Environ Monit Assess (2023) 195:1180
https://doi.org/10.1007/s10661-023-11646-3

RESEARCH

AQI prediction using layer recurrent neural network model: a new approach

Shadab Ahmad · Tarique Ahmad

Received: 20 April 2022 / Accepted: 25 July 2023 / Published online: 10 September 2023

Abstract The air quality index (AQI) prediction is important to evaluate the effects of air pollutants on human health. The airborne pollutants have been a major threat in Delhi both in the past and coming years. The air quality index is a figure, based on the cumulative effect of major air pollutant concentrations, used by Government agencies, for air quality assessment. Thus, the main aim of the present study is to predict the daily AQI one year in advance through three different neural network models (FF-NN, CF-NN and LR-NN) for the year 2020 and compare them. The models were trained using AQI values of previous year (2019). In addition to main air pollutants like PM_{10}/$PM_{2.5}$, O_3, SO_2, NOx, CO and NH_3, the non-criteria pollutants and meteorological data were also included as input parameter in this study. The model performances were assessed using statistical analysis. The key air pollutants contributing to high level of daily AQI were found to be $PM_{2.5}$/PM_{10}, CO and NO_2. The root mean square error (RMSE) values of 31.86 and 28.03 were obtained for the FF-NN and CF-NN models respectively whereas the LR-NN model has the minimum RMSE value of 26.79. LR-NN algorithm predicted the AQI values very closely to the actual values in almost all the seasons of the year. The LR-NN performance was also found to be the best in post-monsoon season i.e., October and November (maximum R^2 =0.94) with respect to other seasons. The study would aid air pollution control authorities to predict AQI more precisely and adopt suitable pollution control measures. Further research studies are recommended to compare the performance of LR-NN model with statistical, numerical and computational models for accurate air quality assessment.

Keywords Air quality index · Air pollution · AQI prediction · Human health · LR-NN · Particulate matter

S. Ahmad
Department of Civil Engineering, Bharat Institute of Engineering and Technology, Hyderabad, Telangana, India

T. Ahmad (✉)
Department of Civil Engineering, College of Engineering, Jazan University, Jazan 45142, Saudi Arabia
e-mail: trahmad@jazanu.edu.sa; tariqueahmadamu@gmail.com

Introduction

Environmental pollution particularly the air pollution is a major problem and continues to pose serious threats to health and well beings of the public globally and especially in metropolises (Angelevska et al., 2021; Elsunousi et al., 2021; WHO, 2016). It is estimated that air pollution causes the death of around 1 in 8 people globally (Elsunousi et al., 2021). In the year 2016, 4.2 million preterm deaths worldwide and 0.6 million in India were primarily attributed to atmospheric pollution (Ghude et al., 2016; Lelieveld et al., 2015; WHO, 2016). In major Indian cities, such as Delhi, some of the highest levels of air pollution concentrations have been recorded (Kumar et al., 2015). Air pollution exceeds the danger limits when unplanned urbanization takes place threatening the human health or decreasing the quality of life (Cetin & Sevik, 2016). With a rapidly growing population and expanding residential, industrial, and transportation infrastructure, Delhi city has been observed to have the poorest air quality in the world. (Kumar et al., 2017; WHO, 2016). According to the studies by Kandlikar and Ramachandaran (2000), inhabitants of Delhi are 12 times more prone to serious health issues in comparison to inhabitants of other cities of the country. The particulate matters ($PM_{2.5}$ and PM_{10}) are found to be the major air pollutants affecting the air quality, public health, and weather condition. The prolonged exposure to particulate matter has reduced the average life expectancy by about 3 years in India and about 6 years Delhi (Ghude et al., 2016). $PM_{2.5}$ is considered as the hazardous air pollutant and its long-term exposure causes serious health issues like cardiac and bronchial problems, preterm deaths and reduced infant weight after birth (Brook et al., 2010; Coker et al., 2016; Lippmann, 2014). The annual average $PM_{2.5}$ concentration in Delhi is observed to be 200% higher than the limit of 40 μg/m^3 as per National Ambient Air Quality Standards (NAAQS) thus, posing serious risks to public health (Singh et al., 2021). North-western dust-laden winds, vehicle emissions, diesel-generator fumes, open construction sites, power plants, road dust, and biomass burning in the city and neighbouring states are the main sources of particulate pollution in Delhi (Kumar et al., 2017; Pant et al., 2015; Saxena et al., 2017; Sharma et al., 2016; Villalobos et al., 2015). The other PM sources include oxidation and condensation of precursor gases and their products in the atmosphere. Other major air pollutants affecting the environmental condition of Delhi include sulphur dioxide (SO_2), carbon monoxide (CO), oxides of nitrogen (mainly NO and NO_2) and ozone (O_3). These gases are emitted from thermal power plants, automobiles, industries etc. Earlier studies have reported that about 80–90% of CO and NO_X are emitted from the road traffic and transport services in Delhi (Gulia et al., 2015; Gurjar et al., 2004; Tyagi et al., 2016). Higher concentration of these pollutants and their prolonged exposure can be life threatening, causing headache, dizziness, breathing problem and even heart attacks (Kunzli et al., 2000). Meteorological factors like wind velocity and wind angle/direction are also responsible for higher concentrations of air pollutants in the ambient air of Indian cities (Guttikunda & Gurjar, 2012; Tiwari et al., 2014; Yadav et al., 2014; Yadav et al., 2016). Winter season in Delhi is marked with low wind speeds, temperature lapse rate inversion, lower mixing height and poor ventilation resulting in negligible dispersion and higher concentration of locally emitted air pollutants (Tiwari et al., 2013). Whereas, during summer, higher temperature along with strong winds, causes the increased dispersion of emitted pollutants and thus resulting in lesser air pollution. Hence, monitoring, modelling, and forecasting of the pollutants in the air is required to be carried out by the environmental agencies.

The air quality index, or AQI, is a statistical measurement that combines the concentrations of different pollutants into a single numerical form and is used to assess the quality of the air we breathe and associated health risks (Zhu et al., 2017). The AQI of an area indicates its air pollution level at a given time (Rahman et al., 2017). It alerts the local authorities as well as inhabitants about the extent of pollution in the ambient air and related health issues. Therefore, AQI prediction in advance is vital to aid the planning of corrective actions and regulations to reduce public health problems. Hence, there is utmost need to develop a consistent AQI prediction model so as to enhance the community cognizance about severe air pollution incidents and pre-informs the susceptible section of human population to minimize their exposure time.

Meteorology and emission source of pollutants are two basic factors affecting the air quality indices and can be used for prediction of spatial and temporal pollutant concentrations and AQI values. Earlier studies have used statistical models for predicting the AQI. However, the statistical models using basic mathematical equations were unable to represent the non-linear relationship between different variables in forecasting (Afzali et al., 2012) as the AQI prediction involves the dynamic processes in which various parameter changes continuously. Artificial neural network (ANN) technique can be used to improve the prediction accuracy with respect to statistical models used previously (Nagendra & Khare, 2005). Different ANN models which have been reported in similar Machine Learning Based (MLB) studies include multilayer perceptron

(MLP) (Durão et al., 2016; Wang & Lu, 2006), back-propagation neural network (BP-NN) (Bai et al., 2016; Chen & Pai, 2015), layer recurrent neural network (LR-NN) (Choi et al., 2017; Yu et al., 2019), radial basis function (RBF) (Iliyas et al., 2013), and adopted neuro-fuzzy inference systems (ANFIS) (Prasad et al., 2016; Shahraiyni et al., 2015). ANN models have been extensively studied and applied in air quality prediction and modelling of environmental systems and are found to outperform the statistical models with superior flexibility, accuracy and efficiency (Azid et al., 2013; Barai et al., 2007; Demuth et al., 2009; Krzysztof & Osowski, 2016; Muhammad et al., 2015; Niharika & Rao, 2014; Rahman et al., 2016; Taneja et al., 2016; Wang et al., 2003). However, it should be kept in mind that a single ANN model developed cannot be effective for each pollutant and for every location. Its performance can be strongly impacted by the choice of its architecture, which includes the quantity and variety of neurons, as well as the choice of a learning algorithm, and each of these factors must be examined separately for each scenario.

Research gap and motivation

The variation in national standards for air quality in different countries have led to the formation of different AQI scales. Thus, apart from the forecast of the air pollutants, presently the AQI prediction problems are appealing to the researchers. Table 1 summarizes the AQI prediction carried out in previous studies. It is clear from Table 1 that the methodologies adopted for AQI prediction varies in several aspects. The first difference is the use of meteorological data in the data sets used to develop AQI models. Arnaudo et al. (2020), included meteorological parameters like humidity, air velocity and temperature along with air pollutants data for developing an AQI model. However, meteorological parameters were not taken into account in the studies of Veljanovska and Dimoski (2018). Another aspect is the use of technique to develop the prediction model. Some researchers (Nimesh et al., 2014; Zhu et al., 2017) have used multiple linear regression and statistical prediction models like ARIMA. Others have studied the MLB modelling techniques. The AQI prediction data are of non-linear nature and vary over time and also tend to be inconsistent, heterogeneous, uncertain and missing. These characteristics of data demonstrate that the models using machine learning techniques, incorporating uncertainty might work well for AQI prediction (Hajek & Olej, 2015). Therefore, AQI prediction modelling needs to be done through optimum machine learning technique which incorporates both air quality and meteorological factors.

In this study, machine learning based prediction models were developed using FF-NN (feed forward back-propagation neural network), CF-NN (cascade forward back-propagation neural network) and LR-NN (layer recurrent neural network) techniques including

Table 1 AQI prediction reported in earlier studies

Reference	Technique				Meteorological data	Data sample size	Location
	Statistical	Hybrid	MLB	MLE			
Hajek and Olej (2015)			√		√		Czech Republic
Qin et al. (2014)			√			1 year	China
Liu et al. (2019)			√	√		9358	China
Castelli et al. (2020)			√		√	2 years	USA
Nimesh et al. (2014)	√					1 year	India
Veljanovska and Dimoski (2018)		√				1 year	Macedonia
Wu and Lin (2019)		√				761	China
Arnaudo et al. (2020)	√		√	√	√	3 years	Italy
Koo et al. (2020)	√		√			6 years	Malaysia
Zhu et al. (2017)	√	√	√		√	10 years	USA
Xi et al. (2015)			√	√	√	3 years	China

MLB Machine learning based, *MLE* Machine learning ensemble the metrological data to predict hourly AQI one year in advance for the year 2020. The models were trained using previous year AQI data. The study was conducted for Okhla Phase II area, New Delhi city, which suffers from some of the worst air quality scenario worldwide owing to a rapidly growing population, expanding residential, industrial, and transportation infrastructure and dust storms from north-western part of India. (Kumar et al., 2015). The subsequent analysis focuses on inter-comparison of developed models using statistical analysis to evaluate the best model performance.

