Contemporary Mathematics
http://ojs.wiserpub.com/index.php/CM/

Research Article

An Improved Weight Optimization of Hybrid Machine Learning Models for Forecasting Daily $PM_{2.5}$ Concentration

Manlika Ratchagit

Program in Statistics and Information Management, Faculty of Science, Maejo University, Chiang Mai, Thailand
E-mail: manlika@mju.ac.th

Received: 12 June 2024; **Revised:** 29 August 2024; **Accepted:** 10 September 2024

Abstract: $PM_{2.5}$ is an air pollutant primarily produced by human activities, including the combustion of fossil fuels, industrial emissions, vehicle exhaust, and more. This issue has emerged as a substantial global concern, particularly in Thailand, where the levels of $PM_{2.5}$ during the summer season have reached hazardous levels. $PM_{2.5}$ forecasting is a widely discussed subject that raises awareness and safeguards individuals against pollution. The novelty of this paper is to estimate the weight of linear and nonlinear hybrid models using a differential evolution algorithm. This approach is used for the minimization of the objective function based on hybrid procedures. The data utilized in this study consists of the daily mean $PM_{2.5}$ concentration (micrograms per cubic meter) obtained from the Pollution Control Department, Ministry of Natural Resources and Environment, Thailand. The data covers the period from January 2014 to June 2023, encompassing a total of 3,468 observations. Three well-known machine learning approaches, namely the artificial neural network, the long short-term memory, and the convolutional neural network, are employed. We then combined the predicted $PM_{2.5}$ obtained from the single machine learning model using linear and nonlinear hybrid procedures. The differential evolution algorithm is utilized to estimate the weight of the hybrid techniques for both scenarios and compare it with state-of-the-art weight approximation. The criteria for evaluating the performance of various hybrid approaches are the performance metrics: the mean absolute error and the median absolute error. The findings of this paper indicate that using a differential evolution algorithm for weight optimization in hybrid procedures outperforms state-of-the-art weight approaches for both linear and nonlinear hybrid models in terms of performance metrics.

Keywords: machine learning, differential evolution algorithm, $PM_{2.5}$, air pollution, optimization

MSC: 68T07, 60G25, 34G25

1. Introduction

In the atmosphere, particulate matter (PM) is a mixture of liquid droplets and solid particles. $PM_{2.5}$ denotes airborne particles that have a size smaller than 2.5 microns. $PM_{2.5}$ is a category of air pollution mainly generated by human activities. It encompasses tiny particles emitted from several sources, including burning fossil fuels, building sites, industrial pollutants, automobile emissions, and etc. $PM_{2.5}$ pollution has been the subject of an increase in the number of articles published over the past decade [1]. The researchers aimed to investigate the effects of $PM_{2.5}$ dust pollution on human health and anticipate harmful $PM_{2.5}$ levels in society [2]. High levels of $PM_{2.5}$ pollutants can raise the risk

DOI: https://doi.org/10.37256/cm.5320245131

of health issues, particularly in the respiratory and cardiovascular systems, such as heart disease, asthma, and low birth weight [3]. Dust pollution, which affects most of the globe, is primarily concentrated in Asia. Thailand experiences one of the most severe levels of air pollution during the summer months of January through April. Thailand's people, particularly in larger cities like Chiang Mai, Khon Kaen, and Samut Sakhon, are experiencing respiratory problems. Chiang Mai is located in the northern region. Khon Kaen is located in the Northeast, whereas Samut Sakhon is in the Central region. To forecast $PM_{2.5}$ concentration, both parametric and non-parametric statistical models were used. The autoregressive integrated moving average (ARIMA), seasonal autoregressive integrated moving average (SARIMA), and double exponential smoothing method (DES) are well-established parametric statistical methods commonly used for single forecasting models [4–6]. Parametric statistical techniques are based on a linear model framework; however, the air pollution time series measurements exhibit a nonlinear pattern [7]. Non-parametric statistical approaches, such as machine learning (ML) techniques, have been used to address this issue of $PM_{2.5}$ dust prediction. These approaches include artificial neural network (ANN) [8], multilayer perceptron (MLP) [9], long short-term memory (LSTM), convolutional neural networks (CNN) [10], bidirectional long short-term memory (BiLSTM) [11], random forest (RF), gradient boosting machine (GBM), and k-nearest neighbor (KNN) [12].

Amnuaylojaroen [13] predicted $PM_{2.5}$ levels using a multivariate linear regression method. The work utilizes three provinces in the Northern region: Chiang Mai, Lampang, and Nan. The criteria for choosing the suitable method are the coefficient of determination (R^2), root mean square error (RMSE), and standard error (SE). Wongrin et al. [14] investigated deep learning techniques and statistical methods to quantify the concentration of $PM_{2.5}$ in Thailand. This study examines 16 stations in Thailand as case studies. The RMSE is the criterion to evaluate the performance of all models. The findings confirm that ARIMA outperforms deep learning approaches in most stations. Saiohai et al. [15] examined the prediction of $PM_{2.5}$ using both multiple linear regression (MLR) and multilayer perceptron (MLP) models. The data utilized in this study are obtained from the Microclimate and Air Pollutants Monitoring Tower station located at Kasetsart University in Bangkok, Thailand. The authors found that MLR outperformed MLP in predicting $PM_{2.5}$. Wanishsakpong et al. [16] examined the architecture of the $PM_{2.5}$ concentration model using the Deep Belief Network (DBN) technique. The data included in this analysis were collected from the air quality monitoring station located at Yupparaj Wittayalai School in Chiang Mai. The results indicate that the optimal DBN structure for $PM_{2.5}$ concentration forecasting is composed of five input nodes and twenty hidden neurons in the initial hidden layer, with an accuracy of 88.4 percent.

