



Relief-Based Feature Selection for ANN-Driven Air Pollution Prediction and School Safety Policy in India

Vibha Yadav · Amit Kumar Yadav ·
Rohit Khargotra

Received: 18 July 2025 / Accepted: 17 October 2025 / Published online: 25 October 2025
© The Author(s) 2025

Abstract Public schools educate diverse socioeconomic groups, making air quality monitoring vital for students' health and learning. These students are vulnerable to pollutants such as NO₂, SO₂, and PM_{2.5}. Therefore air pollutants (AP) prediction helps in children's health protection, learning enhancement, preventive measures, policy and awareness. Predictive models using ANNs, feature selection poses challenges due to data's high dimensionality which increases over fitting risk and model complexity. Irrelevant features introduce noise, thereby reducing accuracy. So exploration of feature selection technique on ANN model accuracy remains an important

area. This study explore relief algorithm (RA) for input selection to assess NO₂, SO₂, and PM_{2.5} prediction accuracy using multilayer feed forward neural network (MLFFNN), radial basis function neural network (RBFNN), generalized regression neural network (GRNN) and multi linear regression (MLR) model. RA ranks variables like wind direction, vertical wind speed, and solar radiation. It significantly enhanced model performance, increasing R² by 82.79% (SO₂) in RA-MLFFNN-1 and 77.37% (NO₂) in RA-MLFFNN-2. However, PM_{2.5} prediction using RA-MLFFNN-3 saw reduced accuracy, indicating model-specific effectiveness. The findings offer key insights for improving school air quality policy.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11270-025-08759-5>.

V. Yadav
School of Agricultural Sciences, IIMT University, Meerut,
Uttar Pradesh-250007, India
e-mail: yvibha3@gmail.com; vibha_agrisc@iimtindia.net

A. K. Yadav
School of Computer Science and Artificial Intelligence,
SR University, Warangal 506371, Telangana, India
e-mail: amit1986.529@rediffmail.com; amitkumaryadav@sru.edu.in

R. Khargotra (✉)
Escuela Politécnica Superior, Universidad Francisco
de Vitoria, Ctra. Pozuelo-Majadahonda Km 1.800,
28223 Madrid, Pozuelo de Alarcón, Spain
e-mail: rohit.khargotra@ufv.es

Keywords Air pollutants · ANN · RA · MLR · RBFNN · GRNN

1 Introduction

Air pollutants (AP), when present in the atmosphere, cause harm to the environment, materials, and public health and can also affect other forms of life (<https://www.who.int>, health-topics, air-pollution. n.d.). All organisms face potential threats when undesirable environmental changes negatively impact their health (Hassan et al., 2023). Industrial Revolution, progress has come at the cost of severe ecological damage, driving climate change and health risks (Akhtar et al., 2025). Pollutant affects children's health (Tas

et al., 2025). The undesirable environmental shift can potentially harm natural resources and the ecosystem. It can modify the biosphere in various ways, both directly and indirectly (Ebrahimi et al., 2024; Nizhelskiy et al., 2025). AP mainly arises from natural occurrences such as forest fires, lightning, and volcanic eruptions, emitting substantial amounts of dust, sulphur and other pollutants (Nandan et al., 2023). It is referred to as natural AP. Artificial or man-made AP, also known as anthropogenic AP, is mainly generated by various human activities such as agricultural operations, urbanization, and the combustion of fossil fuels (Yadav et al., 2024). Accurate monitoring of daily AP is an essential component of environmental research focused on human health. Multiple researchers have offered daily forecasts for air pollution in different regions (Pande et al., 2024; Petrić et al., 2024; Singh & Suthar, 2025).

Liang et al., 2023 proposed an innovative Transformer framework called AirFormer to jointly forecast national air quality in China, covering thousands of locations with unprecedented spatial granularity. The learning process is divided into two stages: a ground-up systematic stage and a top-down stochastic stage. The model lowers prediction errors by 5% to 8% for 72-h forecasts. Chen et al., 2022 implemented a hybrid of convolution neural network (CNN) and transformer model to predict O_3 concentration in Beijing, China. The CNN-transformer model outperforms other models with an RMSE of 16.27 for long-term predictions and 7.75 for near-term predictions. Yu et al., 2023 used multi-head attention-driven neural network design and spatiotemporal (ST)-transformer for predicting $PM_{2.5}$ levels in Los Angeles. Based on an average RMSE of $6.92 (\pm 2.93) \mu g/m^3$, the ST-transformer model was more effective than other models in the research. Iskandaryan et al., 2023 developed a method to forecast the concentration of NO_2 in Madrid, Spain. Their approach uses an attention temporal graph convolutional network that combines attention, a gated recurrent unit, and a graph convolution network. The authors said the suggested model outperforms the temporal graph convolution network, long short-term memory (LSTM), and gated memory cell. Wu et al., 2024 used the LSTM method to predict the concentration of $PM_{2.5}$ in Heilongjiang Province, China. The study's findings suggest that the LSTM neural network model can adequately forecast the levels of air pollutants in long-term series.

Yang & Huang, 2024 used extreme learning machines to forecast the $PM_{2.5}$ level in Xi'an, China. Maximum improvements in MAE (82.64%), RMSE (79.21%), and MAPE (84.72%) were achieved with the suggested model. To predict the NO_x - O_3 concentrations in Hong Kong, Lin et al., 2024 used a computational fluid dynamics-based neural network. The study concludes that the backpropagation neural network presents high NO concentration and wind speed prediction accuracy. Lin et al., 2025 used the LSTM model to forecast the levels of AP concentrations in Taipei, Taiwan. The LSTM model, which combines the generalized additive model with Bayesian network results, produces outstanding performance. The first and second-order air pollutants, including NO , CO , NO_2 , O_3 , NO_x , PM_{10} , and $PM_{2.5}$, have correlation coefficients (R^2) greater than 0.9 and 0.8, respectively. Monteiro et al., 2024 present a study using artificial neural networks (ANNs) to predict AP produced during direct biomass combustion in northeast Brazil. The study used 116 worldwide biomasses as input data, overcoming northeast Brazil's lack of experimental resources. The ANNs were optimized by k-fold cross-validation, and accuracy evaluations were performed using statistical errors. Accuracy parameters showed strong statistical performance with high emission prediction accuracy. Deviations between expected and actual values for biomass emissions in northeastern Brazil ~0.01% (CO), 0.14% (NO), 0.05% (SO_2), and 10–6% (CO_2). Convergent cross-mapping is used in the research by Liponhay et al., 2024 to deduce the time-delayed causal network for particulate matter (PM_{10} , $PM_{2.5}$, and PM_1), volatile organic compounds, and local meteorological variables in Baguio City, Philippines. Pollutants respond to changes in air quality with a delayed time lag, and wind direction is a significant factor influencing air quality immediately. After a day, the temperature takes control, and after eight days on average, pollutants have an effect. The research sheds light on the relationship between regional weather patterns and air quality, offering essential insights into mitigation strategies and environmental management.

Zhang et al., 2023a introduce a novel deep learning-driven model called GATBL-learning for AP diffusion trend prediction for North China. The model uses an attention-based graph network and encoder-decoder model to extract spatial and temporal information from each city in a region. The model extracts

time series features and results spatiotemporal data. Seasonality discipline and training seasonal aspects are considered to enhance the model's flexibility further. According to the data, the model is more accurate and efficient in predicting AP. Karimi et al., 2023a used machine learning algorithms to predict CO₂, CO, O₃, NO₂, PM_{2.5}, SO₂, and PM₁₀ levels in Siraf City. The simulation accuracy for CO₂ (12.91%), NO₂ (0.94%), CO (12.54%), O₃ (3.05%), SO₂ (6.7%), PM₁₀ (4.62%), and PM_{2.5} (12.47%) was enhanced by the established equations.

The study by Ravindiran et al., 2023 examines the effect of air pollution in Chennai, India, using machine learning models like XGBoost, random forest, Bagging Regressor, and LGBM Regressor. Data from 2017 to 2022 indicates the peak air quality index (AQI) in January and the lowest in August. PM_{2.5} has the most significant impact on AQI. The XGBoost model showed strong performance, while the Light GBM model was less effective. The study by Ren & Ji, 2023 presents a novel method using a Takagi–Sugeno fuzzy model to predict AP (particulate matter and ozone pollution emitted by industry) in Shanghai, China. The Takagi–Sugeno model can understand the spatial–temporal interdependence and local trend pattern in time series data about air quality that several variables affect. Luo & Gong, 2023 propose an ARIMA-WOA-LSTM model to accurately predict AP (CO, NO₂, O₃, PM_{2.5}, PM₁₀, SO₂) in Shijiazhuang, China, ensuring efficient management. The model employs the WOA-LSTM model to make nonlinear predictions after extracting linear data from air pollution data. Finding the optimal LSTM hyperparameters is done using the whale method. This model is head and shoulders above the competition regarding pollutant predictions, model predictions, and forecast stability. Additionally, compared to the standalone model, the integrated one performs better.

