



A novel multi-model data-driven ensemble approach for the prediction of particulate matter concentration

Ibrahim Khalil Umar² · Vahid Nourani^{1,2} · Hüseyin Gökçekuş²

Received: 23 December 2020 / Accepted: 22 April 2021 / Published online: 3 May 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Accuracy in the prediction of the particulate matter (PM_{2.5} and PM₁₀) concentration in the atmosphere is essential for both its monitoring and control. In this study, a novel neuro fuzzy ensemble (NF-E) model was proposed for prediction of hourly PM_{2.5} and PM₁₀ concentration. The NF-E involves careful selection of relevant input parameters for base modelling and using an adaptive neuro fuzzy inference system (ANFIS) model as a nonlinear kernel for obtaining ensemble output. The four base models used include ANFIS, artificial neural network (ANN), support vector regression (SVR) and multilinear regression (MLR). The dominant input parameters for developing the base models were selected using two nonlinear approaches (mutual information and single-input single-output ANN-based sensitivity analysis) and a conventional Pearson correlation coefficient. The NF-E model was found to predict both PM_{2.5} and PM₁₀ with higher generalization ability and least error. The NF-E model outperformed all the single base models and other linear ensemble techniques with a Nash-Sutcliffe efficiency (NSE) of 0.9594 and 0.9865, mean absolute error (MAE) of 1.63 µg/m³ and 1.66 µg/m³ and BIAS of 0.0760 and 0.0340 in the testing stage for PM_{2.5} and PM₁₀, respectively. The NF-E could improve the efficiency of other models by 4–22% for PM_{2.5} and 3–20% for PM₁₀ depending on the model.

Keywords Ensemble modelling · Air quality · Sensitivity analysis · ANFIS

Introduction

Air pollution is one of the major environmental challenges affecting the health condition of many people living in the

urban areas due to increased industrial activities and urbanization. About 91% of the world population are believed to be exposed to polluted air causing premature death of almost 4.2 million people annually (WHO (World Health Organization)

Highlights

- Accuracy in the prediction of PM_{2.5} and PM₁₀ concentration is essential for both its monitoring and control.
- Performance of 4 data-driven models for prediction of PM_{2.5} and PM₁₀ was evaluated and compared.
- For an enhanced modelling performance, a neuro fuzzy ensemble approach was developed.
- The ensemble model could improve the efficiency of other models by 4–22% for PM_{2.5} and 3–20% for PM₁₀ depending on the model.
- Nonlinear dependencies between the potential inputs and the particulate matter were used for dominant input selection.

Responsible Editor: Marcus Schulz

✉ Ibrahim Khalil Umar
ikdotumar@gmail.com

Vahid Nourani
vnourani@yahoo.com; nourani@tabrizu.ac.ir

Hüseyin Gökçekuş
huseyin.gokcekus@neu.edu.tr

¹ Center of Excellence in Hydroinformatics, Faculty of Civil Engineering, University of Tabriz, Tabriz, Iran

² Faculty of Civil and Environmental Engineering, Near East University, Near East Boulevard, Via Mersin, 99138 Nicosia, North Cyprus, Turkey

2018). Particulate matters (PM_{2.5} and PM₁₀), ozone (O₃), nitrogen dioxide (NO₂), carbon monoxide (CO) and sulphur dioxide (SO₂) were identified as the most hazardous ambient air pollutants (Uzoigwe et al. 2013). PM_{2.5} acts as the major indicator for the air quality monitor system (Van Donkelaar et al. 2006). These toxic substances can be breathed into the lungs and distribute throughout the body as blood circulates due to their extremely small volume. Additionally, the increase of PM_{2.5} and PM₁₀ concentration can lead to declining visibility, contributing to adverse impacts on the transportation industry (Sun and Li 2020). These problems can effectively be reduced by careful application of a good urban air quality management (UAQM). The fundamental elements of the UAQM involve clear description of objectives and standards, well-designed monitoring system, reliable air quality modelling, emission inventory, source apportionment, health exposure assessment, control strategies and public participation (Gulia et al. 2020). A reliable air quality model can provide required information for analysis and management of the air quality parameters which will help stakeholders in decision making regarding issues related to UAQM budget and selection of potential mitigation measures required to reduce the pollution crisis and public health (Suleiman et al. 2019). The factors influencing the concentrations of the air pollutants can be classified into traffic-related factors, background concentration, meteorological and geographical factors (Cai et al. 2009).

Various mathematical models for advection and reactions of the air pollutants were proposed for forecasting the time-varying concentration of air pollutants in urban areas, e.g., steady-state Gaussian plume models. However, the diversity and complexity of the processes (physical and chemical) involved in both formation and transportation of the air pollutants in the urban areas make the application of these models very challenging or impossible in some situation. This is because large database and good understanding of the formation processes are required for application of the empirical methods, and in some cases the data is not available or insufficient (Arhami et al. 2013).

Motivated with the efficiency of artificial intelligence (AI)-based models in the prediction of complex processes in the fields of engineering, several AI-based models were developed for the prediction of air quality parameters. For instance, Arhami et al. (2013) developed an ANN model for the prediction of hourly criteria pollutants (NO_x, NO₂, NO, O₃ and PM₁₀) in an urban environment using wind direction, wind speed, relative humidity and air temperature as input variables. Suleiman et al. (2016) applied both ANN and boosted regression trees (BRT) to predict the concentration of PM_{2.5}, PM₁₀ and particle number count (PNC) at Marylebone road in London. The BRT model demonstrated higher efficiency over the ANN model. Azeez et al. (2019) integrated GIS into a hybrid model combining ANN and the

correlation-based feature selection (CFS) algorithm for prediction of vehicular CO emissions. For comparison, Mehdipour et al. (2018) applied three different AI methods namely Bayesian network (BN) and decision tree (DT) support vector machines (SVM) for prediction of PM in Tehran. The model input parameters were temperature, precipitation, wind speed, nebulosity, relative humidity, sunshine, O₃, PM₁₀, SO₂, NO₂ and CO. The SVM has demonstrated higher prediction capability than both BN and DT. Krishan et al. (2019) used meteorological data, transport emissions, traffic data and air quality parameters to model hourly concentration of air quality indicators in Delhi, India using the long short-term memory (LSTM) approach. The AI models have demonstrated a high accuracy in the prediction of the air quality parameters (Cai et al. 2009). This is because the AI models are capable of handling multivariate inputs, nonlinearity and uncertainty of complex processes without requiring prior assumptions between the input parameters.

Although the mentioned AI models (ANN, SVM, ANFIS, etc.) provide higher prediction capability than both empirical and conventional multilinear regression (MLR) models, it is known that different models may lead to different outcomes for a particular problem depending on the conditions. Therefore, combining the outputs of the different models through an ensemble approach will provide outputs with lesser error variance compared to the single models (Nourani et al. 2019). The ensemble approach combines the unique features of the constituent models to come out with a better pattern of the presented database (Sharghi et al. 2018). The objective of this study is to present and also apply a novel neuro fuzzy ensemble (NS-E) technique for improved performance in the prediction of PM_{2.5} and PM₁₀. The objectives could be achieved in three steps. First, selection of the dominant input parameters relevant in the prediction of the PM_{2.5} and PM₁₀. Secondly, development of 4-single black box models (ANN, ANFIS, SVM and MLR). Finally, the NS-E models and two linear ensemble models were developed by combining the predicted outputs from the 4 different black box models developed in stage 2. This study presents the first application of the novel NS-E technique for the prediction of the PM_{2.5} and PM₁₀ to the best of the authors' knowledge. The selection of PM_{2.5} and PM₁₀ for conducting the study was based on the strong adverse effect they have on human health as reported by Uzoigwe et al. (2013) and their major role in defining air quality (Sun and Li 2020).

Materials and methods

Data

For purposes of conducting the study, hourly data from the air quality monitoring site along the Marylebone road in central

London was obtained from January 1, 2007 to December 31, 2007. Marylebone was selected for conducting the study due to its high average daily traffic of about 75,000 veh/day (Jones and Harrison 2005) since 64% of the PM was reported to come from the vehicular traffic (European Environment Agency 2012). The monitoring station was located at approximately 1.5 m from the road (southern side of road). Simultaneously with air pollutants (O₃, NO, NO₂, NO_x, CO, SO₂, PM₁₀, PM_{2.5}), traffic data (volumes of buses, cars and taxis, motorcycles, light commercial vehicles, pedal cycles and heavy goods vehicles) and the speed and the meteorological data (wind speed, wind direction and temperature) were recorded at the monitoring site (Jones and Harrison 2005). The data was made available for download at the UK air quality data archive (<https://uk-air.defra.gov.uk/data/maryleboneroad>). The traffic data was collected using high accuracy induction tubes for classification and counting buried on each lane. Two tapered element oscillating microbalance (1400AB model) each equipped with a different sampling head were used for monitoring the concentration of the PM_{2.5} and PM₁₀ at the sampling location (Jones and Harrison 2005). The descriptive statistics measured data was presented in Table 1.

Data preparation and performance evaluation

To ensure all input variables receive equal attention in black box models, the data are usually normalized to the same range usually between zero and unity. The normalization makes the data dimensionless during training and prevents overshadowing of the parameters in the lower numeric range by those in the higher numeric range. It also helps in reducing the computational difficulties of the model. In this study, the data are normalized between 0 and 1 using (Nourani et al. 2012):

$$P_i = \frac{P - P_{\min}}{P_{\max} - P_{\min}} \tag{1}$$

where, P_i is the normalized value, P is the measured value, and P_{\max} and P_{\min} are the maximum and minimum measured concentrations, respectively.

