scientific reports

OPEN

Estimating PM$_{2.5}$ utilizing multiple linear regression and ANN techniques

Sumita Gulati[1], Anshul Bansal[2], Ashok Pal[3], Nitin Mittal[4], Abhishek Sharma[5] & Fikreselam Gared[6✉]

The accurate prediction of air pollutants, particularly Particulate Matter (PM), is critical to support effective and persuasive air quality management. Numerous variables influence the prediction of PM, and it's crucial to combine the most relevant input variables to ensure the most dependable predictions. This study aims to address this issue by utilizing correlation coefficients to select the most pertinent input and output variables for an air pollution model. In this work, PM$_{2.5}$ concentration is estimated by employing concentrations of sulfur dioxide, nitrogen dioxide, and PM$_{10}$ found in the air through the application of Artificial Neural Networks (ANNs). The proposed approach involves the comparison of three ANN models: one trained with the Levenberg–Marquardt algorithm (LM-ANN), another with the Bayesian Regularization algorithm (BR-ANN), and a third with the Scaled Conjugate Gradient algorithm (SCG-ANN). The findings revealed that the LM-ANN model outperforms the other two models and even surpasses the Multiple Linear Regression method. The LM-ANN model yields a higher R^2 value of 0.8164 and a lower RMSE value of 9.5223.

Air is an admirable and most valued resource. It is the essential source on this earth that supports all living beings to survive and sustain. Unfortunately, in current years, due to several human exercises our valuable natural resources are getting contaminated. Air pollution is the most substantial environmental concern in almost all parts of the world. Due to the extensive advancements in economy around the globe, air quality evolves into a major issue as the diminishing air quality has incessant and somber effects not only on human health but also on the ecosystem. The WHO stated that around 90% of the world's population is inhaling polluted air (www.who.in). Moreover, the State of the Global Air report (2019), addresses air pollution as the fifth dominant hazard for mortality across the globe. Main pollutants that affect most of the nation comprises of particulate matter (PM), nitrogen dioxide (NO$_2$), lead (Pb), carbon monoxide (CO), ozone (O$_3$), sulphur dioxide (SO$_2$) etc.. Raising levels of these deadly pollutants due to industrial activities, vehicles, construction sites, power plants, and natural processes like volcanoes, forest fires, has considerable brunt on human wellbeing. The deteriorated quality of air results in several types of aversions, cardiovascular diseases, respiratory ill health, etc.[1–3]. Owing to the numerous adverse effects on human well-being, this environmental issue must be considered significantly. Moreover, now a days many nations are collaborating to address the issue of increasing air pollution[4].

The primitive pollutant affecting human health is PM. PM is composed up primarily of smoke, dust and soot or liquid droplets discharged into the environment from industries, vehicles, construction spots etc. The particles having an aerodynamic diameter lower than 2.5 µm are fine, PM$_{2.5}$ and those having a diameter less than 10 µm are known as coarse particles, PM$_{10}$. The pollutant deeply penetrates the respiratory system and the blood streams leading to many health hazards[5,6]. Hence, it is imperative to develop effective tools for monitoring PM levels, disseminating information regarding hazardous concentrations, and providing recommendations for preventive measures to mitigate such levels. Numerous investigations have been carried out, encompassing not only the quantification of PM levels but also the evaluation of potential health hazards associated with heightened PM exposure for the population. These studies greatly enhance our understanding of the complex public health challenges caused by the widespread effects of air pollution[7–16].

[1]Department of Mathematics, S. A. Jain College, Ambala, Haryana 134003, India. [2]Department of Chemistry, S. A. Jain College, Ambala, Haryana 134003, India. [3]Department of Mathematics, Chandigarh University, Gharuan, Mohali 140413, India. [4]University Centre for Research and Development, Chandigarh University, Gharuan, Mohali 140413, India. [5]Department of Computer Engineering and Applications, GLA University, Mathura 281406, India. [6]Faculty of Electrical and Computer Engineering, Bahir Dar Institue of Technology, Bahir Dar University, Bahir Dar, Ethiopia. ✉email: fikreseafomi@gmail.com

Further, investigating the entire involved parameters contributing air pollution is an arduous task. To deal with this, air pollution models are desired to evolve early warnings and command actions and further to examine forthcoming ensuing discharge schemes[17,18]. The increasing role of machine learning in air quality prediction represents a significant leap forward in our ability to monitor and manage environmental health. Machine learning techniques have ushered in a new era of air quality forecasting, allowing us to harness vast amounts of data, including historical air quality information, meteorological data, and even satellite imagery. These algorithms can identify complex patterns and correlations within this data, enabling more accurate predictions of air quality parameters such as PM concentrations, ozone levels, and pollutant concentrations. By providing real-time, high-resolution forecasts, machine learning models empower policymakers, environmental agencies, and the public to make informed decisions, take preventive measures, and mitigate the adverse effects of air pollution on public health and the environment. The growing integration of machine learning into air quality prediction signifies a promising avenue for advancing our understanding of air pollution dynamics and enhancing the quality of life for communities around the world. Among various statistical procedures, Artificial Neural Networks (ANNs) have been demonstrated to be altogether effective for appropriating complex relationships and enhancing forecast accuracy[19–24].

ANNs are computational models inspired by the structure and function of the human brain. They consist of interconnected nodes, or artificial neurons, organized into layers. These networks are used for various machine learning tasks, including pattern recognition, classification, regression, and even more complex tasks like natural language processing and image recognition. In ANNs, information flows through the network, with each neuron processing and transmitting data to the next layer. ANNs have gained widespread popularity due to their ability to handle complex and high-dimensional data, making them a crucial component of modern artificial intelligence and deep learning applications. They have been instrumental in advancing fields such as computer vision, speech recognition, and autonomous systems, among many others. Broadly, different kinds of ANN involve the back-propagation neural network[25,26], multilayer perceptron[27,28], radial basis function[29,30], and adaptive neuro-fuzzy inference systems[31,32].

The primary objective of this study is to assess the performance of ANN trained with different algorithms for predicting $PM_{2.5}$ concentration. Additionally, we have conducted a comparative analysis with the traditional multiple linear regression model (MLR).

