Analysis and Predicition of Air Quality And Its Parameters Of Major Metropolitan Cities In India Using ANN

Varna Vishakar V[1], Ayush Jain[1], Gaurav Sharma[2], Manoj Yadav[2], Shrey Tiwari[2], Rajat Sirohi[2], Mohd Anas[2]

[1] Supervisors, Department of Civil Engineering, KIET Group of Institutions, Delhi-NCR, Ghaziabad.

[2]Undergraduate Scholars, Department of Civil Engineering, KIET Group of Institutions, Delhi-NCR, Ghaziabad

Abstract-

In India's major cities air pollution has grown to be a serious problem for both the environment and human health. In order to estimate the air quality index (AQI) this research article uses artificial neural networks to evaluate the air quality in Delhi, Mumbai, and Chennai three significant metropolises. The AQI is a metric for air quality that shows how dirty the air is in a specific location. In recent years air pollution has grown to be a significant problem in India's main cities. The purpose of this study paper is to evaluate the artificial neural network (ANN)-based air quality index of India's major urban areas. In this study, data on air pollutants (PM10, PM2.5, NO2, SO2, and CO) were gathered from Delhi, Mumbai, and Chennai three of India's largest metropolises. The ANN model was trained and validated using the gathered data. The air quality index of these cities for the year 2021 was subsequently predicted using the ANN model. The study's findings demonstrated that the ANN model is a trustworthy and efficient method for forecasting these cities' air quality indices.

The study examines the industrial, vehicular and home air pollution and tracks trends over time using data from a number of sources, including the Central Pollution Control Board, the Indian Meteorological Department, and the National Air Quality Index. The study also examines how affect air quality because were instituted in response to the COVID-19 pandemic.

Index Terms- Model creation , Multilayer Perceptron(MLP) , Artificial Neural Network(ANN) , Linear Regression ,SO2,NO2, Ozone and RSPM , Air Pollution, Meteorological Information

1 Introduction

The degree to which the air is healthy or clean for people or the environment is known as its quality. When pollutants are released into the atmosphere, it is harmful to both human health and the ecosystem. Poor air quality and health are closely related and the World Health considers air pollution to be the biggest environmental health concern. Governmental employ an air quality index (AQI) to inform the general public of how Bad the air is now or is expected to become. By averaging readings from an air quality sensor, which can rise owing to traffic, forest fires, or other factors that can increase air pollution to various national air quality standards, various nations have their own air quality indices. National Air Monitoring run by the Central Pollution Control Board and State Pollution Control Boards covers 239 cities across India and has more than 341 monitoring stations.

The National Air Quality Index was introduced there in 2014 as part of the Swachh Bharat . The eight factors and six categories of the AQI scheme—recommended by an expert panel and IIT Kanpur—are Good, Satisfactory, Moderately Polluted, Poor, Very Poor, and Severe. Twelve parameters are included in the revised National Ambient Air Quality Standards (CPCB 2009) and it would be ideal for a new method of calculating AQI to take into account as many contaminants as possible from the list of identified pollutants. The choice of parameters however is mostly influenced by the AQI objective(s), data accessibility, averaging period, monitoring frequency, and measuring techniques.

In order to increase public awareness and their involvement in air quality management, a uniform and effective AQI scheme that offers information about each pollutant, creates an overall index, and is specific to the entire nation is required. The difficulties of creating an AQI scale that is exact, clear, and easy to grasp while simultaneously including intricate scientific and medical data must be addressed. order to increase public awareness and their involvement in air quality management, a uniform and effective AQI scheme that offers information about each pollutant, creates an overall index, and is specific to the entire nation is required. The difficulties of creating an AQI scale that is exact, clear, and easy to grasp while simultaneously including intricate scientific and medical data must be addressed.

In order to increase public awareness and their involvement in air quality management, a uniform and effective AQI scheme that offers information about each pollutant, creates an overall index, and is specific to the entire nation is required. The difficulties of creating an AQI scale that is exact, clear, and easy to grasp while simultaneously including intricate scientific and medical data must be addressed.