A novel termite fly optimization (TFO) algorithm is presented to predict AP levels in various locations of India by Rani and Sampathkumar (Rani & Sampathkumar, 2023). By comparing the search space's quantities of contaminants and termites, the TFO algorithm can make human health predictions. The harm to human health varied against the concentrations of pollutants and model forecast AQI. The model is trained using an open-source dataset and then tested using a real-time dataset from Manali, Chennai. The study by Wang et al., 2023 presents

a model that uses an attention mechanism, CNN, and LSTM to predict AP-like O₃ and PM_{2.5} levels and AQI. The model outperforms other models and has a higher correlation coefficient than other models. Sobol-based sensitivity analysis identifies critical influencing factors, with solar irradiance being the most significant for O₃, CO for PM_{2.5}, and particulate matter for AQI. Removing these variables improves model efficiency and reduces computational costs. Zhang et al., 2023b present Dev-LSTM, a deep learning model that extracts spatial feature correlations from AP concentration data, preserving data integrity. It mines temporal feature relationships for accurate predictions. Experimental findings demonstrate that Dev-LSTM surpasses traditional models across multiple metrics. The research by Karimi et al., 2023b investigates the association between vertical meteorological features and air pollution concentrations in Sanandaj, Iran. Four models were created: adaptive Neuro-Fuzzy inference system (ANFIS), multi-linear perceptron (MLP), multi-linear regression (MLR), and generalized regression neural network (GRNN). The study found that vertical meteorological factors mainly impacted changes in PM₁₀ concentration. The GRNN model demonstrated the most significant outcomes. The ANFIS model had the highest level of accuracy in estimating PM_{2.5} concentration. The study's findings indicate that the GRNN and ANFIS models exhibit enhanced performance over other models in forecasting air pollution levels using vertical atmospheric data.

Accurate AP prediction and policy is vital for health, sustainability, and compliance, but remains challenging due to complex chemical reactions and nonlinear interactions among pollutants, weather, and human activities. This research addresses these challenges by proposing a novel application of the RA for input variable selection, aimed at enhancing the predictive performance of Artificial Neural Network (ANN) and Multiple Linear Regression (MLR) models. Unlike traditional approaches that rely on exhaustive input sets, our method systematically identifies the most influential features, thereby improving model accuracy and computational efficiency. Moreover, the study goes beyond technical modeling by offering **practical policy recommendations**. This information is specifically designed for decision-makers and school authorities, allowing them to make

well-informed decisions, such as adjusting outdoor activities according to daily pollution forecasts to enhance the well-being of children.

2 Methodology

2.1 Research Region and Data Gathering

This study uses data from 2016–2017 public school in Nagar Nigam Hazratganj in Lucknow. It is the capital city of Uttar Pradesh, India, having 27.38° North, latitude, and longitude of 81.83° East with ~123 m elevation. The humid subtropical climate with an average temperature variation of 3 to 30° is found here. Daily concentration of meteorological variables like vertical wind speed (VWS), solar radiation (SR), wind speed (WS), and wind direction (WD), as well as the AP (PM_{2.5}, NO₂, and SO₂), have been used to develop ANN models. Data used in the current research was taken from the Central Pollution Control Board site in New Delhi. The values of daily VWS, SR, WS, WD, PM_{2.5}, NO₂, and SO₂ are shown in Fig. S1, Fig. S2, Fig.S3, Fig.S4, Fig. S6 and Fig. S7 (Supplementary material). The PM_{2.5}, NO₂, and SO₂ vary from 8.53 to 334.75, 5.17 to 79.56, and 0.46 to 13.73. The average value of PM_{2.5}, NO₂, and SO₂ are 85.9706, 30.2885, and 3.5179 respectively. The statistics calculated with standard deviation in MATLAB using code $S = \text{std}(A)$ are represented in Table 1, where AP or meteorological variables are marked by A. The standard deviation for wind direction is more whereas it is less for wind speed. PM_{2.5} has high standard deviation indicate individual experience short term spike which is dangerous to health. It means someday value of PM_{2.5} is more due to traffic congestion or nearby industrial emissions

while other days it had much cleaner air. It also indicates that a highly accurate predictive model is required to predict spike in the data.

2.2 RA Based Relevant Input Selection for AP Prediction using ANN Models

The RA is derived from instant-based learning, first created by Kira & Rendell, 1994 and subsequently enhanced by Kononenko, 1994. RA is very effective in accurately determining characteristics. The system identifies and chooses the appropriate characteristics to focus on (Arauzo-Azofra et al., 2004). The algorithm identifies the closest neighbor by randomly selecting a sample from the dataset. The feature values of the closest neighbors are evaluated against those of the sampled instance to adjust the relevance score for each feature. RELIEF looks for two closest neighbors in each instance: one from the same class (nearest hit) and the other from a different class (known as nearest miss) (Kononenko, 1994). The estimate $W[A]$ provided by RELIEF approximates the difference in probability as follows:

$$W[A] = P(\text{different value of } A \setminus \text{nearest instance from a different class}) - P(\text{different value of } A \setminus \text{nearest instance from the same class}).$$

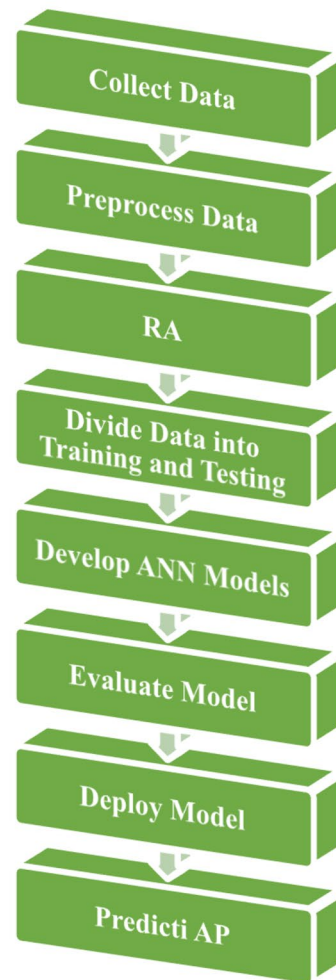
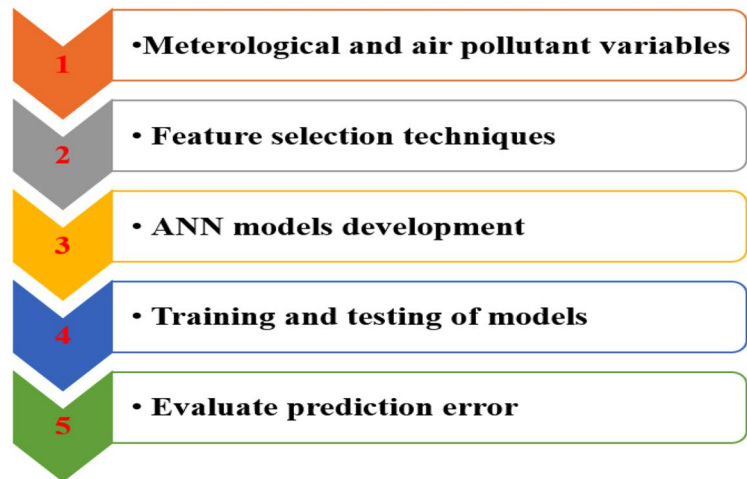
RA was chosen because it preserves the interpretability of original features (unlike Principal Component Analysis), is computationally more efficient than Boruta, and avoids the overfitting risks often associated with wrapper approaches. This balance of efficiency, robustness, and interpretability makes Relief particularly well-suited.

RA selects relevant input from SR, WS, WD, and VWS for AP prediction. ANN utilized three layers: input, hidden, and output. The standard ANN model for AP prediction is shown in Fig. 1 (Qazi et al., 2015). The

Table 1 Descriptive statistics of air pollutants and meteorological variables

AP/meteorological variables	Minimum	Maximum	Average	Standard deviation
PM _{2.5} (µg/m ³)	8.5300	334.7500	85.9706	58.5415
NO ₂	5.1700	79.5600	30.2885	15.6888
SO ₂	0.4600	13.7300	3.5179	2.1402
Solar radiation (W/m ²)	1.5500	155.0400	49.2823	18.1085
Wind speed (m/s)	0.2100	0.8600	0.3608	0.0824
Vertical wind speed (m/s)	−0.1300	2.7800	0.075	0.1843
Wind direction	73.2200	313.4600	199.4960	66.6684