Background

Numerous researchers have engaged in the thorough evaluation of air quality prediction models, with a specific emphasis on the precision of PM concentration predictions across a wide spectrum of scenarios, employing ANN. The use of ANNs for estimating PM concentration has been asserted for the prediction of hourly and daily average concentrations relying on air pollutants and atmospheric data[33,34]. In the Santiago city of Chile, Perez et al.[35] demonstrated estimations of hourly average concentrations of $PM_{2.5}$ several hours before, depending on values attained at a steady site. Further, outcomes acquired employing ANN revealed estimated errors within the extent 30–60%. Moreover, they examined the noise cutback of dataset to enhance predictions as imperative. A comparison of ANN technique with classical regression techniques for PM_{10} and $PM_{2.5}$ estimation was conducted by McKendry[36]. He established that meteorological variables, endurance, and co-pollutant values effectively estimated PM levels. In another investigation, Chelani et al.[37] entrenched an ANN procedure to predict PM_{10} and noxious metals contamination investigated in the Jaipur city of India. Authors were adept at estimating contaminations quite justly. Tecer[38] suggested ANNs to estimate PM levels in Zonguldak Province, Turkey. The outcomes revealed that the suggested technique can effectively be employed to estimate air quality. Pires et al.[39] demonstrated the accomplishment of five linear models to estimate the daily average PM_{10} levels and certified that the size of the dataset is an imperative factor for the estimation of models. Paschalidou et al.[40] employed multilayer perceptron for PM_{10} hourly levels prediction in Cyprus. The prediction revealed that the MLP models displays the best estimation performance. Also, Roy et al.[41] have suggested the utilization of both multiple regression and ANN techniques for analyzing PM levels in different seasons at a vast opencast coal mine in India. The findings indicated that the ANN-based forecasting outperformed the multiple regression models. An online air pollutants predicting ANN technique that utilizes parameters attained through geographic modeling for the district Besiktas, Instanbul was suggested by Kurt & Oktay[42]. This system employs the meteorological parameters, the air pollutants levels and certain area specific attributes as input parameters. The ANN technique was carried out in this study to develop $PM_{2.5}$ concentration prediction model. In Spain, another ANN model for PM_{10} daily levels estimation was suggested that executes the estimation of a 24 h average PM_{10} levels and employs deterministic variables for overall transit of aerosols from arid areas[43]. An innovative approach was employed to forecast $PM_{2.5}$ and PM_{10} levels in major Chinese cities. This approach integrated a feedforward ANN model with a rolling criterion to capture input data patterns and a cumulative generating conduct of gray model to reduce data sample unpredictability[44]. The prediction procedure relied mainly on the daily values of $PM_{2.5}$ and PM_{10} levels and on a few atmospheric parameters. With an aim to analyze the impact of exposure to PM_{10} on health and to estimate PM_{10} levels using ANN another study was conducted in Yasuj city[45]. The daily average values of PM_{10} as well as the climatic data was utilized in this analysis. In general, amongst all the machine learning approaches, ANN has been proven to be the most favorable approach of the researchers. This study examined ANN technique with varying training functions to establish the most effective model for $PM_{2.5}$ estimation.

Methodology
Study area and air quality data
In India, central pollution control board (CPCB) is the pinnacle institution that investigates and monitors air quality. This institution supervises air pollution with the support of its abundant stations extended in nearly

every city. Air quality across the country is systematically monitored through a combination of Manual and Continuous Ambient Air Quality monitoring stations. At present, this network comprises a total of 1257 monitoring stations. Manual monitoring activities are undertaken at 883 stations, encompassing 378 cities and towns distributed across 28 States and 7 Union Territories. Simultaneously, continuous monitoring is carried out at 374 stations, situated in 190 cities and towns across 27 States and 4 Union Territories. To facilitate the monitoring of air pollutants, the responsibility is shared with various entities such as the State Pollution Control Boards (SPCB), Pollution Control Committees (PCC), and other reputable institutions. The CPCB collaborates with these organizations to ensure the uniformity and consistency of air quality data while offering technical and financial support. It identifies and calculates pollutants as well as atmospheric parameters. Moreover, the monitoring of air pollutants is enforced with the support of SPCB, PCC, and several other reputed organizations. CPCB work with these assisting institutes to provide uniform, consistent air quality data[46]. The data generated through manual and continuous monitoring integrated for the year 2021 has been taken for this study involving the annual average values of SO_2, NO_2, PM_{10} and $PM_{2.5}$ (in μg/m³) as shown in Fig. 1.

Modeling and opting suitable input variables

The observed levels of air pollutants PM_{10}, $PM_{2.5}$, NO_2, and SO_2 were investigated with an objective to frame an air pollution estimation model. The specific dataset was sourced from CPCB for the year 2021. Figure 2 visually represents the relationships among SO_2, NO_2, PM_{10}, and $PM_{2.5}$ levels. We observed a positive correlation among all these variables, signifying their relevance to the study. Notably, the maximum correlation values were found to be 0.31 for $PM_{2.5}$ with SO_2, 0.61 with NO_2, and 0.83 with PM_{10}. Consequently, SO_2, NO_2, and PM_{10} were selected as the input variables for the $PM_{2.5}$ air pollution estimation model.

Estimating PM$_{2.5}$

Multiple linear regression (MLR) model

Multiple linear regression (MLR) is a statistical technique used to assess the relationships between a single dependent variable and two or more independent variables. The method works by fitting a linear equation to the data, with coefficients representing the contribution of each independent variable to the dependent variable. The model aims to find the best-fitting line through the data points, which minimizes the sum of the squared differences between the observed and predicted values. MLR is a valuable tool for uncovering complex associations and understanding the underlying factors that influence a particular phenomenon. The formula for expressing the output dependent variable y in terms of independent variables x_1, x_2, …, x_n is as follows:

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n + \varepsilon, \tag{1}$$

where n = number of observations, α_0 = the y intercept, α_n = coefficient of the independent variable x_n and ε = model error.

In this particular investigation, $PM_{2.5}$ is taken as a dependent variable, while SO_2, NO_2 and PM_{10} are considered as independent variables. The MLR model computes the coefficients α_1, α_2,…,α_n using the least square method.
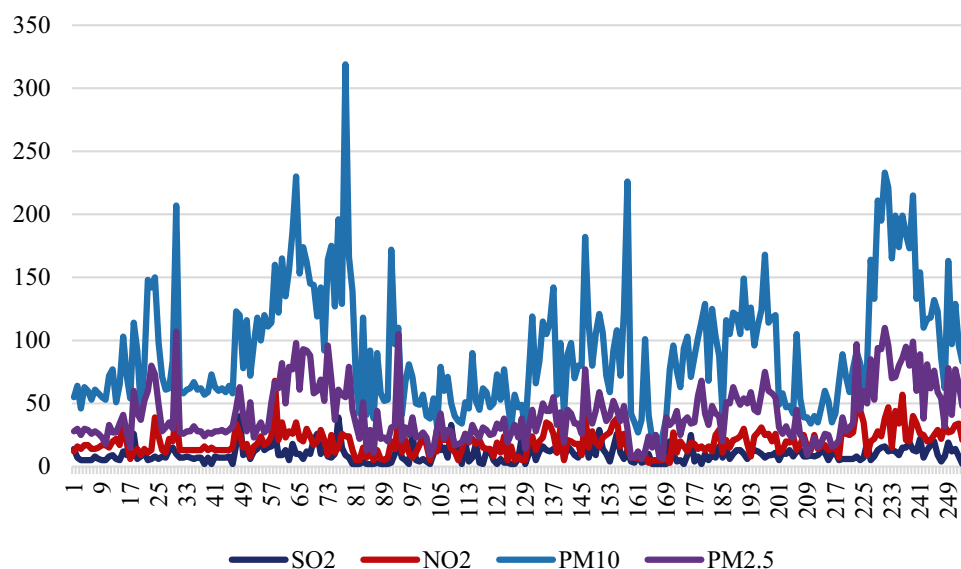