AQI Category, Pollutants and Health Breakpoints

AQI Category (Range)	PM_{10} (24hr)	$PM_{2.5}$ (24hr)	NO_2 (24hr)	O_3 (8hr)	CO (8hr)	SO_2 (24hr)	NH_3 (24hr)	Pb (24hr)
Good (0–50)	0–50	0–30	0–40	0–50	0–1.0	0–40	0–200	0–0.5
Satisfactory (51–100)	51–100	31–60	41–80	51–100	1.1–2.0	41–80	201–400	0.5–1.0
Moderate (101–200)	101–250	61–90	81–180	101–168	2.1–10	81–380	401–800	1.1–2.0
Poor (201–300)	251–350	91–120	181–280	169–208	10–17	381–800	801–1200	2.1–3.0
Severe (301–400)	351–430	121–250	281–400	209–748	17–34	801–1600	1200–1800	3.1–3.5
Hazardous (401+)	430+	250+	400+	748+	34+	1600+	1800+	3.5+

Fig. 1 . AQI Category Performance Of NAR Neural Network

2 Problem Statement

The introduction of chemicals, particulates, or biological materials into the atmosphere that cause discomfort, disease, or death in humans, harm other living organisms like food crops, or harm the natural environment or built environment, is known as air pollution. Indeed, one of the major environmental issues in industrial and cosmopolitan cities is air pollution. Therefore, it's crucial to anticipate pollution and prevent these issues. of the most intriguing and difficult challenges is predicting air pollution using data mining, thus we provide the prediction algorithms used to provide next day and next month air in order to solve the issue.

3 METHODOLOGY

Artificial neural networks, or have been proposed as an alternative model for time series forecasting because of their flexibility in capturing nonlinear time series data. We used a multi-layer (MLP) for this investigation, which consists of processing elements known as neurons. An MLP typically consists of numerous layers of nodes, with the first layer serving as an input layer where outside data is brought in. The output layer, which is the final and topmost layer, is where the problem solution is found. One or more intermediary layers known as the hidden layers stand between the input layer and the output layer. Three layers of simple processing units (input, hidden, and output) coupled by linkages make up the MLP. The strength of each connection is shown by arrows from the input to the hidden layer and from the hidden layer to the output layer. This strength is quantified as weight. Each unit applies a transfer function or activation function to the data in the input layer and hidden layer before distributing the outcome to the output layer. The logistic function is frequently employed as the activation function. the aid of the technique, the ANN model was trained. (RMSE), (MAE), and (R2) metrics were used to assess the performance of the ANN model. The AQI of Delhi, Mumbai, and Chennai were then predicted using the trained ANN model.

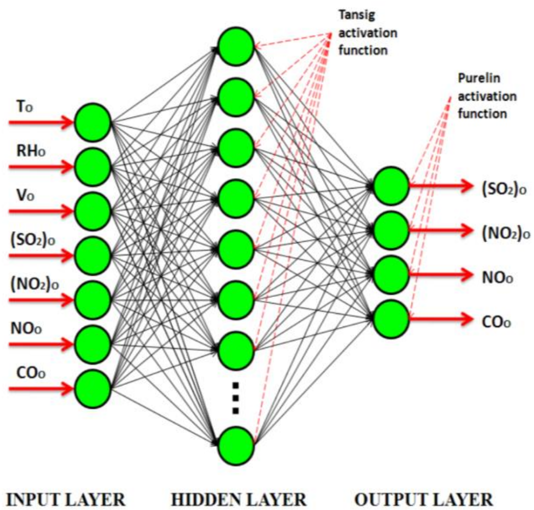

Fig. 2 .Layout Of ANN

3.1 Forecasting of Ambient air Quality in Mumbai by neural network

In this study, two National Air Quality Monitoring (NAMP) stations in Mumbai, India, were used to forecast the weekly air quality using artificial neural networks (ANN). At the industrial and residential stations in and , respectively, the levels of suspended particulate matter (RSPM) were assessed. On historical data from 2009 to 2016, nonlinear (NAR) neural network models were trained. After proper training, the study demonstrated that NAR neural networks are efficient at forecasting ambient air quality with low Mean Squared Error (MSE) values. Due to its link to significant medical diseases like lung cancer and respiratory disorders, air pollution is a major cause for concern. The World Health advises particulate matter (PM) concentrations in the air since they have been associated with increased rates of illness and mortality. Due to its rapid expansion and expanding car population, India's air quality is getting worse. The Control of Pollution Act 1981 was approved by the Central Government, and the Central Pollution Control Board (CPCB) is in charge of gathering and disseminating technical and statistical data about air pollution. The study reveals that employing neural networks for air quality prediction yields better results than statistical methods. and industry have resulted in worsening air quality.

