decreased stamina and even death. (National Air Quality Index(2015))

AQI is calculated using equation 1 as shown below:

$$I_P = [\{\frac{(I_{HI} - I_{LO})}{(B_{HI} - B_{LO})}\} * (C_P - B_{LO})] + I_{LO}$$

……….e

quation 1

where,

B_{HI}= Breakpoint concentration greater or equal to given concentration.

B_{LO}= Breakpoint concentration smaller or equal to given concentration.

I_{HI} =AQI value corresponding to B_{HI}.

I_{LO} = AQI value corresponding to B_{LO}.

Cp = Pollutant concentration. (National Air Quality Index(2015))

2. Data Collection And Methodology

2.1 Data Collection

The air quality data is collected from Karnataka State Pollution Control Board (KSPCB) Bangalore, Karnataka. **Five years of monthly average** has been collected for the Air Quality Parameters like Lead, NH_3, PM_{10}, $PM_{2.5}$, NO_2, SO_2 for the following stations Export promotional Industrial park IPTL White Field road, Central silk board Hosur road, Amco Batteries mysore road. Swan Silk Pvt. Ltd. Peenya from 2014 to 2018. From Indian Meterological Department **Five years of monthly average** data has been collected for the rainfall, temperature, relative humidity, wind velocity for the Bangalore city from 2014 to 2018.

2.2 Software Used

SPSS version 25 was used in this study. SPSS is a software package that includes a number of statistical packages for social science research. (Priyanka Gogoi 2020).

2.3 Methodology

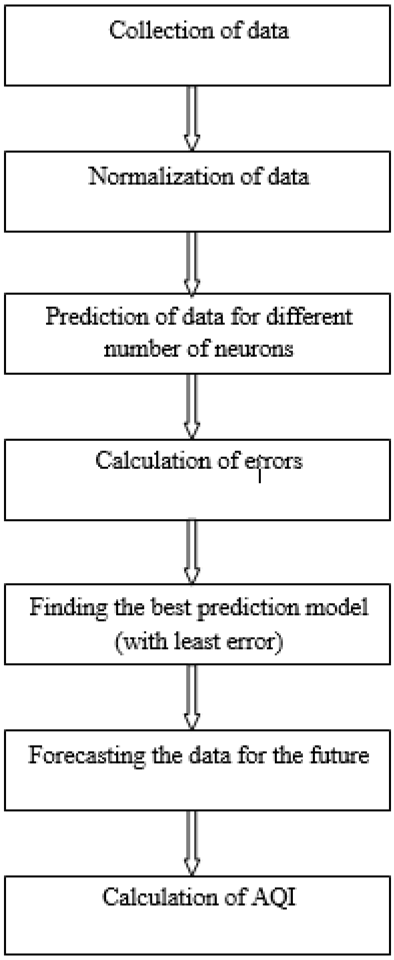

Fig 2.1: Methodology Flow chart

Data Transformation

Normalization of samples: The data model is susceptible to overflows in the network because of irregularities in the values or weights. To remove these irregularities the range transformation method of normalizing all the values in the range of [0, 1] is applied. The range normalization function is as shown in equation 2:

$$X_{normalization} = \frac{(Xi - Xmin)}{(Xmax - Xmin)}$$ ....equation 2

Where, X normalization is the normalized value, Xi is the ith value passed, and Xmin and Xmax are the minimum and maximum value for Xi value.

Prediction

Multilayer Perceptron is used to predict data. MLP is known for its straightforward structure and easy-to-understand operating principles. It has many flaws, including a lengthy training phase, local optimums, and, most importantly, a lack of structure optimization theory. Trial and error is used to calculate the number of hidden layers and the number of neurons in each hidden layer. As a result, MLP optimization is largely problem-specific, and manual experimentation is often needed, limiting the model's applicability. An MLP model with just one hidden layer and one neuron in the hidden layer has the simplest structure. (Dahe Jiang et, al 2013).

Partitions: For developing the ANN model, the data were divided into two sets. 70% of the data for the whole training set, follows with 30% of the whole data for testing. (Azman Azid et, al 2013).

Architecture: For this study automatic architecture is used so that required number of neurons can be assigned. Automatic architecture selection uses the default activation functions for the hidden and output layers.

Number of Hidden Layers: A multilayer perceptron can have one or two hidden layers. For this study one hidden layer is used with the number of hidden neurons ranging from 2,4,5,6,7,8,9,10,12,14,16,24,32. The output with the neuron number which has less error is considered as to be best output.

Training: Batch training is often preferred because it minimizes the total error. It is most useful for "smaller" datasets. Gradient descent is used as optimization algorithm, The intial learning rate and the momentum are chosen as 0.3 and 0.2 respectively for this project.

Calculation of Errors

To evaluate the prediction results, the following error measures were considered: Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Relative Error (RE)as shown in equation 3, 4 and 5.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|P_i - M_i|$$ ....Equation 3

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - M_i)^2}$$ ....Equation 4

$$RE = \frac{1}{n}\sum_{i=1}^{n}\frac{|P_i - M_i|}{M_i}$$ ....Equation 5

where n denotes the number of data points in the test dataset, and Pi and Mi denote the predicted and measured values for the ith hour, respectively. Different layers of Neural Network (NN) models for pollutant concentration were built in this project. For the purpose of evaluating the best performing model, these models were all trained on the same dataset. With the smallest difference between the expected and measured values, the neuron layer with the lowest error rates will have the best prediction rate. Excel is used to measure these errors. With the smallest difference between the expected and measured values, the neuron layer with the lowest error rates will have the best prediction rate. For future forecasting, these expected values with lower errors are used. The neuron layer with the lowest error rate would have a smaller difference between expected and measured values, resulting in the best prediction rate. Future forecasting is based on these predicted values with lower errors.

Forecasting

Forecasting is done using forecast option in SPSS software. Forecasting is done from the year 2019 to 2023.

3. Results and Discussions

This chapter deals with all the results and discussions obtained in the present study. It deals with the prediction and forecasting of air quality parameters in the study area using Multilayer perceptron. The details of results obtained along with their discussions are described below.

3.1 Export promotional Industrial park, IPTL, WF road

Monthly average of air quality parameters like Lead, NH_3, PM_{10}, $PM_{2.5}$, NO_2, SO_2, were collected for the study area. The parameter which has to be predicted is taken as dependent variable, the other variables are taken as independent variables. These dependent and independent variables are taken as input in input layer. Different number of neurons are taken in hidden layers like 2, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 24, 32 for identifying the best performing model. Different number of neuron is assigned for every trail and in total 13 outputs are obtained for one air quality parameter

Error Calculation

The error functions like MSE, MAE, RE were calculated for various hidden neurons and are presented in the below Table 3.1. It can be seen from the table that different hidden number of neurons has different error rates and the neuron layer with the lowest error rates will have the smaller difference between the predicted and the measured value leading to best prediction rate.

	No. of neurons	2	4	5	6	7	8	9	10	12	14	16	24
	RMSE	0.14495	0.17272	0.1767	0.1363	0.1532	0.14286	0.1452	0.15106	**0.1235**	0.11856	0.1509	0.12656
	MAE	0.092	0.0977	0.107	0.0877	0.1000	0.0984	0.0923	0.1042	**0.07798**	0.0781	0.0873	0.08421
Lead	RE	37.89	36.92	43.788	37.112	44.509	45.114	46.461	43.85	**2.4786**	43.008	41.043	43.968
	RMSE	0.1858	0.3131	0.1609	0.13636	0.14479	0.1758	0.1808	0.1463	0.1658	**0.14606**	0.1529	0.1515
NH_3	MAE	0.1351	0.10048	0.1168	0.0985	0.10356	0.1564	0.1363	0.1044	0.124	**0.1061**	0.1121	0.1129
	RE	115.55	59.0616	71.62	58.471	61.366	86.25	114.42	62.7109	76.618	**55.745**	78.322	84.934
	RMSE	0.1533	**0.1235**	0.1454	0.1443	0.1529	0.14800	0.1549	0.0169	0.15587	0.14265	0.1574	0.1433
	MAE	0.1051	**0.1031**	0.098	0.104	0.1073	0.0952	0.1161	0.104	0.1084	0.0942	0.1108	0.0996
$N0_2$	RE	14.256	**0.1235**	11.8631	12.069	13.7	13.252	13.18	12.36	13.85	11.3065	14.082	9.968
	RMSE	0.5382	0.14628	0.1278	0.1144	0.1358	**0.12545**	0.3681	0.1281	0.1139	0.1194	0.1203	0.1329
	MAE	0.0879	0.1149	0.0979	0.088	0.1196	**0.0981**	0.312	0.0944	0.0884	0.0944	0.0921	0.1042
PM2.5	RE	8.6732	11.2505	11.0018	9.0132	11.3875	**0.1152**	16.681	8.63	5.0745	9.01	10.061	11.016
	RMSE	0.13875	0.11312	0.1573	0.1415	0.0794	0.174	0.0937	**0.1954**	0.1541	0.0822	0.1244	0.153
	MAE	0.10348	0.08512	0.1179	0.1011	0.066	0.1267	0.074	**0.1372**	0.11108	0.0618	0.094	0.1191
PM10	RE	53.349	44.31	64.253	47.088	36.656	62.75	41.331	**6.1412**	57.031	33.134	44.99	64.003
	RMSE	0.31041	0.2822	0.2936	0.323	0.306	0.3185	**0.3**	0.27633	0.31582	0.31568	0.303	0.3073
	MAE	0.241	0.2103	0.2377	0.2583	0.2448	0.2673	**0.234**	0.21359	0.25876	0.271	0..2430	0.233
SO_2	RE	17.064	16.799	16.588	17.465	17.186	17.11	**16.491**	17.39	17.566	17.1083	17.2017	16.955

Table 3.1 : Table showing errors rates for different number of neurons for Export promotional Industrial park, IPTL, WF road.