Figure 1. Values of air pollutant variables for the year 2021.
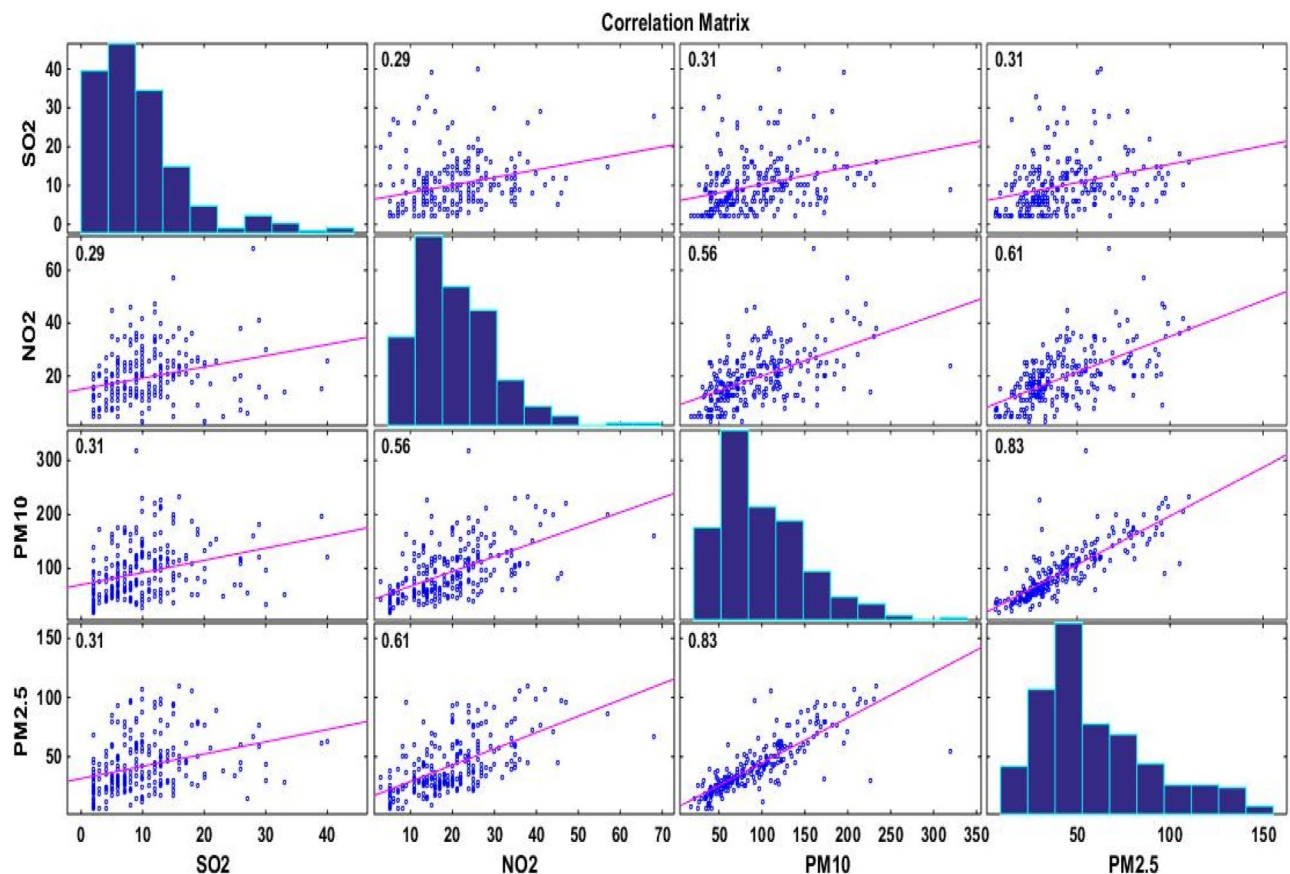
Figure 2. Correlation matrix of air pollutants in India for the year 2021.

Proposed ANN models

In the present analysis, the nftool of MATLAB (Version R2014b) was used and executed on a system equipped with an Intel HD Graphics card, a 17-inch display, 4 GB of memory, an Intel 11th generation i5 430 M processor, and a 512 GB SSD[47–49]. The ANN were trained using NO_2, SO_2, and PM_{10} as input variables and $PM_{2.5}$ as the target variable. The neural network architecture consisted of 20 neurons in the hidden layer, as depicted in Fig. 3. For the initial training phase, the dataset was partitioned into training (70%), validation (10%), and testing (20%) subsets. ANNs employ various training algorithms to adjust the network's parameters (weights and biases) in order to minimize errors and improve performance. In this study, the network underwent training using three distinct algorithms sequentially: Levenberg–Marquardt, Bayesian Regularization, and Scalar Conjugate training algorithms.

<u>Performance metrics.</u>   The assessment and differentiation of the MLR and the three proposed Artificial Neural Network (ANN) models are carried out by examining the Root Mean Square Error (RMSE) and Coefficient of Determination (R^2). These metrics are defined as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (o(t) - p(t))^2},$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(o(t) - p(t)\right)^2}{\sum_{i=1}^{n} \left(o(t) - \frac{1}{n} \sum_{i=1}^{n} o(t)\right)^2},$$

where, n = number of observations, o(t) = actual value of the variable , p(t) = predicted value of the variable.

Simulation results and discussion

This study focuses on the estimation of $PM_{2.5}$ concentrations using annual average data of SO_2, NO_2, and PM_{10} for the year 2021 as input parameters. The performance evaluation of the models was based on the Root Mean Square Error (RMSE) and the coefficient of determination (R^2). These metrics provided insights into the effectiveness of the MLR model and the three ANN models: ANN trained using the Levenberg–Marquardt algorithm (LM-ANN), the Bayesian Regularization algorithm (BR-ANN), and the Scaled Conjugate Gradient algorithm (SCG-ANN).
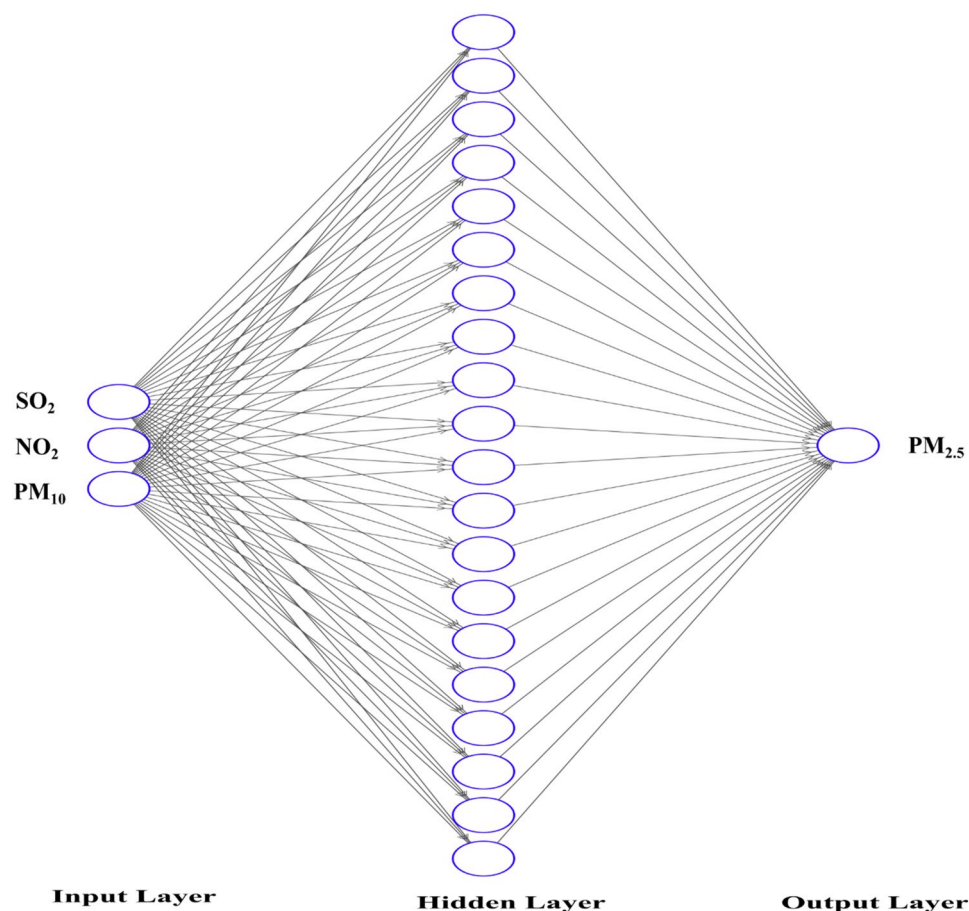
Figure 3. The structure of the ANN layers.