Three statistical performance measures were used for evaluating the performance and efficiency of the models developed for predicting the PM_{2.5} and PM₁₀. The statistical performance measures are Nash-Sutcliffe efficiency (NSE) which measures the model’s goodness of fit, mean absolute error (MAE) which evaluates the absolute mean error of the models and bias (BIAS) which reflects how much the computed value deviates from the observed value. Legates and McCabe Jr (1999) suggested that one absolute error measure and one goodness of fit measures can sufficiently evaluate the performance of prediction models. The performance criteria were computed using Eqs. 2–4, respectively (Nourani and Fard Sayyah 2012). The model’s accuracy can be interpreted based on the NSE values as very good ($0.75 < NSE \leq 1$), good ($0.65 < NSE \leq 0.75$), satisfactory ($0.50 \leq NSE \leq 0.65$) and unsatisfactory ($NSE < 0.50$) (Moriassi et al. 2007). The closer MAE and BIAS values approach 0, the better the model’s prediction.

$$NSE = 1 - \frac{\sum_{i=1}^n (P_{obs_i} - P_{pre_i})^2}{\sum_{i=1}^n (P_{obs_i} - \bar{P}_{obs_i})^2} \quad -\infty < DC \leq 1.0 \tag{2}$$

$$MAE = \frac{\sum_{i=1}^n |P_{obs_i} - P_{pre_i}|}{n} \tag{3}$$

$$BIAS = \frac{\sum_{i=1}^n (P_{obs_i} - P_{pre_i})}{\sum_{i=1}^n (P_{pre_i})} \tag{4}$$

where n represents the number of observations, \bar{P}_{obs} is the mean value of the observed value, P_{obs} is the observed value and P_{pre} is the predicted value.

Table 1 Descriptive statistics of the data

	S	CLS1	CLS2	CL3	CLS4	CLS5	CLS6	WD	WS	T	O ₃	NO	NO ₂	NO _x	SO ₂	CO	PM ₁₀	PM _{2.5}
Mean	42.98	2.57	186.08	4.07	5.02	9.43	24.72	185.92	0.72	12.32	17.11	113.43	100.07	273.21	6.34	0.76	43.69	20.83
Standard deviation	11.07	4.77	117.95	3.07	5.64	5.63	9.13	100.27	0.60	6.25	18.11	102.82	54.00	207.34	5.55	0.46	22.04	11.43
Minimum	12.60	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.8	0.00	0.00	0.00	0.00	0.00	0.00	-3.00	0.00
Maximum	78.90	71.00	641.00	25.00	76.00	32.00	60.00	359.70	4.50	29.8	120.0	781.0	329.00	1492.0	67.0	3.00	183.00	118.0
Percent f missing data	0.00	0.00	0.00	0.00	0.00	0.00	0.00	8.68	8.68	8.63	1.42	2.00	2.00	2.00	1.32	4.10	2.19	4.34

S average speed, CLS1 volumes of buses, CLS2 volume of cars and taxis, CLS3 volume of motorcycles, CLS4 volume of light commercial vehicles, CLS5 volume of pedal cycles, CLS6 volume of heavy goods vehicles

Proposed methodology

The study was conducted in three major steps as shown in Fig. 1. In the first step, the most relevant input parameters for developing the base models were selected using a single-input single-output neural network. In the second step, ANN, ANFIS, SVR and MLR models were proposed for the estimation of the air quality parameters ($PM_{2.5}$, PM_{10}). Finally, the NS-E model and two linear ensemble models combining the outputs of the proposed base models (ANN, ANFIS, SVR, and MLR) were proposed for enhanced performance in the estimation of $PM_{2.5}$ and PM_{10} .

The notion behind developing ensemble models is for achieving the following benefits: (i) Sometimes, it is difficult to select an appropriate model for modelling a particular time series problem; by the ensemble approach, the difficulty in model selection has been removed since the nonlinear ensemble models are capable of providing a result that is even better than that of the best base model (Nourani et al. 2020a). (ii) In certain real-life processes that possess both linear and nonlinear characteristics, neither linear nor nonlinear models can do well in the prediction since errors in the linear pattern could be inherited and magnified by the nonlinear models and vice versa. By combining the outputs of linear models (MLR) and the nonlinear models (ANN, SVR, ANFIS), the linear and the nonlinear patterns in the data could be captured effectively (Nourani et al. 2019). (iii) There is no particular model

to perfectly investigate a certain process as approved by Sharghi et al. (2018). This is due to the complex nature of real-world problems whereby a unique model may not be able to identify a distinct pattern of a particular process.

Selection of relevant input parameters

The performance of all black models depends on the selection of appropriate input variables. Imposing many input parameters into the model will increase the complexity of the model, decrease the computational accuracy and increase the time required to train the model (Ahmed and Pradhan 2019). On the other hand, an insufficient number of input parameters will also result in poor performance of the model. Therefore, an optimum number of input parameters is required in developing a model with high estimation accuracy. Traditionally, the Pearson correlation matrix is used for selecting dominant input parameters, but the method had been criticized because correlations are built on linear relationships and most of real-life processes are complex and nonlinear in nature (Nourani et al. 2014). In view of that, single-input single-output nonlinear sensitivity analysis trained with feed forward neural network and a mutual information (MI) measure which computes the statistical dependency between variables based on entropy function were used in addition to the Pearson correlation matrix for determining the relevant input parameters. The single-

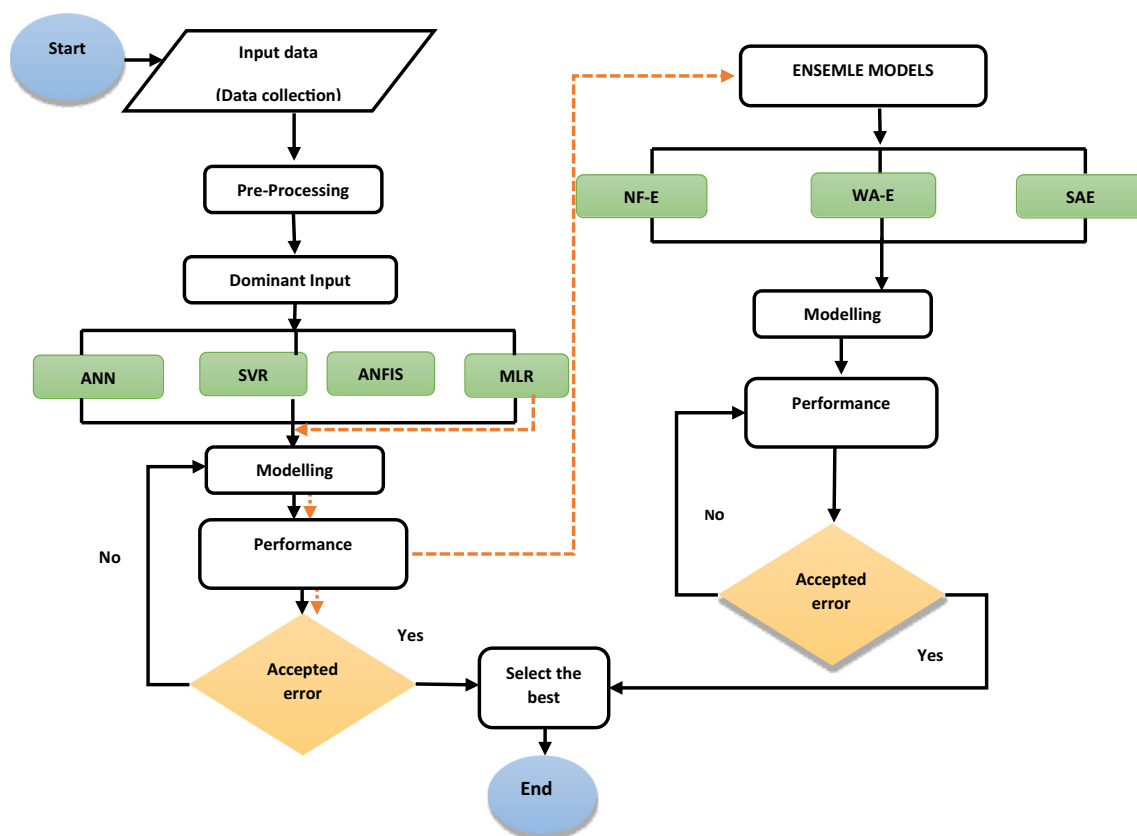


Fig. 1 Schematic of the proposed methodology for air quality parameters

input single-output nonlinear linear sensitivity analysis was evaluated using the NSE of the ANN model in the verification stage.