From the table 3.1, it can be seen that different parameters have achieved best result at different number of neurons. Best prediction model for lead is obtained when 12 number of neurons are used in the hidden layer because of the lesser error rate having RMSE, MAE, RE, 0.1235, 0.077986, 2.4786 respectively. NH_3 achieved its best prediction model when used 14 number of hidden neurons having RMSE, MAE, RE, 0.14606, 0.1061, 55.745 respectively likewise NO_2 at 4 number of hidden neurons having error rates RMSE, MAE, RE 0.1235, 0.1031, 0.1235 respectively. $PM_{2.5}$ at 8 number of hidden neurons having error rates RMSE, MAE, RE 0.12545, 0.0981, 0.1152 respectively. PM_{10} at 10 number of hidden layers having RMSE, MAE, RE 0.1954, 0.1372, 6.1412 respectively, and then SO_2 at 9 number of hidden neurons having RMSE, MAE, RE 0.3, 0.234, 16.49 respectively. we can conclude, to yield the best prediction for the Neural Network (NN) This architecture will yield the best result because of its lesser error rate when compared to other hidden neuron layers. The graphs of the Neural Network (NN) with best prediction models that is with the least errors are shown below.

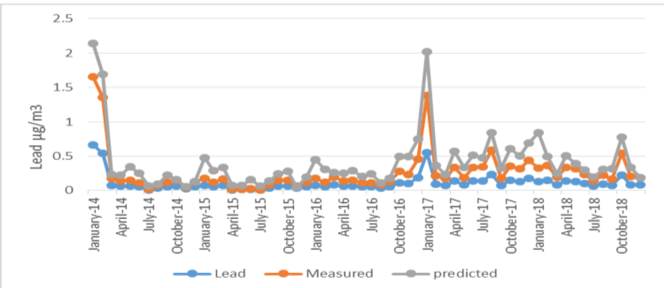

Fig3.1: Graph showing variation between measured and predicted values for Lead.

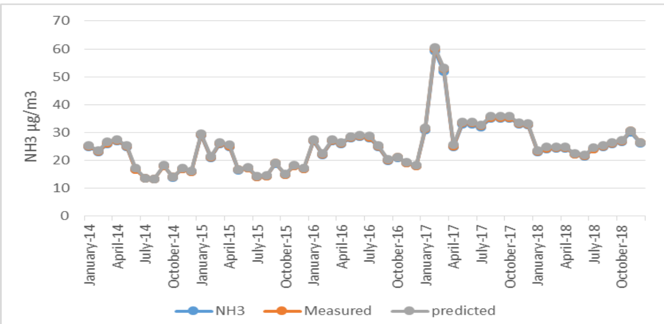

Fig3.2: Graph showing variation between measured and predicted values for NH_3.

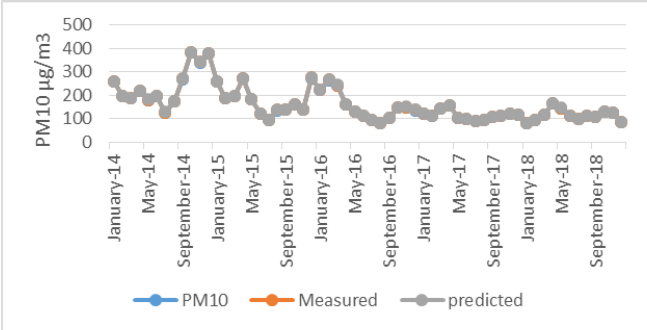

Fig3.3: Graph showing varation between measured and predicted values for PM_{10}.

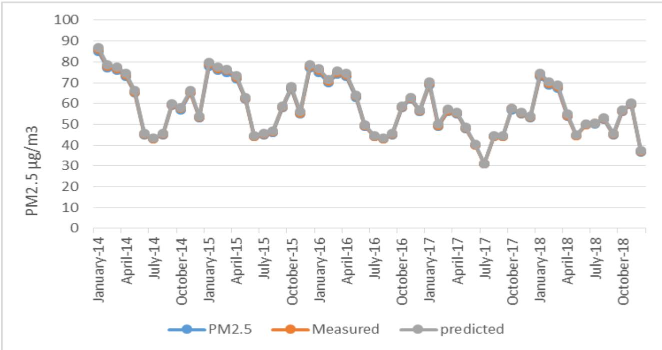

Fig3.4: Graph showing variation between measured and predicted values for $PM_{2.5}$.

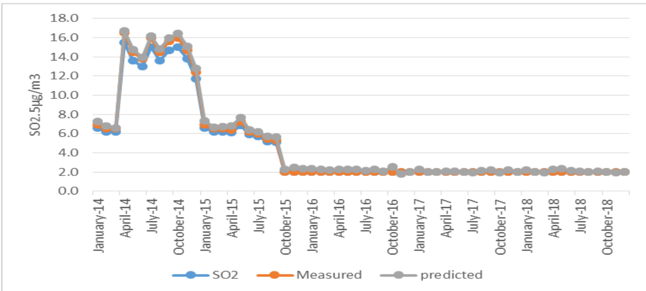

Fig3.5: Graph showing variation between measured and predicted values for SO_2.

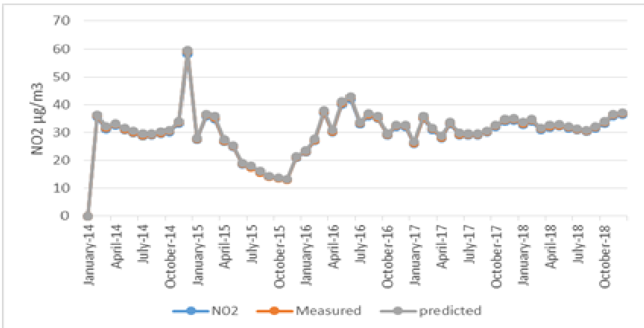

Fig3.6: Graph showing varation between measured and predicted values for NO_2

Forecasting for Export promotional Industrial park, IPTL, WF road

Output obtained from the above mentioned best prediction models are used for the forecasting for the future. Data is collected from 2014 to 2018, the forecasting is done for the next five years that is for 2019 to 2023. The below mentioned graphs from 3.7 to 3.12 shows the forecast of

Export promotional Industrial park, IPTL, White Field road for the lead, NH_3, PM_{10}, $PM_{2.5}$, NO_2, SO_2.

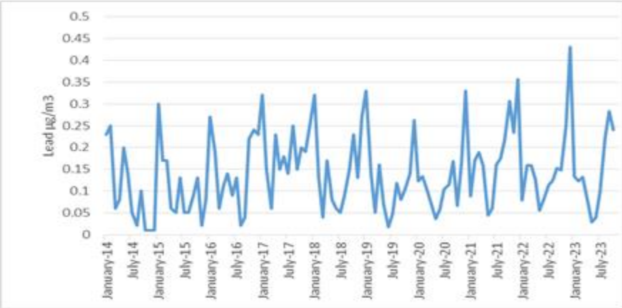

Fig3.7: Graph showing observed and forecasted values for Lead.

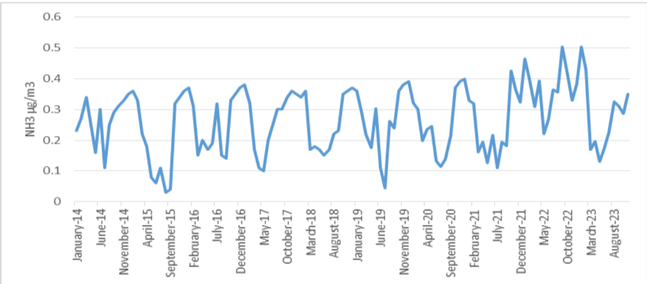

Fig3.8: Graph showing observed and forecasted values forNH_3

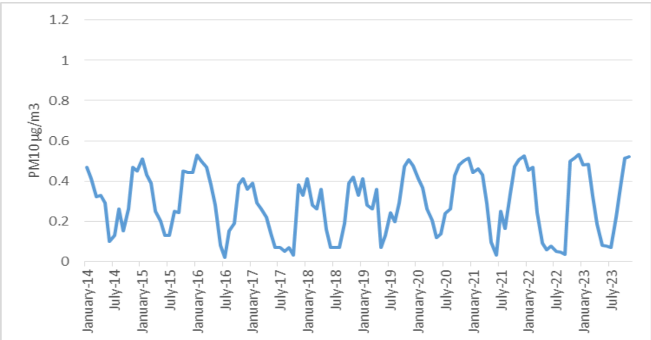

Fig3.9: Graph showing observed and forecasted values for PM_{10}.