The experiments entail a comparison between the MLR model and three ANN models. The results of this comparison, specifically the evaluation metrics RMSE and R^2, are presented in Table 1 for reference.

The results revealed that the LM-ANN model outperformed the others, yielding the lowest RMSE of 9.5223 compared to 9.6555, 11.0165, and 11.7585 for BR-ANN, SCG-ANN, and MLR, respectively. Furthermore, the $PM_{2.5}$ concentration estimated by the LM-ANN model demonstrates a strong correlation with observed values, with an R^2 of 0.8164. In contrast, BR-ANN exhibits an R^2 value of 0.8118, while SCG-ANN yields 0.7551, and MLR results in 0.7201.

Correlation amongst observed and estimated LM-ANN, BR-ANN SCG-ANN and MLR models are illustrated in Fig. 4. Additionally, Figs. 5, 6, and 7 represent the regression plots for the LM-ANN, BR-ANN, and SCG-ANN models, respectively. Moreover, Fig. 8 provides a time series illustration of the detected and estimated $PM_{2.5}$ values for the suggested models.

Notably, the results highlighted the superior performance of the LM-ANN model in comparison to the other models, signifying its enhanced capability in estimating $PM_{2.5}$ concentrations.

The current investigation offers a comprehensive exploration of the effectiveness of various ANN techniques when applied to the air quality modeling. This research not only sheds light on the adequacy of different ANN methodologies but also delves into their relative strengths and weaknesses in the context of air quality modeling. Consistency and size of data used, alongwith upholding of identical controlling factors for training and testing data are some of the limitations of proposed ANN models. By examining these diverse ANN approaches, we gain a deeper understanding of how they perform and contribute to the field of air quality modeling.

Model	RMSE	R^2
LM-ANN	**9.5223**	**0.8164**
BR-ANN	9.6555	0.8118
SCG-ANN	11.0165	0.7551
MLR	11.7585	0.7201

Table 1. Statistical error indices. Bold values indicate the best value from others.
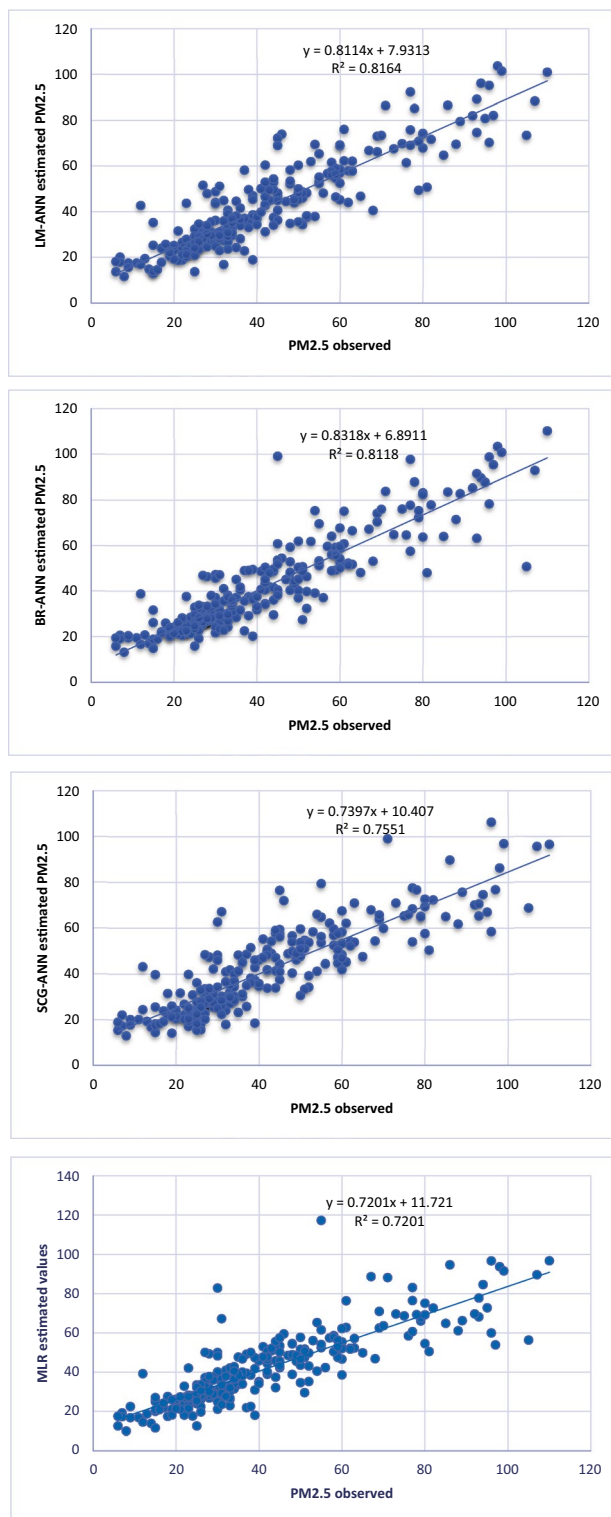
Figure 4. Correlation amongst observed and estimated PM$_{2.5}$.