Fig. 1 Standard AP prediction pipeline



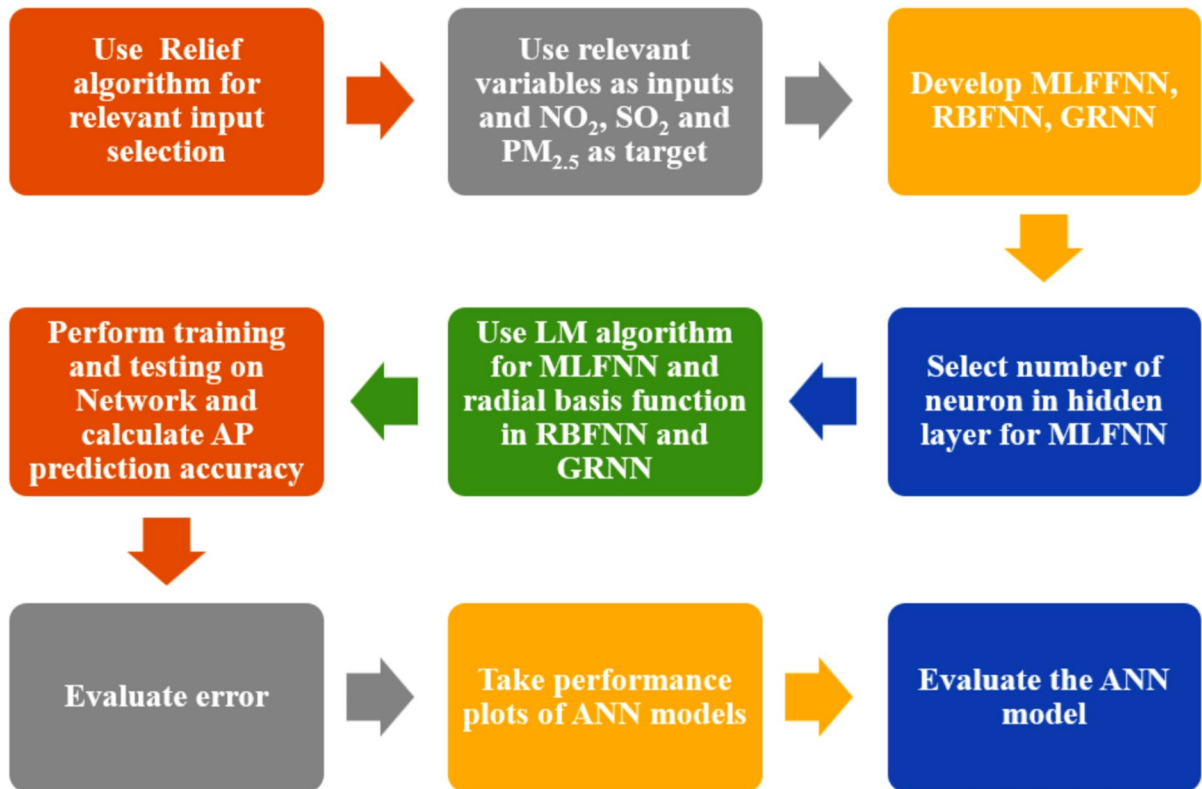


Fig. 2 Proposed algorithm for relevant input variable selection for ANN-based AP prediction

implemented ANN models are Multi-Layer Feedforward neural network (MLFFNN) with LM algorithm, Radial Basis Function Neural Network (RBFNN), and Generalized Regression Neural Network (GRNN). The proposed Relief-ANN algorithm used in this study is shown in Fig. 2.

2.3 ANN Prediction Models Accuracy Estimation

To evaluate AP prediction mode accuracy, mean bias error (MBE, or bias), mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), coefficient of determination (R^2) are used. These are defined as follows (Chaturvedi, 2024; Dawar et al., 2024; Yadav et al., 2025):

$$MAE = \left(\frac{1}{N} \sum_{i=1}^N |AP_{i(PR)} - AP_{i(actual)}| \right) \quad (1)$$

$$MBE = \frac{1}{N} \sum_{i=1}^N (AP_{i(PR)} - AP_{i(actual)}) \quad (2)$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (AP_{i(PR)} - AP_{i(actual)})^2 \quad (3)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (AP_{i(PR)} - AP_{i(actual)})^2} \quad (4)$$

$$R^2 = 1 - \frac{(RMSE)^2}{\text{var}(AP_{i(actual)})} \quad (5)$$

where $AP_{i(PR)}$ is the predicted value of air pollutant, $AP_{i(actual)}$ is the measured value of air pollutants, and n is the number of data samples.

The R^2 (coefficient of determination) quantifies the accuracy of ANN models in predicting trends in observed data. This is a comparison between the variability of the mistakes and the variability of the data being modelled. To get a flawless prediction, R^2 must

equal to 1. The RMSE quantifies the dispersion of the mistakes. MAE and RMSE indicate the magnitude of errors and may be seen as the 1-norm and 2-norm of the errors, respectively. The bias, also known as the MBE, is the average discrepancy between the predicted values of an ANN model and the actual values, indicating the presence of systematic errors in the model's tendency to either underestimate or overestimate the forecasts. MAE provides the mean value of the absolute magnitude of prediction errors.

3 Results and Discussion

The ANN model was developed with a **3–8–1 architecture**, using a **tansig activation function in the hidden layer** and a **linear function in the output layer**. Training was carried out with the **Levenberg–Marquardt algorithm** and **Mean Squared Error (MSE)** as the performance metric. For convergence, the **MLFFNN trained within 9–15 epochs**, while the **RBF network trained for 350 epochs**. All input and output data were **normalized to [0, 1]** before training and demoralized afterward. The dataset was divided into **70% training, 15% validation, and 15% testing**, with **early stopping** applied to prevent over fitting. To ensure reproducibility, a **fixed random seed** was used throughout the training process.

Table 2 Average Air pollutant monthly variation

Month	NO ₂	SO ₂	PM _{2.5}
January	27.12	1.64	24.98
February	30.03	3.49	27.11
March	23.71	2.18	35.84
April	40.52	2.65	105.05
May	41.02	4.68	157.21
June	40.72	4.32	153.36
July	34.85	2.70	136.94
August	37.73	3.79	97.93
September	46.56	5.38	113.88
October	5.71	3.17	78.38
November	16.48	4.42	63.06
December	24.11	4.70	49.99
Maximum	46.56	5.38	157.21
Minimum	5.71	1.64	24.98

3.1 Features of Air Pollution Levels

This study sought to develop a novel estimation model for a public school in Lucknow, using the concentration of AP and meteorological emissions. The observed average concentration with monthly variation of NO₂, SO₂, and PM_{2.5} from 2016 to 2017 is represented in Table 2. NO₂ has maximum value in September and minimum in October. SO₂ is highest in January and lowest in September, while PM_{2.5} peaks in May and drops to its minimum in January. PM_{2.5} shows the strongest seasonal variation with summer peaks, whereas NO₂ and SO₂ stay moderate with smaller seasonal spikes.

3.2 Modeling and Choosing Pertinent Input Factors for Prediction

AP used for the prediction of NO₂, SO₂, and PM_{2.5} are taken here. Key input variables are chosen through the relief attribute evaluator, as shown in Table 3. Wind direction determines movement of air pollutants from source to public school. Solar radiation affects pollutant formation and dispersion, so it is highly ranked by RA showing strong correlation with SO₂ and PM_{2.5}. To predict NO₂, SO₂, and PM_{2.5}, three relevant input variables are selected as per rank order 1, 2, and 3. The architecture of MLFFNN, RBFNN and GRNN are shown in Figs. 3, 4 to Fig. 5.

Based on relevant input variables, nine ANN models (RA-MLFFNN-1, RA-RBFNN-1, RA-GRNN-1, RA-MLFFNN-2, RA-RBFNN-2, RA-GRNN-2, RA-MLFFNN-3, RA-RBFNN-3, RA-GRNN-3) are developed. MLFFNN-1, RBFNN-1, GRNN-1, RA-MLFFNN-1, RA-RBFNN-1, RA-GRNN-1 are developed for prediction of SO₂, MLFFNN-2, RBFNN-2, GRNN-2, RA-MLFFNN-2, RA-RBFNN-2,

Table 3 Prediction of NO₂, SO₂, and PM_{2.5} using rank selection

Meteorological variables	Rank of input variables for prediction of:		
	NO ₂	SO ₂	PM _{2.5}
Solar radiation	3	2	2
Wind speed	4	4	3
Wind direction	1	1	1
Vertical wind speed	2	3	4