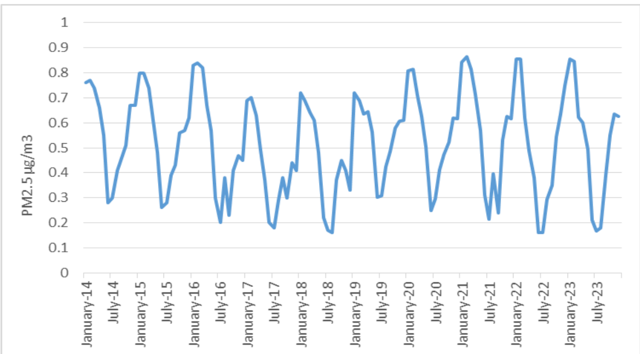

Fig3.10: Graph showing observed and forecasted values for$PM_{2.5}$

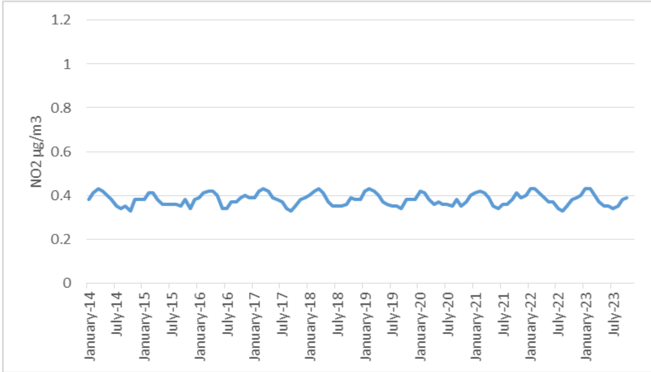

Fig3.11: Graph showing observed and forecasted values for NO_2.

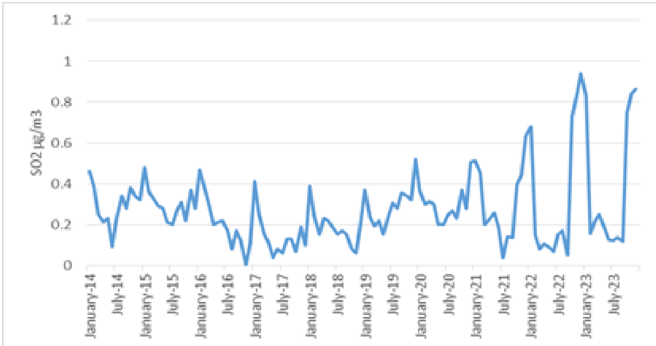

Fig3.12: Graph showing observed and forecasted values for SO_2.

SL. No	parameters	R^2 value
1	Lead	0.814
2	NH_3	0.714
3	PM_{10}	0.955
4	$PM_{2.5}$	0.855
5	NO_2	0.955
6	SO_2	0.953

Table 3.2: Table showing R^2 values for the forecasted data

DISCUSSIONS

1. Figures 3.1 to 3.6 shows the graphs of measured vs predicted for the air quality parameters. The graphs are plotted only for the best prediction model, that is neuron that is giving the less error.

2. From graphs it can observed that the predicted values were very close to the observed pollutant concentrations both in trend and pattern. It was found that all the developed ANN prognostic models perform absolutely well.

3. Figures 3.7 to 3.12 shows the forecast of Lead, NH_3, PM_{10}, $PM_{2.5}$, NO_2, SO_2 from 2019 to 2023. It can be observed that all the parameters are showing

increase in trend. The percentage increase values for Lead, NH_3, PM_{10}, $PM_{2.5}$, NO_2, SO_2 are 10%, 11.5%, 16.33%, 4.3%, 0.3%, 17.78% respectively.

4. The Air quality index values are calculated for forecasted data, The AQI values are 95, 107, 112 for the year 2021, 2022, 2023 respectively. These AQI values are increasing when compared to previous years and fall under the moderate range.

5. The Table 3.2 shows the R^2 values for the forecasted values, The closer the value is to one, the better and more accurate is the prediction.

6. The reason for increase in the concentration for parameters may be increase in the number of vehicles, setting up of industries as it is the second most polluted area in Bangalore next to Silk Board Junction according to Karnataka State Pollution Control Board's survey in 2016.

3.2 Central silk board, Hosur road.

Monthly average of air quality parameters like Lead, NH_3, PM_{10}, NO_2, SO_2, are collected for the study area. The parameter which has to be predicted is taken as dependent variable, the other variables are taken as independent variables. These dependent and independent variables are taken as input in input layer. Different number of neurons are taken in hidden layers like 2, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 24, 32 for identifying the best performing model.

Error calculation

The error functions were calculated for various hidden neurons and are presented in the below Table 3.3. It can be seen from the table that different hidden number of neurons has different error rates and the neuron layer with the lowest error rates will have the smaller difference between the predicted and the measured value leading to best prediction rate.

	No of neurons	2	4	5	6	7	8	9	10	12	14	16	24	32
lead	RMSE	0.25087	0.2054	0.213	0.2167	0.2122	0.2108	0.2249	0.2297	0.2247	0.00118	**0.2209**	0.24	0.24135
	MAE	0.219	0.1695	0.1596	0.1842	0.1767	0.1754	0.1854	0.1988	0.18429	0.1866	**0.1828**	0.2093	0.1977
	RE	137.23	1059.019	1298.7	1047.23	1057.89	986.181	1291.418	1135.66	1151.78	1171.785	**58.23**	1029.246	1257.86
NH_3	RMSE	0.010211	0.15217	0.15855	0.1707	**0.15921**	0.17019	0.1714	0.16508	0.1523	0.1559	0.165	0.19441	0.01697
	MAE	0.11873	0.09664	0.10615	0.09847	**0.1017**	0.11024	0.1029	1.6508	0.0966	0.100401	0.1019	0.1187	0.110307
	RE	122.34	95.879	120.79	92.294	**66.8**	106.68	120.76	125.98	102.119	105.262	114.45	122.34	116.29
	RMSE	0.2399	0.2401	0.2507	0.2708	0.224	0.2331	0.234	0.2395	0.2372	0.2239	0.229	0.2122	**0.2194**
	MAE	0.1842	0.1812	0.19022	0.2062	0.1681	0.1781	0.1796	0.184	0.18596	0.17459	0.1749	0.1709	**0.16**
PM_{10}	RE	41.353	39.45	43.244	54.54	32.35	30.99	40.575	38.268	38.81	32.22	39.477	41.88	**28.538**
	RMSE	0.18322	0.172	0.148	0.3396	0.1488	0.16165	0.1419	0.1812	0.1767	0.2	0.2	0.2026	**0.2**
	MAE	0.1431	0.1376	0.1098	0.12577	0.121194	0.1274	0.1095	0.1439	0.1388	0.1276	0.1618	0.1611	**0.14546**
$N0_2$	RE	14.61	16.567	14.685	16.764	15.845	15.99	13.511	15.26	13.959	14.82	16.619	20.093	**5.383**
	RMSE	0.3543	0.3179	**0.3309**	0.3407	0.38205	0.3393	0.3738	0.3223	0.3577	0.3721	0.3418	0.3859	0.3515
	MAE	0.3267	0.2537	**0.2832**	0.3112	0.3344	0.2934	0.3335	0.2589	0.3285	0.3415	0.308	0.3289	0.3084
SO_2	RE	17.74	13.651	**13.134**	15.543	21.96	17.61	22.661	18.516	18.0992	17.94	17.9	17.76	15.401

Table 3.3 : Table showing errors for different number of neurons for Central silk board, Hosur road.

From the table 3.3, it can be seen that different parameters have achieved best result at different number of neurons. Best prediction model for lead is obtained when 16 number of neurons are used in the hidden layer because of the lesser error rate having RMSE, MAE, RE, 0.2209, 0.1828, 58.23 respectively. NH_3 achieved its best prediction model when used 7 number of hidden neurons having RMSE, MAE, RE, 0.15921, 0.1017, 66.8 respectively likewise NO_2 at 32 number of hidden neurons having error rates RMSE, MAE, RE 0.2, 0.14546, 5.383 respectively. PM_{10} at 32 number of hidden layers having RMSE, MAE, RE 0.2194, 0.16, 28.538 respectively, and then SO_2 at 5 number of hidden neurons having RMSE, MAE, RE 0.3309, 0.2832, 13.134 respectively. we can conclude, to yield the best prediction for the Neural Network (NN) This architecture will yield the best result because of its lesser error rate when compared to other hidden neuron layers. The graphs of the Neural Network (NN) with best prediction models that is with the least errors are shown below from figure 3.13 to 3.17.