Conclusion and future scope

Particulate matter (PM) is a major air pollutant known to have detrimental impacts on human health. This study involved the predictive analysis of PM$_{2.5}$ levels by utilizing Artificial Neural Network (ANN) models based on
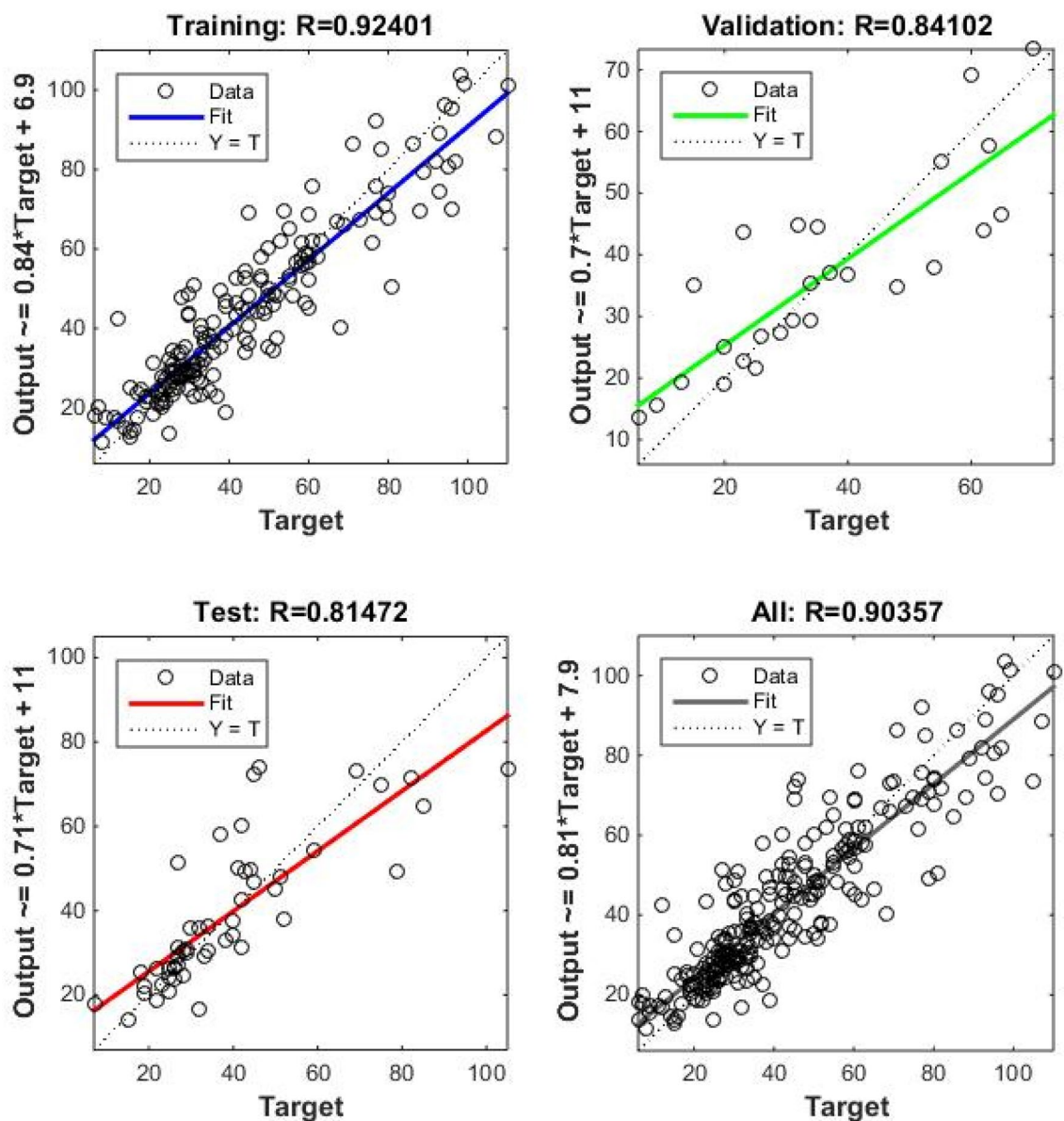
Figure 5. Regression plot of LM-ANN.

data concerning SO_2, NO_2, and PM_{10} concentrations. The three distinct ANN models, namely LM-ANN, BR-ANN, and SCG-ANN were applied to India's air quality dataset for the year 2021 sourced from the CPCB. The error metrics, specifically Root Mean Square Error (RMSE) and R-squared (R^2) were employed to assess the performance of these models. The findings demonstrated that the LM-ANN model exhibited superior performance compared to the other two ANN models and the Multiple Linear Regression (MLR) model. Moreover, these models have the potential to alert the public when PM concentration surpasses its prescribed level. Furthermore, the suggested models can be deployed to forecast real-time air quality trends using historical data, making them valuable tools for proactive planning and management of air pollution concerns. In summary, this ANN modeling approach offers a practical solution for governmental agencies to address air pollution issues and formulate effective strategies for mitigating their impact.

Regarding future research directions, we aim to extend our investigations into air pollution by integrating daily and hourly data, thereby enabling a more exhaustive analysis of pollution levels across diverse urban areas. Additionally, doe to the prominent performance exhibited by the LM-ANN model, there is potential for further enhancements to fine-tune its capabilities in air quality prediction.
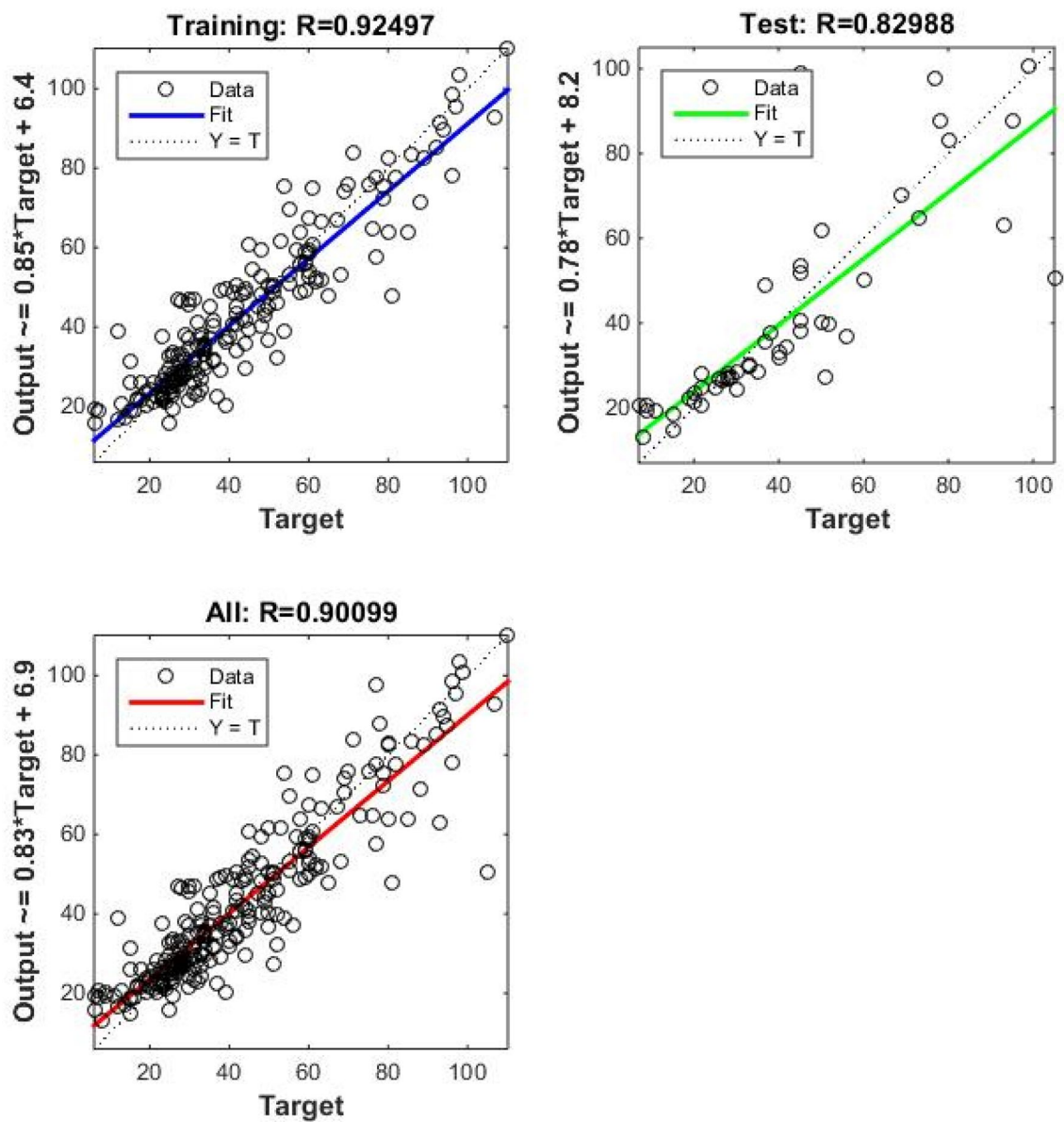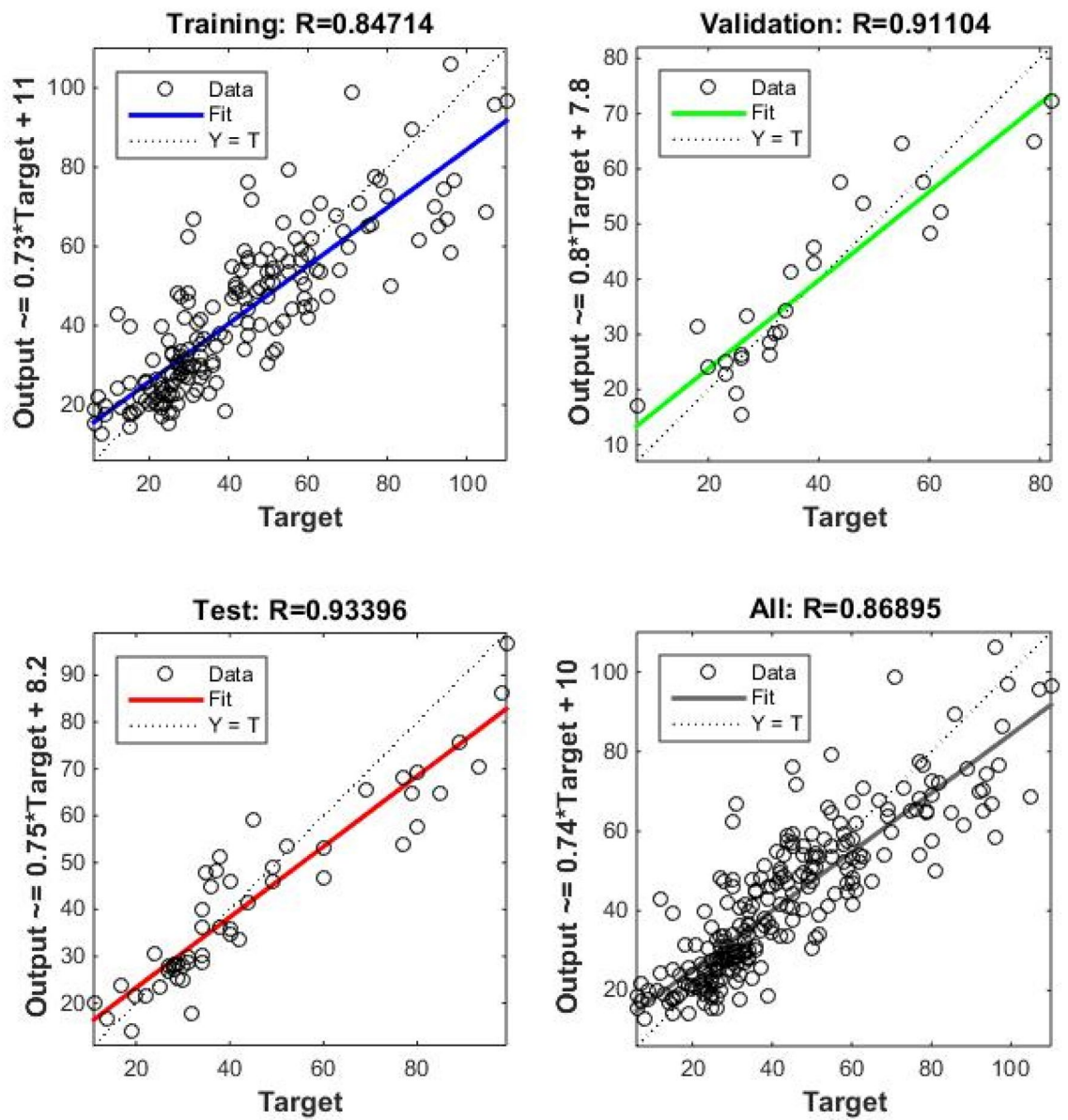
Figure 6. Regression plot of BR-ANN.