Main contributions of the present study are as follows: (1) Not only the air pollutants data but also the meteorological data are assessed as input parameters in the development of ANN models for AQI prediction. (2) In addition to basic machine learning models such as (MLP/FF-NN) and cascade model (CF-NN), a layer recurrent model (LR-NN) is used in this study for AQI prediction. LR-NN model has not been used so far for AQI prediction as observed through literature survey. (3) Prediction performance of all ANN models (FF-NN, CF-NN and LR-NN) are also compared. (4) It has been found that the LR-NN model used for AQI prediction gives superior prediction performance in comparison to traditional ANN models.

Methodology

Study area

Delhi city (28.61°N, 77.23°E), the national capital territory (NCT) of India is reported to be the most polluted city in the whole world (Hoshyaripour et al., 2016). It is among the biggest metropolitan cities of India with a huge population load of 31.18 million (WPR, 2021). The vehicular population in Delhi city in 2021 was reported to be 11.89 million with 1.749 km of road length per 100 km^2 (Statista, 2022). Being a metro-city of developing country, it confronts with enormous problems of traffic management, congestion and subsequent failure in productivity. Land locked from all sides, it has little possibilities of pollutant dispersion ending up in huge pollution events (Masood et al., 2018). With unprecedented increase in construction activities, large number of industries, increase in residential and vehicular population, there are limited probabilities of air pollution clearance, resulting in high level of air contamination in city-centre. The location of the work area is displayed in the Fig. 1.

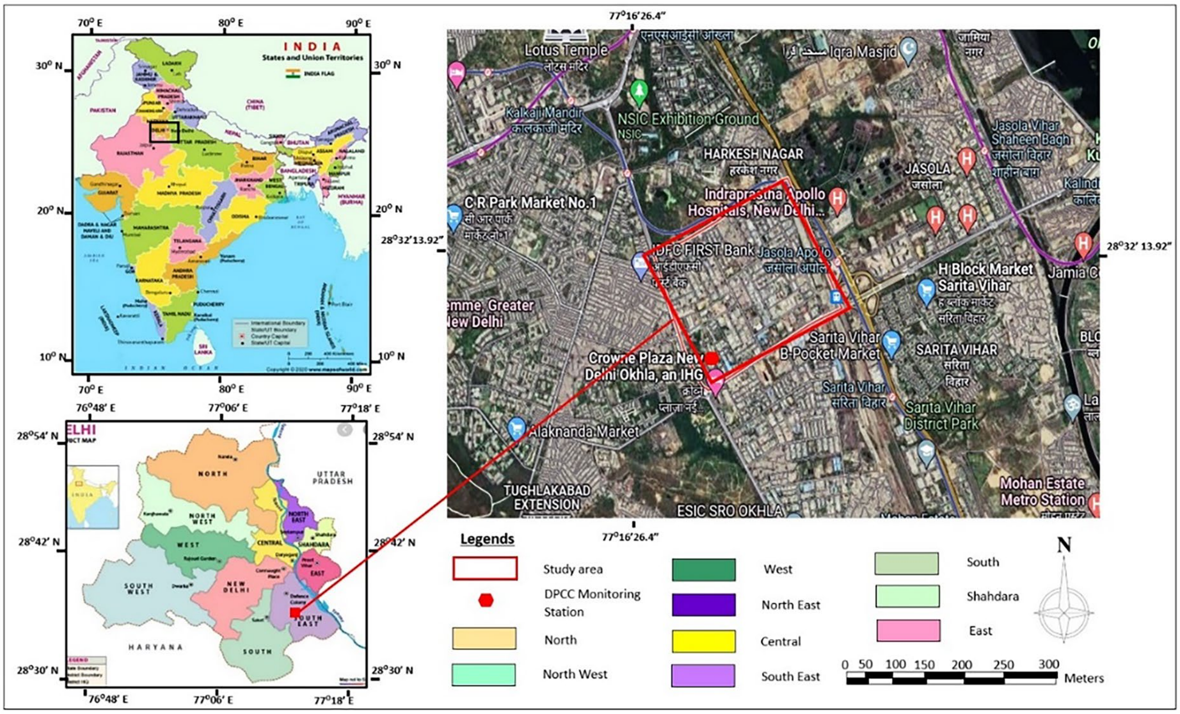

Fig. 1 Map showing the CPCB monitoring station located in Okhla Phase II, south-east Delhi (Map not to scale)

Table 2 Input meteorological conditions and observed criteria air pollutants

Variable	Unit	Range	Mean	SD
$PM_{2.5}$	μg/m^3	[8.15–551.9]	107.68	94.95
PM_{10}	μg/m^3	[15.52–652.03]	214.46	128.17
NO	μg/m^3	[1.85–242.14]	42.45	44.85
NO_2	μg/m^3	[10.53–101.67]	39.71	16.45
NO_x	ppb	[8.07–248.16]	55.97	44.18
Benzene	μg/m^3	[0.44–12.92]	4.63	2.36
Toluene	μg/m^3	[1.65–79.51]	33.45	13.36
RH	%	[16.32–99]	61.86	19.206
WS	m/s	[0.16–3.17]	1.04	0.43
WD	°	[63.94–331.25]	189.48	65.26
SR	W/mt^2	[4.42–218.31]	81.4	40.35
BP	mm Hg	[971.31–997.48]	984.81	7.09
AT	°C	[5.8–39.15]	25.67	7.78
SO_2	μg/m^3	[6.43–26.95]	12.58	3.35
NH_3	μg/m^3	[12.01–99.23]	33.21	14.44
CO max 8-hr	μg/m^3	[0.67–6.79]	2.04	1.04
O_3 max 8-hr	μg/m^3	[6.35–189.8]	49.35	28.04

Data source

Air pollution and meteorological data were procured from the online data monitoring portal of Central Pollution Control Board (CPCB) for a period of two years (January 2019–December 2020) (CPCB, 2021). The Continuous Ambient Air Quality Monitoring System (CAAQM-CPCB) (27°31′51.4560“ N, 077°16’16.4280” E) is located in front of *Maa Anandmayee Marg Road*, in Okhla Phase 2 locality as shown in Fig. 1. The sub-locality Okhla Phase 2, Okhla Industrial Area located in south- east Delhi has a population of about 66,820 and plan area of about 1.99 km^2. The daily average data of $PM_{2.5}$, PM_{10}, NO_2, NO, NO_x, CO, SO_2, O_3, NH_3, Benzene and Toluene and meteorological variables like WS (wind speed), WD (wind direction), SR (solar radiation), AT (atmospheric temperature), BP (barometric pressure), and RH (relative humidity) for the same period were used as primary input parameters in the ANN models development. The general statistics of the data taken into consideration in this study from Jan 2019 – Dec 2020 is shown in Table 2.

Air quality index

An air quality index (AQI) represents the level of air quality adopted and implemented by air pollution regulatory authority like CPCB in India. The objective of development of air pollution standards is to set a parameter for public health protection from air pollution toxicity and to mitigate the air pollutant hazards. It also helps the local and national agencies in adopting pollution control measures. AQI calculation converts weighted values of individual air pollutant concentrations into a particular numeral or a set of numerals. The final value of AQI is the maximum of all sub-index values calculated for six air pollutants: $PM_{2.5}$, PM_{10}, NO_2, SO_2, CO, O_3, NH_3 (CPCB, 2014). AQI values are categorized as good (0–50, green), satisfactory (51–100, light green), moderate (101–250, yellow), poor (251–350, light pink), very poor (351–430, red), and severe (430+, brown) as shown in Table 3.

The calculation of the AQI is accomplished according to Eq. (1) and (2):

$$I_p = \frac{I_{Hi} - I_{Lo}}{BP_{Hi} - BP_{Lo}}\left(C_p - BP_{Lo}\right) + I_{Lo} \tag{1}$$

$$\text{AQI} = \text{Max (Ip) (where; p} = 1,\ 2,\ \ldots,\ \text{n; denotes n number of air pollutants)} \tag{2}$$

Where,

Ip = Sub-index for pollutant Cp, based on 'linear segmented principle'.

Cp = Concentration of pollutant p,

BP_{Hi} = Break Point that is ≥ Cp,

BP_{Lo} = Break Point that is ≤ Cp,

I_{Hi} = AQI value referring to BP_{Hi},

I_{Lo} = AQI value referring to BP_{Lo};

If I_{Lo} is greater than 50, subtract one from I_{Lo}.

AQI was calculated from January – December 2019 at DPCC (Delhi Pollution Control Committee) monitoring stations Okhla Phase II, New Delhi (DPCC, 2020) as shown in Table 3.

Proposed AQI modelling methodology

The main objective of this study was to make an early forecast of daily AQI for a period of one year

Table 3 Indian AQI breakpoints and corresponding mass concentration

AQI Category	Range	PM_{10} 24-hr (μg/m³)	$PM_{2.5}$ 24-hr (μg/m³)	NO_2 24-hr (μg/m³)	O_3 8-hr (μg/m³)	CO 8-hr (mg/m³)	SO_2 24-hr (μg/m³)	NH_3 24-hr (μg/m³)
Good	0-50	0-50	0-30	0-40	0-50	0-1.0	0-40	0-200
Satisfactory	51-100	51-100	31-60	51-80	51-100	1.1-2.0	41-80	201-400
Moderate	101-250	101-250	61-90	81-180	101-168	2.1-10	81-380	401-800
Poor	251-350	251-350	91-120	181-280	169-208	10.1-17	381-800	801-1200
Very poor	351-430	351-430	121-250	281-400	209-748	17.1-34	801-1600	1201-1800
Severe	430+	430+	250+	400+	748+	34+	1600+	1800+

(covering all five seasons) using neural network models like multi-layer perceptron (MLP) or feed forward back-propagation (FF-NN), cascade forward back-propagation (CF-NN), and layer recurrent (LR-NN) neural network models. The AQI was initially calculated using daily average air pollutant concentrations. The models were initially trained based on 2019 air quality data and simulated for 2020 air quality data to predict daily AQI. The model with best statistical performance was finally selected.