Black box models

FFNN

FFNN is among the most commonly used ANN models employed for capturing the nonlinearity and complex interaction between the predictor and response parameters (Jahani and Mohammadi 2019). FFNN gets the name from the manner in which information is transmitted, that is information only flows in the forward direction (Ghaffari et al. 2006). This type of ANN acquires its acceptance due to its simplicity in modelling and capturing nonlinear pattern in complex problems (Rumelhart et al. 1986). The suitability of the model to learn from experience without the need to necessarily identify the physical connection between the predictor and explained variables makes it effective and vital in modelling complex processes in many engineering fields (Kumar et al. 2014). In the FFNN, an interactive link between neurons is used to process the information and establish a relationship rather than build any complex mathematical model. The most widely used algorithm for training the FFNN is the backpropagation algorithms. To train the FFNN model, some attuned weights are initialized and multiplied by the inputs, the cumulative results then passed through the transfer function to handle the nonlinear pattern in the data before giving out the output values (Ghaffari et al. 2006). The architecture of the FFNN as shown in Fig. 2 consists of one input layer and one output layer connected by an intermediary hidden layer(s). All the nodes in any layer are only connected to the nodes of the immediately succeeding layer (Kim and Singh 2014). The general expression for the ANN model is given by

$$y_i = \sum_{j=1}^n w_{ji}x_j + b_{i0}. \tag{5}$$

SVR

The SVM learning was first proposed by Vapnik (1998) and was proved capable of providing a reasonable and acceptable solution to the prediction, classification, pattern recognition and regression problems. It is one of the data-driven machine learning approaches. The two useful functions of the SVM models that differentiate it with other machine learning approaches like the ANN are minimization of structural risk and statistical learning theory. The SVR which is one of the SVM-based models is employed for the nonlinear regression problems that consider the minimization of operational risk as its objective function rather than error minimization between

the predicted and measured values that is considered in other data-driven models like ANN models. In SVR, the data is first mapped into a linear regression which is then passed through a nonlinear kernel that captures the nonlinearity pattern in the data. For more details, the readers are referred to Wang et al. (2015) and Nourani et al. (2020b) about SVR modelling. Figure 3 gives the general structure of the SVR model. The SVR equation can be expressed as (Wang et al. 2015):

$$f(x, \alpha_i, \alpha_i^*) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) K(x, x_i) + b \tag{6}$$

where x represents the input vector, α_i and α_i^* are the Lagrange multipliers, $k(x_i, x_j)$ is the kernel function performing the nonlinear mapping into feature space and b is bias term. Gaussian radial basis function (RBF) kernel is the most commonly used kernel in the SVR and is given as

$$k(x_1, x_2) = \exp\left(-\gamma \|x_1 - x_2\|^2\right) \tag{7}$$

where, γ is the kernel parameter.

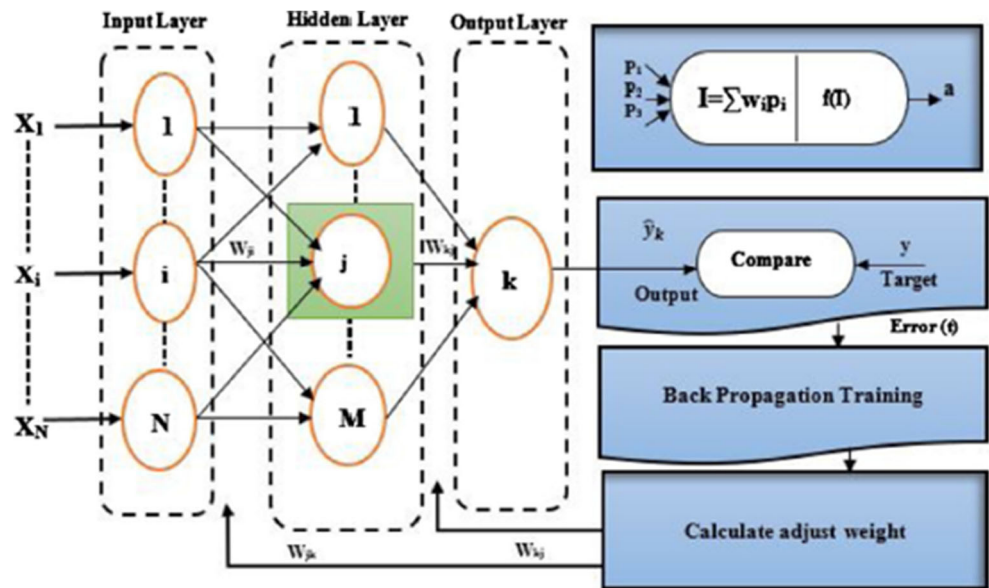
ANFIS

ANFIS is a hybrid model fused by Jang in 1993 for overcoming the limitations of both the ANN and the FIS. It combines both power of the fuzzy logic in dealing with the uncertainties and learning ability of the ANN. The ANFIS model is built on a fuzzy logic definition, and the system parameters are optimized automatically by the ANN unlike in the fuzzy system where the system parameters are manually tuned (Rai et al. 2015). The ANFIS proved to be a good useful tool for approximation problems due to its adaptive capability and flexibility in dealing with uncertainties and ability in processing huge noisy data from complex and dynamic systems (Çaydaş et al. 2009). The architecture of the ANFIS model (Fig. 4) consists of five layers constructed like a multi-layer feed forward neural network. The layers are named according to their operative function (Codur et al. 2017). ANFIS uses backpropagation algorithm for learning the parameters of membership functions and conventional least-squares estimator for estimating the parameter of the first-order polynomial of the Takagi-Sugeno fuzzy model. The overall output of the ANFIS system can be expressed as a linear combination of the consequent parameters (Çaydaş et al. 2009).

MLR

The most commonly used method for the prediction and analysis of engineering problems is the MLR. It helps understand the linear dependency between the predictor and the dependent variables. It explores the interaction between the variables and describes the relationship between them by keeping the independent variables fixed and varying one (Doğan and

Fig. 2 Structure of the three-layer FFNN (Wang et al. 2015)



Akgüngör 2013). The dependent variable y and n regressor variables can be correlated by (Elkiran et al. 2018):

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_ix_i + \xi. \quad (8)$$

In Eq. 8, x_i represents the value of the i^{th} predictor, b_i stands for the coefficient of the i^{th} predictor, b_0 is the constant of regression and ξ is the error term.

Ensemble approach

Ensemble approach is a machine learning approach used to merge the process of multiple predictors for an enhanced performance of the prediction process (Sharghi et al. 2018). The ensemble approach could either be linear or nonlinear (Raj

Kiran and Ravi 2008). In the linear ensemble approach, simple average (SA), weighted average (WA) or weighted median (WM) are used to ensemble the results obtained by individual predictor models, while in the nonlinear approach, nonlinear kernels such as ANFIS, ANN, SVR, etc. are used to obtain the nonlinear average of the results obtained by the individual base models. The input layer of the ensemble technique is fed by the outputs of the considered models, each considered as one input variable (Nourani et al. 2018). The use of the ensemble approach for prediction, clustering and classification in the several engineering fields proved to provide higher accuracy than the individual models (Shtein et al. 2019; Nourani et al. 2020a). For the nonlinear ensemble approach employed for PM_{10} and $PM_{2.5}$ prediction in this study, an ANFIS model was trained using the *gbell* function and a

Fig. 3 Conceptual architecture of SVM algorithm

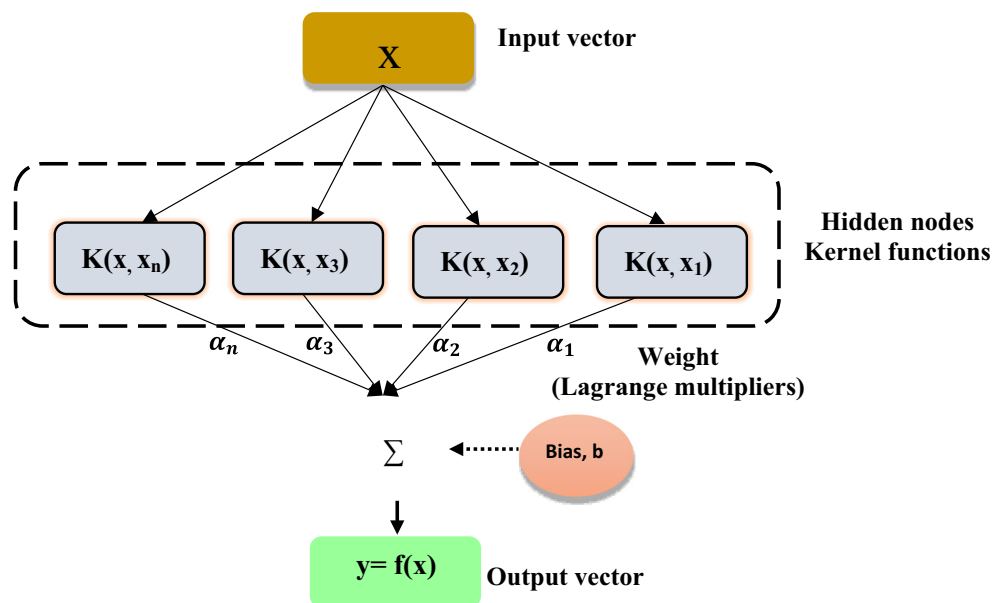
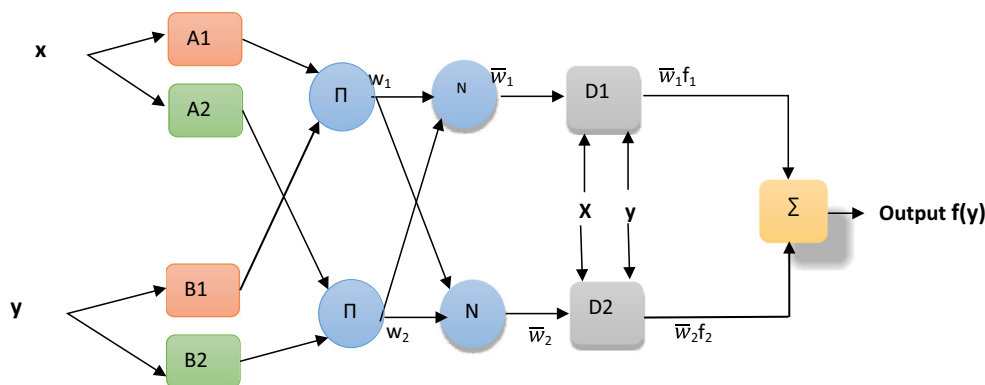


Fig. 4 First-order type Sugeno FIS and ANFIS model structure (Jang 1993)



hybrid algorithm for nonlinear averaging of the values predicted using the base models. The predicted PM_{10} and $PM_{2.5}$ obtained using the four base models (ANFIS, ANN, SVR, MLR) were fed into the input layer of the ANFIS model, and the corresponding PM concentrations were obtained.