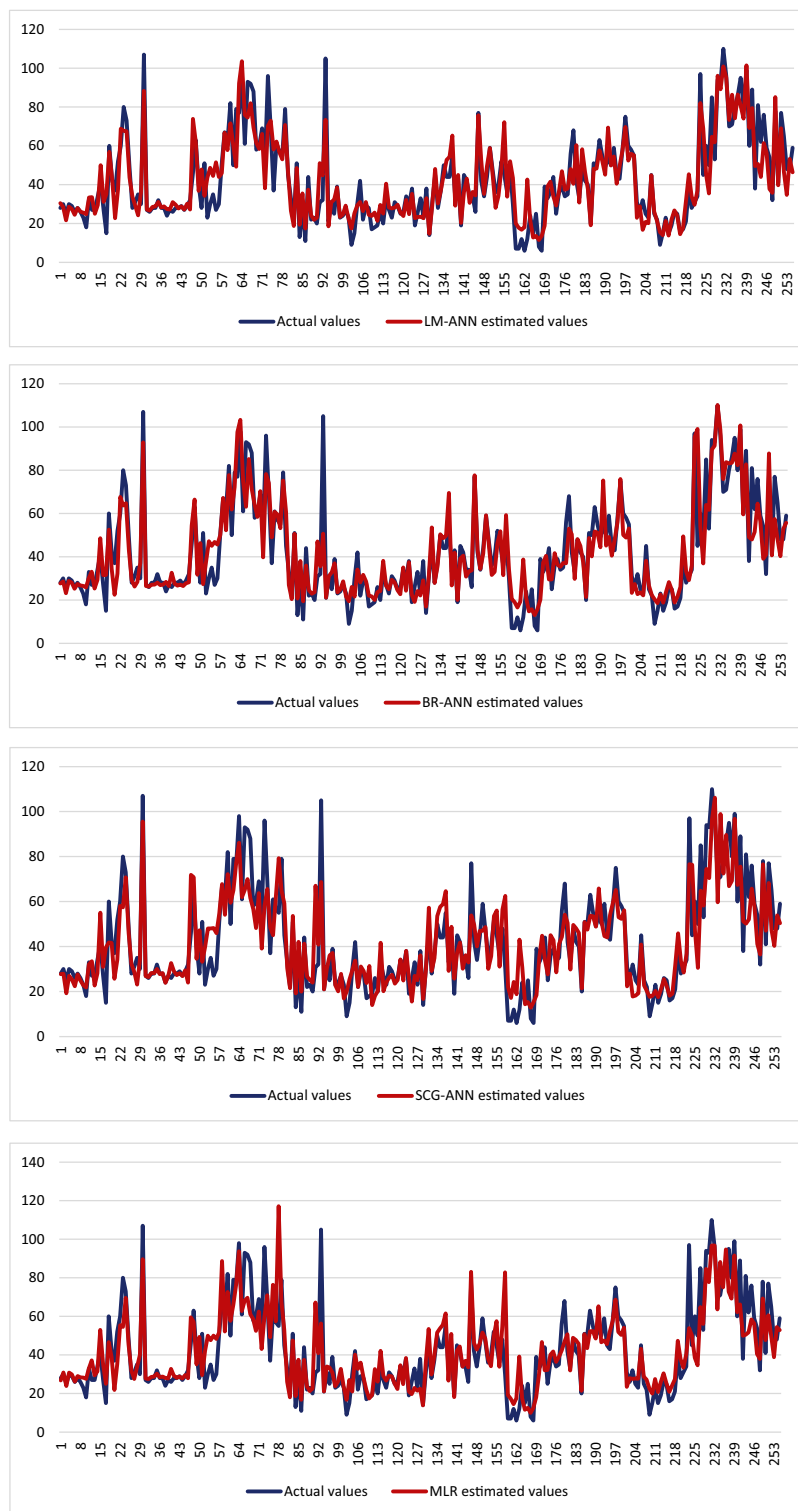**Figure 7.** Regression plot of SCG-ANN.