The modelling was performed in three phases:

At first, the daily AQI was estimated for the year 2019 using daily average concentrations of $PM_{2.5}$, PM_{10}, O_3, NO_2, CO, SO_2 and NH_3 pollutants except for CO where daily 8-hour maximum concentration was taken.

1. Next, the ANN models viz. feed forward back-propagation neural network (FF-NN/MLP), cascade feed forward back-propagation neural network (CF-NN) and layer recurrent neural network (LR-NN) were trained using daily average air pollutant concentration for the year 2019. The daily average concentrations of air pollutants $PM_{2.5}$, PM_{10}, NO_2, NO, NO_x, NH_3, SO_2, O_3, CO, Benzene and Toluene as well as meteorological parameters RH, WS, WA, SR, BP and AT were taken as model *Input values* and the calculated daily AQI from step 1 as model *Target values*.
2. Finally, the daily air quality index for the year 2020 were predicted or simulated using the ANN models with daily average concentrations of air pollutants $PM_{2.5}$, PM_{10}, NO_2, NO, NO_x, NH_3, SO_2, O_3, CO, Benzene and Toluene as well as meteorological parameters RH, WS, WA, SR, BP and AT of year 2020 (covering all four seasons) as model *Input values* and the predicted daily AQI as model *Output values*.

The proposed study started by collecting data from CPCB website for the period of 2019–2020 (CPCB, 2021). After data collection, machine learning models were selected to predict the daily average AQI.

The model with best statistical performance was finally selected based on the parameters like coefficient of determination (R^2), mean error (ME), mean absolute error (MAE), normalized mean square error (NMSE), root mean square error (RMSE), fractional bias (FB), geometric mean bias (MG), geometric variance (VG), fractional variance (FS), Pearson's corelation coefficient (R), index of agreement (d), factor-of-two (Fa2).

The proposed methodology is shown in Fig. 2. The detailed description of the proposed method is in the following sub-sections.

Artificial neural network

A neural network employs artificial neurons, which are the tiniest data processing units (Sadorsky, 2006). ANNs are employed for the time series prediction and are the most famous machine learning programs (Zhang et al., 1998). These models are developed using an assembly of neurons which are inter-connected with several other neurons of the neural network. The input and output layers of these models are visible layers while in-between the two visible layers lies a hidden layer. The initial layer is always an input layer and the last layer is of output. The

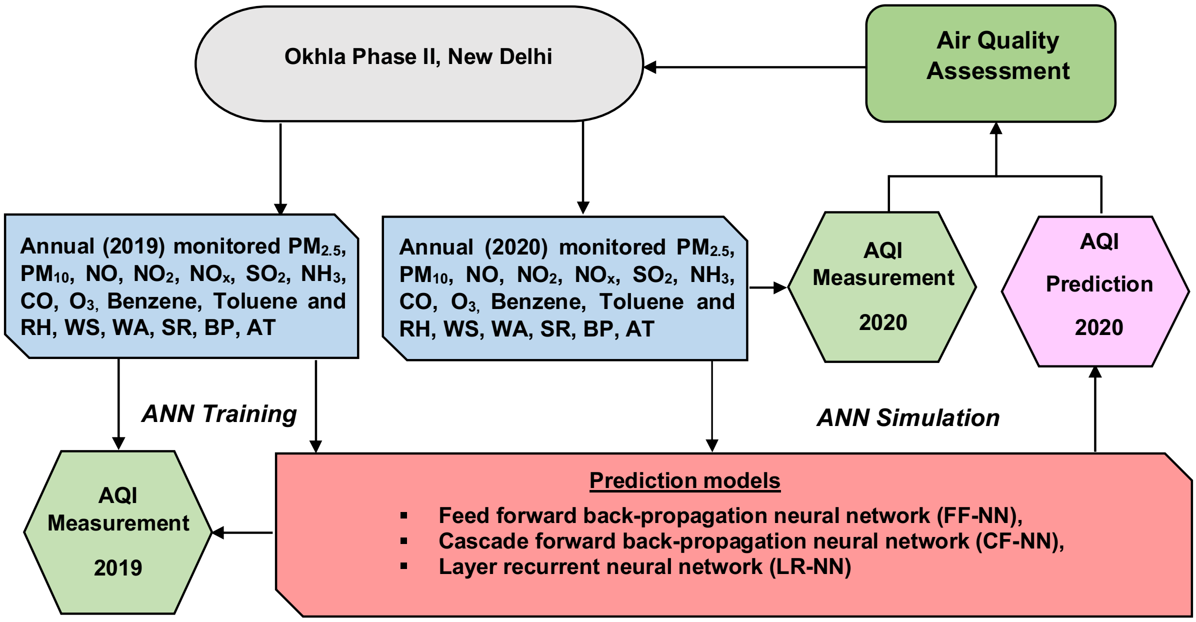

Fig. 2 Proposed AQI modelling approach

neural network prediction performance and complexity improves with the increase in number of neurons present in the hidden layer. The output calculated from such a neuron is calculated as:

$$y = f\left(\sum_{i=1}^{M} X_i W_{k_i} - \theta\right) \quad (3)$$

Where, X_i *are* number of inputs, θ *is* the threshold, W_i *are* the synaptic weights, and f *is* the activation function. The intensity of the connections is regulated by the synaptic weights depending upon the relationship. Many different activation functions are used like linear, sigmoid, hyperbolic tangent, etc.

The most basic ANN is FF-NN/MLP. The basic architecture of FF-NN/MLP, CF-NN and LR-NN models are described in the following sections.

Feed forward back-propagation neural network

A feed-forward back propagation neural network (FF-NN) or multi-layer perceptron (MLP) has been widely used due to its simple neural network structure (Podnar et al., 2002). In this network type, the error signal is transmitted across the network in the backward direction. The free variables of the network are adjusted so as to reduce the error statistically. Back propagation is generated by generalizing the gradient descent with momentum weight and bias learning function to multi-layer networks along with non-linear differentiable transfer functions. Biases and weights are adjusted with different gradient descent algorithms. The estimation of gradient is done by back-propagation of the calculations from the output layer to the first hidden layer. With appropriate training, the back-propagation network can generalize and generate satisfactory outputs from inputs analogous to training inputs. The back propagation learning algorithm is easy to implement and mathematically proficient. The two-layer FF-NN diagram with 17 inputs, 25 neurons in hidden layer and 1 output layer used for this study is displayed in Fig. 3.

Cascade feed forward back-propagation neural network

Cascade feed forward back-propagation neural network (CF-NN) is identical to FF-NN as shown in Fig. 4 with 17 inputs, 20 neurons in the hidden layer and 1 in the output layer used for this study. The first hidden layer receives weights from the inputs and each succeeding layer get weights from the input as well as all preceding layers. All layers include biases and the last layer of neuron in the network is the output (Hedayat et al., 2009). Though feed forward neural networks can practically learn any relationship between input and output variables, the CF-NNs with more complex layer structure might quickly learn more intricate input-output relationships and a comprehensive description of CF-NN can be explored in the previous literature (Haykin, 2009; Wozniak et al., 2015).

Layer recurrent neural network

The layer recurrent neural network (LR-NN) is an efficient network system. It effectively predicts air quality index since it takes a local feedback from output layer to preceding hidden layers along with a time delay during training process. The feedback loop exists within each layer except in the last. Figure 5 displays LR-NN neural network diagram with 17 inputs, 15 neurons in hidden layer and 1 output layer with connection feedback. Basically, during neural network training, the output signal coming from the recurrent neural network is connected to the output of the previous hidden layer. The total of outputs is used as an argument of the transfer function to refine the outputs in the next iterations. The LR-NN function is shown in Eq. 4, where u(k) and v(k) represent the input and output values in layers. $W_{u,i}$ and $W_{v,j}$ demonstrates weights within u and v, respectively. The final results in the form of output y(k) are calculated from Eq. (5), where f(v(k)) is considered as the transfer function.

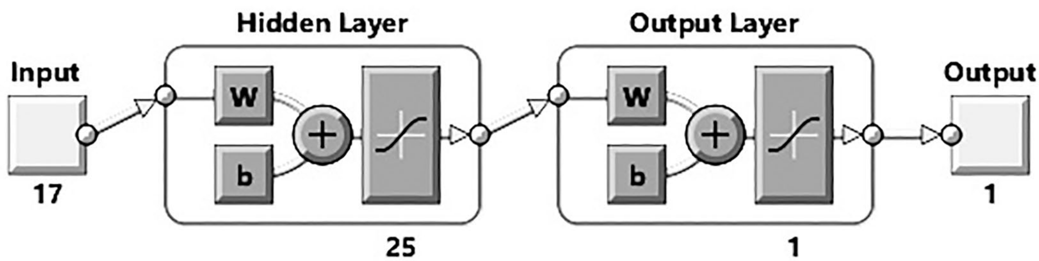

Fig. 3 Structural diagram of FF-NN

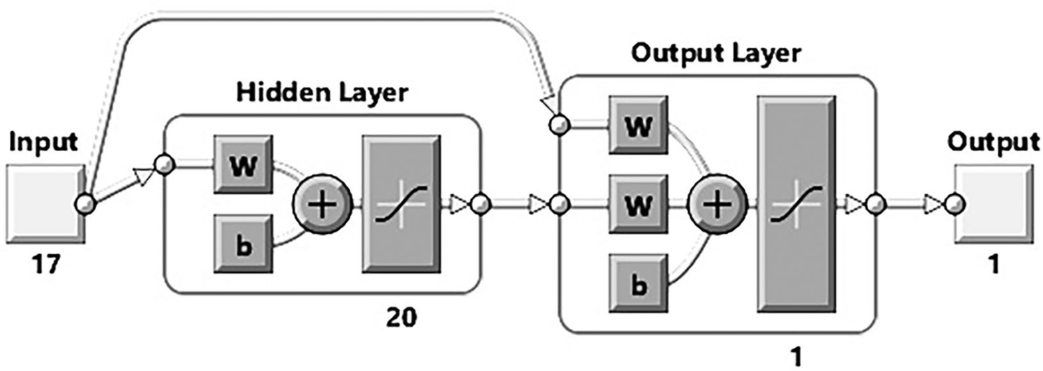

Fig. 4 Structural diagram of CF-NN

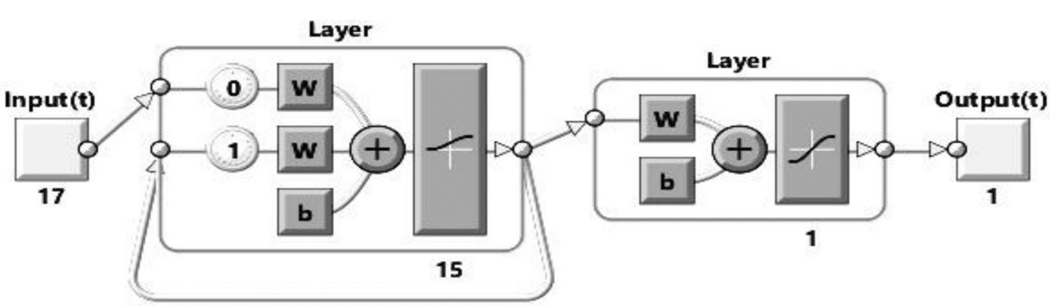

Fig. 5 Structural diagram of LR-NN

$$v(k+1) = \sum_{i=0}^{n} W_{u,i}(k)u(k) + \sum_{j=0}^{m} w_{v,j}(k)v(k) \quad (4)$$