For comparing the performance of the nonlinear ensemble technique, two linear ensembles, SA-ensemble (SA-E) and WA-ensemble (WA-E), were also developed for the prediction of both PM_{10} and $PM_{2.5}$. In the SA-E, the arithmetic means of the predicted PM_{10} and $PM_{2.5}$ concentrations were computed using Eq. 9. In the WA-E, the predicted PM_{10} and $PM_{2.5}$ concentrations are computed by giving distinct weights to the outputs of the base models based on their relative importance. The weight is assigned based on the relative significance (NSE value) of the output. The WA-E is expressed by Eq. 10:

$$\bar{P} = \frac{1}{N} \sum_{i=1}^{n_m} P_i \tag{9}$$

$$\bar{P} = \sum_{i=1}^{n_m} w_i P_i \tag{10}$$

in which \bar{P} shows the outcome of ensemble technique, n_m is the number of models used ($n_m = 4$) and P_i stands for the outcome of the i^{th} method (i.e. ANN, ANFIS, SVR and MLR); w_i is the applied weight on the output of the i^{th} model and is determined by

$$w_i = \frac{NSE_i}{\sum_{i=1}^{n_m} NSE_i} \tag{11}$$

NSE_i is the performance efficiency of the i^{th} base model.

Results and discussions

Selection of relevant input parameters

Accuracy in selecting relevant input parameters for developing black box models is crucial since the accuracy and complexity of the model depend heavily on the models' structure. In view of that, two nonlinear measures (single-input single-output

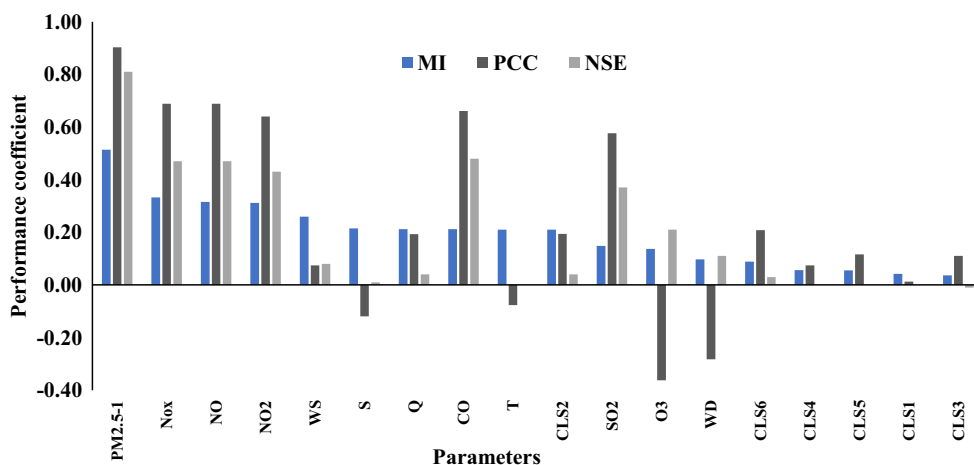
sensitivity analysis evaluated by NSE and MI value between the parameters) were used for obtaining the dominant input parameters. PCC values between the potential input parameters and the responses ($PM_{2.5}$ and PM_{10}) were also computed for incorporating parameters that have a strong linear relationship with the PM into the models. The relevance of the parameter increases as the performance coefficient value of MI, NSE and PCC approaches 1. The parameters having an MI value > 0.2 or NSE value > 0.4 value or PCC values > ±0.5 were considered relevant and hence included in the models. Based on the set criteria, $PM_{2.5-1}$, NO_x , NO, NO_2 , CO, SO_2 , WS, S, Q, T and CLS2 were found to be relevant in the prediction of the $PM_{2.5}$. From Fig. 5, it is clear that the background level of the $PM_{2.5}$ ($PM_{2.5-1}$) has the highest relevance in the prediction of the $PM_{2.5}$ with an MI, NSE and PCC values of 0.51, 0.81 and 0.90, respectively. These findings were supported by several studies; for example Suleiman et al. (2016) found background $PM_{2.5}$ to be the most relevant factor in predicting the $PM_{2.5}$ in Marylebone, London followed by NO. NO_2 and NO_x were also identified to be the second most important factors in the prediction of the $PM_{2.5}$ after vehicle emission (Suleiman et al. 2019). Yazdi et al. (2020) found average city wide $PM_{2.5}$ and the average wind speed to be the most relevant parameters in the prediction of $PM_{2.5}$ with 66.75% and 6.36% contributions, respectively.

Figure 6 showed that, PM_{10-1} , NO_x , NO, NO_2 , CO, SO_2 , WS, S, Q, T and CLS2 were also found to be the most important factors in the prediction of PM_{10} in the study area with PM_{10-1} being the most significant followed by NO_x , NO and NO_2 concentrations. Background concentration was identified to be the most relevant factor due to positive autocorrelation existing in the PM_{10} time series (Paschalidou et al. 2011). Other air pollutants, such as NO, NO_2 , CO and SO_2 , were also reported by Whalley and Zandi (2016) to provide a good prediction of PM_{10} when combined with metrological parameters like T and WS.

Base (single) models

In the second phase of the study, the dominant input parameters for the prediction of $PM_{2.5}$ and PM_{10} obtained in stage 1

Fig. 5 MI, PCC and NSE coefficients indicating the relationship between the input parameters and PM_{2.5}



were used for the development of three AI-based models (ANN, ANFIS and SVR). The model’s efficiencies were evaluated using NSE, MAE and BIAS. The best model is considered as the model with the highest NSE value, least MAE value and BIAS closer to 0. Matlab 2019a was employed for development of all the models. The validation of the models was done using a 10-fold cross validation technique. According to Nourani et al. (2020b), obtaining an optimal structure is essential for any ANN-based models, as such several ANN models trained with Levenberg Marquardt algorithm and sigmoid transfer function to predict PM_{2.5} and PM₁₀ were developed by changing the number of hidden neurons (8–23) using the 11 dominant input parameters. The range for the number of hidden neurons in the ANN was selected based on the range $(2n^{1/2} + m)$ to $(2n + 1)$ given by Fletcher and Goss (1993) where n is the number of input neurons and m is the number of neurons in the output layer. The optimum ANN model was obtained with 12 and 14 hidden neurons for PM_{2.5} and PM₁₀, respectively. For the ANFIS model, a matlab code was developed, and several models using the hybrid optimization algorithm were trained with different membership functions where the best model was

obtained using the “*gbell*” membership function. The SVR on the other hand was trained with a radial basis function (RBF). The RBF kernel was selected for the SVR model due to a fewer number of parameters to be calibrated than the polynomial and the sigmoid kernel functions. Sharghi et al. (2018) also hinted that the RBF kernel mostly provides better performance than the polynomial and the sigmoid kernels. For comparison, a linear model (MLR) was also used for the PM prediction. The result of the best models was given in Table 2.

From Table 2, it can be seen that all the AI-based models give a very good performance in the PM_{2.5} prediction based on the NSE values (>0.75) in both training and testing stages. The result also demonstrated the higher prediction capability of the ANFIS model with NSE, MAE and a BIAS value of 91.03%, 2.26 μg/m³ and 0.09, respectively in the testing stage. The ANN model ranked second in terms of prediction efficiency and lastly SVR with an NSE and MAE values of 85.86%, 80.41% and 3.02 μg/m³, 3.79 μg/m³, respectively. Scatter plots between observed, and the computed values in training (Fig. 7) and testing stages (Fig. 8) show that the data was more compacted along the bisector line of the ANFIS plot indicating higher

Fig. 6 MI, PCC and NSE coefficients indicating the relationship between the input parameters and PM₁₀

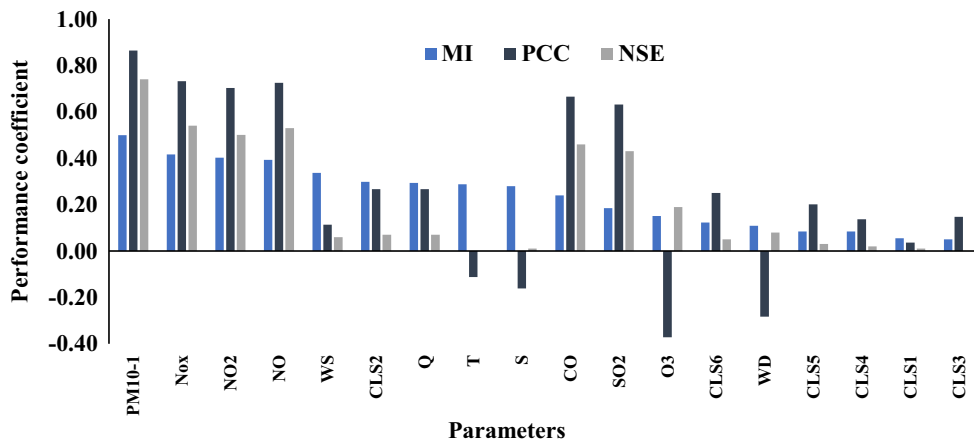


Table 2 Results of the base models for the PM_{2.5} and PM₁₀

Parameter	Models	Training			Testing		
		NSE	MAE (µg/m ³)	BIAS	NSE	MAE (µg/m ³)	BIAS
PM _{2.5}	ANN	0.8894	2.6047	0.1331	0.8586	3.0230	0.1217
	SVR	0.8152	3.1118	0.1590	0.8041	3.7994	0.1530
	ANFIS	0.9353	2.1361	0.1091	0.9103	2.4659	0.0993
	MLR	0.7808	3.8158	0.1950	0.7380	4.2261	0.1701
PM ₁₀	ANN	0.8445	6.1388	0.1460	0.7986	6.6695	0.1362
	SVR	0.8619	5.6156	0.1336	0.8144	6.0285	0.1231
	ANFIS	0.9701	2.9837	0.0710	0.9540	3.0343	0.0620
	MLR	0.8168	6.4011	0.1523	0.7832	6.9738	0.1424

goodness of fit by the ANFIS model. The higher performance of the ANFIS model compared to other models is due to the combined power of the ANN model and the fuzzy logic in prediction. The stability of the models

assessed by comparing the NSE values of the models in the training and testing stages found the SVR model to be more stable with 1.1% decrease in the NSE values followed by ANFIS (2.5%). The high stability of the SVR model