In addition, hybrid methods have become widespread in the forecasting area. The hybrid structure is composed of parallel and series structures. The linear and nonlinear combination methods are part of a parallel framework, which presents a challenge for researchers in their search for optimal weight optimization. Furthermore, the series structure is a combination of linear and nonlinear models, such as ARIMA-LSTM [17], ARIMA-ANN, ARIMA-SVM [18], and LSTM-AR [19]. The structure of the hybrid system is shown in Figure 1.

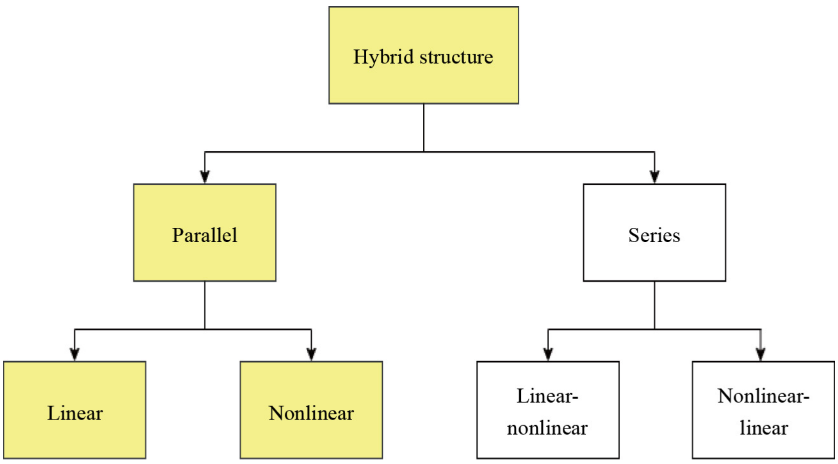


Figure 1. The hybrid framework

In this work, we focus on parallel hybrid structures only because improving weight estimation based on the parallel hybrid framework is still challenging for many scholars to search for the optimal weight. Linear hybrid approaches were initially proposed by Bates and Granger [20]. The basic idea of the linear combination model is to combine the predicted values obtained from two different techniques based on various weight approaches. Several studies have shown that the hybrid forecast is more efficient than the individual technique [21]. For instance, AH Mohamed [22] compared the performance of single and hybrid models for international trade in Egypt. The hybrid models used in this study are linear and nonlinear combination methods. The findings suggest that the Altavilla and Ciccarelli method, a hybrid technique, performs better than other combination methods for forecasting the value of imports and exports. It also demonstrates greater accuracy compared to the best individual forecasting approaches for the international total trade in Egypt dataset and the total exports and imports dataset. Hu et al. [23] investigate the relationship between combined and single-model forecasts of the demand for China and Taiwan tourism demand. They proposed the nonadditive combination method by using the fuzzy integral to integrate single-model forecasts obtained from individual grey prediction models for tourism demand forecasting. The finding shows that the proposed NA-FCM performed well. Furthermore, the accuracy of combined forecasts obtained by the NA-FCM with different model combinations was significantly superior to the average accuracy of single-model forecasts. Li et al. [24] introduce a novel combination forecasting model using the MIDAS regression model and the machine learning MIDAS models. The data utilized in this work are the rate of return growth of weekly NBP natural gas and daily carbon prices, coal prices, crude oil prices, and FTSE-100 index in the European market. It can be concluded that the novel Combination-MIDAS-ELM model outperforms another combination model in forecasting.

As mentioned above, hybrid techniques are presented to improve the accuracy of the single method. The linear and nonlinear hybrid procedures are employed to integrate the well-known ML methods. To our knowledge, there has been limited research on weight estimation for both linear and nonlinear parallel hybrid procedures. This motivated us to integrate two ML methods using different weight optimization techniques of the hybrid procedures. The primary objectives of this paper are as follows:

- To improve weight optimization in linear and nonlinear hybrid models using a differential evolution algorithm
- To compare the linear and nonlinear weight optimization of the hybrid technique
- To develop a hybrid method using popular ML procedures

The remaining sections of this paper are organized as follows. In Section 2, the methodology is presented. The data is shown in Section 3. Section 4 contains the empirical findings. The paper is finally concluded and discussed in Section 5.

2. Methodology

In this work, the linear and nonlinear hybrid procedures are examined using three well-known machine learning techniques, including the artificial neural network (ANN), long short-term memory (LSTM), and convolution neural network (CNN).

2.1 *Machine learning (ML)*

Machine learning (ML) is a specialized area within computer science and artificial intelligence (AI) that focuses on utilizing data and algorithms to enable AI to emulate how humans learn, thereby enhancing its accuracy. The popularity of machine learning has increased in comparison to parametric statistical models due to its capacity to address nonlinear and multidimensional issues that are appropriate for real-world datasets [25].

2.1.1 *Artificial neural network (ANN)*

A conventional artificial neural network (ANN) comprises three interconnected layers: the input, hidden, and output. The number of units in the input and output layers is determined by the size of the input and output data. The input units in an artificial neural network (ANN) take raw data from the external environment and pass it on to the network without performing any computational operations. Instead, these units transmit information to the hidden units. The hidden nodes process data from input to output units [26]. The architecture of the ANN is depicted in Figure 2.