$$y(k) = f(v(k)) \quad where : f(v(k)) = \frac{1}{1 + \exp\left(-(v(k))\right)} \quad (5)$$

To assess the model performances during the training stage, twelve statistical estimation criteria, R^2, MG, VG, R, d, Fa2, FS, RMSE, ME, MAE, FB and NMSE were adopted (Equations are given in Appendix). These criterions help in better evaluating the capabilities of the three models for the AQI prediction.

Correlation analysis

The assessment of air quality is based on the air pollutant concentration present in the atmosphere. There are numerous factors that influence the air quality and these factors also affect each other. In order to evaluate the correlation between the "two dusts and nine gases" concentration and the six meteorological parameters, we use Eq. (9) (in Appendix) to find the Pearson's correlation coefficient (R) between them, as shown in Table 9.

Results and discussion

The selected network structure, learning, training and transfer function used for each ANN model is displayed in Table 4. The three-layer neuron network structure (input-hidden-output) used for FF-NN, CF-NN and LR-NN models is 17–25-1, 17–20-1 and 17–15-1 respectively. All models were fed with same number of input variables. The number of neurons in the hidden layer for each model was selected after several simulations till the minimum error and gradient was obtained. The logistic sigmoid (LOGSIG) activation transfer function was used in first hidden layer and the second hidden output layer used linear transfer function for LR-NN and FF-NN models. For CF-NN model, the hyperbolic tangent sigmoid (TANSIG) transfer function was used. The training function that updates bias and weight values is according to Levenberg-Marquardt (TRAINLM) and the learning is according to gradient descent with momentum weight and bias learning function (LEARNGDM) for all models. The network diagram for FF-NN, CF-NN and LR-NN models is displayed in Fig. 3, Fig. 4 and Fig. 5.

Table 4 Network structure, learning, training and transfer function used for each ANN model

Network Type	FF-NN	CF-NN	LR-NN
Neuron Structure	17–25-1	17–20-1	17–15-1
Learning Algorithm	LEARNGDM	LEARNGDM	LEARNGDM
Training Function	TRAINLM	TRAINLM	TRAINLM
Transfer Function	LOGSIG	TANSIG	LOGSIG

The performance of models with statistical results is represented in Table 5. The LR-NN represents the best AQI prediction for year 2020 out of the three NN models. The model showed maximum R^2 value of (0.963). The R^2 provides a quantification of how well the actual concentrations are simulated by the model, based upon the amount of total deviations of model predictions from the measured data (Draper & Smith, 1998; Glantz & Slinker, 1990). The MG value from LR-NN model was estimated to be 1.03, representing ideal agreement with measured AQI. An "ideal" MG would give $MG = 1$. The $MG \geq 1$ signifies that model over predicted while for $MG \leq 1$ means under-prediction. The VG value was 1.022, the minimum value in all three models. The VG characterizes the scatter in the model forecast readings near to an average forecast reading. A perfect agreement between predicted and measured data-sets is observed for $0.7 \leq MG \leq 1.3$ and $VG \leq 1.6$ (Baumann-Stanzer & Piringer, 2011). The R value was found close to ideal value of 1 (0.981) for LR-NN with respect to other two models demonstrating best prediction correlation. R is a measurement of how the variations in predicted values follow the variations in measured values. In real, it is an evaluation of a straight-line correlation, for a plot of predicted values v/s measured values. It has a range of $-1 < R < 1$, where ± 1 indicates an ideal correlation and 0 indicates no correlation. It does not estimate correlation in real numerical values between monitored and forecasted data. The model predicted most accurately as indicated by the highest d value (0.989) estimation. The d value emphasizes the extent of closeness in precise assessment

Table 5 Performance statistics of FF-NN, CF-NN and LR-NN models

Statistical Parameter		FF-NN	CF-NN	LR-NN
R^2	$(0 \leq R^2 \leq 1)$ Ideal Value 1	0.943	0.961	0.963
MG	Ideal Value 1	0.964	1.056	1.030
VG	VG < 1.6	1.041	1.023	1.022
R	Ideal Correlation ±1	0.971	0.980	0.981
d	$(0 \leq d \leq 1)$ Ideal Value 1	0.984	0.833	0.989
Fa2	$(0 \leq Fa2 \leq 1)$ Ideal Value 1	1	1	1
FS	$(-2 \leq FS \leq 2)$ Ideal Agreement FS = 0	−0.071	−0.035	−0.038
RMSE	Ideal Value 0	31.858	28.033	26.797
ME	Ideal Value 0	0.060	−0.044	−0.019
MAE	Ideal Value 0	0.156	0.110	0.108
FB	$(-0.7 \leq FB \leq 0.7)$ Ideal Value 0	0.008	−0.052	−0.043
NMSE	$(NMSE \leq 0.5)$ Ideal Value 0	0.026	0.021	0.019

of measured variable with the predicted variable. Its range lies between 0 and 1, where 1 represents perfect relation between the measured and modelled, while, 0 represents no relation at all (Willmott, 1982). The Fa2 for all models were 100% within limits, with an ideal value of 1. The Fa2 factor is defined as the fragment of readings in which the ratio Pi/Oi is found within the limits of $0.5 \leq (Pi/Oi) < 2$. It has an extent of $0 < Fa2 < 1$, where $Fa2 = 1$ indicates that all predictions are in the limits of 0.5 to 2 and $Fa2 = 0$ indicates that none of the predictions are in this limit. The FS estimates the correlation within the deviation of the predicted values and the deviation of the measured values. Its range is $-2 < FS < 2$, where (−2) signifies maximum under-prediction in variance and (+2) signifies maximum over-prediction while $FS = 0$ denotes ideal agreement. In the case of LR-NN, FS value is close to 0 (−0.038) indicating very slight under-prediction of variance with measured AQI. The RMSE value of LR-NN was found to be the lowest (26.797) in comparison to others. The accuracy of predicted outputs with respect to measured is indicated by RMSE. It is the most significant parameter for evaluating the prediction efficiency of model. The lower the RMSE value, the better is the model prediction efficiency. It represents the model performance in terms of its deviations and accurateness. Similarly MAE and NMSE value for LR-NN was also estimated to be the least with a value of 0.108 and 0.019. The MAE is the mean value of the absolute errors. The MAE shows an anticipated extent of error from the mean of forecasted values. The NMSE calculates the correlation between each measured and forecasted values point wise (micro statistics). It has a range of 0 to infinity, where 0 signifies perfect relation between measured and forecasted data. The performance of model in representing actual data is assumed to be satisfactory if the NMSE value is below 0.5 (Dubey et al., 2013; Raducan & Stefanescu, 2012). The FB value of the model was nearest to ideal value of 0 (−0.043), indicating negligible under-estimation of the AQI values. The FB is basically a statistical indicator which signifies that the predicted values are higher or over-estimated if $FB > 0$, similarly if $FB < 0$, the predicted values are lower or under-estimated than the observed values. A FB value = 0 signifies ideal relation between observed and forecasted values. The performance of model is acceptable if $-0.7 < FB < 0.7$ (Dubey et al., 2013). The graph displaying the daily modelled AQI from LR-NN model and estimated daily AQI for the year 2020 is shown in Fig. 6.

As observed from Fig. 6, the AQI values are found to be in hazardous range during winter season, moderate in summers and satisfactory in monsoon season. The AQI values fall in very poor category (351–430) from October to January and severe category (430+) in October and November 2020. In October and November, each year AQI of Delhi city is observed to be severe. The monsoon season ends during the month of October in north-western part of India. During monsoon season, the prevalent easterly winds coming from Bay of Bengal, carry significant moisture bringing rainfall in these landscapes. Once the monsoon ends, the major winds are north-western carrying dust storms from Rajasthan and at times