3.1.1 DATA ACQUISITION AND PREPROCESSING

The NAMP data, which were used to produce the RSPM data for this investigation, were provided by the two ambient air quality monitoring stations in Mumbai, located in and . This information is available thanks to the Indian government's Open Government Data Platform.

Weekly RSPM data from January 2008 to December 2014 were used in this work's time series . Each station has a total of 342 observations, which are weekly RSPM level readings.

NO. of Neurons	Tapped delay lines	Mean square error
6	1	25.5789
7	1	25.3720
8	1	25.5475
9	1	25.0763
10	1	25.0139
11	1	28.2685
12	1	25.2609
6	2	25.3449
7	2	26.6375
8	2	26.2028
9	2	27.1262
10	2	26.2553
11	2	25.8295
12	2	25.0524
6	3	26.1219

TABLE. 1 .Performance of Nar Neural Network For Residential RSPM Data

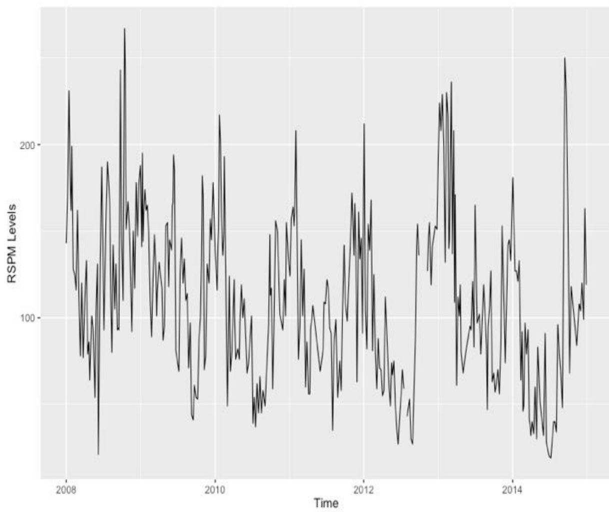

Fig. 3 .**Time Series plot of RSPM**.

3.2 Forecasting of Ambient air Quality in Delhi Zone by neural network

To forecast Delhi's air quality index (AQI), we used a number of linear and non-linear regression techniques, including random forest and artificial neural networks. The primary pollutants taken into account included PM2.5, PM10, CO, NO2, SO2, and O3. With the exception of New Delhi, evaluations using RMS error, mean absolute error, and R2 showed that support vector regression and artificial neural networks were the most efficient. By examining atmospheric and surface parameters including wind speed and temperature, we sought to forecast PM2.5 concentration hourly in various parts of Delhi. Our suggested regression model, which incorporates adaptive boosting, performs better than earlier research.