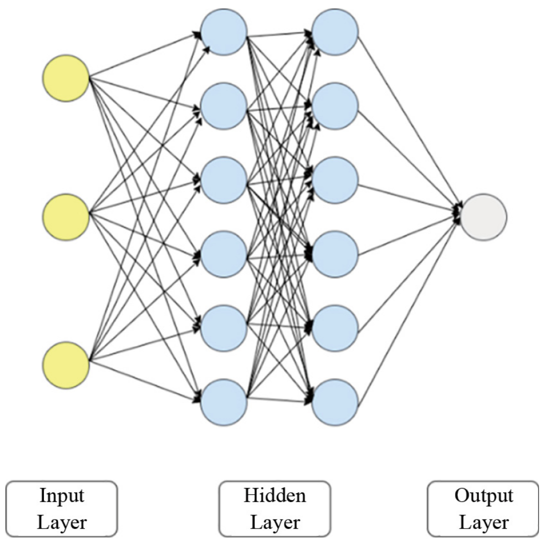


Figure 2. ANN diagram [27]

2.1.2 *Long short-term memory (LSTM)*

The purpose of LSTM is to address the limitations of the current recurrent neural network (RNN). LSTM is a type of neural network that produces its output by processing the input in a forward direction. An LSTM comprises three main components: an input gate, a forget gate, and an output gate. The primary objective of these gates is to execute the processing of information obtained from the memory units [28]. Figure 3 displays the interior cell of the LSTM.

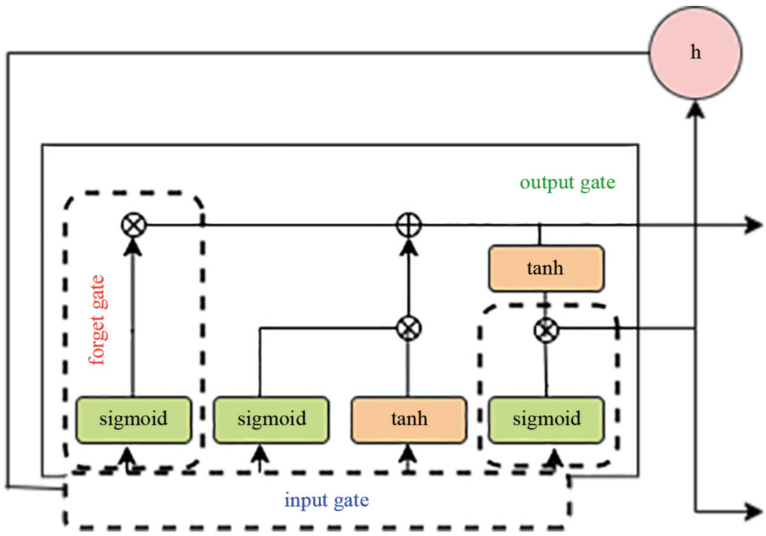


Figure 3. LSTM cell [26]

2.1.3 *Convolution neural network (CNN)*

CNN is a form of neural network with several layers, including convolutional, max-pooling, and fully connected MLP layers [28]. Figure 4 illustrates the architecture of CNN.

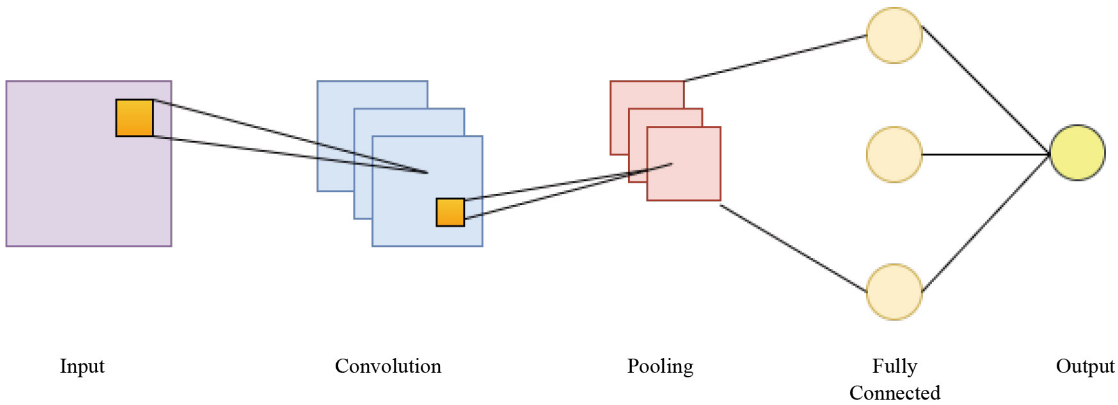


Figure 4. Architecture of CNN [5]

The process to obtain predicted $PM_{2.5}$ pollution from each ML model is displayed in Figure 5.