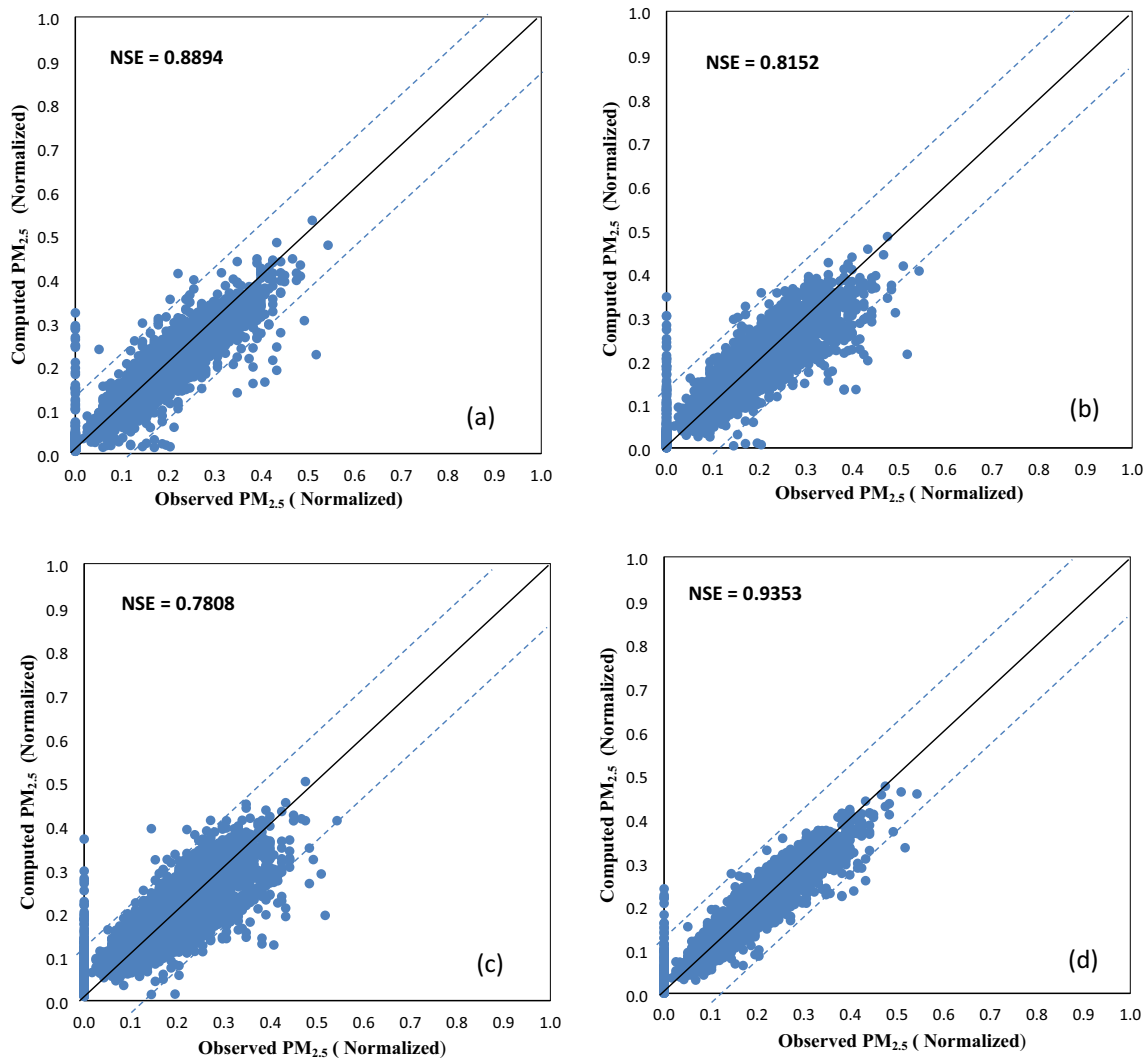


Fig. 7 Scatter plots between observed and computed PM_{2.5} in the training phase for **a** ANN, **b** SVR, **c** MLR and **d** ANFIS

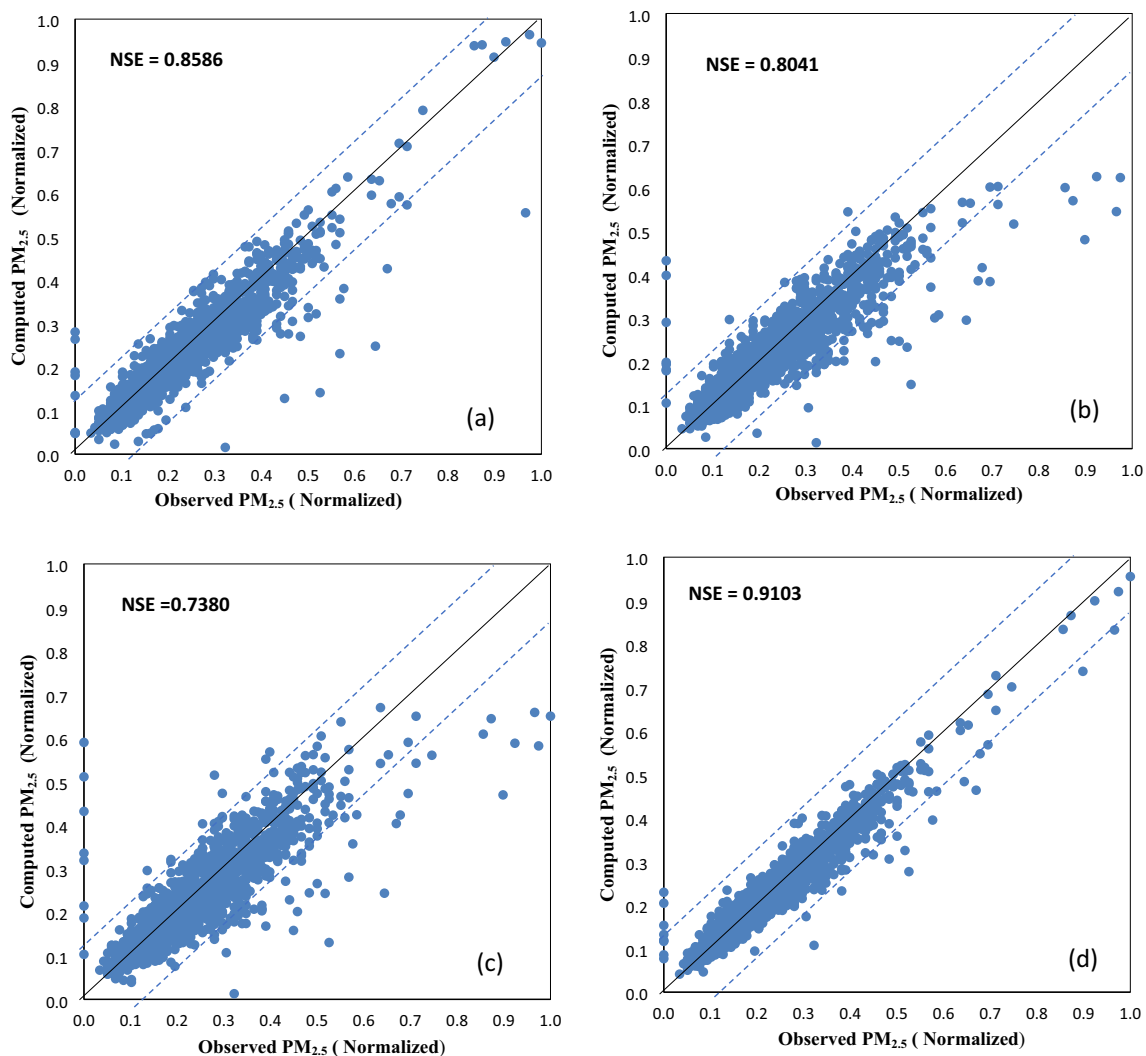


Fig. 8 Scatter plots between observed and computed $PM_{2.5}$ in the testing phase for **a** ANN, **b** SVR, **c** MLR and **d** ANFIS

in prediction has been reported by Fan et al. (2018). Compared to MLR, all the AI-based models have shown higher performance than the MLR with an improved performance of 17%, 12% and 6.6% for ANFIS, ANN and SVR, respectively. The superiority of the ANFIS model over ANN and SVR in $PM_{2.5}$ prediction was also reported by Yeganeh et al. (2017).

All the models including the MLR have shown a very good accuracy in PM_{10} prediction with an NSE value >0.75 in the testing stage. The results indicated higher performance of the ANFIS model (NSE = 95.40% and MAE = 3.03) in the testing stage followed by the SVR model (NSE = 81.44% and MAE = 6.03) and finally ANN. Figures 9 and 10 have also indicated better goodness of fit of the ANFIS model. The ANFIS model was found to be more stable with a NSE decrease of 1.6% between the training and testing stages. The high accuracy of the ANFIS model in predicting PM_{10} in this study was supported by the study conducted by Prasad et al. (2016).

Comparing the performance of ANFIS, ANN and SVR models with MLR has shown an improved performance of 17%, 3.1% and 1.5%, respectively.