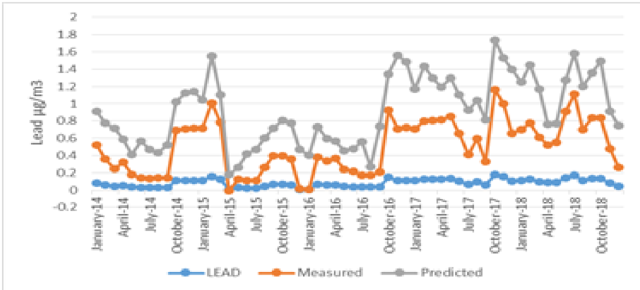

Fig3.13: Graph showing varation between measured and predicted values for Lead.

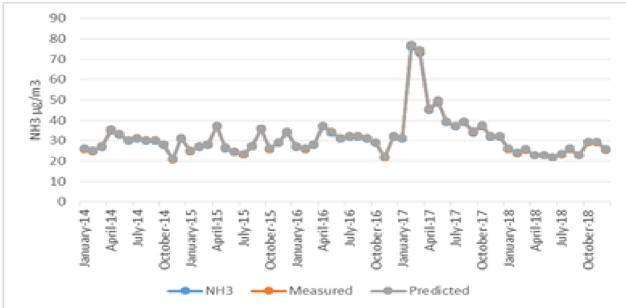

Fig3.14: Graph showing varation between measured and predicted values for NH_3.

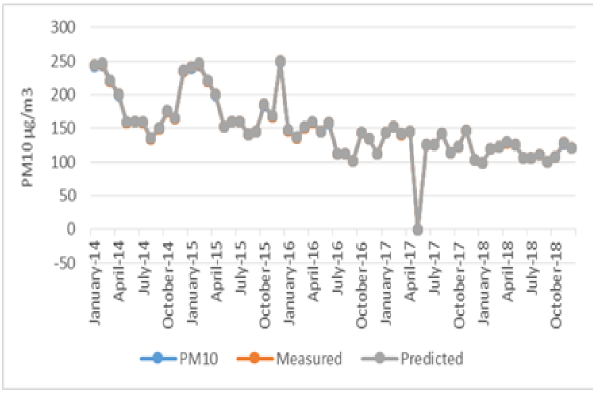

Fig3.15: Graph showing varation between measured and predicted values for PM_{10}.

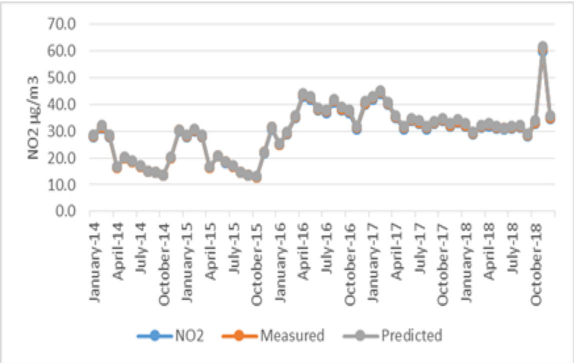

Fig3.16: Graph showing varation between measured and predicted values for NO_2

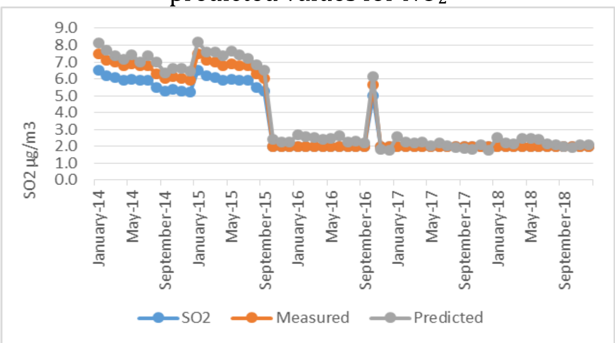

Fig3.17: Graph showing varation between measured and predicted values for SO_2.

Forecasting for Central silk board, Hosur road.

Output obtained from the above mentioned best prediction models are used for the forecasting for the future. Data is collected from 2014 to 2018, the forecasting is done for the next five years that is for 2019 to 2023. The below mentioned graphs shows the forecast of Amco Batteries, Mysore road for the Lead, NH_3, PM_{10}, NO_2, SO_2,

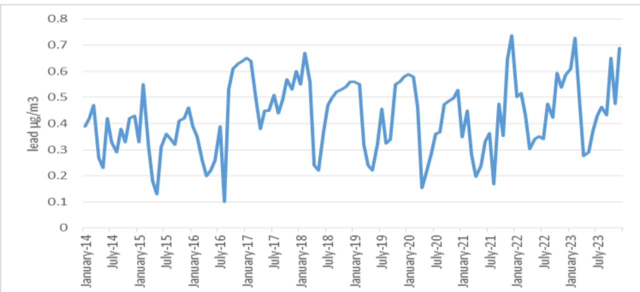

Fig3.18: Graph showing observed and forecasted values for Lead

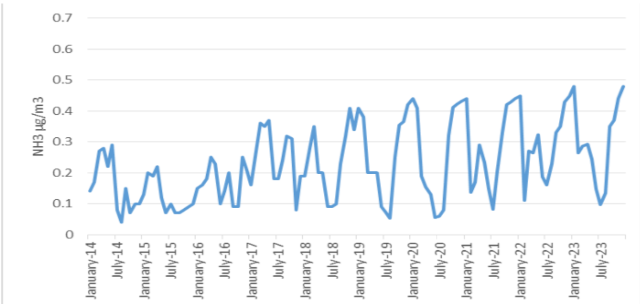

Fig3.19: Graph showing observed and forecasted values forNH_3

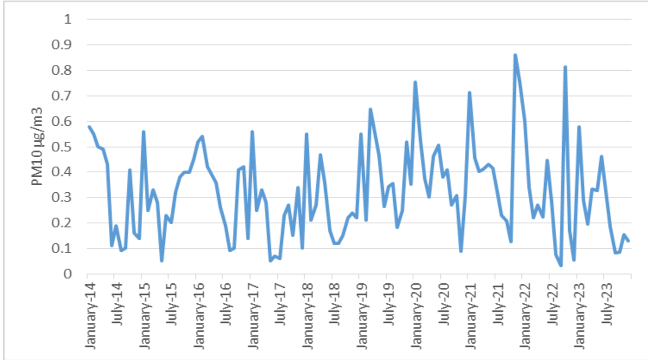

Fig3.20: Graph showing observed and forecasted values for PM_{10}.

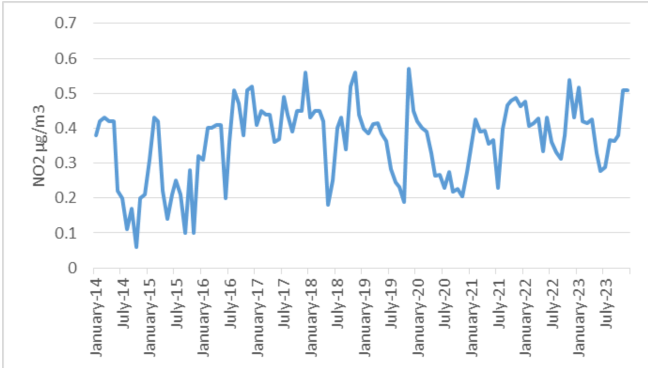

Fig3.21: Graph showing observed and forecasted values for NO_2.

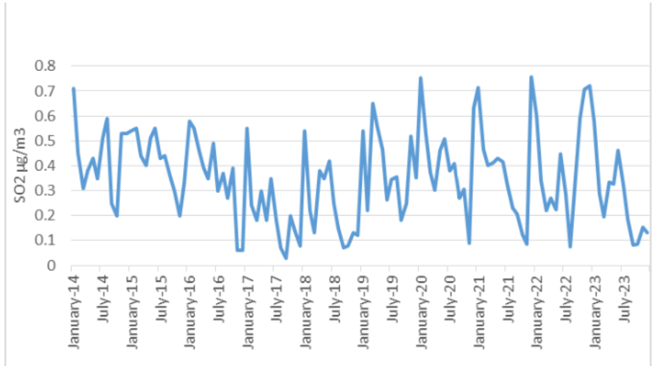

Fig3.22: Graph showing observed and forecasted values for SO_2

SL. No	parameters	R^2 value
1	Lead	0.714
2	NH_3	0.654
3	PM_{10}	0.879
4	NO_2	0.659
5	SO_2	0.801

Table 3.4: Table showing R^2 values for the forecasted data.