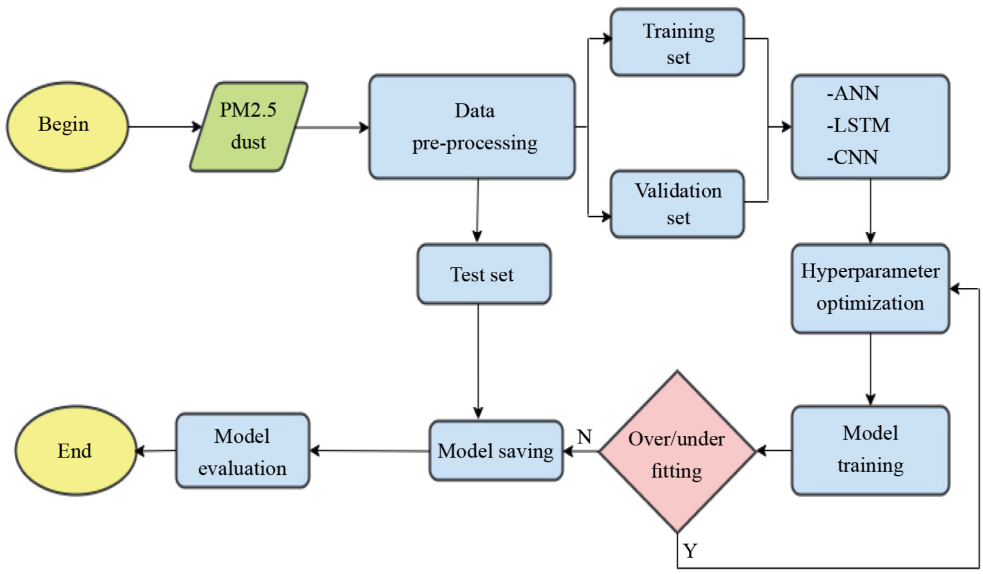


Figure 5. The flowchart of ML procedure

This flowchart starts with obtaining the daily $PM_{2.5}$ concentration from the Pollution Control Department, Ministry of Natural Resources and Environment, Thailand. The data are cleaned up, and any missing values are filled in using a splines interpolation process. After that, we separated the data into three categories: training, validation, and testing. The Sherpa algorithm searches for the hyperparameters in machine learning procedures. Next, the data are trained and validated using the hyperparameter from the previous step. To evaluate overfitting and underfitting, the performance metrics of the loss function between the training and validation sets are compared. The hyperparameter of the ML model is stored if the data does not exhibit any overfitting or underfitting issues. Finally, the optimal hyperparameter of each ML technique is employed to acquire the predicted $PM_{2.5}$ values.

2.2 *Hybrid techniques*

This work investigates the linear and nonlinear combination forecasts using several weight estimations. The differential evolution algorithm estimates the weight for two cases and compares it with the state-of-the-art weight optimization. The eight hybrid techniques, including linear and nonlinear combinations, are listed below [7, 29].

2.2.1 *Simple average method (AVG)*

$$\hat{y}_{ct} = 0.5\hat{y}_{1t} + 0.5\hat{y}_{2t}, \tag{1}$$

where $\hat{y}_{ct}$ is the hybrid forecast at time t, $\hat{y}_{1t}$ and $\hat{y}_{2t}$ are the predicted $PM_{2.5}$ pollution obtained from the first and the second ML procedures, respectively.

2.2.2 *Variance No Covariance (VAR-NO-CORR)*

$$\hat{y}_{ct} = w_1\hat{y}_{1t} + w_2\hat{y}_{2t}, \tag{2}$$

here $w_1 = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}$, $w_2 = 1 - w_1$, σ_1^2, σ_2^2 are the variances derived from the first and the second ML methods, respectively.

2.2.3 *Discount mean square forecast error (DMSFE)*

$$\hat{y}_{ct} = w_1\hat{y}_{1t} + w_2\hat{y}_{2t}, \tag{3}$$

where $w_k = \frac{[\sum_{t=1}^{n} \phi^{n-t-1}(y_t-\hat{y}_{kt})^2]^{-1}}{\sum_{t=1}^{2}[\sum_{t=1}^{n} \phi^{n-t-1}(y_t-\hat{y}_{kt})^2]^{-1}}$, $k = 1, 2$ and ϕ is a discount factor with $0 < \phi \leq 1$.

2.2.4 *Simple average with differential evolution algorithm (AVG-DE)*

$$\hat{y}_{ct} = w'_1\hat{y}_{1t} + w'_2\hat{y}_{2t}, \tag{4}$$

where w'_1, w'_2 are the weights obtained from the DE technique.

2.2.5 *Geometric mean (GM)*

$$\hat{y}_{ct} = [\hat{y}_{1t}]^{w_1}[\hat{y}_{2t}]^{w_2}, \tag{5}$$

where $w_1 = w_2 = 0.5$.

2.2.6 *Harmonic mean (HM)*

$$\hat{y}_{ct} = \frac{\hat{y}_{1t}\hat{y}_{2t}}{w_1\hat{y}_{1t} + w_2\hat{y}_{2t}}, \tag{6}$$

here $w_1 = w_2 = 0.5$.

2.2.7 *Geometric mean with differential evolution algorithm (GM-DE)*

$$\hat{y}_{ct} = [\hat{y}_{1t}]^{w'_1}[\hat{y}_{2t}]^{w'_2}. \tag{7}$$

2.2.8 *Harmonic mean with differential evolution algorithm (HM-DE)*

$$\hat{y}_{ct} = \frac{\hat{y}_{1t}\hat{y}_{2t}}{w'_1\hat{y}_{1t} + w'_2\hat{y}_{2t}}. \tag{8}$$

2.3 *Differential evolution algorithm*

A population-based technique for resolving global optimization issues is the differential evolution algorithm. The concept was initially proposed by Storn and Price in 1996 [30]. The pseudocode for the differential evolution algorithm is presented in Algorithm 1.

Algorithm 1. DE algorithm

Input: NP (Population size), D (Dimension), F (Scaling factor), CR (Cross rate), MAXI (Maximum number of iterations), I (Iteration),

Output: $\mathbf{X}_i^{\text{best}}$ (Best solution)

$I \leftarrow 1$

for $i = 1$: NP do

$X_i^I = X_{min}^I + rand(0, 1) \cdot (X_{max}^I - X_{min}^I), \quad i = 1, 2, \ldots, NP$

end for

while ($MAE_{min} > \varepsilon$ or $I <$ MAXI) **do**

for $i = 1$: NP do

randomly select where $R_1 \neq R_2 \neq R_3 \neq i$

$V_i^{I+1} = X_{R1}^I + F \cdot (X_{R2}^I - X_{R3}^I)$

if $rand(0, 1) \leq CR$ then

$U_i^{I+1} \leftarrow V_i^{I+1}$

else

$U_i^{I+1} \leftarrow X_i^I$

end if

import the actual (Y_j) and predicted $PM_{2.5}$ concentration obtained from ML model 1 and 2 ($\hat{Y}_j^{M_1}$, $\hat{Y}_j^{M_2}$), respectively where $j = 1, 2\ldots, N$