Figure 8. Comparison of the observed and estimated values of PM$_{2.5}$.

Data availability
The data set utilized for this study is available on https://cpcb.nic.in/.

References

1. Kampa, M. & Castanas, E. Human health effects of air pollution. *Environ. Pollut.* **151**(2), 362–367. https://doi.org/10.1016/j.envpol.2007.06.012 (2008).
2. Qin, G. & Meng, Z. Effects of sulfur dioxide derivatives on expression of oncogenes and tumor suppressor genes in human bronchial epithelial cells. *Food Chem. Toxicol.* **47**(4), 734–744. https://doi.org/10.1016/j.fct.2009.01.005 (2009).
3. Iordache, S., Dunea, D. & Bøhler, T. Current status of citizens protection against the risk of air pollution in urban areas. In *Methods to Assess the Effects of Air Pollution with Particulate Matter on Children's Health (in Romanian)* (eds Iordache, S. *et al.*) 1–44 (MatrixRom, 2014).
4. Mehmood, K., Saifullah, Iqbal, M., Rengel, Z. & Abrar, M. M. Pakistan and India collaboration to improve regional air quality has never been more promising. *Integr. Environ. Assess. Manag.* **16**(5), 549–551. https://doi.org/10.1002/ieam.4292 (2020).
5. Yang, B., Guo, J. & Xiao, C. Effect of PM$_{2.5}$ environmental pollution on rat lung. *Environ. Sci. Pollut. Res.* **25**(36), 36136–36146. https://doi.org/10.1007/s11356-018-3492-y (2018).
6. Baker, K. R. & Foley, K. M. A nonlinear regression model estimating single source concentrations of primary and secondarily formed PM2.5. *Atmos. Environ.* **45**(22), 3758–3767. https://doi.org/10.1016/j.atmosenv.2011.03.074 (2011).
7. Goudarzi, G. *et al.* Health risk assessment on human exposed to heavy metals in the ambient air PM 10 in Ahvaz, southwest Iran. *Int. J. Biometeorol.* **62**, 1075–1083. https://doi.org/10.1007/s00484-021-02172-3 (2018).
8. Faraji Ghasemi, F. *et al.* Levels and ecological and health risk assessment of PM 2.5-bound heavy metals in the northern part of the Persian Gulf. *Environ. Sci. Pollut. Res.* **27**, 5305–5313. https://doi.org/10.1007/s11356-019-07272-7 (2020).
9. Tahery, N. *et al.* Estimation of PM 10 pollutant and its effect on total mortality (TM), hospitalizations due to cardiovascular diseases (HACD), and respiratory disease (HARD) outcome. *Environ. Sci. Pollut. Res.* **28**, 22123–22130. https://doi.org/10.1007/s11356-020-12052-9 (2021).
10. Dastoorpoor, M. *et al.* Exposure to particulate matter and carbon monoxide and cause-specific Cardiovascular-Respiratory disease mortality in Ahvaz. *Toxin Rev.* **40**(4), 1362–1372. https://doi.org/10.1080/15569543.2020.1716256 (2021).
11. Moradi, M., Mokhtari, A., Mohammadi, M. J., Hadei, M. & Vosoughi, M. Estimation of long-term and short-term health effects attributed to PM 2.5 standard pollutants in the air of Ardabil (using Air Q+ model). *Environ. Sci. Pollut. Res.* https://doi.org/10.1007/s11356-021-17303-x (2022).
12. Shahriyari, H. A. *et al.* Air pollution and human health risks: mechanisms and clinical manifestations of cardiovascular and respiratory diseases. *Toxin Rev.* **41**(2), 606–617. https://doi.org/10.1080/15569543.2021.1887261 (2022).
13. Mohammadi, M. J. *et al.* Cardiovascular disease, mortality and exposure to particulate matter (PM): A systematic review and meta-analysis. *Rev. Environ. Health* https://doi.org/10.1515/reveh-2022-0090 (2022).
14. Borsi, S. H. *et al.* Health endpoint of exposure to criteria air pollutants in ambient air of on a populated in Ahvaz City, Iran. *Front. Public Health* **10**, 869656. https://doi.org/10.3389/fpubh.2022.869656 (2022).
15. Abbasi-Kangevari, M. *et al.* Effect of air pollution on disease burden, mortality, and life expectancy in North Africa and the Middle East: A systematic analysis for the Global Burden of Disease Study 2019. *Lancet Planet. Health* **7**(5), e358–e369. https://doi.org/10.1016/S2542-5196(23)00053-0 (2023).
16. Nezhad, M. E., Goudarzi, G., Babaei, A. A. & Mohammadi, M. J. Characterization, ratio analysis, and carcinogenic risk assessment of polycyclic aromatic hydrocarbon compounds bounded PM10 in a southwest of Iran. *Clin. Epidemiol. Glob. Health* **24**, 101419. https://doi.org/10.1016/j.cegh.2023.101419 (2023).
17. El-Shahawy, M. A. Prediction of air-pollution episodes. *Bound. Layer Meteorol.* **104**(2), 319–329. https://doi.org/10.1023/A:1016052013540 (2002).
18. Mehmood, K. *et al.* Predicting the quality of air with machine learning approaches: Current research priorities and future perspectives. *J. Clean. Prod.* https://doi.org/10.1016/j.jclepro.2022.134656 (2022).
19. Boznar, M., Lesjak, M. & Mlakar, P. A neural network-based method for short-term predictions of ambient SO$_2$ concentrations in highly polluted industrial areas of complex terrain. *Atmos. Environ. B Urban Atmos.* **27**(2), 221–230. https://doi.org/10.1016/0957-1272(93)90007-S (1993).
20. Gardner, M. W. & Dorling, S. R. Neural network modelling and prediction of hourly NOx and NO$_2$ concentrations in urban air in London. *Atmos. Environ.* **33**(5), 709–719. https://doi.org/10.1016/S1352-2310(98)00230-1 (1999).
21. Hadjiiski, L. & Hopke, P. Application of artificial neural networks to modeling and prediction of ambient ozone concentrations. *J. Waste Manag. Assoc.* **50**(5), 894–901. https://doi.org/10.1080/10473289.2000.10464105 (2000).
22. Chaloulakou, A., Grivas, G. & Spyrellis, N. Neural network and multiple regression models for PM10 prediction in Athens: A comparative assessment. *J. Air Waste Manag. Assoc.* **53**(10), 1183–1190. https://doi.org/10.1080/10473289.2003.10466276 (2003).
23. Kolehmainen, M., Martikainen, H. & Ruuskanen, J. Neural networks and periodic components used in air quality forecasting. *Atmos. Environ.* **35**(5), 815–825. https://doi.org/10.1016/S1352-2310(00)00385-X (2001).
24. Nagendra, S. M. & Khare, M. Modelling urban air quality using artificial neural network. *Clean Technol. Environ. Policy* **7**(2), 116–126. https://doi.org/10.1007/s10098-004-0267-6 (2005).
25. Chen, L. & Pai, T. Y. Comparisons of GM (1, 1), and BPNN for predicting hourly particulate matter in Dali area of Taichung City, Taiwan. *Atmos. Pollut. Res.* **6**(4), 572–580. https://doi.org/10.5094/APR.2015.064 (2015).
26. Bai, Y., Li, Y., Wang, X., Xie, J. & Li, C. Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions. *Atmos. Pollut. Res.* **7**(3), 557–566. https://doi.org/10.1016/j.apr.2016.01.004 (2016).
27. Wang, D. & Lu, W. Z. Forecasting of ozone level in time series using MLP model with a novel hybrid training algorithm. *Atmos. Environ.* **40**(5), 913–924. https://doi.org/10.1016/j.atmosenv.2005.10.042 (2006).
28. Durao, R. M., Mendes, M. T. & Pereira, M. J. Forecasting O$_3$ levels in industrial area surroundings up to 24 h in advance, combining classification trees and MLP models. *Atmos. Pollut. Res.* **7**(6), 961–970. https://doi.org/10.1016/j.apr.2016.05.008 (2016).
29. Lu, W. Z., Wang, W. J., Wang, X. K., Yan, S. H. & Lam, J. C. Potential assessment of a neural network model with PCA/RBF approach for forecasting pollutant trends in Mong Kok urban air, Hong Kong. *Environ. Res.* **96**(1), 79–87. https://doi.org/10.1016/j.envres.2003.11.003 (2004).
30. Iliyas, S. A., Elshafei, M., Habib, M. A. & Adeniran, A. A. RBF neural network inferential sensor for process emission monitoring. *Control Eng. Pract.* **21**(7), 962–970. https://doi.org/10.1016/j.conengprac.2013.01.007 (2013).
31. Shahraiyni, H. T., Sodoudi, S., Kerschbaumer, A. & Cubasch, U. A new structure identification scheme for ANFIS and its application for the simulation of virtual air pollution monitoring stations in urban areas. *Eng. Appl. Artif. Intell.* **41**, 175–182. https://doi.org/10.1016/j.engappai.2015.02.010 (2015).
32. Prasad, K., Gorai, A. K. & Goyal, P. Development of ANFIS models for air quality forecasting and input optimization for reducing the computational cost and time. *Atmos. Environ.* **128**, 246–262. https://doi.org/10.1016/j.atmosenv.2016.01.007 (2016).
33. Maier, H. R. & Dandy, G. C. Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications. *Environ. Model. Softw.* **15**(1), 101–124. https://doi.org/10.1016/S1364-8152(99)00007-9 (2000).
34. Maier, H. R., Morgan, N. & Chow, C. W. Use of artificial neural networks for predicting optimal alum doses and treated water quality parameters. *Environ. Model. Softw.* **19**(5), 485–494. https://doi.org/10.1016/S1364-8152(03)00163-4 (2004).
35. Pérez, P., Trier, A. & Reyes, J. Prediction of PM$_{2.5}$ concentrations several hours in advance using neural networks in Santiago, Chile. *Atmos. Environ.* **34**(8), 1189–1196. https://doi.org/10.1016/S1352-2310(99)00316-7 (2000).