The results obtained show that both PM_{10} and $PM_{2.5}$ could be modelled with minimum error using the ANFIS model. Higher MAE values in PM_{10} models compared to $PM_{2.5}$ are due to the higher data range and standard deviation of the PM_{10} data compared to the $PM_{2.5}$ data. Except for the ANN model, the PM_{10} models have higher NSE and lower BIAS than the $PM_{2.5}$ models indicating higher accuracy of PM_{10} models. Although ANFIS showed higher prediction accuracy in terms of NSE, the MAE is high and needs to be minimized.

Ensemble techniques

The ensemble modelling technique was employed to combine the advantages of the individual models for improved accuracy in prediction. The ANFIS model being the most robust base

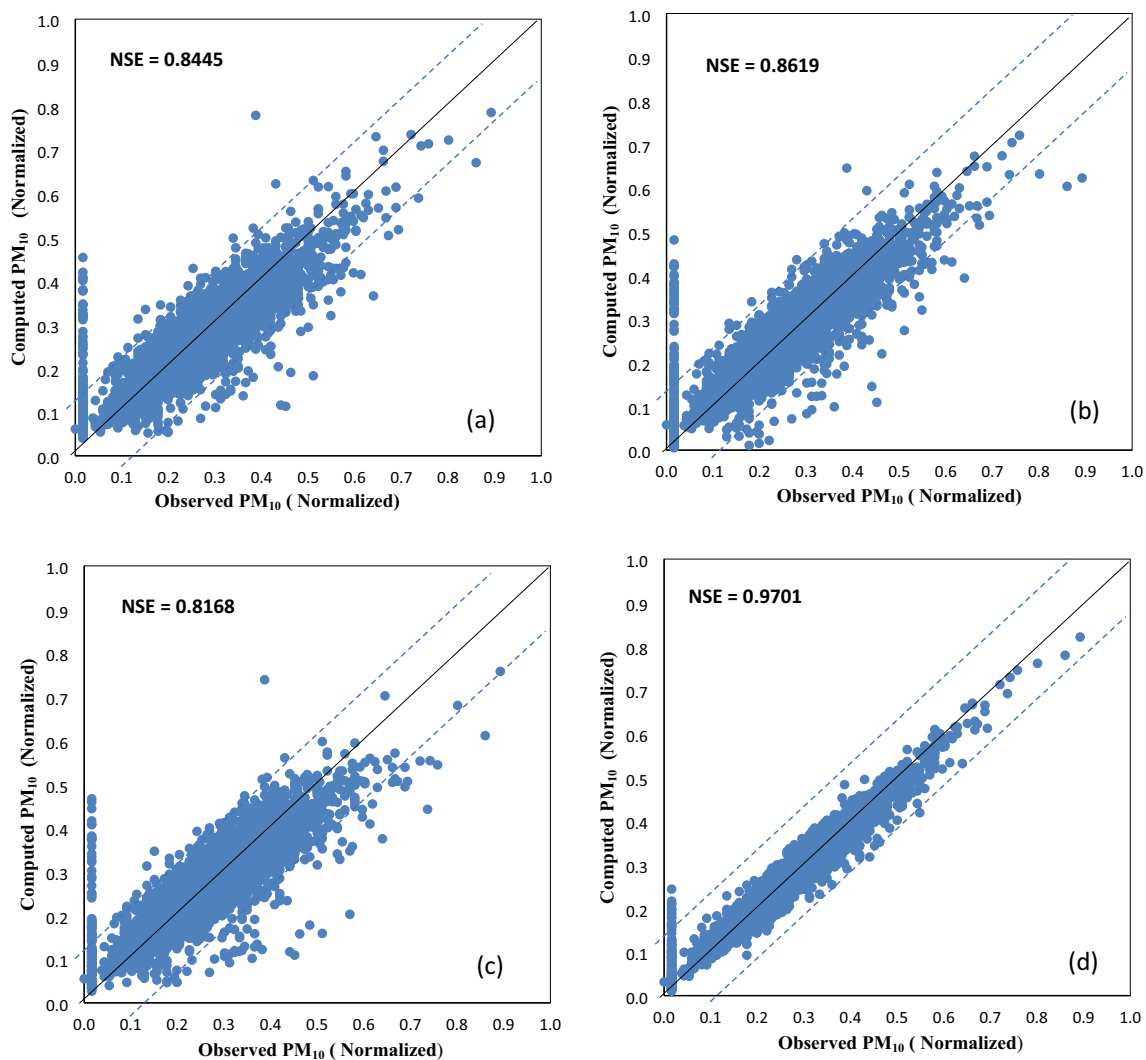


Fig. 9 Scatter plots between observed and computed PM_{10} in the training phase for **a** ANN, **b** SVR, **c** MLR and **d** ANFIS

model in this study was used for nonlinear averaging of the predicted $PM_{2.5}$ and PM_{10} for enhanced prediction. The NF-E model for both $PM_{2.5}$ and PM_{10} was trained using the “*gbell*” function and a hybrid training algorithm. WA-E and SA-E were also developed for comparing the performance of the NF-E. Only results of the best models reported using the single models were used in the ensemble approach. The ensemble result was given in Table 3. It can be seen that NF-E performed better than all the ensemble models given NSE values of 0.9594 and 0.9865 in the testing stage for the $PM_{2.5}$ and PM_{10} , respectively. The WA-E and SA-E gave NSE values lower than the best single model (ANFIS); this is because in any linear averaging, the resulting value is always lower than the highest number (Nourani et al. 2020a). The accuracy of the ensemble models was compared by a radar plot (Figs. 11 and 12), and the result demonstrated a higher accuracy of the NF-E with the smallest NSE change between the training and testing stages. The modelling results

of the models (single and ensemble) were further compared using a Taylor diagram (Figs. 13 and 14) which is a comprehensive tool for comparing models’ performances using three statistical measures (RMSE, R and standard deviation). In the Taylor diagram, the azimuthal position gives the correlation between the actual and the computed values. The RMSE values are directly proportional to the distance between the observed and the predicted fields having the same unit with the standard deviation. For any increase in correlation, the value of the RMSE is decreased. The standard deviation of the pattern increases with increasing radial distance measured from the origin (Taylor 2001). A model is said to be a perfect model by a reference point when its correlation coefficient is 1 (Yaseen et al. 2018). If the standard deviation of the computed values is greater than the standard deviation of the observed values, then it may lead to overestimation and vice versa; hence, standard deviation close to that of the actual data is always preferred. From Figs. 13 and 14, it is clear that NF-E

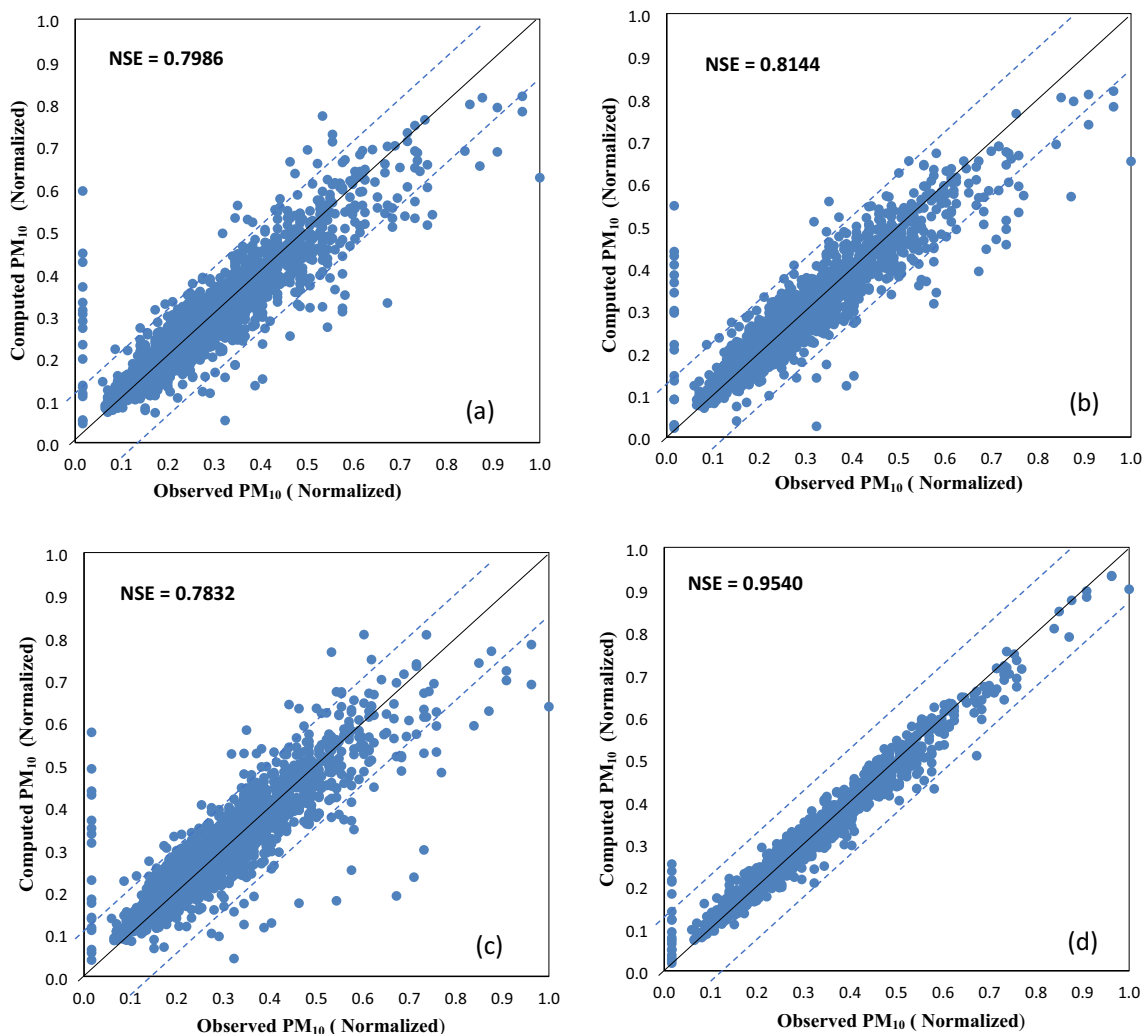


Fig. 10 Scatter plots between observed and computed PM₁₀ in the testing phase for **a** ANN, **b** SVR, **c** MLR and **d** ANFIS

outperformed all in the models in the prediction of PM_{2.5} and PM₁₀ with highest *R* values and lowest RMSE value and value of standard deviation closer to that of the actual data. The improved performance of prediction models in the prediction of PM_{2.5} and PM₁₀ using the ensemble technique has been proved by several studies including the study by