for $j = 1$: N do

$F_i^M(j) = U_i^{I+1}\hat{Y}_j^{M_1} + (1 - U_i^{I+1})\hat{Y}_j^{M_2}$

$F_i(j) = X_i^I\hat{Y}_j^{M_1} + (1 - X_i^I)\hat{Y}_j^{M_2}$

$E^M(j) = |F_i^M(j) - Y(j)|$

$E(j) = |F_i(j) - Y(j)|$

$SUM(E_m) = SUM(E_m) + E^M(j); \quad SUM(E) = SUM(E) + E(j)$

end for

$MAE(m) = SUM(E_m)/N; \quad MAE = SUM(E)/N$

if $MAE(m) < MAE$ then

$X_i^{I+1} \leftarrow U_i^{I+1}; \quad MAE(i) \leftarrow MAE(m)$

else

$X_i^{I+1} \leftarrow X_i^I; \quad MAE(i) \leftarrow MAE$

end if

end for

$MAE_{min} \leftarrow min(MAE); \quad X_i^{best} \leftarrow X_i^I; \quad I = I + 1$

end while

Algorithm 1 begins by inputting several parameters, including the population size (NP), crossover rate (CR), scaling factor (F), and constant term (ε). Next, we create a population by randomly selecting values from a uniform distribution between 0 and 1. The range of values is determined by the search space's minimum and maximum values for the decision parameter. The mutation process selects three individuals (X_{R1}, X_{R2}, X_{R3}) from the population set of NP elements, which $R_1 \neq R_2 \neq R_3 \neq i$. The scaling factor, F, is a user-defined constant with the value $F \in [0, 1]$. The mutant vectors $(\mathbf{V}_i)$ are then obtained by applying all values. The trial vectors $(\mathbf{U}_i)$ are created by combining the parameters of the target vectors $(\mathbf{X}_i)$ with the mutant vectors using a predetermined crossing probability (CR). The selection technique is used in the DE algorithm for the next phase. The predicted $PM_{2.5}$ of the test set is imported to determine the fitness of the target and trial vectors. The fitness of the sample vectors is compared to the corresponding target vectors to determine the optimal solution. Next, we compute the forecasted $PM_{2.5}$ values (F_i) for each machine learning model. After that, the differences between the observed and predicted $PM_{2.5}$ are computed. The mean absolute error (MAE) determines the optimum weight. The optimization procedure ends when the MAE falls below a specific constant term (ε) or the iteration count reaches its limit. Finally, the best weight is determined and used in the hybrid forecasting model to get the predicted values. The structure of the hybrid forecasting model is presented in Figure 6.

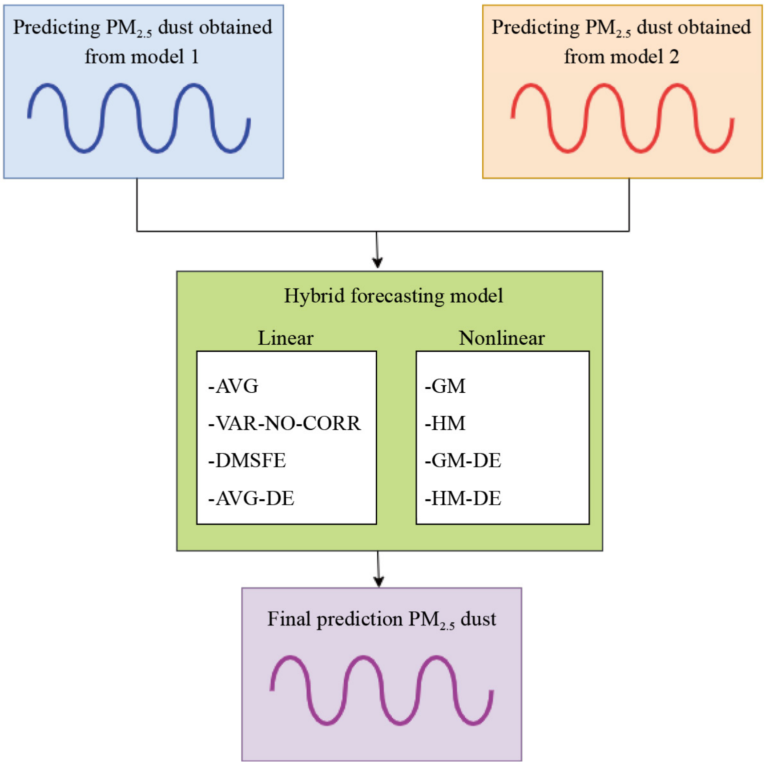


Figure 6. The structure of the hybrid forecasting model using several weight estimations

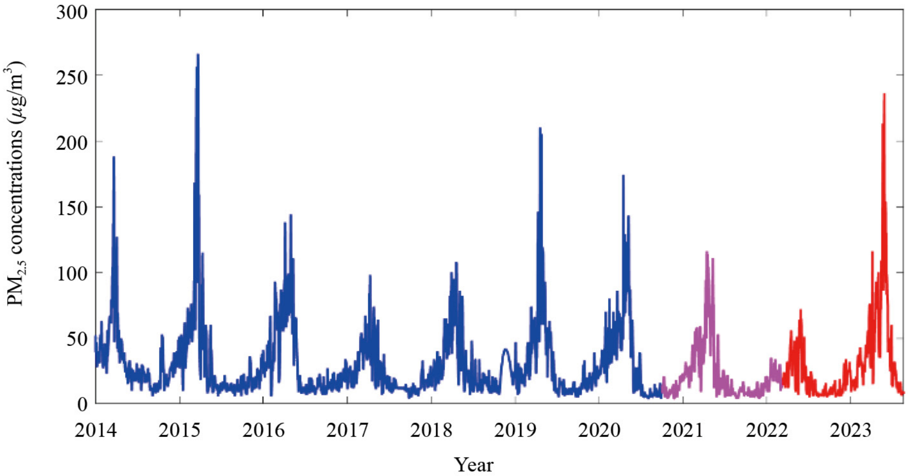
3. Data

The primary data utilized in this paper consists of the mean daily $PM_{2.5}$ concentration from January 2014 to June 2023. This dataset comprises a total of 3,468 data points collected from three monitoring stations located in Thailand. The primary data are from the Pollution Control Department, Ministry of Natural Resources and Environment, Thailand [31]. Next, we use a spline interpolation method to complete any missing observations. The data are then divided into three separate sets: training, validation, and test. The initial 70% of the data, which serves as the training set from January 2014 to August 2020, comprises 2,428 observations. Subsequently, 15% of the data collected between September 2020 and January 2022 is employed to verify the network's performance, resulting in 520 observations. The remaining 520 observations (15%) are utilized as the test set. Figure 7 illustrates the patterns of $PM_{2.5}$ particulate pollution at three stations in Thailand, which are separated by training, validation, and test sets.