36. McKendry, I. G. Evaluation of artificial neural networks for fine particulate pollution (PM10 and PM2.5) forecasting. *J. Air Waste Manag. Assoc.* **52**(9), 1096–1101. https://doi.org/10.1080/10473289.2002.10470836 (2002).
37. Chelani, A. B., Gajghate, D. G. & Hasan, M. Z. Prediction of ambient PM10 and toxic metals using artificial neural networks. *J. Air Waste Manag. Assoc.* **52**(7), 805–810. https://doi.org/10.1080/10473289.2002.10470827 (2002).
38. Tecer, L. H. Prediction of SO₂ and PM concentrations in a coastal mining area (Zonguldak, Turkey) using an artificial neural network. *Polish J. Environ. Stud.* **16**(4), 633–638 (2007).
39. Pires, J. C. M., Martins, F. G., Sousa, S. I. V., Ferraz, M. C. M. A. & Pereira, M. C. Prediction of the daily mean PM10 concentrations using linear models. *Am. J. Environ. Sci.* **4**(5), 445. https://doi.org/10.3844/ajessp.2008.445.453 (2008).
40. Paschalidou, A. K., Karakitsios, S., Kleanthous, S. & Kassomenos, P. A. Forecasting hourly PM10 concentration in Cyprus through artificial neural networks and multiple regression models: Implications to local environmental management. *Environ. Sci. Pollut. Res.* **18**(2), 316–327. https://doi.org/10.1007/s11356-010-0375-2 (2011).
41. Roy, S., Adhikari, G. R., Renaldy, T. A. & Jha, A. K. Development of multiple regression and neural network models for assessment of blasting dust at a large surface coal mine. *J. Environ. Sci. Technol.* **4**, 284–301 (2011).
42. Kurt, A. & Oktay, A. B. Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks. *Expert Syst. Appl.* **37**(12), 7986–7992. https://doi.org/10.1016/j.eswa.2010.05.093 (2010).
43. de Gennaro, G. *et al.* Neural network model for the prediction of PM10 daily concentrations in two sites in the Western Mediterranean. *Sci. Total Environ.* **463**, 875–883. https://doi.org/10.1016/j.scitotenv.2013.06.093 (2013).
44. Fu, M., Wang, W., Le, Z. & Khorram, M. S. Prediction of particular matter concentrations by developed feed-forward neural network with rolling mechanism and gray model. *Neural Comput. Appl.* **26**(8), 1789–1797. https://doi.org/10.1007/s00521-015-1853-8 (2015).
45. Fallahizadeh, S., Kermani, M., Esrafili, A., Asadgol, Z. & Gholami, M. The effects of meteorological parameters on PM10: Health impacts assessment using AirQ+ model and prediction by an artificial neural network (ANN). *Urban Clim.* **38**, 100905. https://doi.org/10.1016/j.uclim.2021.100905 (2021).
46. Bhardwaj, R. & Pruthi, D. Evolutionary techniques for optimizing air quality model. *Procedia Comput. Sci.* **167**, 1872–1879. https://doi.org/10.1016/j.procs.2020.03.206 (2020).
47. He, J. *et al.* Numerical model-based artificial neural network model and its application for quantifying impact factors of urban air quality. *Water Air Soil Pollut.* **227**, 1–16. https://doi.org/10.1007/s11270-016-2930-z (2016).
48. Maleki, H. *et al.* Air pollution prediction by using an artificial neural network model. *Clean Technol. Environ. Policy* **21**, 1341–1352. https://doi.org/10.1007/s10098-019-01709-w (2019).
49. Cakir, S. & Sita, M. Evaluating the performance of ANN in predicting the concentrations of ambient air pollutants in Nicosia. *Atmos. Pollut. Res.* **11**(12), 2327–2334. https://doi.org/10.1016/j.apr.2020.06.011 (2020).

Author contributions

S.G. Conceptualization, Methodology, Writing—original draft. A.B. Investigation, Writing—review & editing, Supervision. A.P. Methodology, Writing—review & editing, Supervision. N.M. Validation, Writing—review & editing, Software. A.S. Conceptualization, Writing—review & editing. F.G. Writing—review & editing.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to F.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.