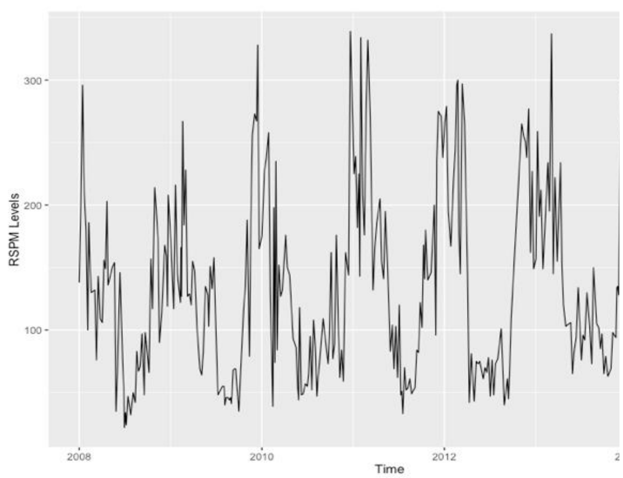

Fig. 4 .**Time Series plot of RSPM levels at Parel station**

Pollutant	PM 2 .5			PM 10			03/NO:/CO SOz		
Parameter	MSE	MAE		MSE	MAE		MSE	MAE	
SGD	0.5295	0.5202	0.4222	0.4124	0.45424	0.52444	0,6246	0.56244	0.32483
RER	0.5527	0.59222	0,44222	).5429	0.44269	0.52445	0,6422	0.64291	0.32405
DTR	0.4624	0.-54264	0.54432	0.3423	0.39492	0.62453	0.5257	0.5548	0,42140
MLP	0.5123	0.51122	0,49422	0.4224	().43024	0.-52408	0.6429	0.-58426	0.32941

SVR	0.4972	0.46122	0.61427	0,5248	0.40421	0.49422	0.5452	0,54281	0,42494
GBR	0.4254	0.46204	0.60422	0.4243	0.32044	0.582549	0.5422	0.-54482	0,42417
ABR	0.4482	0.47211	0.54926	0.4428	0.39929	0.5244	0.5442	0.44172	0.42467
	0.6320	0.62112	0.34429	0.8427	0.81242	0.18022	0.9242	0.74822	0.02424

TABLE.NO. 2 .Anand vihar

Pollutant	PM 2.5			PM IO			03/NO:/CO/S02		
Parameter	MSE	MAE		MSE	MAE		MSE	MAE	
SGD	0,3281	0.42320	068291	0.5249	0.42237	0.59298	0.7626	0.58028	0.26373
RER	0,3202	0,43952	0,66128	0,4275	0,44280	0,48269	0.7625	0,55228	0,20391
DTR	0,3221	0,42496	0,67183	0.4222	0.41239	0,61282	0.6421	0,58126	0,26384
MLP	0.3214	0.42264	0.66106	0,4767	0.43216	0.62203	0,6426	0.55811	0,38361
SVR	0,2256	0,32566	0,76260	0,4705	0,40102	0,62243	0,6722	0,51118	0,38210
GBR	0.4122	0,32551	0,67215	0.4703	0,37112	0,66213	0.6521	0,471273	0,35262
ABR	0.2729	0,32422	0.71226	0.8783	0.38174	0,64247	1.-5923	0.51227	0,37021

TABLE . 3 .Punjabi bagh

Pollutant	PM 2.5			PM 10			03/N02/CO/S02		
Parameter	MSE	MAE		MSE	MAE		MSE	MÅE	
	0.3435	0.42705	0.65746	0.4037	0.40082	0.46651	0.5470	0.56240	0.31126
SGD	0.3286	0.44977	0.65522	0.5014	0.40981	0.41684	0.6501	0.54277	0.23099
RER	0.3234	0.42854	0.65546	0.4037	0.40082	0.45161	0.5470	0.56240	0.30126
DIR	0.2220	0.4222	0.62555	0.4032	0.40618	0.41461	0.5747	0.56299	0.41196
MLP	0.2564	0.3727	0.69575	0.4029	0.30769	0.31049	0.5211	0.50253	0.48102

SVR	0.29267	0.36227	0.68494	0.5062	0.40779	0.32772	0.5277	0.48260	0.47137
GBR	0.2724	0.36242	0.6965	0.4006	0.40905	0.40858	0.5377	0.52217	0.48141
ABR	0.4620	0.42205	0,65275	0.6297	0.60545	0.30049	1.2250	0.2579	-0.2143