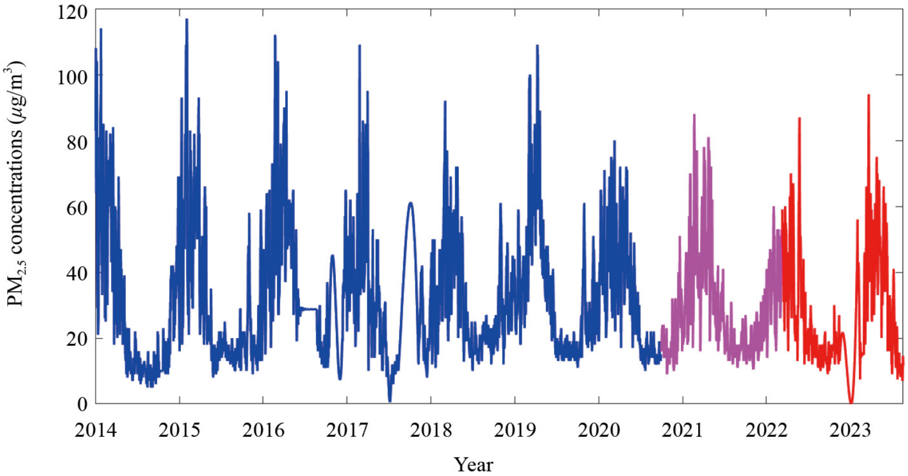
According to Figure 7, the blue, pink, and red lines denote the original $PM_{2.5}$ particles, as well as the training, validation, and test sets. It is obvious that $PM_{2.5}$ concentration for all three stations is non-stationary as they contain seasonal effects. The augmented Dickey-Fuller test is applied to confirm that all observed $PM_{2.5}$ pollution is non-stationary, as shown in Table 1.

Table 1. Stationarity test (Augmented Dickey-Fuller test)

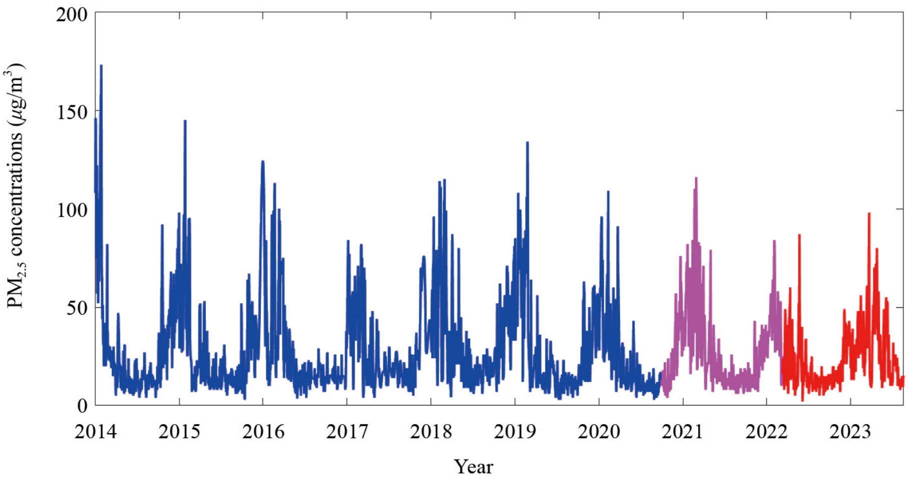
Station	Province	ADF Statistic	P-value
36T	Chiang Mai	-2.3745	0.1491
46T	Khon Kaen	-1.9210	0.3222
27T	Samut Sakhon	-2.1579	0.2299



(a) Station 36T



(b) Station 46T



(c) Station 27T

Figure 7. The pattern of $PM_{2.5}$ pollution in Thailand: (a) Station 36T, (b) Station 46T, and (c) Station 27T

According to the stationarity test results, the p-value exceeds the significance level ($p > 0.05$). This suggests that the initial $PM_{2.5}$ from the three locations is non-stationary. Table 2 displays the descriptive statistics for the daily average of $PM_{2.5}$ dust pollution.

Table 2. Descriptive statistics of the daily average $PM_{2.5}$ concentration ($\mu g/m^3$)

Station	Province	n	Missing	Range	Mean	Median	S.D.
36T	Chiang Mai	3,468	151	262	29.35	20	27.35
46T	Khon Kaen	3,468	436	116	25.54	16	18.88
27T	Samut Sakhon	3,468	388	171	26.95	19	22.11

From Table 2, the maximum range of $PM_{2.5}$ dust falls in the Northern part of Thailand (Chiang Mai). The daily average $PM_{2.5}$ concentration for all three provinces is between 25.54 and $29.35\mu g/m^3$. The lowest descriptive statistics are found in Khon Kaen where the values of range, mean, median, and standard deviation are 116, 25.54, 16, and 18.88 $\mu g/m^3$, respectively. Before applying our data to ML procedures, the MinMaxScaler method is utilized to transform data between 0 and 1. The performance measure used in this work is scale-dependent because we compare numerous methods with the same dataset [32]. Consequently, the performance metrics used in this work are the mean absolute error (MAE) and the median absolute error (MdAE). The explicit formulas are as follows:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|e_t|, \tag{9}$$

$$MdAE = \text{median}(|e_t|), \tag{10}$$

where e_t is an error at time t, then $e_t = y_t - \hat{y}_t$, y_t is the observed $PM_{2.5}$ dust, $\hat{y}_t$ are the predicted $PM_{2.5}$ concentration obtained from ML models, and n is the number of observations.