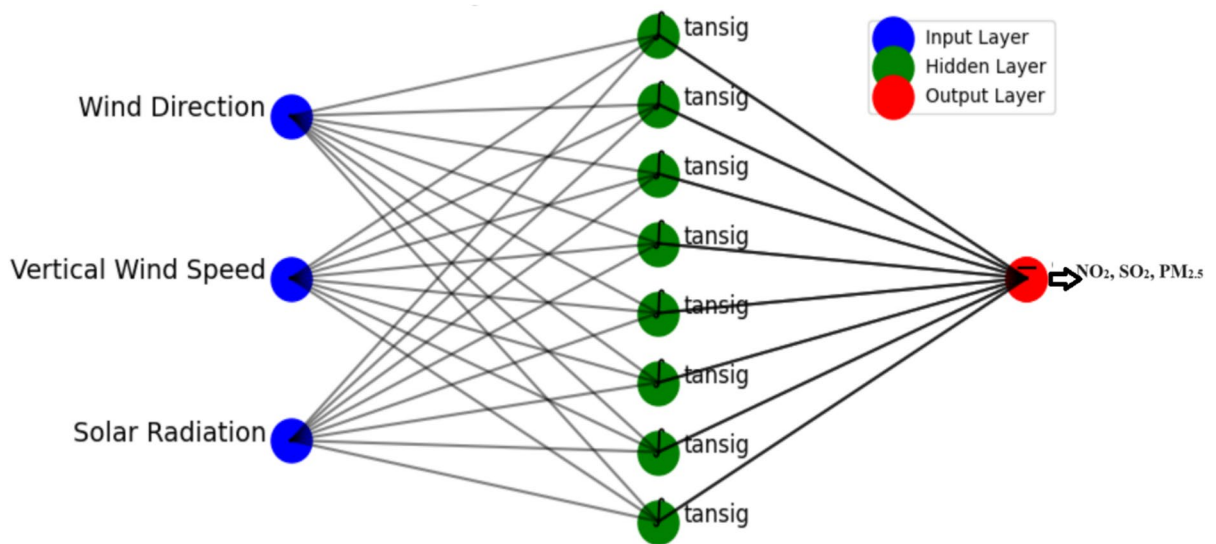


Fig. 3 MLFFNN architecture

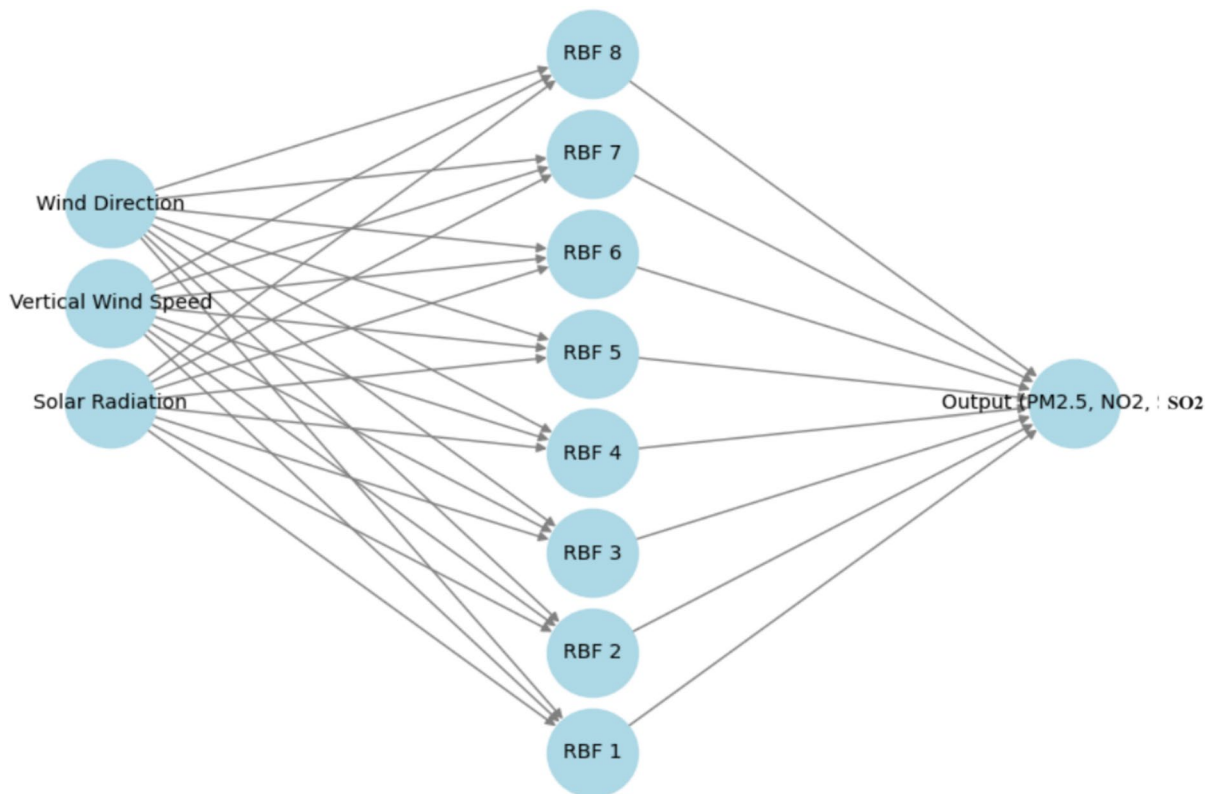


Fig. 4 GRNN architecture

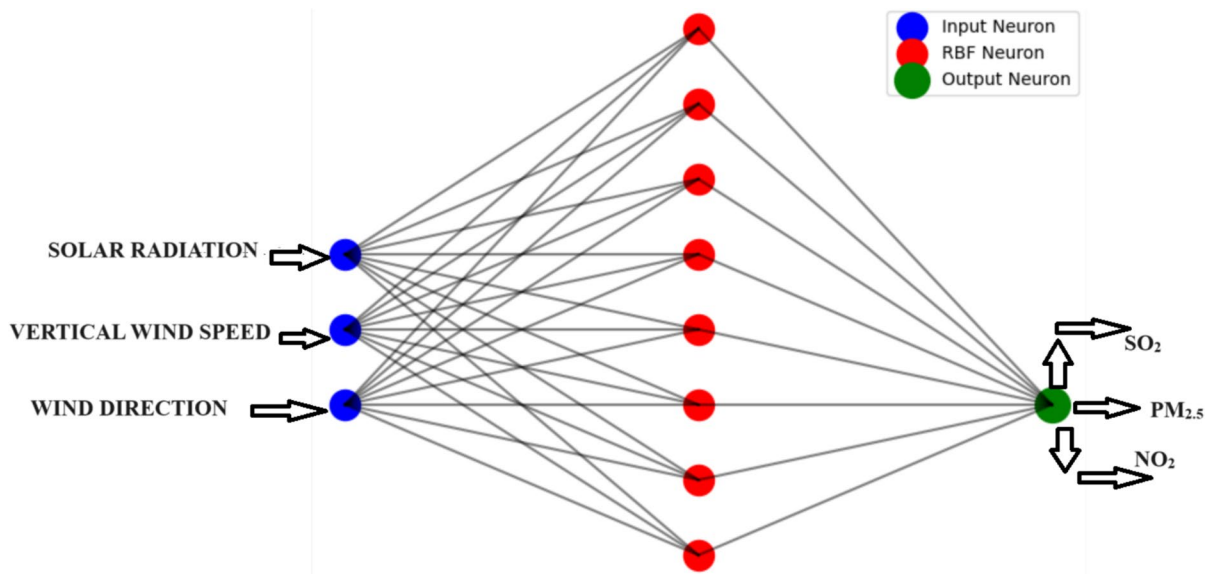
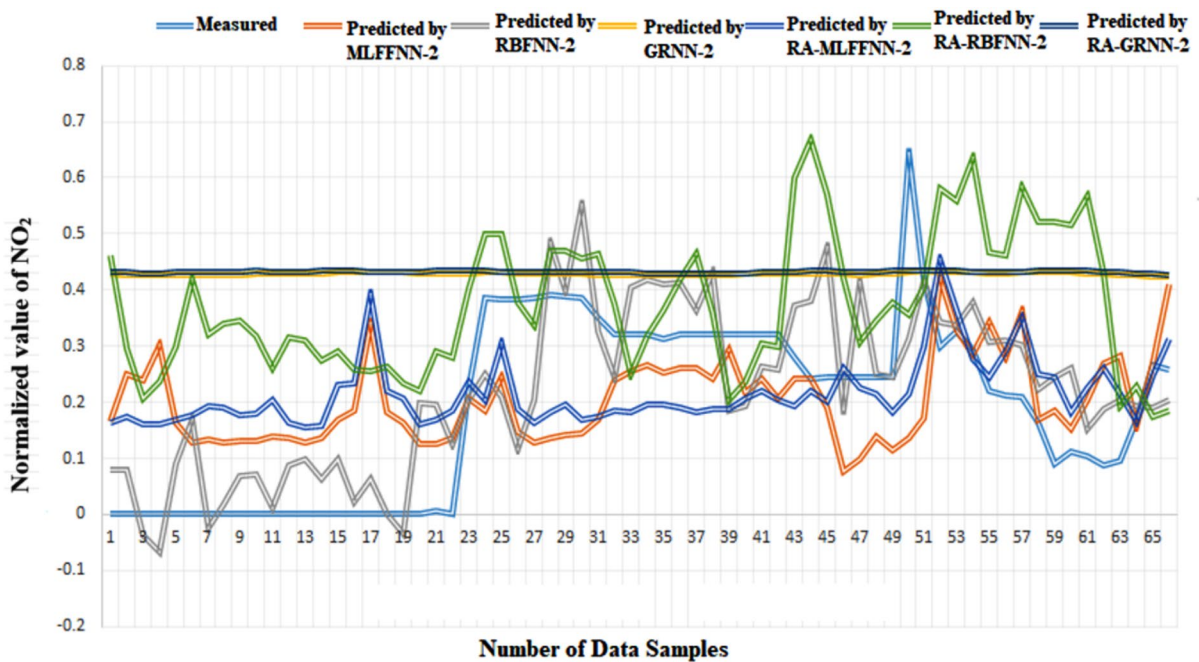


Fig. 5 RBFNN architecture

Fig. 6 Comparison between measured and predicted NO_2

RA-GRNN-2 are developed for prediction of NO_2 . MLFFNN-3, RBFNN-3, GRNN-3, RA-MLFFNN-3, RA-RBFNN-3, RA-GRNN-3 are developed for prediction of $\text{PM}_{2.5}$. The input variables to MLFFNN-1,

RBFNN-1, GRNN-1, MLFFNN-2, RBFNN-2, GRNN-2, MLFFNN-3, RBFNN-3, GRNN-3 are SR, WS, WD, and VWS. The performance plot of MLFFNN-1, RBFNN-1, GRNN-1, MLFFNN-2,