TABLE . 4 .RK Puram

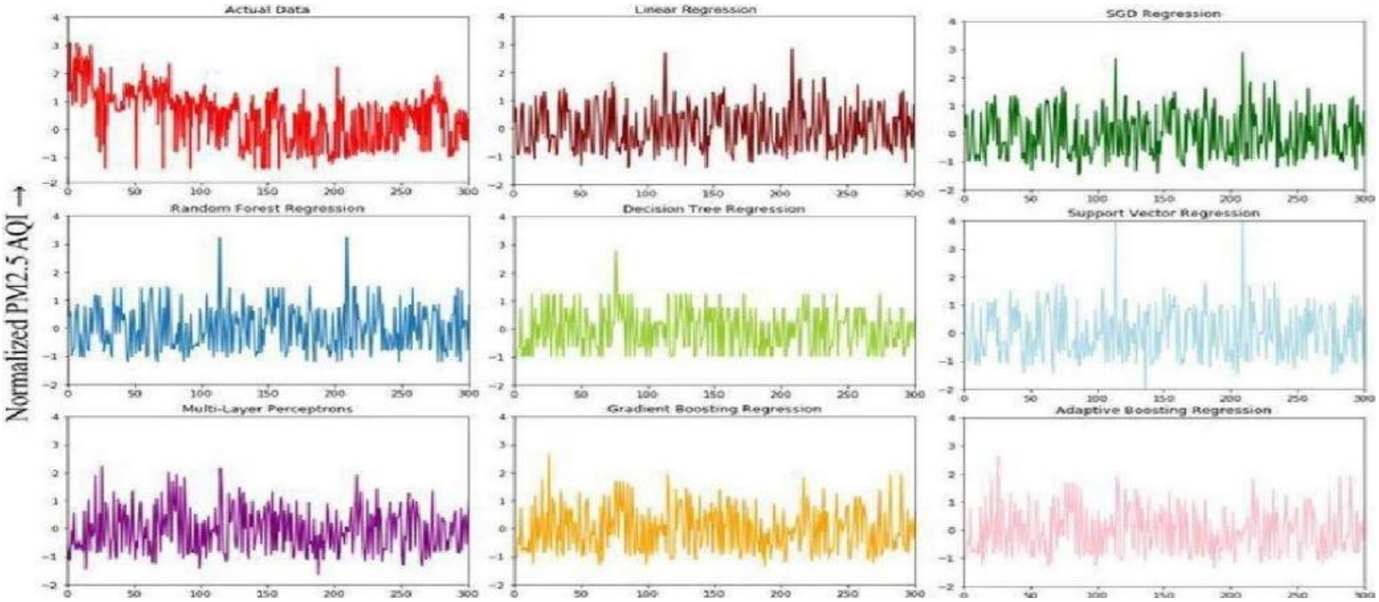

Fig. 5. .Test datasets Mean Square Error (MSE), Mean Absolute error (MAE) Methods

3.3 Forecasting of Ambient air Quality in Chennai Zone by neural network

This study uses data from Chennai, India, an important cultural, commercial, and educational hub with a population of around 10 million, for air pollution monitoring and forecasting. In and around Chennai, the primary air pollutants areSO2,PM10,PM2.5,03,C0,NO2. According to the Table and Figure

Air Pollu-tant	Units	Range	Mean	Standard Deviation	Variance
CO	mg/m3	0-2	39.58	16.77	284.66
NO2	µg/m3	0-80	21.08	9.21	84.65
Ozone(O3)	µg/m3	0-180	23.74	12.52	154.35
PM2.5	µg/m3	0-60	70.09	52.22	2716.29
SO2	µg/m3	0-80	9.25	8.45	64.82

TABLE. 5 . Statistical measure of Chennai data samples

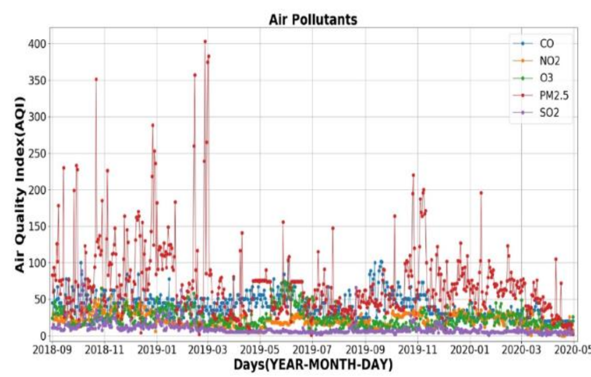

Fig. 6 . Multilinear Regression (MLR) model

Using regression (MLR), this study examines the linear association between air pollutants and increased AQI using 607 data points, 122 for testing and 486 for training. The linear segmented technique is used to determine the for a certain pollutant concentration. The aggregated index is then created by adding all of the together using an equation. The Central Pollution Control Board (CPCB)'s recommended AQI measures for India are to calculate the air quality index's range and its associated health consequences. Regression and ARIMA time series models are in the study to forecast and predict AQI, and error indices are used to assess each model's effectiveness.