The percentage improvement (PIM) is another criterion used as an evaluation indicator for comparing the predicted results of hybrid and individual models. The PIM is formulated as[6]:

$$PIM = \frac{MAE_{best} - MAE_c}{MAE_{best}} \times 100\%, \tag{11}$$

where MAE_{best} is the best single model in terms of MAE and MAE_c is the MAE obtained from the combination method.

4. Experimental results

All experiments are carried out using Python programming on the Google Colaboratory (Google Colab) platform. The input data are univariate $PM_{2.5}$ time series with a time lag of 7 [33]. Following the completion of the data cleaning process, the Sherpa algorithm is employed to search for the most optimal hyperparameters. Determining the parameter search space is essential before implementing the Sherpa method in our model. The following parameters are established for the entire experiment: stochastic gradient descent (SGD) is the optimizer, and mean square error (MSE) is the loss function. We penalized weight parameters with a coefficient of 0.1 for L2 regularization to solve the overfitting problem. The parameter search space activation functions for ML algorithms are sigmoid, tanh, relu, and softmax. The learning rates vary from 0.001 to 0.1, and the batch sizes are 32, 64, 128, 256, and 512. The optimal hyperparameters for all ML procedures are presented in Table 3.

Table 3. The final hyperparameters of three ML procedures

Model	Parameter	Station 36T	Station 46T	Station 27T
ANN	Hidden unit 1	94	149	248
	Hidden unit 2	213	195	120
	Learning rate	0.039859	0.009217	0.001181
	Activation function	relu	relu	relu
LSTM	LSTM(layer 1)	121	122	119
	Learning rate	0.002608	0.006018	0.003561
	Activation function	relu	relu	tanh
CNN	Conv1D	10	128	10
	Kernel size	2	2	2
	Hidden unit 1	74	78	59
	Learning rate	0.042326	0.002597	0.008710
	Activation function	relu	relu	tanh

Next, we evaluate the performance metrics on both the training and validation sets. Table 4 displays the mean absolute error (MAE) and median absolute error (MdAE) of three machine learning (ML) models.

Table 4. Comparison of the performance metrics between training and validation sets

Model	Station	MAE		MdAE	
		Training	Validation	Training	Validation
ANN	36T	5.8679	3.9816	3.2301	2.3602
	46T	5.7793	5.1602	3.5631	3.4331
	27T	7.3059	7.0381	4.5875	4.8421
LSTM	36T	9.8989	8.3935	7.0118	7.3218
	46T	7.5668	6.8929	5.1403	5.2875
	27T	9.6160	8.7782	6.4260	5.6637
CNN	36T	6.9515	4.6911	3.8021	2.7066
	46T	7.4774	6.7798	4.8757	4.8739
	27T	8.4651	8.1693	5.0947	5.0147

Table 5. Comparisons of the performance metrics between single and hybrid techniques

Model	Station 36T		Station 46T		Station 27T	
	MAE	MdAE	MAE	MdAE	MAE	MdAE
ANN	6.0603	3.2957	5.1848	3.0824	6.3307	4.5640
LSTM	11.0429	8.4065	6.8436	4.8548	8.0350	6.0305
CNN	6.7029	3.5524	6.6781	4.6445	7.2285	4.6008
ANN+LSTM						
•AVG	7.7202	4.9116	5.8252	4.0313	6.9910	5.1930
•VAR-NO-CORR	9.1423	6.3429	5.9531	4.0987	7.2402	5.5152
•DMSFE	6.1517	3.2258	5.5030	3.5194	6.7197	4.8015
•AVG-DE	6.1208	3.2151	**5.0229**	**2.5315**	6.2535	4.2003
• GM	7.5925	4.4898	5.7959	3.9335	6.9691	5.1430
•HM	7.5145	4.2987	5.7695	3.8717	6.9482	5.0270
• GM-DE	6.0556	3.1536	5.0330	2.7086	6.2356	4.1378
•HM-DE	**6.0399**	**3.0811**	5.0358	2.7856	**6.2302**	**4.0777**
ANN+CNN						
•AVG	6.0709	3.1150	5.7402	3.7828	6.6780	4.5841
•VAR-NO-CORR	6.0875	3.0576	5.8043	3.8552	6.7236	4.5484
•DMSFE	6.0896	3.0622	5.5096	3.5238	6.7392	4.5676
•AVG-DE	6.0500	3.0999	5.0490	2.6347	6.2456	4.4281
• GM	6.0615	3.0334	5.7007	3.7260	6.6709	4.5847
•HM	6.0635	3.1216	5.6669	3.6767	6.6640	4.5751
• GM-DE	6.0447	3.0842	5.0527	2.7493	6.3145	4.4435
•HM-DE	6.0499	3.0788	5.0529	2.8073	6.4313	4.5278
LSTM+CNN						
•AVG	8.1635	5.1037	6.2473	4.6672	7.5100	5.3932
•VAR-NO-CORR	9.2506	6.2715	6.7482	4.6491	7.5823	5.4940
•DMSFE	6.7000	3.4635	6.7364	4.6307	7.3617	5.0382
•AVG-DE	6.5957	3.5980	6.7289	4.6055	7.2183	4.3649
• GM	8.0208	4.6280	6.7392	4.6659	7.4874	5.3646
•HM	7.9104	4.1777	6.7362	4.6646	7.4656	5.3481
• GM-DE	6.5720	3.5977	6.7259	4.5937	7.2089	4.3351
•HM-DE	6.5592	3.5376	6.7230	4.5883	7.2082	4.3047
PIM (%)	**0.34%**	**6.51%**	**3.12%**	**17.87%**	**1.59%**	**10.66%**