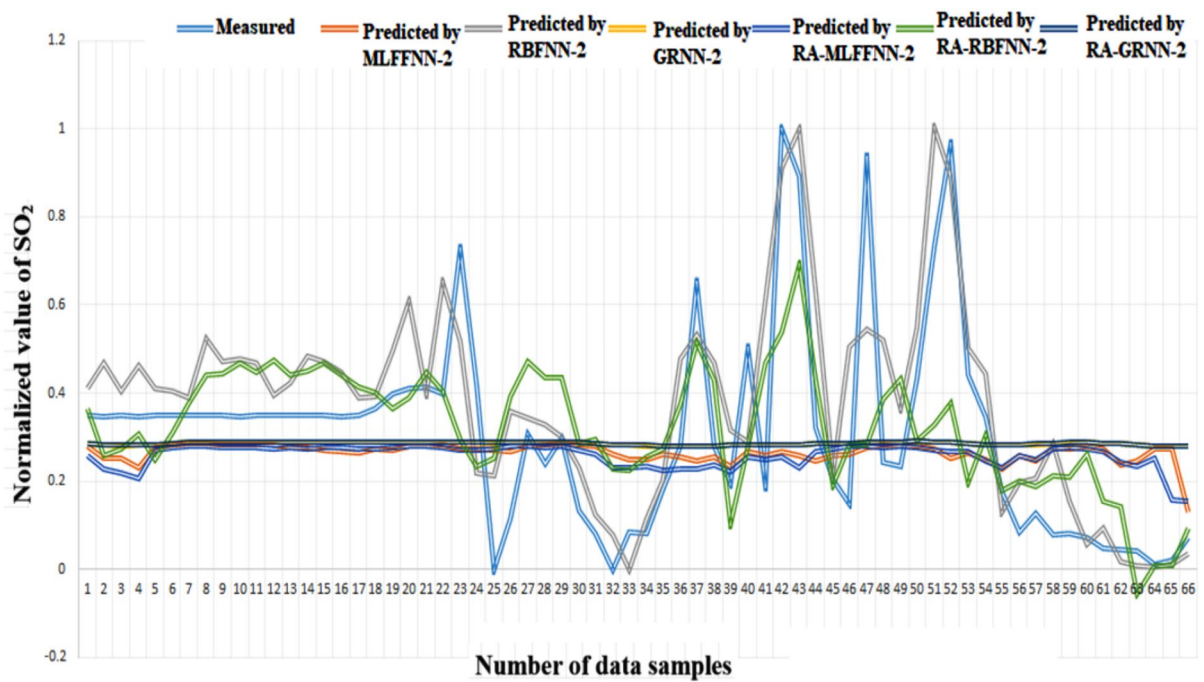


Fig. 7 Comparison between measured and predicted SO_2

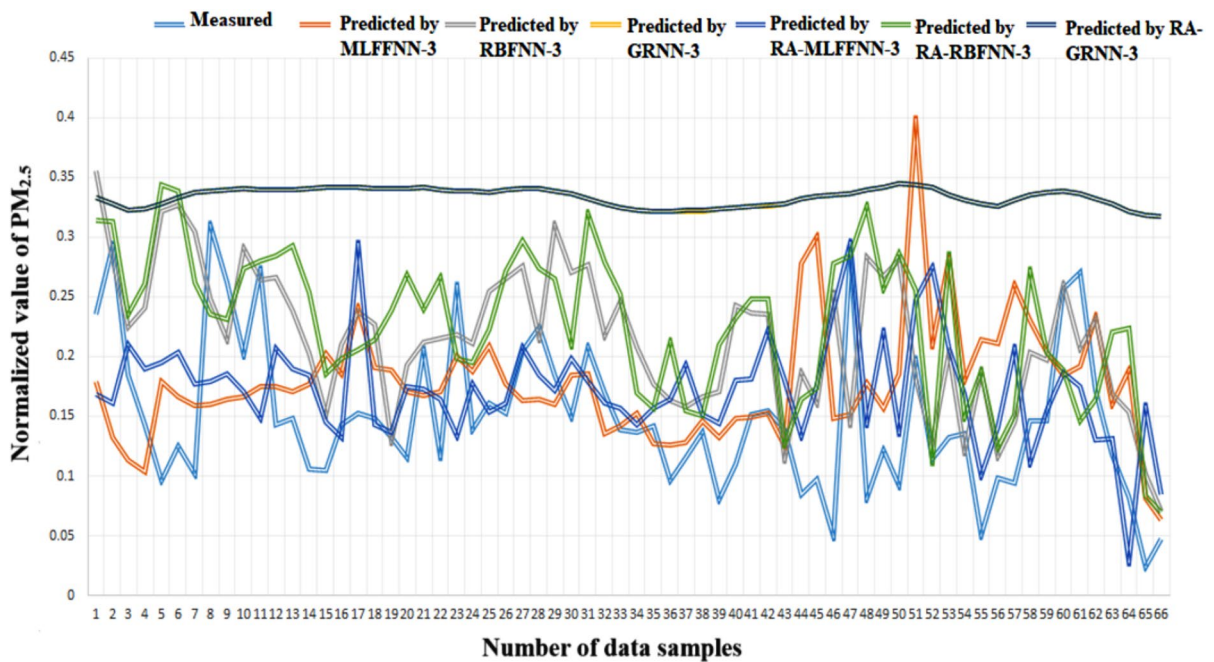


Fig. 8 Comparison between measured and predicted $\text{PM}_{2.5}$

RBFNN-2, GRNN-2, MLFFNN-3, RBFNN-3, GRNN-3, RA-MLFFNN-1, RA-RBFNN-1, RA-GRNN-1, RA-MLFFNN-2, RA-RBFNN-2, RA-GRNN-2, RA-MLFFNN-3, RA-RBFNN-3, RA-GRNN-3 are shown in Fig. S8, Fig. S9, Fig. S10, Fig. S11, Fig. S12, Fig. S13, Fig. S14, Fig. S15, Fig. S16, Fig. S17, Fig. S18, Fig. S19, respectively (Supplementary material). It is found that the MSE decreases

with an increase in the number of epochs. For further validation of ANN model prediction, Figs. 6, 7 to Fig. 8 represents the compared and measured values using different ANN models.

The error analysis of ANN models for SO₂, NO₂, and PM_{2.5} prediction are shown in Tables 4, 5 to Table 6, respectively. Table 4 shows that the R² value of RBFNN-1 is higher than other models. Relevant

Table 4 Error analysis of ANN models for SO₂ prediction

ANN Models	RMSE	MAE	MSE	MBE	R ²	Remarks
MLFFNN-1	0.2340	0.1647	0.00548	−0.0495	0.0059	Very low R ² ; weak explanatory power
RBFNN-1	0.0825	0.0567	0.0068	−0.0194	0.8961	Best performance overall (lowest errors, highest R ²)
GRNN-1	0.2726	0.2276	0.00743	−0.1133	0.0068	High error; R ² near zero suggests poor fit
RA-MLFFNN-1	0.2313	0.1653	0.0535	−0.0519	0.0343	High MSE with slightly better R ² vs. original
RA-RBFNN-1	0.2088	0.1543	0.0436	−0.0775	0.3929	Substantial drop in R ² after applying Relief
RA-GRNN-1	0.2725	0.2275	0.0742	−0.1130	0.0070	Minor improvement in R ² value

Table 5 Error analysis of ANN models for NO₂ prediction

ANN Models	RMSE	MAE	MSE	MBE	R ²	Remark
MLFFNN-2	0.1618	0.1348	0.0262	0.0151	−0.0391	Negative value of R ² shows poor prediction accuracy. It highlights under fit or over fits of data. It means the model's predictions are so poor that the mean of the observed data would give a better fit than the model
RBFNN-2	0.0890	0.0666	0.0079	−0.0197	0.7266	Best overall performance; high accuracy and strong model generalization
GRNN-2	0.2507	0.1909	0.0629	0.1878	−0.0012	Weak model with high errors and almost no predictive power (R ² ≈ 0)
RA-MLFFNN-2	0.1596	0.1385	0.0255	0.0268	0.0093	Slight improvement in MSE over MLFFNN-2 but still underperforms overall
RA-RBFNN-2	0.2086	0.1699	0.0435	0.1335	0.0671	Performance worsens significantly after relief; not recommended
RA-GRNN-2	0.2532	0.1935	0.0641	0.1911	−0.0017	Worst performer; high bias and negligible predictive capability

Table 6 Error analysis of ANN models for PM_{2.5} prediction

ANN Models	RMSE	MAE	MSE	MBE	R ²	Remark
MLFFNN-3	0.0836	0.0674	0.0070	0.0278	−0.5754	Poor Model fit due to negative R ² which shows over fitting or learning noise
RBFNN-3	0.1901	0.1678	0.0361	−0.0253	0.2882	Relatively high errors but acceptable R ² ; some generalization but low accuracy
GRNN-3	0.1430	0.1357	0.0205	0.1324	0.0769	Low R ² and high bias (MBE) indicate poor generalization
RA-MLFFNN-3	0.0705	0.0568	0.0050	0.0225	−0.1302	Input selection helped reduce error, but not overall model reliability
RA-RBFNN-3	0.1889	0.1683	0.0357	−0.0133	0.2878	Almost identical to RBFNN-3. Relief input selection has negligible impact on this model
RA-GRNN-3	0.1431	0.1358	0.0205	0.1325	0.0766	No improvement over GRNN-3, identical errors, suggesting Relief did not influence this model's behavior

input selection by RA only increases the R^2 value of RA-MLFFNN-1, RA-GRNN-1 compared to MLFFNN-1, GRNN-1 model for SO_2 Prediction by 82.79% and 2.94%. **RA** did not consistently improve performance. In RA-RBFNN-1, the R^2 dropped sharply (from 0.8961 to 0.3929), indicating that variable selection may have removed relevant features or altered model structure detrimentally. It means RBFNNs are sensitive to the **choice and number of input features**, and in certain cases, the feature subset selected by Relief may reduce the model's capacity to generalize, especially when redundant but complementary features are removed.

Table 5 and Table 6 show that relevant input selection by RA increases the R^2 value of RA-MLFFNN-2 and RA-MLFFNN-3 by -123.78% and 77.37% in comparison to only MLFFNN-2, MLFFNN-3 model, respectively, showing relevant input selection by RA is prominent in MLFNN for prediction of NO_2 and $PM_{2.5}$.