DISCUSSIONS

- Figures 3.13 to 3.17 shows the graphs of measured vs predicted for the air quality parameters. The graphs are plotted only for the best prediction model, that is neuron that is giving the less error.
- From graphs we can observe that the measured line and the predicted line are close in both trend and pattern, which indicates the least error between the measured and predicted data leading to best prediction model.
- Figures 3.18 to 3.22 shows the forecast of Lead, NH_3, PM_{10}, NO_2, SO_2 from 2019 to 2023. We can observe that all the parameters are showing increase in trend. The percentage increase values for Lead, NH_3, PM_{10}, NO_2, SO_2 are 6.3%, 8.6%, 16.2%, 8.5%, 9.71% respectively.
- The Table 3.4 shows the R^2 values for the forecasted values, The closer the value is to one, the better and more accurate is the prediction.
- Figures 3.7 to 3.12 shows the forecast of Lead, NH_3, PM_{10}, $PM_{2.5}$, NO_2, SO_2 from 2019 to 2023. It can be observed that all the parameters are showing increase in trend. The percentage increase values for Lead, NH_3, PM_{10}, $PM_{2.5}$, NO_2, SO_2 are 10%, 11.5%, 16.33%, 4.3%, 0.3%, 17.78% respectively.
- The Air quality index values are calculated for forecasted data, The AQI values are 121, 163, 201 for 2021, 2022, 2023 respectively
- The AQI values for 2021 and 2022 falls under moderate category, yet shows increasing trend compared to previous years. The AQI value for 2023 is 201 which falls under poor quality.
- Central silk board junction is also one of the two most polluted areas in the city, recording high levels of residual suspended particulate matter well above the national permissible limit; the high pollution level has been attributed to the traffic congestion at the junction leading to the increase in the AQI values.

3.3 Amco Batteries, mysore road.

Monthly average of air quality parameters like Lead, NH_3, $PM_{2.5}$, NO_2, SO_2, are collected for the study area. The parameter which has to be predicted is taken as dependent variable, the other variables are taken as independent variables. These dependent and independent variables are taken as input in input layer. Different number of neurons are taken in hidden layers like 2, 4, 5, 6, 7 , 8, 9, 10, 12, 14, 16, 24, 32 for identifying the best performing model.

Error calculations

The error functions were calculated for various hidden neurons and are presented in the below Table 3.5. It can be seen from the table that different hidden number of neurons has different error rates and the neuron layer with the lowest error rates will have the smaller difference between the predicted and the measured value leading to best prediction rate.

	Number of Neurons	2	4	5	6	7	8	9	10	12	14	16	24	32
lead	RMSE	0.2294	0.2420	0.2417	0.2653	0.2585	**0.2141**	0.227	0.2305	0.2387	0.2725	0.2446	0.2370	0.255
	MAE	0.1658	0.1891	0.1964	0.2129	0.2001	**0.1515**	0.1782	0.1683	0.1807	0.2076	0.1941	0.1899	0.189
	RE	8.2323	11.885	12.266	15.33	12.923	**6.4829**	9.633	9.960	10520	11.229	11.312	13.525	11.081
NH_3	RMSE	0.254	**0.1793**	0.1718	0.1767	0.1819	0.1657	0.1941	0.1738	0.1767	0.176	0.198	0.179	0.176
	MAE	0.2048	**0.1242**	0.116	0.1242	0.1163	0.1142	0.1323	0.1217	0.1239	0.1259	0.134	0.13	0.125
	RE	26.45	**24.48**	101.22	92.152	61.852	77.33	63.752	96.628	98.92	89.22	88.74	92.76	84.36
$N0_2$	RMSE	0.1703	0.1714	0.1854	0.1560	0.1726	0.1763	**0.1734**	0.1730	0.1704	0.0004	0.1859	0.17	0.158
	MAE	0.1188	0.39	0.1255	0.1068	0.1215	0.1189	**0.1227**	0.1219	0.1118	0.1193	0.1279	0.114	0.108
	RE	91.152	93.699	61.75	79.404	89.49	84.268	**8.033**	77.584	94.813	90.014	96.112	96.21	84.4
$PM_{2.5}$	RMSE	0.2121	0.2010	0.1775	0.2156	0.2174	0.1811	0.1764	**0.1553**	0.157	0.213	0.221	0.2429	0.220
	MAE	0.1694	0.1551	0.1393	0.1717	0.1756	0.1402	0.1362	**0.1199**	0.1203	0.171	0.1827	0.2013	0.524
	RE	44.248	40.148	42.13	43.754	41.91	40.416	36.054	**19.089**	30.693	41.95	23.71	47.556	39.74
SO_2	RMSE	0.318	0.3024	**0.2581**	0.2782	0.2811	0.3175	0.3082	0.3109	0.2044	0.3148	0.3183	0.2949	0.295
	MAE	0.2542	0.231	**0.2035**	0.20684	0.2106	0.2255	0.2383	0.2359	0.1487	0.2334	0.2355	0.2274	0.214
	RE	15.54	29.433	**12.798**	14.504	17.047	19.239	13.984	29.372	14.11	21.17	17.48	14.319	13.73

Table 3.5 : Table showing errors for different number of neurons for Amco Batteries, Mysore Road.

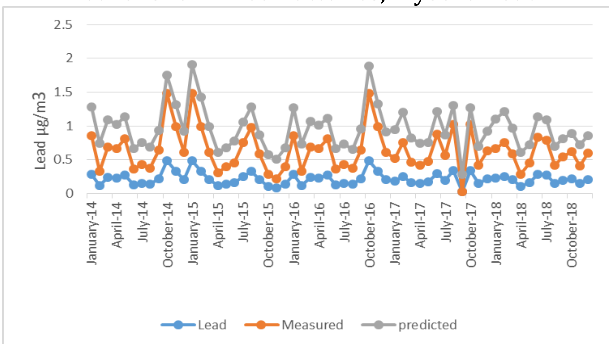

Fig3.23: Graph showing varation between measured and predicted values for Lead.

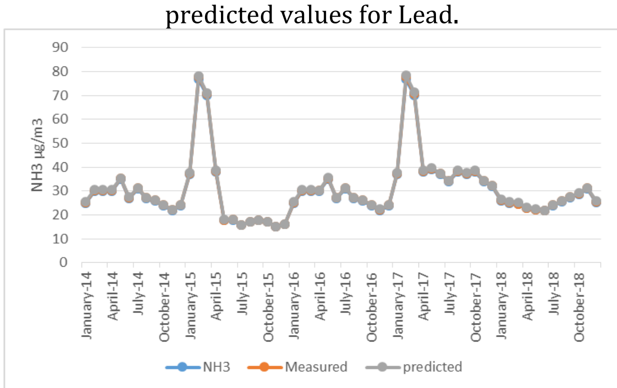

Fig3.24: Graph showing varation between measured and predicted values for NH_3.

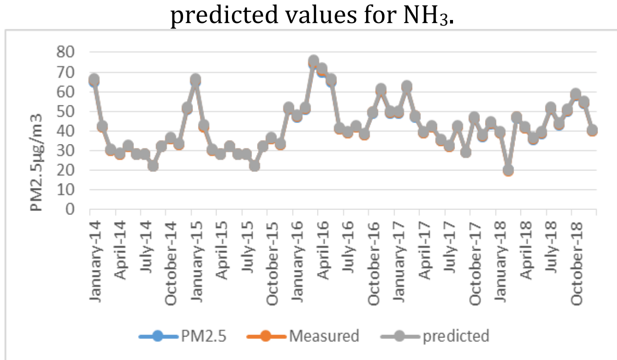

Fig3.25: Graph showing variation between measured and predicted values for $PM_{2.5}$.

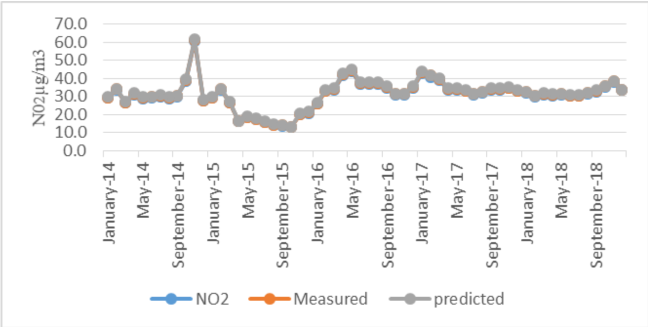

Fig3.26: Graph showing variation between measured and predicted values for NO_2.

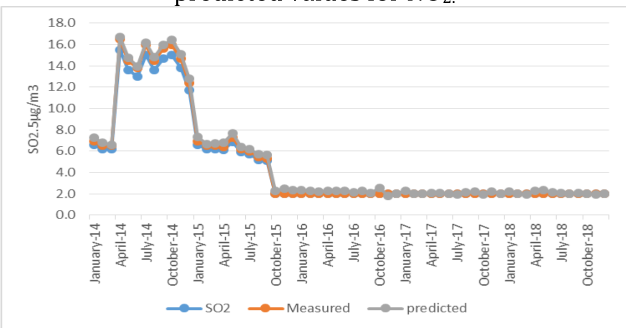

Fig3.27: Graph showing variation between measured and predicted values for SO_2.

Forecasting for Amco Batteries, mysore road.