Shtein et al. (2019) using the generalized additive model ensemble model. An improved performance in the prediction of PM_{2.5} concentration using BP-NN ensemble (Feng et al. 2019) and feature extraction and stacking-driven ensemble (Sun and Li 2020) supported the findings of this study. Maciag et al. (2019) also found the clustering-based ensemble

Table 3 Results of the ensemble modelling

Parameter	Models	Training			Testing		
		NSE	MAE (µg/m ³)	BIAS	NSE	MAE(µg/m ³)	BIAS
PM _{2.5}	NF-E	0.9762	1.4875	0.0658	0.9594	1.6342	0.0760
	WA-E	0.8859	2.6991	0.1379	0.8581	3.1142	0.1254
	SA-E	0.8817	2.7423	0.1401	0.8539	3.1639	0.1274
PM ₁₀	NF-E	0.9923	1.7348	0.04127	0.9865	1.6624	0.0340
	WA-E	0.8946	5.0610	0.1204	0.8605	5.4304	0.1109
	SA-E	0.8902	5.1514	0.1226	0.8554	5.5338	0.1130

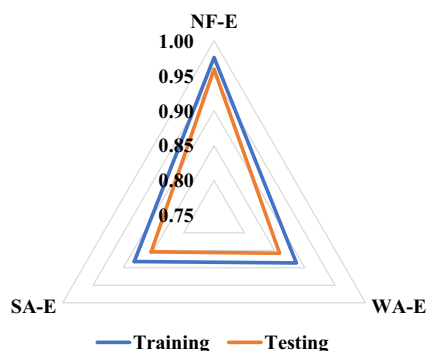


Fig. 11 Radar plots comparing the NSE values of the PM_{2.5} ensemble models in the training and testing stages

to improve the prediction accuracy of PM₁₀ concentration in London. The higher performance of the ensemble approach is due to its ability to combine the unique advantage of each of the base models.

Conclusion

The study proposed a novel nonlinear ensemble approach for the prediction of PM_{2.5} and PM₁₀ concentration in Marylebone, London. The NF-E involves three main stages, that is relevant input selection via MI, PCC and sensitivity analysis, single modeling and lastly ensemble modeling. Findings from the sensitivity analysis revealed NO_x, NO, NO₂, CO and SO₂ to be the most relevant air pollutants in the prediction of PM_{2.5} and PM₁₀ concentration after the background concentration, while the most relevant metrological parameters were found to be WS and T. Q and CLS2 traffic were found to be the most important traffic-related parameters. The result of the ensemble models revealed higher prediction accuracy of the NF-E than all the models (linear ensemble and single models) which depending on the model could enhance the performance of the base models by 4–22% and 3–20% for PM_{2.5} and PM₁₀, respectively at the testing stage. Higher prediction accuracy demonstrated by the

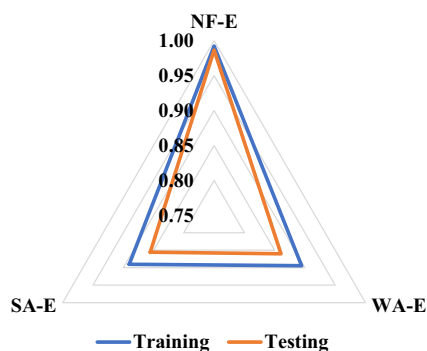


Fig. 12 Radar plots comparing the NSE values of the PM₁₀ ensemble models in the training and testing stages

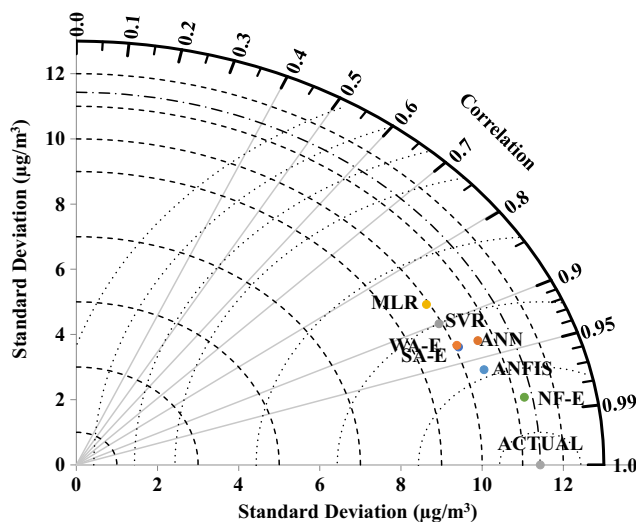


Fig. 13 Taylor diagram representing different statistical parameters of the PM_{2.5} models

proposed methodology was due to the careful selection of the relevant input parameters in the single modelling stage and combining the unique features of the four base models in the ensemble stage. Although the NF-E estimated both the PM_{2.5} and PM₁₀ concentration with high accuracy, careful selection of the base models to be used for the ensemble could be a major limitation of the methodology, since the efficiency of the ensemble models depends heavily on the results obtained using the base models. In other words, including the result of a model with a low performance could result to a lower prediction accuracy of the ensemble model. The efficiency of the proposed methodology could be compared with other advanced models such as the emotional neural network and linear-nonlinear hybrid models in future studies.

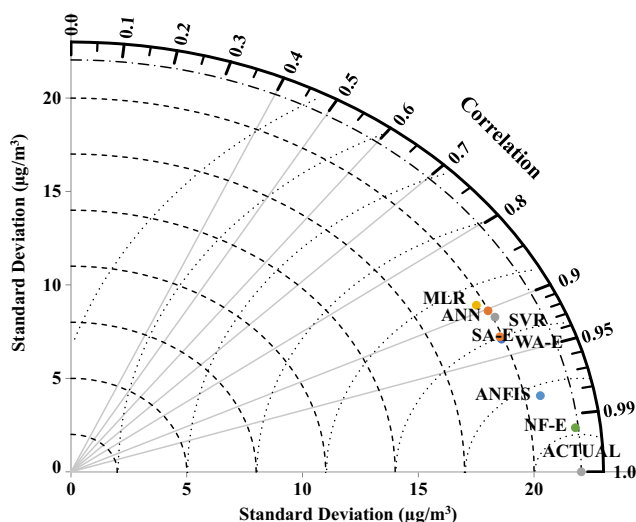


Fig. 14 Taylor diagram representing different statistical parameters of the PM₁₀ models

Code availability Matlab 2019a was used for conducting the study and is available.

Author contribution The authors' contribution to the paper is as follows: study conception and design: VN, IKU, HG; analysis and interpretation of results: IKU, VN; draft manuscript preparation: VN, IKU, HG. All authors reviewed the results and approved the final version of the manuscript.

Data availability Data for the study is available and can be found at <https://uk-air.defra.gov.uk/data/maryleboneroad>.

Declarations

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Conflict of interest The authors declare no competing interests.