From Table 5 it is found that RBFNN-2 is the most suitable model for NO_2 prediction and RA not universally improve performance and model having negative or near-zero R^2 values are not suitable for NO_2 prediction. From Table 6, RA improves RMSE, MAE and MSE in MLFFNN but generalization capability is not improved. RBFNN present more robust in prediction of $PM_{2.5}$.

3.3 Multiple Linear Regression (MLR) Model

For every observation, the MLR model's output Y is connected to the input X . The central premise of this model is that the conditional mean function is linear, which means:

$$E(Y/X) = \alpha + \beta_1 X_1 + \dots + \beta_p X_p \quad (6)$$

where α is intercept and β_p are coefficients or slopes.

In this study, NO_2 , SO_2 , and $PM_{2.5}$ are the dependent variables, and meteorological variables (SR , WS , VWS , WD) are the independent variables. The objective of the MLR model is to use a robust fitting approach to estimate the constants $\beta_1, \beta_2, \dots, \beta_p$. Six MLR models ($MLR_1, MLR_2, MLR_3, MLR_4, MLR_5, MLR_6$) are developed using a robust fitting method. Six MLR models developed based on the same database from Fig. S1 to Fig. S7 (Supplementary material) are given in Eq. 6 to Eq. 11.

$$MLR_1 : SO_2 = 1.26417 - 0.33769SR + 0.1168VWS + 0.086997WS + 0.11845WD \quad (7)$$

$$MLR_2 : SO_2 = 1.27283 - 0.3256SR + 0.14414VWS + 0.12073WD \quad (8)$$

$$MLR_3 : NO_2 = 1.72932 - 0.33202SR - 0.18315VWS - 1.2516WS - 0.0029443WD \quad (9)$$

$$MLR_4 : NO_2 = 1.49988 - 0.42699SR + 0.083843VWS + 0.067065WD \quad (10)$$

$$MLR_5 : PM_{2.5} = 1.45092 - 0.9802SR + 0.76867VWS - 0.043617WS + 0.2275WD \quad (11)$$

$$MLR_6 : PM_{2.5} = 1.51728 - 0.97509SR - 0.33907WS + 0.22266WD \quad (12)$$

where SR is solar radiation, VWS is vertical wind speed, WS is wind speed, WD is wind direction.

Table 7 illustrates the error analysis of these models. It is found that relevant input selection by the RA increases the value of R^2 for prediction of SO_2 , NO_2 , and $PM_{2.5}$. The comparison between a measured value of daily NO_2 , SO_2 , and $PM_{2.5}$ and the MLR model's predicted values are represented in Fig. 9. It

Table 7 Error analysis of MLR models

Models	RMSE	MAE	MBE	R^2	Remark
MLR_1	0.2666	0.2207	-0.1140	0.0624	Very weak fit with low R^2 ; high bias (under prediction); limited reliability
MLR_2	0.2717	0.2240	-0.1267	0.0666	Similar to MLR_1 weak performance and negative bias; not suitable
MLR_3	0.1728	0.1427	-0.0206	-0.0684	Moderate errors but negative R^2 suggests predictions worse than the mean
MLR_4	0.2610	0.2038	0.2007	0.0103	High bias (overestimation), near-zero R^2 ; fails to explain data variance
MLR_5	0.1303	0.1008	0.0759	-2.5487	Despite low errors, extremely negative R^2 suggests the model is misleading
MLR_6	0.1330	0.1072	0.0864	-2.2329	Similar to MLR_5 good error metrics overshadowed by very poor R^2

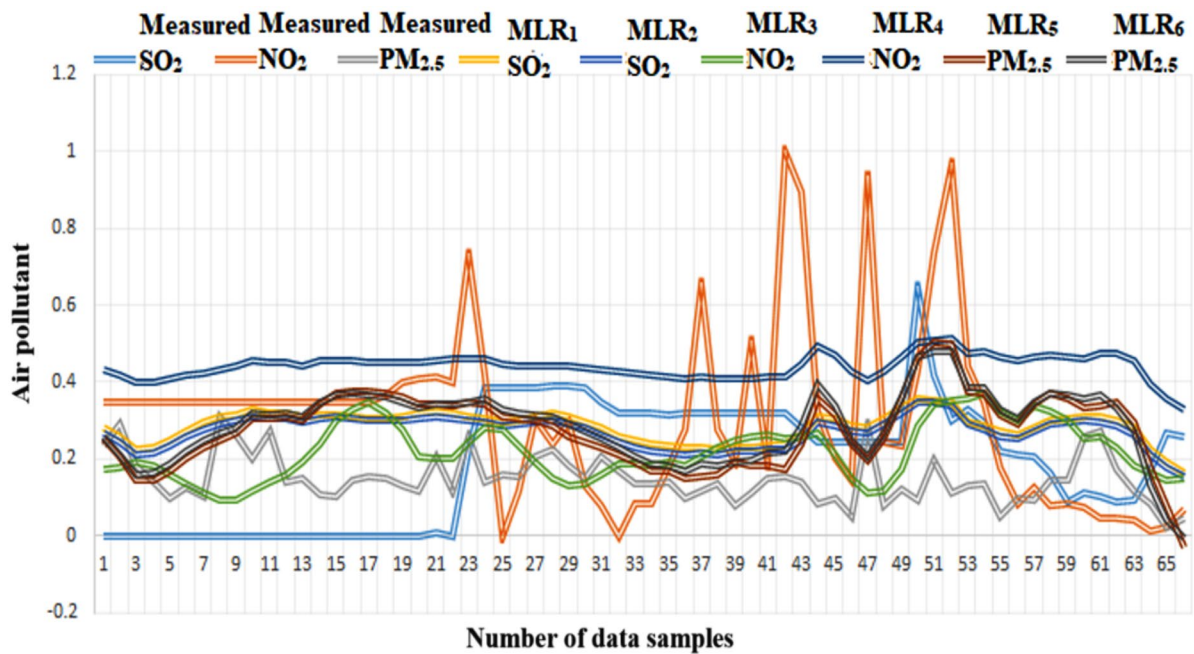


Fig. 9 Comparison between MLR models for AP prediction

Table 8 Comparison of results with previous study

Study	Error
Present	R = 89.61% and RMSE = 0.0825
(Chitranshi et al., 2015)	R < 90%
(Ahmad & Ahmad, 2023)	R = 94%
(Hossain, 2014)	RMSE ranges from 1.89 to 3.39
(He et al., 2016)	R range from 71 to 83%
(Prasad et al., 2016)	RMSE range from 0.12 to 0.48
(Cakir & Sita, 2020)	R = 84% and RMSE = 5.90
(Shams et al., 2021)	R range from 54 to 87%
(Guo et al., 2023)	R = 97.83%
(Elangasinghe et al., 2014)	R = 89%
(Gulati et al., 2023)	RMSE = 9.5223 and R = 82%
(Mohammadi et al., 2024)	R varies from 86.1% to 90.1%
(Masood et al., 2023)	RMSE = 30.325
(Zaini et al., 2022)	RMSE = 7.4250

is found that the value after prediction is much closer to the measured one.

Table 8 presents the performance comparison of the present study's MLR model with previous studies (Ahmad & Ahmad, 2023; Cakir & Sita, 2020; Chitranshi et al., 2015; Elangasinghe et al., 2014; Gulati et al., 2023; Guo et al., 2023; He et al., 2016;

Hossain, 2014; Masood et al., 2023; Mohammadi et al., 2024; Prasad et al., 2016; Shams et al., 2021; Zaini et al., 2022). This shows the superior predictive capability of the MLR model. The research attains a high R value of 89.61% and a low RMSE of 0.0825, surpassing several previous studies. (Cakir & Sita, 2020) introduced ANN models to predict NO₂ concentration levels in Nicosia, a city on the island of Cyprus. The authors conclude that the proposed ANN models predict the NO₂ concentration with an R-value of 84% and an RMSE of 5.90. (Shams et al., 2021) conducted a study comparing the performance of MLP and regression models in predicting SO₂ concentration in Tehran, Iran. The results showed MLP performed significantly better, with an R-value of 90% and an RMSE of 0.42. The study by (Guo et al., 2023) used ANN models to predict hourly PM_{2.5} concentration in Chongqing City, China. After comparing thirteen training functions, the authors concluded the best function for predicting PM_{2.5} concentrations was the one with an R-value of 97.83% with an RMSE of 1.2271. (Elangasinghe et al., 2014) used five machine learning models, including ANN, to predict PM_{2.5} concentrations in Bengaluru City, India, with the ANN model showing the highest R-value of 89%. (Gulati et al., 2023) reported an

RMSE of 9.5223 and an R-value of 82%, indicating that the model exhibited satisfactory performance yet allowed for enhancement. (Mohammadi et al., 2024) documented R values between 86.1% and 90.1%, which closely resemble the findings of our investigation. Conversely, (Masood et al., 2023) reported a significantly elevated RMSE of 30.325, signifying inadequate prediction ability, whereas (Zaini et al., 2022) encountered a substantial mistake with an RMSE of 7.4250. In contrast, the current study demonstrates superior accuracy and reliability in forecasting air pollution.