Output obtained from the above mentioned best prediction models are used for the forecasting for the future. Data is collected from 2014 to 2018, the forecasting is done for the next five years that is for 2019 to 2023. The below mentioned graphs shows the forecast of Amco Batteries, Mysore road for the Lead, NH_3, $PM_{2.5}$, NO_2, SO_2.

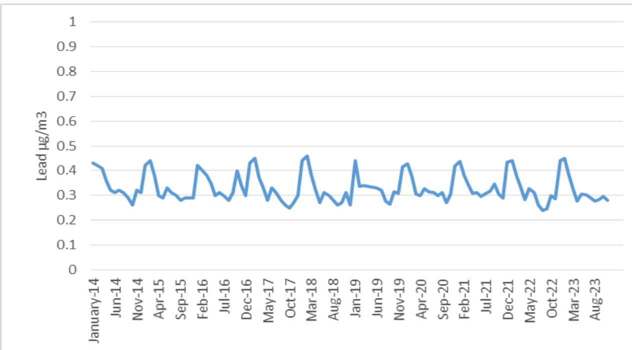

Fig3.28: Graph showing observed and forecasted values for Lead.

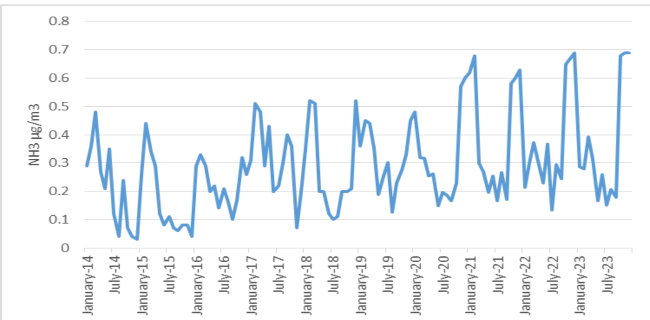

Fig3.29: Graph showing observed and forecasted values for NH_3.

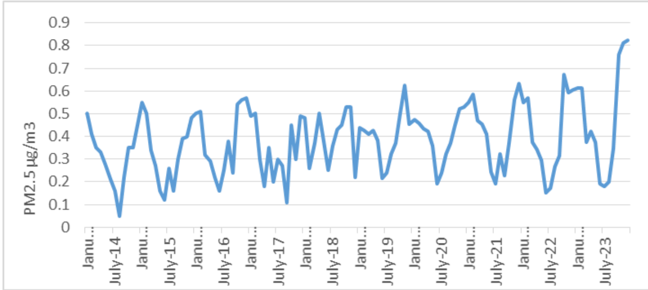

Fig3.30: Graph showing observed and forecasted values for $PM_{2.5}$

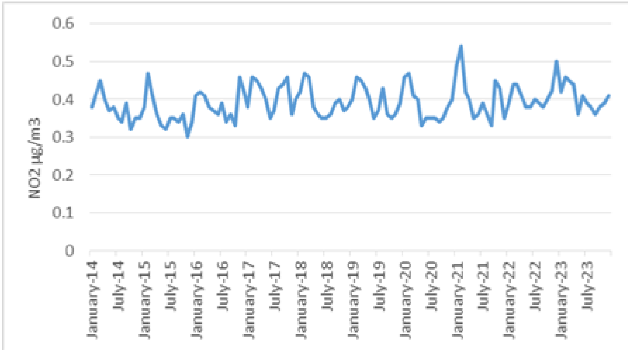

Fig3.31: Graph showing observed and forecasted values forNO_2

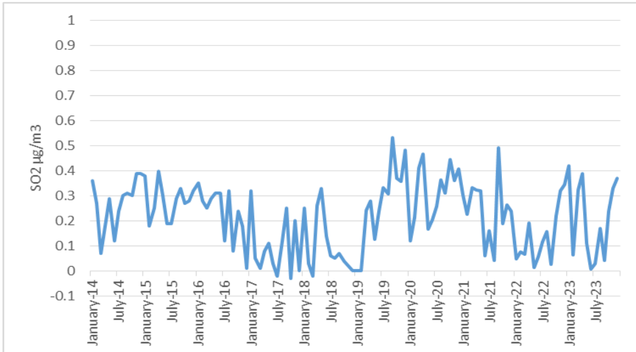

Fig3.32: Graph showing observed and forecasted values for SO_2.

SL. No	parameters	R^2 value
1	Lead	0.89
2	NH_3	0.912
3	$PM_{2.5}$	0.814
4	NO_2	0.645
5	SO_2	0.879

Table 3.6: Table showing R^2 values for the forecasted data.

DISCUSSIONS

- Figures 4.23 to 4.27 shows the graphs of measured vs predicted for the air quality parameters. The graphs are plotted only for the best prediction model, that is neuron that is giving the less error.
- From graphs it can observed that the predicted values were very close to the observed pollutant concentrations both in trend and pattern. It was found that all the developed ANN prognostic models perform absolutely well.
- Figures 3.28 to 3.32 shows the forecast of Lead, NH_3, PM_{10}, NO_2, SO_2 from 2019 to 2023. We can observe that all the parameters are showing increase in trend. The percentage increase in values for Lead, NH_3, $PM_{2.5}$, NO_2, SO_2 are 1.14%, 22%, 15.4%, 1.4%, 36% respectively.
- The Table 4.6 shows the R^2 values for the forecasted values, The closer the value is to one, the better and more accurate is the prediction.
- The Air quality index values are calculated for forecasted data, The AQI values are 95, 97, 102, for 2021, 2022, and 2023 respectively.
- The AQI values are showing increase in trend. The AQI values for 2021 and 2022 falls under satisfactory condition, for 2023 the AQI is 102 and falls under moderate condition.
- The reason for increase in the concentration for parameters may be increase in the number of vehicles in the study area Mysore road as it is one of the busiest roads in Bangalore as it connects Bangalore to Mysore leading to more traffic movement.

3.4 Swan Silk Pvt. Ltd ,Peenya

Monthly average of air quality parameters like Lead, NH_3, PM_{10}, $PM_{2.5}$, NO_2, SO_2, are collected for the study area. The parameter which has to be predicted is taken as dependent variable, the other variables are taken as independent variables. These dependent and independent variables are taken as input in input layer. Different number of neurons are taken in hidden layers like 2,4,5,6,7,8,9,10,12,14,16,24,32 for for identifying the best performing model.

	Number of neurons	2	4	5	6	7	8	9	10	12	14	16	24	32
lead	RMSE	**0.2748**	0.1989	0.2121	0.2557	0.2242	0.2234	0.2443	0.2411	0.2132	0.2186	0.2226	0.24272	0.2707
	MAE	**0.2214**	0.1662	0.168	0.1943	0.1854	0.1784	0.1945	0.1862	0.1715	0.1733	0.17	0.1839	0.219
	RE	**17.484**	31.9	26.489	30.429	32.556	28.506	29.696	30.596	25.981	25.631	24.475	27.705	36.428
NH_3	RMSE	0.1954	0.1752	0.1841	**0.1848**	0.1801	0.1769	0.1705	0.1889	0..1820	0.1780	0.1798	0.1807	0.1738
	MAE	0.139	0.1177	0.1276	**0.12**	0.1214	0.1245	0.1197	0.3161	0.1288	0.1245	0.1242	0.1218	0.1186
	RE	65.603	59.779	53.616	**48.310**	53.321	64.452	60.6	52.472	58.472	60.843	55.57	59.65	51.229
PM_{10}	RMSE	0.2225	**0.1729**	0.2024	0.1586	0.199	0.1653	0.2221	0.2072	0.2104	0.2146	0.1961	0.2043	0.2102
	MAE	11.212	**0.1419**	0.1725	0.1226	0.1656	0.1288	0.1889	0.1767	0.1732	0.1747	0.1629	0.172	0.173
	RE	522.05	**95.043**	443.95	165.63	407.81	119.22	487.19	447.52	551.48	442.94	461.107	514.43	501.43
$PM_{2.5}$	RMSE	0.1706	0.0066	0.1659	0.1533	0.1584	0.1731	0.1032	0.1661	0.1069	0.1683	**0.1076**	0.1747	0.1489
	MAE	0.1379	0.1303	0.1299	0.1165	0.1272	0.1395	0.078	0.1293	0.0819	0.1332	**0.0769**	0.1344	0.117
	RE	27.67	57.423	58.908	57.094	63.78	67.242	34.166	65.156	37.86	60.1912	**27.654**	74.192	61.054
$N0_2$	RMSE	0.2278	0.2318	0.2312	0.2127	0.2405	0.2065	0.22798	0.2086	**0.2319**	0.2347	0.221	0.2418	0.2294
	MAE	0.1577	0.1697	0.1641	0.149	0.1861	0.157	0.1609	0.1485	**0.1669**	0.1751	0.1584	0..1799	0.1669
	RE	51.849	47.031	49.916	42.1798	49.001	39.236	39.236	44.191	**37.115**	53.234	49.731	44.703	47.144
SO_2	RMSE	0.2972	0.2873	0.2536	0.26919	0.3047	0.3038	0.3349	0.2826	0.294	0.2812	0.3203	**0.2988**	0.2974
	MAE	0.2263	0.6258	0.1996	0.1932	0.2323	0.2148	0.2612	0.2086	0.2175	0.2142	0.2256	**0.2297**	0.2175
	RE	17.433	13.534	13.906	15.389	17.744	19.326	17.604	16.056	17.269	14.572	20.775	**12.22**	15.713