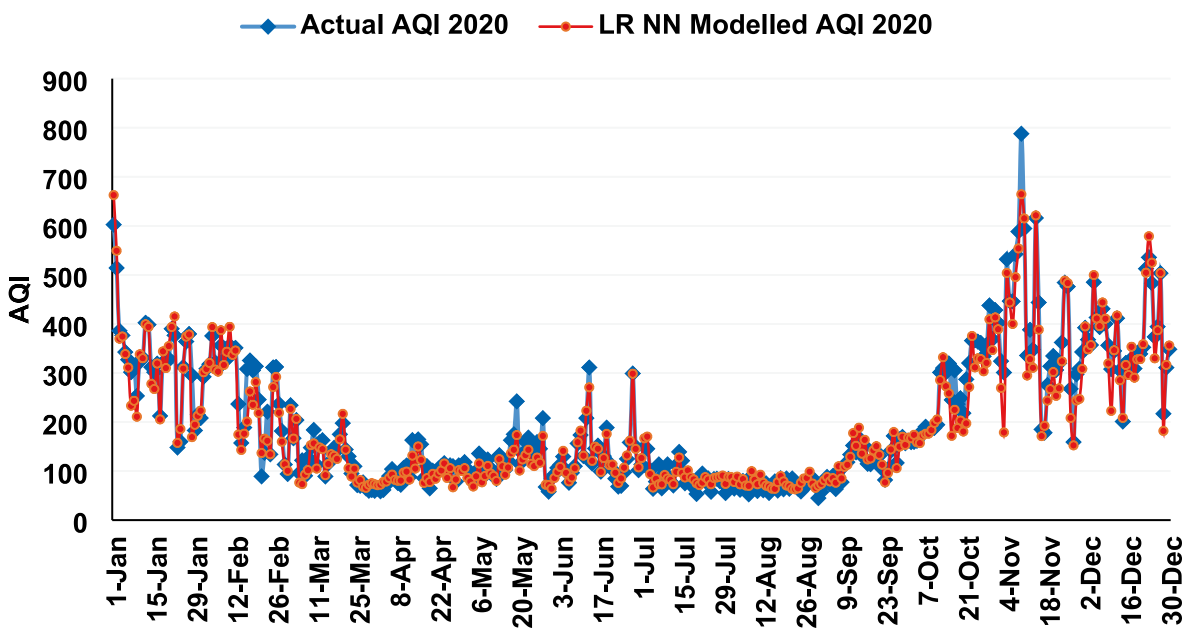

Fig. 6 Average daily AQI predicted by LR-NN model and actual AQI for the year 2020

Pakistan and Afghanistan. As per the peer reviewed studies conducted by scientists of National Physical Laboratory, about 72% of winds in the region is from northwest direction while 28% from eastern Indo-Gangetic plains (Joshi, 2020). Additionally, the drop in ambient temperature during winter season is another cause of higher air pollution levels. During winter season inversion of atmospheric temperature lapse rate causes trap of air pollutants near to ground. Low wind speeds along with lapse rate inversion reduce dispersion process considerably ending up very high level of air pollutants in the city environment. With external air pollutant sources like dust storms and farm fires added into the already deteriorated air quality of the city, the air pollution level further increases. Vehicular emissions are another major source of high ambient pollution in winters. As per recent study, vehicular emissions contribute about 30% of total $PM_{2.5}$ during winters (Bhandarkar, 2013).

The percentage of season wise descriptors of AQI is shown in Table 6, which indicates 'poor' (25%, 36%, 43%) in summer, post-monsoon and winters while 'very poor' (14.7%, 33%) and 'severe' (19.6%,

Table 6 Percentage of season wise AQI description for the year 2019

S.No.	Index values	Description	Spring (Feb-Mar), (%)	Summer (Apr-June), (%)	Monsoon (Jul-Sep), (%)	Post-monsoon (Oct-Nov), (%)	Winter (Dec-Jan), (%)
1	0–50	Good	–	–	1%	–	–
2	51–100	Satisfactory	34%	22%	66%	–	–
3	101–250	Moderate	63.7%	48.3%	32.6%	29.5%	13%
4	251–350	Poor	2%	25%	–	36%	43%
5	351–430	Very poor	–	5%	–	14.7%	33%
6	430+	Severe	–	–	–	19.6%	11%

11%) in post-monsoon and winter season respectively. The post-monsoon and winter have the highest percentage. The reason of this high percentage may be the frequently occurring calm winds and low boundary layer heights in winter and accumulation of dust particles (RSPM and SPM) in summer, which is originating from Rajasthan, situated in west to Delhi City. It is also noticeable that most of the times the winds are blowing from west to northwest directions in summer. The percentage of 'moderate' and 'poor' descriptors is observed in post-monsoon and monsoon seasons, which may be due to the washing out of the pollutants by precipitation. The AQI of weekdays and weekends has also been analysed to assess the effect of vehicular traffic. Saturday and Sunday are weekends and remaining days (Monday– Friday) are weekdays. The differences between the seasonal average of AQIs of weekdays and weekends are found to be very small as (118.4, 122.4); (93.3, 91.1); (322.9, 346); (189.3, 174.4) and (337.8, 372.7) in summer, monsoon, post-monsoon, spring and winter, respectively, which are considered negligible.

Therefore, weekdays and weekends are treated the same in the model. The forecasting of daily AQI has trained using neural network models on the seasonal basis for the period 2019 and validated through the daily AQI of 2020. Firstly, the architectures for the different seasons of summer, monsoon, post-monsoon and winter have been developed using the neural network technique on the basis of daily data of 2019 as already discussed above in the Methodology section. Hence, the forecasted values of daily AQI in the year 2020 have been compared statistically with the observed values of the same year as shown in Table 7. Although, the normalized mean square errors (NMSE) are found to be almost same in all the seasons while minimum root mean square error (RMSE) in monsoon season. The same table also reflects that the model is predicting the AQI satisfactorily in all seasons while most accurate prediction in post-monsoon season as per the values of coefficient of correlation (R).

Table 7 Performance of LR-NN model with observed AQI in the year 2020

S.No.	Season	2020 validation period			
		RMSE	NMSE	Correlation coefficient	Fractional bias
1.	Spring	5.27	0.02	0.92	−0.05
2.	Summer	4.56	0.01	0.88	−0.07
3.	Monsoon	3.65	0.02	0.83	0.04
4.	Post monsoon	6.59	0.02	0.94	−0.09
5.	Winter	7.91	0.03	0.63	0.00

The architecture of the LR-NN model in all five seasons, based on the transferred data of the year 2020, has been developed and evaluated statistically. The model's daily forecasting is compared with observed AQI values in the form of scatter plots in five different seasons for the years 2020 as shown in Fig. 7. The coefficients of determination (R^2) have values of 0.92, 0.88, 0.83, 0.94 and 0.63 in spring, summer, monsoon, post monsoon and winter season respectively, while the critical values for correlation coefficient at the 99-percentile confidence interval using student t-test for spring, summer, monsoon, post-monsoon and winter are 27.3, 27.9, 22.3, 30.7 and 10.1 respectively. The highest t-value is observed in post-monsoon season indicating maximum significance between predicted and observed AQI values.

The high, low and episodic air pollution events have been taken into account through observed concentrations of pollutants, which have been used for estimation of AQI at particular stations. The episodic air pollution events are occurring only on some special occasions, e.g., Diwali, it is only a two day festival in the month of Oct/Nov in every year and people play with firecrackers, which emit more pollutants in the atmosphere. The episodic air pollution events have not been considered in the present study as its occurrences are very limited. However, the characteristics of high and low air pollution events are by different ranges of AQI as (301–500) and (0–300), respectively. The performance of both the events have been analysed statistically, which show the RMSE and the correlation coefficient for high air pollution events are (6.2, 0.8); (0, 0); (0, 0); (6.9, 0.96) and (7.33, 0.83) in spring, summer, monsoon, post-monsoon, and winter respectively. Although, RMSE and the correlation coefficient for low air pollution events are (4.62, 0.93), (0, 0); (0. 0); (5.92, 0.81) and (9.77, 0.47) in spring, summer, monsoon, post-monsoon and winter respectively, which indicate that the model is performing satisfactory in training the data for post monsoon season during high air pollution event and

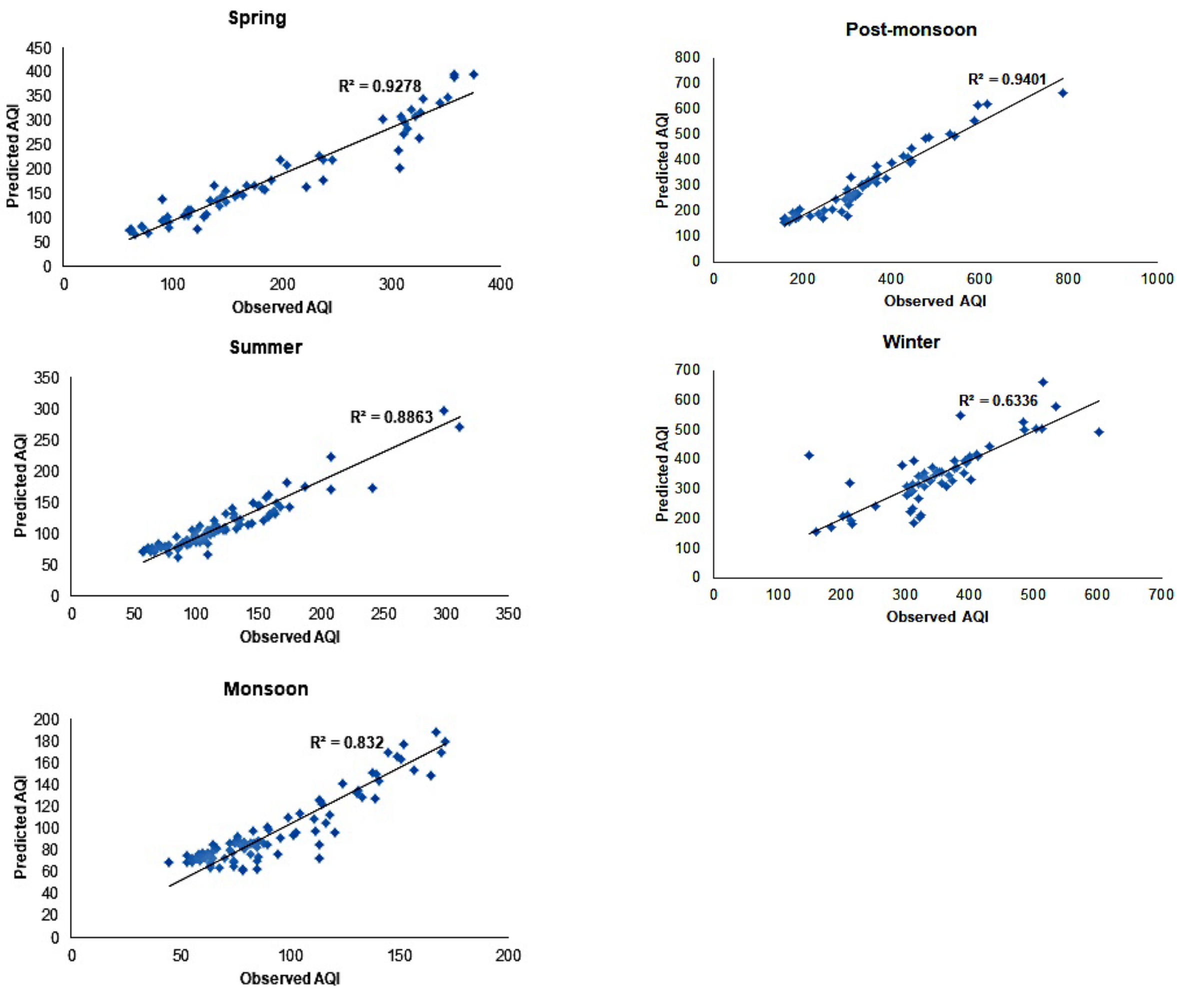

Fig. 7 Scatter plots between observed and predicted AQI for the year 2020

spring season during low air pollution event. Therefore, the same procedure has been followed in validation of the model through the daily AQI of year 2020 and shown in the form of scatter plots in Fig. 7 for different seasons. Figures 7 shows that forecasted and observed AQI in the year 2020 have the maximum R^2 (0.94) value in post-monsoon season.