3.4 Policy Implications and Recommendations

RA-ANN model provides early warnings of pollutants (NO_2 , SO_2 , and $\text{PM}_{2.5}$) to assist policymakers and school authorities to adjust outdoor activities for ensuring children's safety. Predicted AP level provides public school closure decisions on high risky days. Schools can implement air filtration systems in classrooms by identifying pollutant trends predicted by models. It provides prior knowledge regarding restricting industry activity near schools. Policymakers can identify high risk of AP with respect to time and mandate strict emission controls in concerned areas. Model predicted AP value can be integrated into the National Clean Air Programme (NCAP) policies for improving air quality in urban areas. Models will be supportive for environmental litigation and compliance enforcement to take decisions on industry for violating AP norms. Government and schools can use this model to educate parents and students to take preventive measures such as wearing masks to reduce outdoor exposure. Deployment of model in mobile application can inform school persons about risk of AP. RA-ANN based AP models optimize computational efficiency making it a cost-effective alternative to expensive sensors. The model can be scaled for the benefit of school with limited resources.

This study highlights significant ramifications for legislators seeking to improve school air quality monitoring and mitigation initiatives, particularly in urban and high-pollution areas. Schools are delicate locations where kids spend extended periods of time exposed to contaminants; therefore, special air quality regulations are needed to safeguard their health and education. The following are the main policy suggestions derived from the work:

3.4.1 Targeted Surveillance Systems Employing Predictive Models

The study illustrates the effectiveness of the RA in enhancing air quality forecast by precise input selection, with wind direction, vertical wind speed, and sun radiation identified as essential characteristics. Policymakers must emphasize the establishment of specialized air quality monitoring systems at educational institutions that incorporate real-time meteorological data. Employing models such as MLFFNN and MLR with optimized input selection can improve predictive accuracy. This method can assist policymakers in developing adaptive air monitoring systems customized to the specific meteorological circumstances of each school.

3.4.2 Protocols for Interventions Based on Data

The research indicates that prediction accuracy enhances with suitable variable selection, especially for $\text{PM}_{2.5}$, NO_2 , and SO_2 -air pollutants recognized for their impact on respiratory health and cognitive function in children. Educational institutions necessitate data-informed intervention strategies grounded in short-term (daily) air quality predictions. For example, when forecasts suggest elevated pollution levels, schools may adjust outside activities, introduce temporary interior air filtration strategies, or stagger activity schedules.

3.4.3 Economic Surveillance for Schools with Low Resources

One major barrier is the cost of installing and maintaining sophisticated air quality sensors, particularly for educational institutions located in underprivileged areas. There is an opportunity to democratize access to air quality management with data-driven, economical monitoring systems, as evidenced by the effectiveness of more basic models like MLR combined with meteorological inputs.

3.4.4 Including Predictive Monitoring in the Requirements for Air Quality

Most current air quality recommendations prioritize threshold-based assessments over predictive monitoring. Policymakers can handle air pollution exposure

in a proactive rather than reactive manner by incorporating prediction models into air quality rules for educational facilities. This research highlights the value of predictive measures particularly the use of wind speed, direction, and solar radiation to predict pollution surges. Therefore, the policymakers should modify national air quality legislation to include predictive monitoring standards for schools allowing for preventative measures before pollution limits are exceeded.

3.4.5 Public Awareness and Education Campaigns

The effectiveness of air quality monitoring in educational institutions depends on educating stakeholders about the importance of monitoring and mitigation strategies in addition to technical deployment. This means raising community knowledge of air quality risks and solutions as well as teaching school staff how to evaluate predictive data. So, the Governments should put in place school-based educational programs that teach about air quality and the health effects it has including instruction on how to use and react to real-time monitoring equipment and predictive analytics.

3.4.6 Coordination of Policies for Long-Term Air Quality

Collaboration between a variety of stakeholders, such as local authorities, educational boards, and environmental agencies, is required for the development of predictive models. Predictive analytics can be used to inform comprehensive public health decisions, and collaborative frameworks can help with the installation, upkeep, and real-time operation of monitoring systems. Therefore, to improve sustainable air quality monitoring and mitigation programs at educational institutions, develop multi-stakeholder policy frameworks that involve local governments, environmental organizations, and school administrators.

Based on discussions following policy are recommended:

- (i) Establish real-time predictive air quality systems in public schools utilizing advanced data-driven algorithms to deliver early alerts and mitigation strategies during instances of heightened pollution.
- (ii) Affordable options include for monitoring pollutant using PM_{2.5}/PM₁₀ sensors (PMS5003, \$15–\$25), SO₂ sensors (SGX PS1 SO₂–50, ₹6,680.96/~\$80–90), NO₂ sensors (Fermion MEMS, ₹825/~\$10–15), temperature and humidity sensors (DHT11/DHT22, \$5–\$10), and light intensity sensors (LDRs, <\$5). Formulate and execute adaptive intervention strategies based on daily air quality forecasts to reduce student exposure during high-risk intervals.
- (iii) Encourage educational institutions to install reasonably priced air quality monitors, backed by simplified forecasting algorithms that make use of key meteorological data.

4 Conclusions

The present study plays a significant role in applying the RA, which identifies and selects the most influencing variable. Different models, such as Multilayer Feedforward Neural Networks (MLFFNN), Radial Basis Function Neural Networks (RBFNN), Generalized Regression Neural Networks (GRNN), and Multiple Linear Regression (MLR), have been developed to predict daily NO₂, SO₂, and PM_{2.5}. The RA identifies preferential order as wind direction, vertical wind speed, and solar radiation as relevant inputs for prediction of daily NO₂; wind direction, solar radiation, and vertical wind speed as relevant inputs for prediction of daily SO₂; and wind direction, wind speed and solar radiation as relevant inputs to predict daily PM_{2.5}. MLFFNN, RBFNN, GRNN, and MLR models are developed with and without relevant input variables to check authentication. It is found that relevant input variable selection has a more prominent effect on MLFFNN and MLR models. This means that the coefficient of determination increases, but relevant input selection decreases the coefficient of determination value for RBFNN. It is concluded that the RA is more efficient and predicts better than MLR. Policy implications and recommendations are useful for industry and government. The limitations are this study is based on data from selected public school, which is limited data and lack generalizability to other

geographical or institutional contexts. RA presents model-specific effectiveness. This research can be explored in the future incorporating multi-year and multi-school data by applying empirical mode decomposition in artificial neural network models for the daily prediction of air pollutants.

Acknowledgements The authors would like to acknowledge central pollution control board New Delhi India for providing database for this study.

Author Contributions V. Y.: Conceptualization, methodology, data analysis, writing-original draft preparation, and review.

A. K. Y.: Conceptualization, Data collection, analysis, writing-original draft preparation, and manuscript review.

R.K.: Literature review, data validation, and manuscript editing.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Data Availability The data used in this study is publicly available and can be accessed from the Central Pollution Control Board (CPCB), New Delhi, India. The data can be downloaded from <https://airquality.cpcb.gov.in/ccr/#/caaqm-dashbord-all/caaqm-landing/caaqm-data-availability>

Declarations

Ethics Approval Not applicable.

Consent for Publication Not applicable.

Competing Interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ahmad, S., & Ahmad, T. (2023). AQI prediction using layer recurrent neural network model: A new approach. *Environmental Monitoring and Assessment*, 195, 1180.
- Akhtar, M., Majeed, H., Iftikhar, T., & Ahmad, K. (2025). Climate friendly MOFs synthesis for drug delivery systems by integrating AI, intelligent manufacturing, and quantum solutions in industry 6.0 sustainable approach. *Toxicology Research*, 14, tfaf011.
- Arauzo-Azofra, A., Benitez, J. M., & Castro, J. L. (2004). A Feature Set Measure Based on Relief: RACS, 2004, 104–109.
- Cakir, S., & Sita, M. (2020). Evaluating the performance of ANN in predicting the concentrations of ambient air pollutants in Nicosia. *Atmospheric Pollution Research*, 11, 2327–2334.
- Chaturvedi, P. (2024). Air quality prediction system using machine learning models. *Water, Air, and Soil Pollution*, 235, 578.
- Chen, Y., Chen, X., Xu, A., Sun, Q., & Peng, X. (2022). A hybrid CNN-transformer model for ozone concentration prediction. *Air Quality, Atmosphere & Health*, 15, 1533–1546.
- Chitranshi, S., Sharma, S. P., & Dey, S. (2015). Spatio-temporal variations in the estimation of PM10 from MODIS-derived aerosol optical depth for the urban areas in the Central Indo-Gangetic Plain. *Meteorology and Atmospheric Physics*, 127, 107–121.
- Dawar, I., Singal, M., Singh, V., Lamba, S., & Jain, S. (2024). Air quality prediction using machine learning models: A predictive study in the Himalayan City of Rishikesh. *SN Computer Science*, 5, 1025.
- Ebrahimi, A. A., Bazar, M., & Zakeri, H. R. (2024). Investigating the impact of urban-environmental factors on air pollutants: A land use regression model approach and health risk assessment. *Environment Geochemical Health*, 46, 313.
- Elangasinghe, M. A., Singhal, N., Dirks, K. N., & Salmond, J. A. (2014). Development of an ANN-based air pollution forecasting system with explicit knowledge through sensitivity analysis. *Atmospheric Pollution Research*, 5, 696–708.
- Gulati, S., Bansal, A., Pal, A., Mittal, N., Sharma, A., & Gared, F. (2023). Estimating PM_{2.5} utilizing multiple linear regression and ANN techniques. *Scientific Reports*, 13, Article 22578.
- Guo, Q., He, Z., & Wang, Z. (2023). Prediction of hourly PM_{2.5} and PM₁₀ concentrations in Chongqing City in China based on artificial neural network. *Aerosol and Air Quality Research*, 23, Article 220448.
- Hassan, M. A., Mehmood, T., Liu, J., Luo, X., Li, X., Tanveer, M., Faheem, M., Shakoar, A., Dar, A. A., & Abid, M. (2023). A review of particulate pollution over Himalaya region: Characteristics and salient factors contributing ambient PM pollution. *Atmospheric Environment*, 294, Article 119472.