3.3.1 MULTILINEAR REGRESSION

More than one linear regression Using MLR, the AQI is predicted. According to correlation coefficients, it was concluded that there isn't a strong correlation between any two air contaminants. The MLR model is trained using Python and training data.The MLR equation is contained in Equation. X1 = 0.5570 + 0.3647 YI1, 0.1792 YI2, 0.0092 YI3, 0.9138 YI4, and 0.1199 YI5.

	SO2	PM2.5	O3	NO2	CO
SO2	1.0020	0.2342	0.0641	0.2545	0.2025
PM2.5	0.2370	1.0022	0.1852	0.2578	0.1254
O3	0.0670	0.1877	0.1005	0.1485	0.1158
NO2	0.2970	0.2841	0.1485	0.0010	0.0934
CO	0.2063	0.1242	0.1176	0.0934	1.0010

TABLE. 6 . Correlation coefficients of air pollutants.

Model Parameters	Values
Y-Intercept(x0)	-0.5560
X1	0.3657
X2	-0.1692
X3	0.0082
X4	0.9128
X5	0.1189

TABLE. 7 . Coefficients of MLR equation.

Figure displays the AQI prediction made using training data. Using previously unseen test data and the MLR's acquired model parameters, the MLR was put to the test.

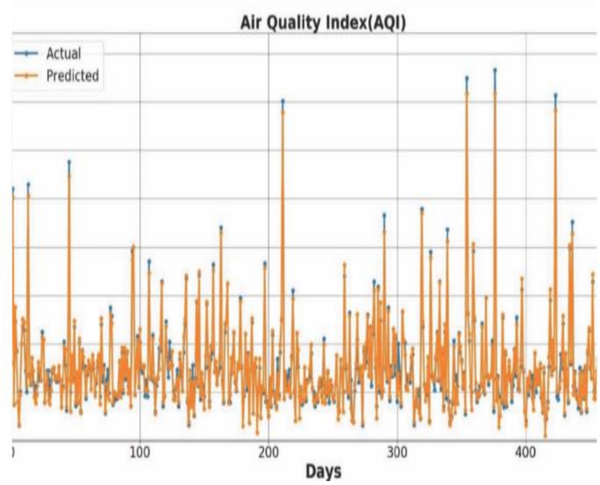

Fig. 7 . Prediction of AQI using training data.

The curve for AQI analysis using test data is shown in the figure above. The data sample used to train the MLR algorithm is randomly selected and makes up 80% (486 samples) of the aggregate data. It was concluded from the above Figure that the MLR algorithm predicts AQI quite accurately. The derived model parameters of the same procedure are validated using the hidden data. Figure below displays the AQI prediction's validation findings.

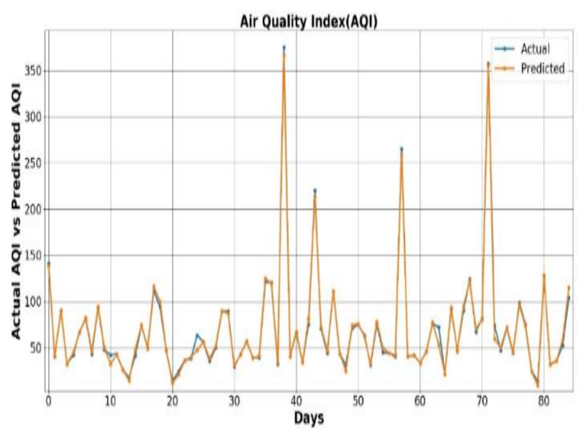

Fig. 8 . Prediction of AQI using Test data

It was deduced that the MLR algorithm predicts AQI with greater accuracy. 21% (119 data samples) of the total data set make up the test data set. The test data are provided separately to the algorithm and are not used to train it rather they are the MLR algorithm's precision. Table contains the estimated quantitative performance indices.