The statistical analysis of the model's validation in 2020 has been shown in Table 5, which reveals that the model is performing satisfactory with respect to NMSE in summer, monsoon, spring, post-monsoon and winter in decreasing order. According to the fractional bias (FB) the model prediction is satisfactory for all seasons ($0.7 \leq FB \leq -0.7$) (Misra et al., 2013).

To check the percentage contribution of each input parameter (from both air quality and meteorological data) given to ANN models on estimated daily AQI, regression analysis was performed between each parameter and daily AQI of the year 2020 (Table 8). The significance of relation between each individual input parameter with daily AQI was assessed by p value, t-test, R^2 and percentage contribution of each t-value. The relation is acceptable if p value is less than 0.5 and t-value less than −2.21 or greater than +2.21. The results are presented in Table 5. It was found the out of the six criteria air pollutants ($PM_{2.5}$, PM_{10}, CO, NO_2, SO_2 and O_3) prescribed by USEPA, the particulates/dust ($PM_{2.5}$/PM_{10}) have a total contribution of 40% on daily AQI, while remaining pollutants except ozone (NO_2, CO and SO_2) have contributed 20% in total. The ozone (O_3) has no contribution (0%) on daily AQI. The hazardous air

Table 8 Relative significance (p value and t-value) and relative importance (%) of input variable on AQI for year 2020

	AQI vs. Input Parameters	p value	t-value	R^2	% Effect
1	$PM_{2.5}$ (μg/m^3)	$9.9778e^{-207}$	67.123	0.925	19.74
2	PM_{10} (μg/m^3)	$1.6136e^{-204}$	66.114	0.923	19.43
3	Benzene (μg/m^3)	$3.0319e^{-120}$	35.538	0.776	10.14
4	NO_2 (μg/m^3)	$4.2063e^{-105}$	31.289	0.729	8.84
5	NO_x (ppb)	$2.27466e^{-89}$	27.138	0.669	7.58
6	CO (μg/m^3) 8 hr-max	$2.73869e^{-76}$	23.845	0.61	6.58
7	NO (μg/m^3)	$5.49088e^{-75}$	23.523	0.603	6.48
8	BP (mmHg)	$7.35539e^{-46}$	16.458	0.427	4.33
9	AT (°C)	$7.31065e^{-45}$	−16.21	0.42	4.26
10	SO_2 (μg/m^3)	$2.70856e^{-36}$	14.096	0.353	3.61
11	SR (W/mt^2)	$2.26421e^{-35}$	−13.86	0.346	3.54
12	Toluene (μg/m^3)	$1.704e^{-33}$	13.387	0.33	3.40
13	WD (°)	$2.29432e^{-09}$	6.1299	0.093	1.19
14	WS (m/s)	$6.43136e^{-05}$	−4.044	0.043	0.56
15	NH_3 (μg/m^3)	0.001449703	3.2092	0.027	0.30
16	RH (%)	0.077224679	−1.772	0.008	0.00
17	O_3 (μg/m^3) 8 hr. max	0.15086345	−1.44	0.005	0.00

Table 9 Pearson linear correlation coefficients between air pollutant concentrations and meteorological variables

	PM2.5	PM10	CO	NO2	SO2	OZONE	NOx	NO	Benzene	NH3	Toluene	WD	WS	BP	AT	SR	RH
PM2.5	1	0.93	0.77	0.81	0.55	-0.05	0.80	0.77	0.86	0.13	0.57	0.24	-0.22	0.63	-0.62	-0.55	-0.05
PM10		1	0.79	0.87	0.65	0.03	0.84	0.8	0.85	0.17	0.64	0.29	-0.26	0.55	-0.5	-0.5	-0.21
CO			1	0.77	0.58	0.08	0.82	0.8	0.75	0.05	0.61	0.24	-0.26	0.5	-0.44	-0.48	-0.15
NO2				1	0.6	-0.01	0.88	0.82	0.81	0.00	0.74	0.42	-0.38	0.55	-0.52	-0.59	-0.15
SO2					1	0.21	0.61	0.59	0.52	0.08	0.43	0.3	-0.09	0.47	-0.27	-0.08	-0.58
OZONE						1	0.01	0.02	-0.07	-0.09	0.11	-0.03	-0.1	-0.16	0.33	0.21	-0.41
NOx							1	0.99	0.86	0.05	0.76	0.28	-0.32	0.55	-0.52	-0.53	-0.09
NO								1	0.84	0.06	0.74	0.24	-0.29	0.52	-0.5	-0.49	-0.07
Benzene									1	0.13	0.66	0.28	-0.23	0.6	0.63	-0.63	0.03
NH3										1	0.00	-0.2	0.04	-0.1	0.01	0.13	0.07
Toluene											1	0.03	-0.51	0.2	-0.21	-0.51	0.01
WD												1	0.12	0.4	-0.31	-0.11	-0.34
WS													1	0.15	-0.12	0.31	-0.11
BP														1	-0.89	-0.49	-0.03
AT															1	0.66	-0.26
SR																1	-0.45
RH																	1

1.00 0.90 0.80 0.70 0.60 0.50 0.40 0.30 0.20 0.10 0

pollutant (benzene) and non-criteria pollutants (NO_x and NO) have a total contribution of about 25% on daily AQI.

Table 9 shows the Pearson linear correlation coefficients (indicated as matrix colour block diagram) between six criteria, five non-criteria air pollutant concentrations and six meteorological variables. It can also be seen from Table 9, that except for O_3, RH and NH_3, all other variables have significant correlations with each other, indicating that the factors affecting each pollutant concentrations are very complex. The correlation coefficient between $PM_{2.5}$ concentration and PM_{10} concentration is as high as 0.93, indicating a high positive correlation between the two, and the correlation coefficient between atmospheric temperature (AT) and $PM_{2.5}$ is −0.62, which indicates that the higher the temperature, the lower the $PM_{2.5}$ concentration. Table 9 is also a matrix colour block diagram between the concentration of "two dusts and nine gases" and six meteorological parameters, which visually shows the correlation coefficients between the variables. The matrix colour block represents the absolute value of the correlation coefficient. As the colour becomes darker, the value of the correlation coefficient gradually increases.

LR-NN modelling methodology for daily AQI prediction has not been performed till date. The modelling results prove LR-NN superiority over other conventional modelling techniques namely FF-NN and CF-NN. The modelling technique would aid air pollution regulatory agencies in forecasting yearly AQI. It will assist in timely pre-planning, control and regulation of policies to save public health. It shall also pre-inform susceptible section of locals about severe air pollution incidents to minimize their exposure time.

Conclusions

Air pollution is one of the major issues of urban cities across the world. Therefore, a steady and precise AQI prediction model is needed to pre-inform the susceptible section of the urban community and concerned authorities about adverse ambient air condition. Among different models used in this study, the LR-NN model predicted the AQI value very close to the actual AQI. It is observed that the LR-NN algorithms performed better than other algorithms with minimum values of ME, MAE, FB, NMSE, RMSE, FS and maximum values of R^2, MG, R, d respectively. The LR-NN model demonstrated the quickest convergence in this study. The model produced best results with same number of input variables. The key air pollutants contributing to high level of daily AQI were $PM_{2.5}$/PM_{10}, CO and NO_2. The statistically error analysis of LR-NN model predictions with previous year AQI data for five seasons of year showed that models performed satisfactorily with best in post-monsoon season relative to other seasons (maximum R^2 =0.94). The Levenberg–Marquardt back-propagation learning algorithm with logistic sigmoid activation transfer function in the hidden layer and linear transfer function in output layer was found to be best in all the five seasons. The uniqueness of the proposed modelling methodology is due to the conjugate gradient descent with momentum algorithm used in training to estimate the optimal weights and biases. Thus, LR-NN modelling technique is an effective tool for forecasting daily AQI one year in advance. The study would aid air pollution control authorities in formulating suitable pollution control measures. It would also assist in providing necessary details to general public to safeguard their health and take necessary preventive actions.

LR-NN is also a time series model similar to other models and do not contain the information of air pollutant sources and dispersion parameters. However, this limitation can be overcome, if above statistical models could be combined with air dispersion model for air quality prediction in more comprehensive studies and at different locations. Further research studies are recommended to compare the performance and effectiveness of LR-NN model with statistical, numerical and computational models to enable authorities to select a suitable toolkit for the precise decision making under the adverse effects of urban air pollution.

Acknowledgements The authors would like to acknowledge the Central Pollution Control Board (CPCB), Delhi for the availability of air pollution and meteorological data on the website. The authors also express gratitude to Professor Farhan A. Kidwai and Professor Kafeel Ahmad, Department of Civil Engineering, Jamia Millia Islamia, New Delhi, India, for their technical and moral support.

Authors' contributions All authors have contributed to this study. Conceptualization: [Shadab Ahmad]; Methodology: [Shadab Ahmad], Formal analysis and investigation: [Shadab Ahmad]; Writing - original draft preparation: [Shadab Ahmad and Tarique Ahmad]; Writing - review and editing: [Tarique ahmad].

Funding All the authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability All the data and materials support their published claims and comply with field standards.

Declarations

Ethical approval Present research work does not involve human or animals and therefore approval from any ethics committee is not required. The work has been accomplished with research ethics.

Consent to participate No human or animal participation therefore any statement of consent is not required.