- He, J., Yu, Y., Xie, Y., Mao, H., Wu, L., Liu, N., & Zhao, S. (2016). Numerical model-based artificial neural network model and its application for quantifying impact factors of urban air quality. *Water, Air, and Soil Pollution*, 227, 1–10.
- Hossain, A. M. K. (2014). Predictive Ability Improved Neural Network Models to Stimulate Pollutant Dispersion. *International Journal of Atmospheric Science*, 141923, 1–12.
- <https://www.who.int/health-topics/air-pollution>
- Iskandaryan, D., Ramos, F., & Trilles, S. (2023). Graph neural network for air quality prediction: A case study in Madrid. *IEEE Access*, 11, 2729–2742.
- Karimi, S., Asghari, M., Rabie, R., & Niri, M. E. (2023a). Machine learning-based white-box prediction and correlation analysis of air pollutants in proximity to industrial zones. *Process Safety and Environmental Protection*, 178, 1009–1025.
- Karimi, F., Amanollahi, J., Reisi, M., & Darand, M. (2023b). Prediction of air quality using vertical atmospheric condition and developing hybrid models. *Advances in Space Research*, 72, 1172–1182.
- Kira K, Rendell, L. A. (1994). A practical approach to feature selection. *Proceedings of the Ninth International Conference on machine learning* (pp. 249–256). <https://sci2s.ugr.es/keel/pdf/algorithm/congreso/kira1992.pdf>
- Kononenko, I. (1994). Estimating attributes: Analysis and extensions of relief. *ECML-94: Proceedings of the European conference on machine learning on machine learning* (pp. 171–182).
- Liang, Y., Xia, Y., Ke, S., Wang, Y., Wen, Q., Zhang, J., & Zimmermann, R. (2023). Airformer: Predicting nationwide air quality in China with transformers. *The Thirty-Seventh AAAI Conference on Artificial Intelligence*, 37(12), 14329–14337.
- Lin, X., Fu, Y., Peng, D. Z., Liu, C., Chu, M., Chen, Z., Yang, F., Tse, T. K. T., Li, C. Y., & Feng, X. (2024). CFD- and BPNN- based investigation and prediction of air pollutant dispersion in urban environment. *Sustainable Cities and Society*, 100, Article 105029.
- Lin, Y. C., Lin, Y. T., Chen, C. R., & Lai, C. Y. (2025). Meteorological and traffic effects on air pollutants using Bayesian networks and deep learning. *Journal of Environmental Sciences*, 152, 54–70.
- Liponhay, M. P., Valerio, A. V., & Monterola, C. P. (2024). Time-delayed causal network analysis of meteorological variables and air pollutants in Baguio city. *Atmospheric Pollution Research*, 15, Article 102095.
- Luo, J., & Gong, Y. (2023). Air pollutant prediction based on ARIMA-WOA-LSTM model. *Atmospheric Pollution Research*, 14, Article 101761.
- Masood, A., Hameed, M. M., Srivastava, A., Pham, Q. B., Ahmad, K., Razali, S. F. M., & Baowidan, S. A. (2023). Improving PM2.5 prediction in New Delhi using a hybrid extreme learning machine coupled with snake optimization algorithm. *Scientific Reports*, 13, 21057.
- Mohammadi, F., Teiri, H., Hajizadeh, Y., Abdollahnejad, A., & Ebrahimi, A. (2024). Prediction of atmospheric PM2.5 level by machine learning techniques in Isfahan, Iran. *Scientific Reports*, 14, Article 2109.
- Monteiro, T. O., BarradasFilho, A. O., Villa-Vélez, H. A., & Cruz, G. (2024). Estimation of the main air pollutants from different biomasses under combustion atmospheres by artificial neural networks. *Chemosphere*, 352, 141484.
- Nandan, A., Mondal, P., Kumar, S., Siddiqui, N. A., Sinha, S., Subramani, S., Singh, A. K., Raja, S., & Hussain, C. M. (2023). Investigation of indoor air pollutants in different environmental settings and their health impact: A case study of Dehradun, India. *Air Quality, Atmosphere & Health*, 16, 2377–2400.
- Nizhelskiy, M., Kazeev, K., Vilkova, V., Fedorenko, A., & Kolesnikov, S. (2025). Impact of gaseous smoke pollutants from modelled fires on air and soil quality. *Carbon Research*, 4, 17.
- Pande, C. B., Kushwaha, N. L., Alawi, O. A., Sammen, S. S., Sidek, L. M., Yaseen, Z. M., Pal, S. C., & Katipoğlu, O. M. (2024). Daily scale air quality index forecasting using bidirectional recurrent neural networks: Case study of Delhi, India. *Environmental Pollution*, 351, Article 124040.
- Petrić, V., Hussain, H., Časni, K., Vuckovic, M., Schopper, A., Andrijić, Ž., Kecorius, S., Madueno, L., Kern, R., & Lovrić, M. (2024). Ensemble machine learning, deep learning, and time series forecasting: Improving prediction accuracy for hourly concentrations of ambient air pollutants. *Aerosol and Air Quality Research*, 24, Article 230317.
- 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.
- Qazi, A., Fayaz, H., Wadi, A., Raj, R. G., Rahim, N. A., & Khan, W. A. (2015). The artificial neural network for solar radiation prediction and designing solar systems: A systematic literature review. *Journal of Cleaner Production*, 104, 1–12.
- Rani, P. A., & Sampathkumar, V. (2023). A novel artificial intelligence algorithm for predicting air quality by analysing the pollutant levels in air quality data in Tamilnadu e-Prime - Advances in Electrical Engineering. *Electronics and Energy*, 5, 100234.
- Ravindiran, G., Rajamanickam, S., Kanagarathinam, K., Hayder, G., Janardhan, G., Arunkumar, P., Arunachalam, S., AlObaid, A. A., Warad, I., & Muniasamy, S. K. (2023). Impact of air pollutants on climate change and prediction of air quality index using machine learning models. *Environmental Research*, 239, Article 117354.
- Ren, Z., & Ji, X. (2023). On prediction of air pollutants with Takagi-Sugeno models based on a hierarchical clustering identification method. *Atmospheric Pollution Research*, 14, Article 101731.
- Shams, S. R., Jahani, A., Kalantary, S., Moeinaddini, M., & Khorasani, N. (2021). The evaluation on artificial neural networks (ANN) and multiple linear regressions (MLR) models for predicting SO₂ concentration. *Urban Climate*, 37, Article 100837.
- Singh, S., & Suthar, G. (2025). Machine learning and deep learning approaches for PM2.5 prediction: A study on urban air quality in Jaipur, India. *Earth Science Informatics*, 18, Article 97.
- Tas, I., Adiguzel, F., Cetin, M., & Cetin, I. Z. (2025). *Evaluation of Soil Pollution in Children's Playground in Nevsehir, Türkiye, using GIS and Statistical Methods*. Cambridge Scholars Publishing.

- Wang, S., Ren, Y., Xia, B., Liu, K., & Li, H. (2023). Prediction of atmospheric pollutants in urban environment based on coupled deep learning model and sensitivity analysis. *Chemosphere*, 331, Article 138830.
- Wu, Z., Tian, Y., Li, M., Wang, B., Quan, Y., & Liu, J. (2024). Prediction of air pollutant concentrations based on the long short-term memory neural network. *Journal of Hazardous Materials*, 465, 1–15.
- Yadav, V., Yadav, A. K., Singh, V., & Singh, T. (2024). Artificial neural network an innovative approach in air pollutant prediction for environmental applications: A review. *Results in Engineering*, 22, Article 102305.
- Yadav, A. K., Kumar, R., Wang, M., Fekete, G., & Singh, T. (2025). Comparative analysis of daily global solar radiation prediction using deep learning models inputted with stochastic variables. *Scientific Reports*, 15, 10786.
- Yang, F., & Huang, G. (2024). An optimized decomposition integration model for deterministic and probabilistic air pollutant concentration prediction considering influencing factors. *Atmospheric Pollution Research*, 15, Article 102144.
- Yu, M., Masrur, A., & Blaszcak-Boxe, C. (2023). Predicting hourly PM_{2.5} concentrations in wildfire-prone areas using a spatio temporal transformer model. *Science of the Total Environment*, 860, Article 160446.
- Zaini, N., Ean, L. W., Ahmed, A. N., Malek, M. A., & Chow, M. F. (2022). PM2.5 forecasting for an urban area based on deep learning and decomposition method. *Scientific Reports*, 12, 17565.
- Zhang, B., Wang, Z., Lu, Y., Li, M., Yang, R., Pan, J., & Kou, Z. (2023a). Air pollutant diffusion trend prediction based on deep learning for targeted season-North China as an example. *Expert Systems with Applications*, 232, Article 120718.
- Zhang, B., Liu, Y., Yong, R., Zou, G., Yang, R., Pan, J., & Li, M. (2023b). A spatial correlation prediction model of urban PM2.5 concentration based on deconvolution and LSTM. *Neurocomputing*, 544, Article 126280.

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