Number of Folds in K-fold Cross-validation	Training Data	Test Data
1	0.000005	0.000031
2	0.000062	0.000626
3	0.000024	0.000663
4	0.000051	0.000994
5	0.000162	0.003458
6	0.000093	0.009376
7	0.000233	0.001519
8	0.000685	0.003730
9	0.000126	0.015223
10	0.000877	0.003785

TABLE. 8 . Error variance for Each Fold in K-fold Cross-Validation.

Performance Indices	Training Data	Testing data
K-fold cross validation	95.99%	92.01%
R-square	1.00	0.96
Mean ab-solyte error	3.0111	3.0121
Root mean square deviation	4.0656	3.9999

TABLE. 9 . Performance indices of Multilinear Regression model.

3.3.2 Time-Series Analysis

The ARIMA time-series analysis is used to forecast the AQI to better understanding. The moving average (MA), auto-regressive (AR), and (I) models make up the ARIMA model. The PACF and ACF are used to calculate the coefficients for the AR and MA models.

The amount of data differentiation, which must be steady, affects the coefficient of the model. The presented data are confirmed to be by the Dickey-Fuller test, necessitating two to estimate a coefficient of 2 for . The AR and MA coefficients for the computed ARIMA model were obtained using the PACF and ACF graphs and are shown in Table.

Model Coefficients	Values
ARP (Auto Regressive Model)(P)	7.9
DM (Differencing Mod-el)(d)	1.9
MAM (Moving Average Model)(q)	1

Table. 10 . ARIMA model coefficients.

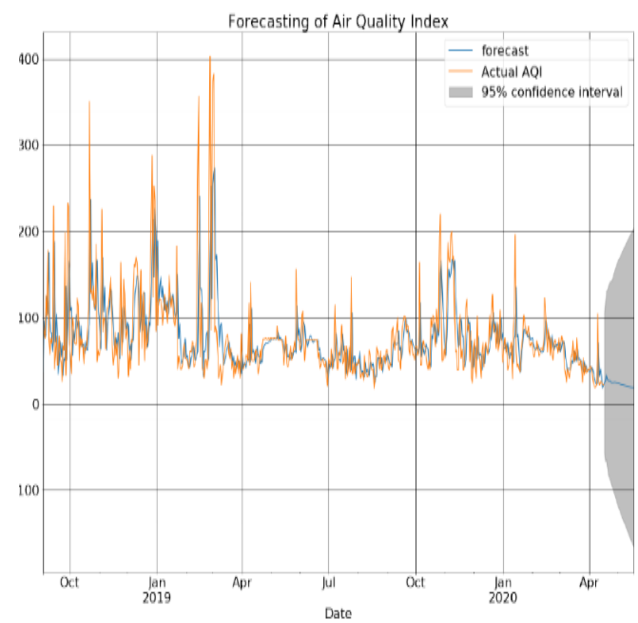

Fig. 9. . Forecasting of AQI using ARIMA model using Training data.

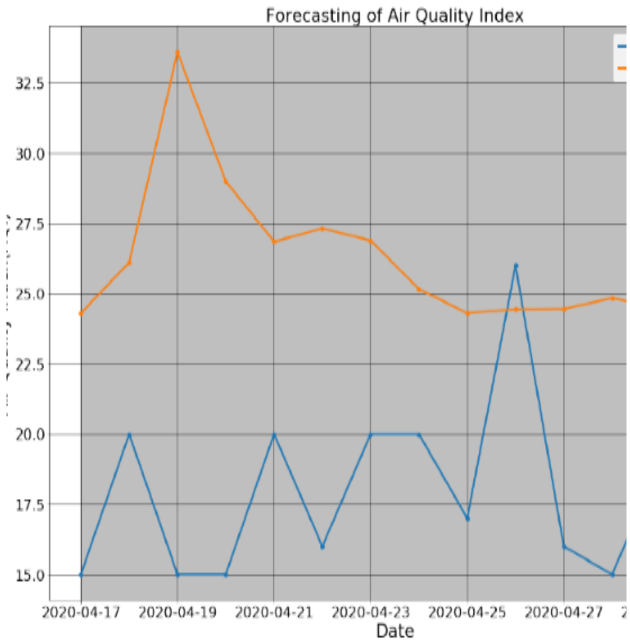

Fig. 10 . Validation of Forecasting of AQI using ARIMA model with test data.