Table 3.7 : Table showing errors for different number of neurons for Swan Silk Pvt. Ltd ,Peenya

From the table 4.7, it can be seen that different parameters have achieved best result at different number of neurons. Best prediction model for lead is obtained when 2 number of neurons are used in the hidden layer because of the lesser error rate having RMSE, MAE, RE, 0.2748, 0.2214, 17.484 respectively. NH_3 achieved its best prediction model when used 6 number of hidden neurons having RMSE, MAE, RE, 0.1848, 0.12, 48.3102 respectively likewise NO_2 at 12 number of hidden neurons having error rates RMSE, MAE, RE 0.2319, 0.16699, 37.115 respectively. $PM_{2.5}$ at 16 number of hidden neurons having error rates RMSE, MAE, RE 0.1076, 0.0769, 27.654 respectively. PM_{10} at 4 number of hidden layers having RMSE, MAE, RE 0.17299, 0.1419, 95.043 respectively, and then SO_2 at 24 number of hidden neurons having RMSE, MAE, RE 0.2988, 0.2297, 12.22 respectively. we can conclude, to yield the best prediction for the Neural Network (NN) This architecture will yield the best result because of its lesser error rate when compared to other hidden neuron layers. The graphs of the Neural Network (NN) with best prediction models that is with the least errors are shown below.

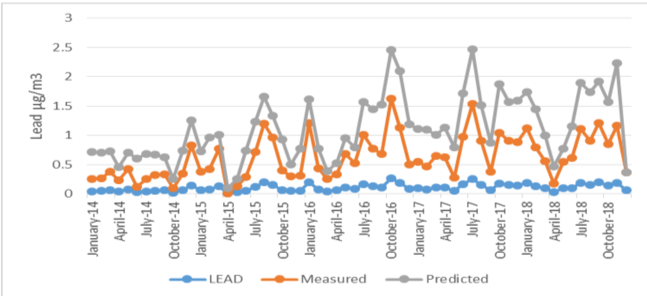

Fig3.33: Graph showing varation between measured and predicted values for Lead.

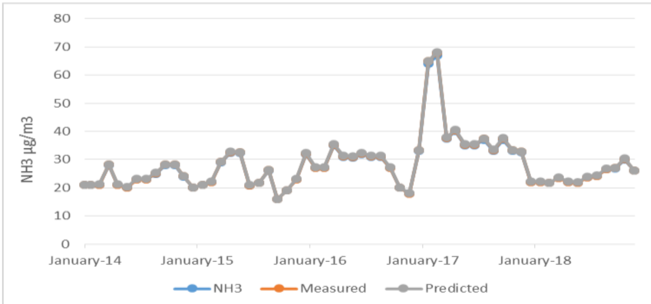

Fig3.34: Graph showing varation between measured and predicted values for NH_3

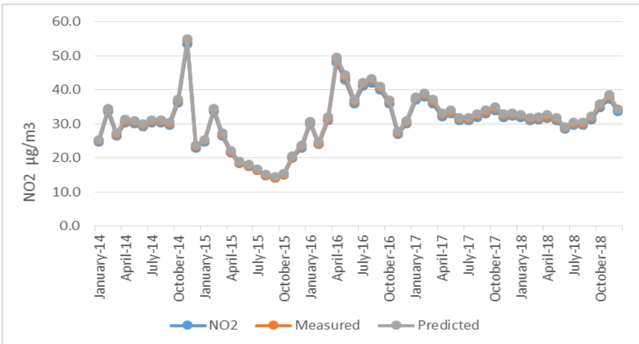

Fig3.35: Graph showing varation between measured and predicted values for NO_2,

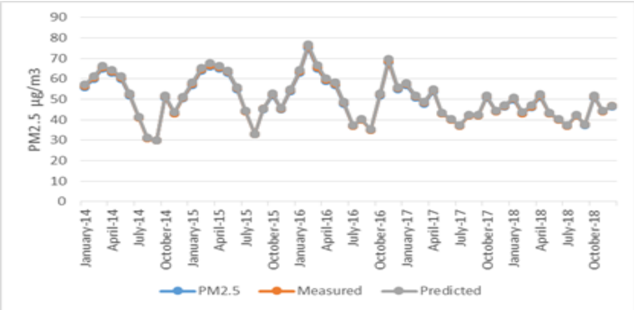

Fig3.36: Graph showing varation between measured and predicted values for $PM_{2.5}$

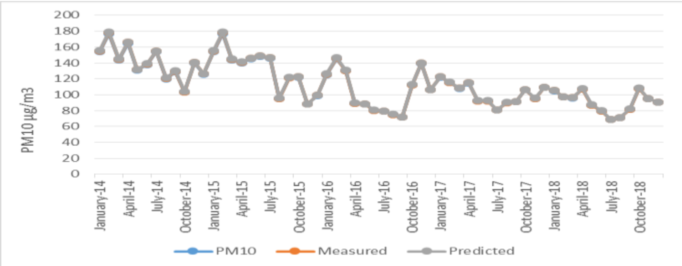

Fig3.37: Graph showing varation between measured and predicted values for PM_{10}.

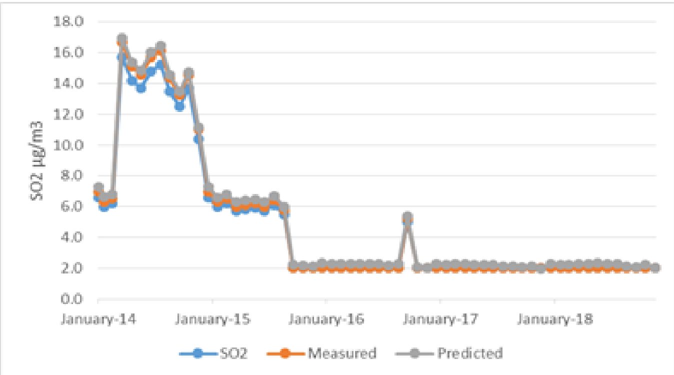

Fig3.38: Graph showing varation between measured and predicted values for SO_2.

Forecasting for Swan Silk Pvt. Ltd ,Peenya

Output obtained from the above mentioned best prediction models are used for the forecasting for the future. Data is collected from 2014 to 2018, the forecasting is done for the next five years that is for 2019 to 2023. The below mentioned graphs shows the forecast for the Lead, NH_3, PM_{10}, $PM_{2.5}$, NO_2, SO_2 Swan Silk Pvt, Ltd. Peenya

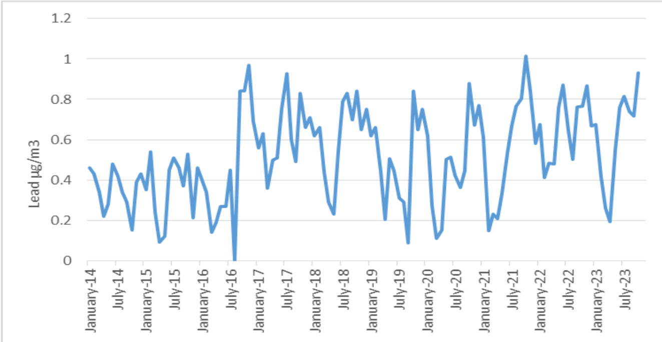

Fig3.39: Graph showing observed and forecasted values for Lead.

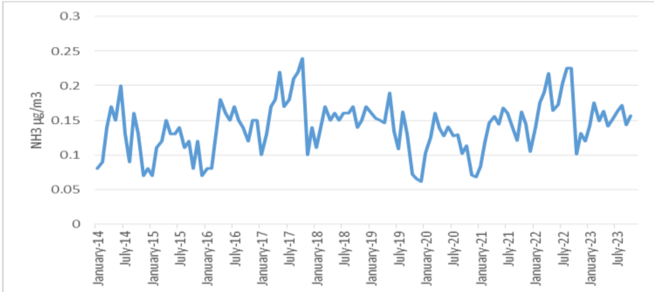

Fig3.40: Graph showing observed and forecasted values for NH_3

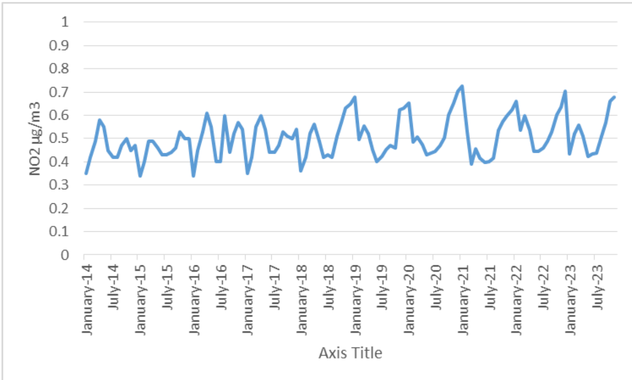

Fig3.41: Graph showing observed and forecasted values for NO_2.