Consent to publish All authors have consent to submit and publish the present research work in the Environmental Science and Pollution Research.

Competing interests All the authors declare that they have no competing interests.

Appendix

The performance of all models were evaluated using statistical indicators like R^2, MG, VG, R, d, Fa2, FS, RMSE, ME, MAE, FB and NMSE. The indicators were computed as given in Eqs. (6)–(17):

$$\boldsymbol{R^2} = \frac{\left(\sum\left[\left(\boldsymbol{X_m} - \overline{\boldsymbol{X_m}}\right) * \left(\boldsymbol{X_p} - \overline{\boldsymbol{X_p}}\right)\right]\right)^2}{\left[\sum\left(\boldsymbol{X_m} - \overline{\boldsymbol{X_m}}\right)^2 * \sum\left(\boldsymbol{X_p} - \overline{\boldsymbol{X_p}}\right)^2\right]} \tag{6}$$

$$\boldsymbol{MG} = \exp\left(\overline{\ln\boldsymbol{X_m}} - \overline{\ln\boldsymbol{X_p}}\right) \tag{7}$$

$$\boldsymbol{VG} = \exp\left(\overline{\ln\boldsymbol{X_m}} - \overline{\ln\boldsymbol{X_p}}\right)^2 \tag{8}$$

$$\boldsymbol{R} = \frac{\sum_{i=1}^{n}\left(\boldsymbol{X_{m_i}} - \overline{\boldsymbol{X_m}}\right) * \left(\boldsymbol{X_{p_i}} - \overline{X_p}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_{m_i} - \overline{X_m}\right)^2 * \left(X_{p_i} - \overline{X_p}\right)^2}} \tag{9}$$

$$d = 1 - \frac{\sum_{i=1}^{n}\left(X_{p_i} - X_{m_i}\right)^2}{\sum_{i=1}^{n}\left(\left|X_{p_i} - \overline{X_m}\right| + \left|X_{m_i} - \overline{X_m}\right|\right)^2} \tag{10}$$

$$\mathrm{Fa2} = \frac{X_p}{X_m}, 0.5 <= \frac{X_p}{X_m} <= 2 \tag{11}$$

$$FS = 2 * \left(\frac{\sigma_{X_m} - \sigma_{X_p}}{\sigma_{X_m} + \sigma_{X_p}}\right) \tag{12}$$

$$RMSE = \sqrt{\frac{1}{n} * \sum_{i=1}^{n}\left(X_{p_i} - X_{m_i}\right)^2} \tag{13}$$

$$ME = \frac{1}{n} * \sum_{i=1}^{n}\left(\frac{X_{p_i} - X_{m_i}}{X_{m_i}}\right) \tag{14}$$

$$MAE = \frac{1}{n} * \sum_{i=1}^{n}\left(\left|\frac{X_{p_i} - X_{m_i}}{X_{m_i}}\right|\right) \tag{15}$$

$$FB = 2 * \left(\frac{\overline{X_m} - \overline{X_p}}{\overline{X_m} + \overline{X_p}}\right) \tag{16}$$

$$NMSE = \frac{\overline{\left(\overline{X_m} - \overline{X_p}\right)^2}}{\overline{X_m} * \overline{X_p}} \tag{17}$$

Where, n the number of data points, X_p is the forecasted AQI value, X_m is the measured AQI value. The input data performance is evaluated, since it signifies the accuracy of prediction by each neural network model. The lowest values of *ME, MAE, FB, NMSE, RMSE, VG, FS* and highest values of R^2, *MG, R and d, Fa2* represents the best model performance.

References

Afzali, M., Afzali, A., & Zahedi, G. (2012). The potential of artificial neural network technique in daily and monthly ambient air temperature prediction. *International Journal of Environmental Science Development, 3*(1), 33–38.

Angelevska, B., Atanasova, V., & Andreevski, I. (2021). Urban air quality guidance based on measures categorization in road transport. *Civil Engineering Journal, 7*(2), 253–267.

Arnaudo, E., Farasin, A., & Rossi, C. (2020). A comparative analysis for air quality estimation from traffic and meteorological data. *Applied Sciences, 10*(13), 4587.

Azid, A., Juahir, H., Latif, M., Zain, S., & Osman, M. (2013). Feed-forward artificial neural network model for air pollutant index prediction in the southern region of peninsular Malaysia. *Journal of Environmental Protection, 4*(12A), 1–10.

Bai, Y., Li, Y., Wang, X., Xie, J., & Li, C. (2016). Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions. *Atmospheric Pollution Research, 7*(3), 557–566.

Barai, S. V., Dikshit, A. K., & Sharma, S. (2007). Neural network models for air quality prediction: A comparative study. In A. Saad, K. Dahal, M. Sarfraz, & R. Roy (Eds.), *Soft computing in industrial applications* (Vol. 39, pp. 290–305). Springer.

Baumann-Stanzer, K., & Piringer, M. (2011). Validation of regulatory micro-scale air quality models: Modelling odour dispersion and built-up areas. *World Review of Science, Technology and Sustainable Development, 8*(2/3/4), 203–213.

Bhandarkar, S. (2013). Vehicular pollution, their effect on human health and mitigation measures. *Vehicle Engineering, 1*(2), 33–40.

Brook, R. D., Rajagopalan, S., Pope, A. C., III, Brook, J. R., Bhatnagar, A., Diez-Roux, A. V., Holguin, F., Hong, Y., Luepker, R. V., Mittleman, M. A., Peters, A., Siscovick, D., Smith, S. C., Jr., Whitsel, L., & Kaufman, J. D. (2010).

Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association. *Circulation, 121*(21), 2331–2378.
Castelli, M., Clemente, F. M., Popovic, A., Silva, S., & Vanneschi, L. (2020). A machine learning approach to predict air quality in California. *Complexity, 2020*, 1–23.
Cetin, M., & Sevik, H. (2016). Change of air quality in Kastamonu city in terms of particulate matter and CO_2 amount. *Oxidation Communications, 39*, 3394–3401.
Chen, L., & Pai, T.-Y. (2015). Comparisons of GM (1, 1), and BPNN for predicting hourly particulate matter in Dali area of Taichung City, Taiwan. *Atmospheric Pollution Research, 6*(4), 572–580.
Choi, E., Schuetz, A., Stewart, W., & Sun, J. (2017). Using recurrent neural network models for early detection of heart failure onset. *Journal of the American Medical Informatics Association, 24*(2), 361–370.
Coker, E., Liverani, S., Ghosh, J. K., Jerrett, M., Beckerman, B., Li, A., Ritz, B., & Molitor, J. (2016). Multi-pollutant exposure profiles associated with term low birth weight in Los Angeles County. *Environment International, 91*, 1–13.
CPCB (Central Pollution Control Board). (2014). National air quality index. CPCB. Retrieved December 22, 2021, from https://app.cpcbccr.com/ccr_docs/FINAL-REPORT_AQI_.pdf.
CPCB (Central Pollution Control Board). (2021). Central Control Room for Air Quality Management - Delhi NCR. CPCB. Retrieved January 15, 2022, from https://app.cpcbccr.com/ccr/#/caaqm-dashboard/caaqm-landing/caaqm-data-availability.
Demuth, H., Beale, M., & Hagan, M. (2009). *Neural networks toolbox manual*. Math Works Inc..
DPCC (Delhi Pollution Control Committee). (2020). Retrieved December 15, 2021, from http://www.dpcc.delhigovt.nic.in/Air40.html.
Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (Third ed.). John Wiley and Sons, Inc.
Dubey, B., Pal, A. K., & Singh, G. (2013). Assessment of vehicular pollution in Dhanbad city using CALINE 4 model. *International journal of geology, Earth and Environmental Sciences, 3*(1), 156–164.
Durão, R. M., Mendes, M. T., & Pereira, M. J. (2016). Forecasting O_3 levels in industrial area surroundings up to 24 hours in advance, combining classification trees and MLP models. *Atmospheric Pollution Research, 7*(6), 961–970.
Elsunousi, A. A. M., Sevik, H., Cetin, M., Ozel, H. B., & Ozel, H. U. (2021). Periodical and regional change of particulate matter and CO_2 concentration in Misurata. *Environmental Monitoring and Assessment, 193*, 707.
Ghude, S. D., Chate, D. M., Jena, C., Beig, G., Kumar, R., & Barth, M. C. (2016). Premature mortality in India due to $PM_{2.5}$ and ozone exposure. *Geophysical Research Letter, 43*, 4650–4658.
Glantz, S. A., & Slinker, B. K. (1990). *Primer of applied regression and analysis of variance. Health Professions Division*. McGraw-Hill.
Gulia, S., Shiva Nagendra, S. M., Khare, M., & Khanna, I. (2015). Urban air quality management–a review. *Atmospheric Pollution Research, 6*(2), 286–304.
Gurjar, B. R., van Aardennea, J. A., Lelievelda, J., & Mohan, M. (2004). Emission estimates and trends (1990–2000) for megacity Delhi and implications. *Atmospheric Environment, 38*, 5663–5681.
Guttikunda, S. K., & Gurjar, B. R. (2012). Role of meteorology in seasonality of air pollution in megacity Delhi, India. *Environmental Monitoring and Assessment, 184*, 3199–3211.
Hajek, P., & Olej, V. (2015). Predicting common air quality index-the case of Czech micro regions. *Aerosol and Air Quality Research, 15*(2), 544–555.
Haykin, S. O. (2009). *Neural networks and learning machines* (3rd ed.). New Jersery.
Hedayat, A., Davilu, H., Barfrosh, A. A., & Sepanloo, K. (2009). Estimation of research reactor core parameters using cascade feed forward artificial neural networks. *Progress in Nuclear Energy, 51*(6), 709–718.
Hoshyaripour, G., Brasseur, G., Andrade, M. F., Gavidia-Calderón, M., Bouarar, I., & Ynoue, R. Y. (2016). Prediction of ground-level ozone concentration in São Paulo, Brazil: Deterministic versus statistic models. *Atmospheric Environment, 145*, 365–375.
Iliyas, S. A., Elshafei, M., Habib, M. A., & Adeniran, A. A. (2013). RBF neural network inferential sensor for process emission monitoring. *Control Engineering Practice, 21*(7), 962–970.
Joshi M. (2020). Explained: Why does air pollution rise in October every year? Retrieved December 18, 2021, from https://indianexpress.com/article/explained/explained-why-does-air-pollution-rise-in-october-each-year-6759030/.
Kandlikar, M., & Ramachandaran, G. (2000). The causes and consequences of particulate air pollution in urban India: A synthesis of the science. *Annual Review of Energy and the Environment, 25*(1), 629–684.
Koo, J. W., Wong, S. W., Selvachandran, G., Long, H. V., & Son, L. H. (2020). Prediction of air pollution index in Kuala Lumpur using fuzzy time series and statistical models. *Air Quality, Atmosphere & Health, 13*(1), 77–88.
Krzysztof, S. A., & Osowski, S. (2016). Data mining methods for prediction of air pollution. *International Journal of Applied Mathematics and Computer Science, 26*(2), 467–478.
Kumar, P., Gulia, S., Harrison, R. M., & Khare, M. (2017). The influence of odd-even car trial on fine and coarse particles in Delhi. *Environmental Pollution, 225*, 20–30.
Kumar, P., Khare, M., Harrison, R. M., Bloss, W. J., Lewis, A. C., Coe, H., & Morawska, L. (2015). New directions: Air pollution challenges for developing megacities like Delhi. *Atmospheric Environment, 122*, 657–661.
Kunzli, N., Kaiser, R., Medina, S., Studnicka, M., Chanel, O., Filliger, P., Herry, M., Horak, F., Jr., Puybonnieux-texier, V., Quenel, P., Schneider, J., Seethaler, R., Vergnaud, J. C., & Sommer, H. (2000). Public-health impact of outdoor and traffic-related air pollution: A European assessment. *The Lancet, 356*, 795–801.
Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., & Pozzer, A. (2015). The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature, 525*, 367–371.