The ARIMA model, test data that have not yet been seen, provided a comparison between the actual value and the projected value shown in Figure. The determined projected values are encircled by a 94% confidence zone. The ARIMA model's quantitative performance evaluation indices for the test data set are calculated and shown in the table 10 .

Performance Indices	Test Data
R-square	0.91
Mean absolute error	7.99990
Root mean square de-viation	8.91999
Average Accuracy of Test Data	80%

TABLE. 11 . Quantitative Performance indices of ARIMA model.

4 *Conclusion*

Effective actions are required to prevent air pollution, which is a major issue in India's main cities. In this study, the air quality index (AQI) for Delhi, Mumbai, and Chennai was forecast using an artificial neural network (ANN) model. Using data on the air pollutants PM10, PM2.5, NO_2, SO_2, and CO for the years 2019 and 2020 the ANN model was trained and verified.

According to the study the ANN model is a trustworthy and useful method for forecasting the AQI of these cities. According to the anticipated AQI readings for 2021. In order to reduce the negative effects of air pollution on both the environment and human health, effective actions must be taken in these cities. Policymakers can use the study's findings to create such measures.

More air contaminants and weather parameters could be added to the ANN model in future study to increase its accuracy. The effectiveness of various pollution control techniques in lowering air pollution in these cities must also be investigated. The study offers a framework that may be used to other cities around the world for employing ANN models to estimate the air quality index (AQI) of India's main metropolitan areas.

Additionally, enhancing these cities' air quality and safeguarding public health depend on the deployment of renewable energy sources, efficient rules and regulations, public awareness campaigns, and transportation networks. Reducing emissions from commercial, transportation, and residential sources is essential. Campaigns for public awareness can inform the public about the value of clean air and the steps they can take to prevent air pollution. The adoption of electric vehicles, the growth of public transportation networks and the promotion of sustainable energy sources like solar and wind power can all contribute to a reduction in emissions from the transportation sector. Promoting greenery and vegetation in urban areas can also help the air quality.

In conclusion, air pollution is a difficult issue in India's big cities that necessitates a diversified response. Researchers and policymakers can get important insights from using ANN models to create strategies that effectively tackle air pollution. The adoption of sustainable energy sources and transport systems, as well as effective rules and regulations, are crucial to enhancing these cities' air quality and safeguarding the general public's health.

5 References

1. Central Pollution Control Board. (2021). National Air Quality Index. Retrieved from http://www.cpcb.nic.in/national-air-quality-index/

2. , M., & , A. (2019). Prediction of air quality index of Delhi using artificial neural network. Journal of Environmental Management, 248, 109319.

3. , V. A., & Singh, S. K. (2021). Analysis of air quality in Mumbai city using artificial neural network. Journal of Air Pollution and Health, 2(1), 1-8.

4. Pal, R., & Pal, N. R. (2019). Artificial neural network approach for forecasting air quality index of Chennai city. Journal of Atmospheric and Solar-Terrestrial Physics, 195, 105137.

5. Singh, S. K., & , M. (2019). Prediction of air quality index of Mumbai using artificial neural network. Journal of Environmental Management, 240, 343-354.

6. , S., & Singh, R. K. (2020). Air pollution in Delhi: An analysis of the status and impact on human health. Environmental Science and Pollution Research, 27(22), 27329-27341.

7. World Health Organization. (2018). Ambient air pollution: A global assessment of exposure and burden of disease. Geneva: World Health Organization.

8., S., Wang, B., , J., & , X. (2018). Urban air pollution: A global health challenge. The Lancet Planetary Health, 2(10), e459-e461.

9. , S. S., & , R. R. (2021). Forecasting air quality index using an artificial neural network in Hyderabad city. Environmental Science and Pollution Research, 28(17), 21591-21602.

10. , B. R., Butler, T. M., & Lawrence, M. G. (2008). Evaluation of emissions and air quality in . Atmospheric Environment, 42(7), 1593-1606.