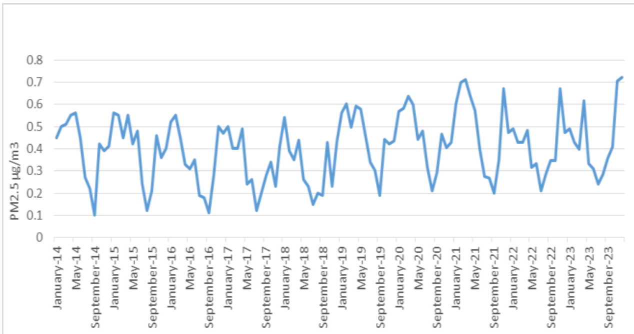

Fig3.42: Graph showing observed and forecasted values for $PM_{2.5}$.

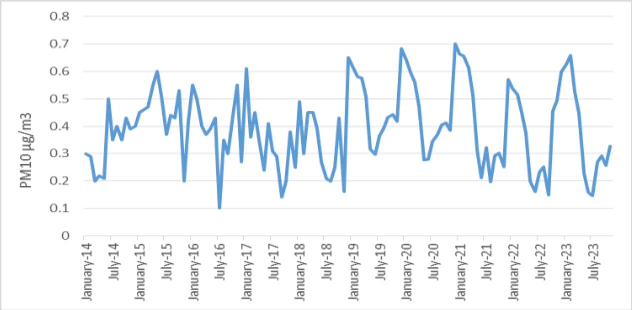

Fig3.43: Graph showing observed and forecasted values for$PM_{10.}$

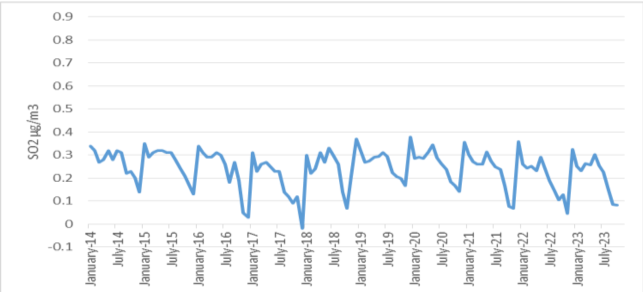

Fig3.44: Graph showing observed and forecasted values for $SO_{2.}$

Table 3.8: Table showing R^2 values of swan silk area

DISCUSSIONS

- Figures 3.33 to 3.38 shows the graphs of measured vs predicted for the air quality parameters. The graphs are plotted only for the best prediction model, that is neuron that is giving the less error.
- From graphs we can observe that the measured line and the predicted line indicates the least error between the measured and predicted data leading to best prediction model.
- Figures 3.39 to 3.44 shows the forecast of Lead, NH_3, PM_{10}, NO_2, SO_2 from 2019 to 2023. We can observe that all the parameters are showing increase in trend. The percentage increase values for Lead, NH_3, PM_{10}, $PM_{2.5}$, NO_2, SO_2 are 10.625%, 2.14%, 6%, 6.2%, 1.33% , 11.36% respectively.
- The Table 3.8 shows the R^2 values for the forecasted values,The closer the value is to one, the better and more accurate is the prediction.
- The Air quality index values are calculated for forecasted data, The AQI values are 87, 99, 105 for 2021, 2022, 2023 respectively.
- The AQI values in the study area are showing increase in trend because the study area represents Industrial area with more number of industries with more emissions.

CONCLUSIONS

- For Export Promotional Industrial park IPTL, WF road, The prediction values with the least errors are used for forecasting. Forecasting has done for five years from 2018 to 2023. The R^2 values for Lead, NH_3, $PM_{2.5}$, NO_2, SO_2 are 0.814, 0.714, 0.955, 0.855, 0.955, 0.953 respectively. The higher the values of R^2, or closest to 1, and the smallest the errors, or closest to zero, the better the model.
- For Export Promotional Industrial park IPTL, WF road, AQI is calculated for the forecasted data for 2019 to 2023 that is 89, 55, 95, 107, 112 for 2019, 2020, 2021, 2022, 2023 respectively. The measured AQI for 2018 is 86 and for 2019 is 56. It can be observed the less error between the measured and forecasted AQI for the the year 2018 and 2019. Which proves the effectiveness of the model.
- The prediction values for Central silk board, Hosur road, with the least errors are used for forecasting. Forecasting is done till 2023. The R^2 values for Lead, NH_3, PM_{10}, NO_2, SO_2 are 0.714, 0.654, 0.879, 0.659, 0.801 respectively. The higher the values of R^2, or closest to 1, and the smallest the errors, or closest to zero, the better the model.
- AQI is calculated for the forecasted data for 2019 to 2023 that is 99, 61, 121, 163, 201 for 2019, 2020, 2021, 2022, 2023 respectively. The measured AQI for 2019 is 97 and for 2020 is 63. It can be observed the less error between the measured and forecasted AQI for the the year 2019 and 2020. Which proves the effectiveness of the model.
- For Amco Batteries Mysore Road, The figures from 3.22 to 3.32 shows the forecasted values of Lead, NH_3, $PM_{2.5}$, NO_2, $SO_{2.}$ The R^2 values for Lead, NH_3, PM_{10}, $PM_{2.5,}$ NO_2, SO_2 are 0.89, 0.912, 0.814, 0.645, 0.879 respectively.
- AQI is calculated for the forecasted data for 2019 to 2023 that is 90, 45, 95, 97, 102 for 2019, 2020, 2021, 2022, 2023 respectively. The measured AQI for 2019 is 88 and for 2020 is 53. It can be observed the less error between the measured and forecasted AQI for the the year 2019 and 2020. Which proves the effectiveness of the model.
- For Swan Silk Pvt, Ltd, Peenya, The R^2 values for Lead, NH_3, PM_{10}, $PM_{2.5,}$ NO_2, SO_2 are 0.785, 0.936, 0.614, 0.729, 0.948, 0.814 respectively.
- AQI is calculated for the forecasted data for 2019 to 2023 that is 73, 50, 87, 99, 105 for 2019, 2020, 2021, 2022, 2023 respectively. The measured AQI for 2019 is 88 and for 2020 is 51. It can be observed the less error between the measured and forecasted AQI for the the year 2019 and 2020. Which proves the effectiveness of the model.
- ANN proves to be an effective model for the purpose of Forecasting.

- AQI values are calculated for Forcasted data for the study area, It can be observed that AQI values are showing increase in trend for all stations and strict actions should be taken to reduce the air pollution in the study area.

- This ANN based forecasting model is fast and less resource intensive, and can be used for generation of daily forecasts. The forecasts generated through the demonstrated model can be used for scientific purposes and also to take short term corrective actions for air quality management in highly polluted cities. The model can be replicated in other cities after the required set up, optimization and validation and become a very useful tool for emergency planning and short term air quality management.

REFERENCES

1. **Azman Azid1, Hafizan Juahir, Mohd Talib Latif, Sharifuddin Mohd Zain, Mohamad Romizan Osman(2013)** "Feed-Forward Artificial Neural Network Model For Air Pollutant Index Prediction In The Southern Region Of Peninsular Malaysia", Journal Of Environmental Protection", 4, 1-10 Published Online

2. **National Air Quality Index(2015)**, Control of Urban Pollution System CUPS/ B2/ 2014-15

3. **Dahe Jianga, Yang Zhanga, Xiang Hua, Yun Zenga, Jianguo Tanb, Demin Shaob(2003),** Progress in developing an ANN model for air pollution index forecast, Atmospheric Environment 387055–7064

4. **Daniel Arkkelin(2014)** Using SPSS To Understand Research And Data Analysis ValparaisoUniversity,Daniel.Arkkelin@Valpo.Edu

5. **Enzo Grossi(2009) ,** Artificial Adaptive Systems and predictive medicine: a revolutionary paradigm shift, Grossi Immunity & Ageing 2010, 7(Suppl 1):S3

6. **Gaganjot Kaur Kang, Jerry Zeyu Gao, Sen Chiao, Shengqiang Lu, And Gang Xie** (**2018)** "Air Quality Prediction: Big Data And Machine Learning Approaches", International Journal Of Environmental Science And Development, Vol. 9, No. 1

7. **Moustris K.P, P.T.Nastos, I.K.Larissi, And A.G.Paliatsos 4(2012),"** Application Of Multiple Linear Regression Models And Artificial Neural Networks On The Surface Ozone Forecast In The Greater Athens Area ,Greece", Hindawi Publishing Corporation Advances In Meteorology Volume 2012, Article ID 894714, 8 Pages Doi:10.1155/2012/894714

8. **Priyanka Gogoi (2020)** Application of SPSS Programme in the Field of Social Science Research, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-5

9. **Suraya Ghazali1 and Lokman Hakim Ismail (2012),** Air Quality Prediction Using Artificial Neural Network.