References

- Ahmed AA, Pradhan B (2019) Vehicular traffic noise prediction and propagation modelling using neural networks and geospatial information system. *Environ Monit Assess* 191:190. <https://doi.org/10.1007/s10661-019-7333-3>
- Arhami M, Kamali N, Rajabi MM (2013) Predicting hourly air pollutant levels using artificial neural networks coupled with uncertainty analysis by Monte Carlo simulations. *Environ Sci Pollut Res* 20:4777–4789. <https://doi.org/10.1007/s11356-012-1451-6>
- Azeez OS, Pradhan B, Shafri HZM, Shukla N, Lee CW, Rizeei H (2019) Modeling of CO emissions from traffic vehicles using artificial neural networks. *Appl Sci* 9:313. <https://doi.org/10.3390/app9020313>
- Cai M, Yin Y, Xie M (2009) Prediction of hourly air pollutant concentrations near urban arterials using artificial neural network approach. *Transp Res Part D Transp Environ* 14:32–41. <https://doi.org/10.1016/j.trd.2008.10.004>
- Çaydaş U, Hasçalık A, Ekici S (2009) An adaptive neuro-fuzzy inference system (ANFIS) model for wire-EDM. *Expert Syst Appl* 36:6135–6139. <https://doi.org/10.1016/j.eswa.2008.07.019>
- Codur MY, Atalay A, Unal A (2017) Performance evaluation of the ANN and ANFIS models in urban traffic noise prediction. *Fresenius Environ Bull* 26:4254–4260
- Doğan E, Akgüngör AP (2013) Forecasting highway casualties under the effect of railway development policy in Turkey using artificial neural networks. *Neural Comput & Applic* 22:869–877. <https://doi.org/10.1007/s00521-011-0778-0>
- Elkiran G, Nourani V, Abba SI, Abdullahi J (2018) Artificial intelligence-based approaches for multi-station modelling of dissolve oxygen in river. *Glob J Environ Sci Manag* 4:439–450. <https://doi.org/10.22034/gjesm.2018.04.005>
- European Environment Agency (2012) Road Traffic's Contribution to Air Quality in European Cities-ETC/ACM Technical Paper 2012/14. Road traffic's contribution to air quality in European cities ETC/ACM Technical Paper 2012/14 — Eionet Portal (europa.eu). Accessed 20 Oct 2020
- Fan J, Yue W, Wu L, Zhang F, Cai H, Wang X, Lu X, Xiang Y (2018) Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. *Agric For Meteorol* 263:225–241. <https://doi.org/10.1016/j.agrformet.2018.08.019>
- Feng X, Fu TM, Cao H, Tian H, Fan Q, Chen X (2019) Neural network predictions of pollutant emissions from open burning of crop residues: application to air quality forecasts in southern China. *Atmos Environ* 204:22–31. <https://doi.org/10.1016/j.atmosenv.2019.02.002>
- Ghaffari A, Abdollahi H, Khoshayand M et al (2006) Performance comparison of neural network training algorithms in modeling of bimodal drug delivery. *Int J Pharm* 327:126–138. <https://doi.org/10.1016/j.ijpharm.2006.07.056>
- Gulia S, Khanna I, Shukla K, Khare M (2020) Ambient air pollutant monitoring and analysis protocol for low and middle income countries: an element of comprehensive urban air quality management framework. *Atmos Environ* 222:117120. <https://doi.org/10.1016/j.atmosenv.2019.117120>
- Jahani B, Mohammadi B (2019) A comparison between the application of empirical and ANN methods for estimation of daily global solar radiation in Iran. *Theor Appl Climatol* 137:1257–1269. <https://doi.org/10.1007/s00704-018-2666-3>
- Jang JSR (1993) ANFIS: adaptive-network-based fuzzy inference system. *IEEE Trans Syst Man Cybern* 23:665–685. <https://doi.org/10.1109/21.256541>
- Jones AM, Harrison RM (2005) Interpretation of particulate elemental and organic carbon concentrations at rural, urban and kerbside sites. *Atmos Environ* 39:7114–7126. <https://doi.org/10.1016/j.atmosenv.2005.08.017>
- Kim S, Singh VP (2014) Modeling daily soil temperature using data-driven models and spatial distribution. *Theor Appl Climatol* 118:465–479. <https://doi.org/10.1007/s00704-013-1065-z>
- Krishan M, Jha S, Das J, Singh A, Goyal MK, Sekar C (2019) Air quality modelling using long short-term memory (LSTM) over NCT-Delhi, India. *Air Qual Atmos Health* 12:899–908. <https://doi.org/10.1007/s11869-019-00696-7>
- Kumar P, Nigam SP, Kumar N (2014) Vehicular traffic noise modeling using artificial neural network approach. *Transp Res Part C Emerg Technol* 40:111–122. <https://doi.org/10.1016/j.trc.2014.01.006>
- Legates DR, McCabe GJ Jr (1999) Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resour Res* 35:233–241. <https://doi.org/10.1029/1998WR900018>
- Maciąg PS, Kasabov N, Kryszkiewicz M, Bembenik R (2019) Air pollution prediction with clustering-based ensemble of evolving spiking neural networks and a case study for London area. *Environ Model Softw* 118:262–280. <https://doi.org/10.1016/j.envsoft.2019.04.012>
- Mehdipour V, Stevenson DS, Memarianfard M, Sihag P (2018) Comparing different methods for statistical modeling of particulate matter in Tehran, Iran. *Air Qual Atmos Health* 11:1155–1165. <https://doi.org/10.1007/s11869-018-0615-z>
- Moriassi DN, Arnold JG, Van Liew MW et al (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans ASABE* 50:885–900. <https://doi.org/10.1234/590>
- Nourani V, Fard Sayyah M (2012) Sensitivity analysis of the artificial neural network outputs in simulation of the evaporation process at different climatologic regimes. *Adv Eng Softw* 47:127–129
- Nourani V, Kalantari O, Baghanam AH (2012) Two semidistributed ANN-based models for estimation of suspended sediment load. *J Hydrol Eng* 17:1368–1380. [https://doi.org/10.1061/\(ASCE\)JHE.1943-5584.0000587](https://doi.org/10.1061/(ASCE)JHE.1943-5584.0000587)
- Nourani V, RezapourKhanghah T, Baghanam H (2014) Case studies in intelligent computing. Taylor and Francis Group, New York
- Nourani V, Elkiran G, Abba SI (2018) Wastewater treatment plant performance analysis using artificial intelligence—an ensemble

- approach. *Water Sci Technol* 78:2064–2076. <https://doi.org/10.2166/wst.2018.477>
- Nourani V, Elkiran G, Abdullahi J (2019) Multi-station artificial intelligence based ensemble modeling of reference evapotranspiration using pan evaporation measurements. *J Hydrol* 577:123958. <https://doi.org/10.1016/j.jhydrol.2019.123958>
- Nourani V, Gökçekuş H, Umar IK (2020a) Artificial intelligence based ensemble model for prediction of vehicular traffic noise. *Environ Res* 180:108852. <https://doi.org/10.1016/j.envres.2019.108852>
- Nourani V, Gökçekuş H, Umar IK, Najafi H (2020b) An emotional artificial neural network for prediction of vehicular traffic noise. *Sci Total Environ* 707:136134. <https://doi.org/10.1016/j.scitotenv.2019.136134>
- Naschalidou AK, Karakitsios S, Kleanthous S, Kassomenos PA (2011) Forecasting hourly PM10 concentration in Cyprus through artificial neural networks and multiple regression models: Implications to local environmental management. *Environ Sci Pollut Res* 18:316–327. <https://doi.org/10.1007/s11356-010-0375-2>
- Prasad K, Gorai AK, Goyal P (2016) 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>
- Rai AA, Pai PS, Rao BRS (2015) Prediction models for performance and emissions of a dual fuel CI engine using ANFIS. *Sadhana - Acad Proc Eng Sci* 40:515–535. <https://doi.org/10.1007/s12046-014-0320-z>
- Raj Kiran N, Ravi V (2008) Software reliability prediction by soft computing techniques. *J Syst Softw* 81:576–583. <https://doi.org/10.1016/j.jss.2007.05.005>
- Rumelhart DE, Hinton GE, Williams R (1986) Learning representations by backpropagating errors. *Nature* 323:533–536. <https://doi.org/10.1038/324227a0>
- Sharghi E, Nourani V, Behfar N (2018) Earthfill dam seepage analysis using ensemble artificial intelligence based modeling. *J Hydroinf* 20:1071–1084. <https://doi.org/10.2166/hydro.2018.151>
- Shtein A, Kloog I, Schwartz J et al (2019) Estimating daily PM2.5 and PM10 over Italy using an ensemble model. *Environ Sci Technol*. <https://doi.org/10.1021/acs.est.9b04279>
- Suleiman A, Tight MR, Quinn AD (2016) Hybrid neural networks and boosted regression tree models for predicting roadside particulate matter. *Environ Model Assess* 21:731–750. <https://doi.org/10.1007/s10666-016-9507-5>
- Suleiman A, Tight MR, Quinn AD (2019) Applying machine learning methods in managing urban concentrations of traffic-related particulate matter (PM10 and PM2.5). *Atmos Pollut Res* 10:134–144. <https://doi.org/10.1016/j.apr.2018.07.001>
- Sun W, Li Z (2020) Hourly PM2.5 concentration forecasting based on feature extraction and stacking-driven ensemble model for the winter of the Beijing-Tianjin-Hebei area. *Atmos Pollut Res* 11:110–121. <https://doi.org/10.1016/j.apr.2020.02.022>
- Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. *J Geophys Res* 106:7183–7192. <https://doi.org/10.1029/2000JD900719>
- Uzoigwe JC, Prum T, Bresnahan E, Garelnabi M (2013) The emerging role of outdoor and indoor air pollution in cardiovascular disease. *N Am J Med Sci* 5:445–453. <https://doi.org/10.4103/1947-2714.117290>
- Van Donkelaar A, Martin RV, Park RJ (2006) Estimating ground-level PM2.5 using aerosol optical depth determined from satellite remote sensing. *J Geophys Res* 111:1–10. <https://doi.org/10.1029/2005JD006996>
- Vapnik VN (1998) *Statistical learning theory*. Wiley, New York
- Wang W, Xu D, Chau KW, Chen S (2015) Improved annual rainfall-runoff forecasting using PSO–SVM model based on EEMD. *J Hydroinf* 15:1377–1390. <https://doi.org/10.2166/hydro.2013.134>
- Whalley J, Zandi S (2016) Particulate matter sampling techniques and data particulate matter Sampling techniques and data modelling methods. In: *Air quality—measurement and modeling*. INTECH Open Science, pp 29–54
- WHO (World Health Organization) (2018) Global ambient air quality database. <https://www.who.int/airpollution/en/>. Accessed 10 Dec 2020
- Yaseen ZM, Deo RC, Hilal A, Abd AM, Bueno LC, Salcedo-Sanz S, Nehdi ML (2018) Predicting compressive strength of lightweight foamed concrete using extreme learning machine model. *Adv Eng Softw* 115:112–125. <https://doi.org/10.1016/j.advengsoft.2017.09.004>
- Yazdi MD, Kuang Z, Dimakopoulou K, Barratt B (2020) Predicting Fine Particulate Matter (PM_{2.5}) in the Greater London Area : An Ensemble Approach using Machine Learning Methods. *Remote Sens* 12:914. <https://doi.org/10.3390/rs12060914>
- Yeganeh B, Hewson MG, Clifford S, Knibbs LD, Morawska L (2017) A satellite-based model for estimating PM2.5 concentration in a sparsely populated environment using soft computing techniques. *Environ Model Softw* 88:84–92. <https://doi.org/10.1016/j.envsoft.2016.11.017>

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