Lippmann, M. (2014). Toxicological and epidemiological studies of cardiovascular effects of ambient air fine particulate matter ($PM_{2.5}$) and its chemical components: Coherence and public health implications. *Critical Reviews in Toxicology, 44*(4), 299–347.

Liu, H., Li, Q., Yu, D., & Gu, Y. (2019). Air quality index and air pollutant concentration prediction based on machine learning algorithms. *Applied Sciences, 9*(19), 4069.

Masood, A., Ahmad, K., & Ahmad, S. (2018). Urban roadside monitoring, modelling and mapping of air pollution. *Applied Journal of Environmental Engineering Science, 3*(2), 3–2.

Misra, A., Roorda, M. J., & MacLean, H. L. (2013). An integrated modelling approach to estimate urban traffic emissions. *Atmospheric Environment, 73*, 81–91.

Muhammad, S. Y., Makhtar, M., Rozaimee, A., Abdul, A., & Jamal, A. A. (2015). Classification model for air quality using machine learning techniques. *International Journal of Software Engineering and Its Applications*, 45–52.

Nagendra, S. S., & Khare, M. (2005). Modelling urban air quality using artificial neural network. *Clean Technologies Environmental Policy, 7*(2), 116–126.

Niharika, V. M., & Rao, P. S. (2014). A survey on air quality forecasting techniques. *International Journal of Computer Science and Information Technologies, 5*(1), 103–107.

Nimesh, R., Arora, S., Mahajan, K. K., Gill, A., & N. (2014). Predicting air quality using ARIMA, ARFIMA and HW smoothing. *Model Assisted Statistics and Applications., 9*(2), 137–149.

Pant, P., Shukla, A., Kohl, S. D., Chow, J. C., Watson, J. G., & Harrison, R. M. (2015). Characterization of ambient PM2.5 at a pollution hotspot in New Delhi, India and inference of sources. *Atmospheric Environment, 109*, 178–189.

Podnar, D., Koracin, D., & Panorska, A. (2002). Application of artificial neural networks to modeling the transport and dispersion of tracers in complex terrain. *Atmospheric Environment, 36*, 561–570.

Prasad, K., Gorai, A. K., & Goyal, P. (2016). Development of ANFIS models for air quality forecasting and input optimization for reducing the computational cost and time. *Atmospheric Environment, 128*, 246–262.

Qin, S. S., Liu, F., Wang, J. Z., & Sun, B. B. (2014). Analysis and forecasting of the particulate matter (PM) concentration levels over four major cities of China using hybrid models. *Atmospheric Environment, 98*, 665–675.

Raducan, G., & Stefanescu, I. (2012). A qualitative study of air pollutants from road traffic. In S. Kumar & R. Kumar (Eds.), *Air quality–monitoring and modelling* (pp. 19–35). IntechOpen.

Rahman, K. H. A., Bhrugubanda, H., & Sivaraman, V. (2017). HazeEst: Machine learning based metropolitan air pollution estimation from fixed and mobile sensors. *IEEE Sensor Journal, 17*(11), 3517–3525.

Rahman, P. A., Panchenko, A. A., & Safarov, A. M. (2016). Using neural networks for prediction of air pollution index in industrial city. *IOP Conference Series: Earth and Environmental Science, 87*(4), 042016.

Sadorsky, P. (2006). Modeling and forecasting petroleum futures volatility. *Energy Economics, 28*(4), 467–488.

Saxena, M., Sharma, A., Sen, A., Saxena, P., Mandal, T. K., Sharma, S. K., & Sharma, C. (2017). Water soluble inorganic species of PM10 and PM2.5 at an urban site of Delhi, India: Seasonal variability and sources. *Atmospheric Research, 184*, 112–125.

Shahraiyni, H. T., Sodoudi, S., Kerschbaumer, A., & Cubasch, U. (2015). New technique for ranking of air pollution monitoring stations in the urban areas based upon spatial Representativity (case study: PM monitoring stations in Berlin). *Aerosol and Air Quality Research, 15*, 743–748.

Sharma, S. K., Mandal, T. K., Jain, S., Sharma, A., & Saxena, M. (2016). Source apportionment of PM2.5 in Delhi, India using PMF model. *Bulletin of Environmental Contamination and Toxicology, 97*(2), 286–293.

Singh, V., Singh, S., & Biswal, A. (2021). Exceedances and trends of particulate matter (PM2.5) in five Indian megacities. *Science of the Total Environment, 750*, 141461.

Statista. (2022). Number of registered motor vehicles across Delhi in India from 1988 to 2020. Retrieved December 7, 2022 from https://www.statista.com/statistics/1073315/india-registered-number-of-private-cars-in-delhi/

Taneja, S., Sharma, N., Oberoi, K., & Navoria, Y. (2016). Predicting Trends in Air Pollution in Delhi using Data Mining. In India International Conference on Information Processing (IICIP). IEEE, 1–6.

Tiwari, S., Bisht, D. S., Srivastava, A. K., Pipal, A. S., Taneja, A., Srivastava, M. K., & Attri, S. D. (2014). Variability in atmospheric particulates and meteorological effects on their mass concentrations over Delhi, India. *Atmospheric Research, 145-146*, 45–56.

Tiwari, S., Srivastava, A. K., Bisht, D. S., Parmita, P., Srivastava, M. K., & Attri, S. D. (2013). Diurnal and seasonal variations of black carbon and PM2.5 over New Delhi, India: Influence of meteorology. *Atmospheric Research, 125-126*, 50–62.

Tyagi, S., Tiwari, S., Mishra, A., Hopke, P. K., Attri, S. D., Srivastava, A. K., & Bisht, D. S. (2016). Spatial variability of concentrations of gaseous pollutants across the National Capital Region of Delhi, India. *Atmospheric Pollution Research, 7*, 808–816.

Veljanovska, K., & Dimoski, A. (2018). Air quality index prediction using simple machine learning algorithms. *International Journal of Emerging Trends & Technology in Computer Science, 7*(1), 025–030.

Villalobos, A. M., Amonov, M. O., Shafer, M. M., Devi, J. J., Gupta, T., Tripathi, S. N., & Schauer, J. J. (2015). Source apportionment of carbonaceous fine particulate matter (PM2.5) in two contrasting cities across the indo-Gangetic plain. *Atmospheric Pollution Research, 6*(3), 398–405.

Wang, D., & Lu, W.-Z. (2006). Forecasting of ozone level in time series using MLP model with a novel hybrid training algorithm. *Atmospheric Environment, 40*(5), 913–924.

Wang, W., Xu, Z., & Lu, J. W. (2003). Three improved neural network models for air quality forecasting. *Engineering Computations, 20*(2), 192–210.

WHO (World Health Organization). (2016). *Ambient air pollution: A global assessment of exposure and burden of disease*. World Health Organization.

Willmott, C. J. (1982). Some comments on the evaluation of model performance. *Bulletin of the American Meteorological Society, 63*, 1309–1313.

Wozniak, M., Napoli, C., Tramontana, E., & Capizzi, G. (2015). A multiscale image compressor with RBFNN and discrete wavelet decomposition. In International Joint Conference on Neural Networks (IJCNN). IEEE, 1219–1225.

WPR (World Population Review). (2021) Retrieved January 30, 2022, from https://worldpopulationreview.com/.

Wu, Q., & Lin, H. (2019). A novel optimal-hybrid model for daily air quality index prediction considering air pollutant factors. *Science of the Total Environment, 683*, 808–821.

Xi, X., Wei, Z., Xiaoguang, R., Yijie, W., Xinxin, B., Wenjun, Y., & Jin, D. A. (2015). Comprehensive evaluation of air pollution prediction improvement by a machine learning method. In: 10th IEEE International Conference on Service Operations and Logistics, and Informatics, Tunisia. IEEE, 176–181.

Yadav, R., Beig, G., & Jaaffrey, S. N. A. (2014). The linkages of anthropogenic emissions and meteorology in the rapid increase of particulate matter at a foothill city in the Arawali range of India. *Atmospheric Environment, 85*, 147–151.

Yadav, R., Sahu, L. K., Beig, G., & Jaaffrey, S. N. A. (2016). Role of long-range transport and local meteorology in seasonal variation of surface ozone and its precursors at an urban site in India. *Atmospheric Research, 176-177*, 96–107.

Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and network architectures. *Neural Computation, 31*(7), 1235–1270.

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting, 14*(1), 35–62.

Zhu, S., Lian, X., Liu, H., Hu, J., Wang, Y., & Che, J. (2017). Daily air quality index forecasting with hybrid models: A case in China. *Environmental Pollution, 231*, 1232–1244.