According to Table 4, the MAE and MdAE derived from the training and validation sets are nearly identical, providing additional evidence that there is no overfitting issue. This suggests that the performance of each ML model was satisfactory on both the training and validation sets and that it can be employed to forecast $PM_{2.5}$ particulate pollution. Next, we integrate two machine learning techniques with different weight methods for linear and nonlinear models. The DE algorithm estimates weight in hybrid approaches and compares it with the state-of-the-art weight estimation method. The resulting $PM_{2.5}$ pollution predictions are then computed using the weight approximations and compared to the individual forecast for each technique. Table 5 presents the performance metrics for the test set acquired from individual and hybrid techniques.

Table 5 presents the MAE and MdAE of each machine learning model and hybrid forecasting model for both linear and nonlinear hybrid processes. These were calculated by applying the state-of-the-art and DE weights optimization techniques. Regarding performance metrics, DE weight optimization yields superior results to the most recent weight estimation methods for linear and nonlinear hybrid forecasting techniques. Focusing on the MAE values, the percentage improvement of the best hybrid procedure using DE weight optimization over the best single model at stations 36T, 46T, and 27T are 0.34%, 3.12%, and 1.59%, respectively. For MdAE values, the percentage improvements for all three stations are 6.51%, 17.87%, and 10.66%, respectively. It was seen in Table 5, that the DE weight estimation of hybrid approaches improves the accuracy of the individual and the state-of-the-art weight of hybrid techniques. Based on the results, Table 6 displays the prediction model on the test set derived using the appropriate hybrid models of three stations.

Table 6. The suitable prediction model for each station

Station	Method	Hybrid Model
36T	HM-DE	$\hat{y}_{ct} = \frac{\hat{y}_{1t}\hat{y}_{2t}}{1.1143\hat{y}_{1t} - 0.1143\hat{y}_{2t}}$
46T	AVG-DE	$\hat{y}_{ct} = 1.236\hat{y}_{1t} - 0.236\hat{y}_{2t}$
27T	HM-DE	$\hat{y}_{ct} = \frac{\hat{y}_{1t}\hat{y}_{2t}}{1.6119\hat{y}_{1t} - 0.6119\hat{y}_{2t}}$

Here, $\hat{y}_{1t}$ and $\hat{y}_{2t}$ are the predicted $PM_{2.5}$ pollution obtained from the ANN and the LSTM models, respectively.

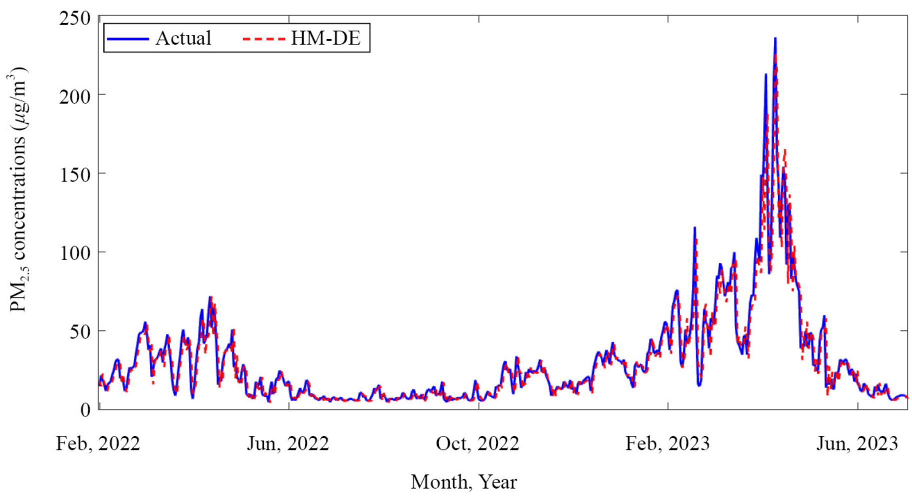


Figure 8. Predicted values of $PM_{2.5}$ pollution obtained from the HM-DE model and its actual values at station 36T

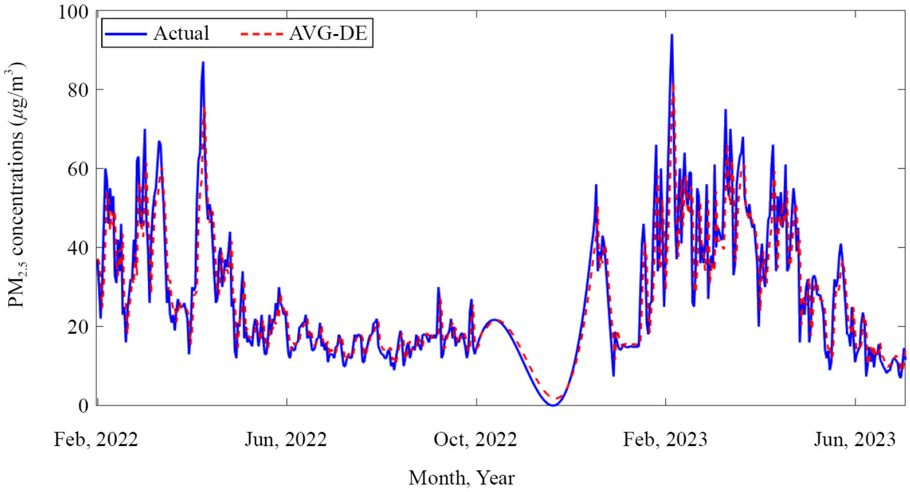


Figure 9. Predicted values of $PM_{2.5}$ pollution obtained from the AVG-DE model and its actual values at station 46T

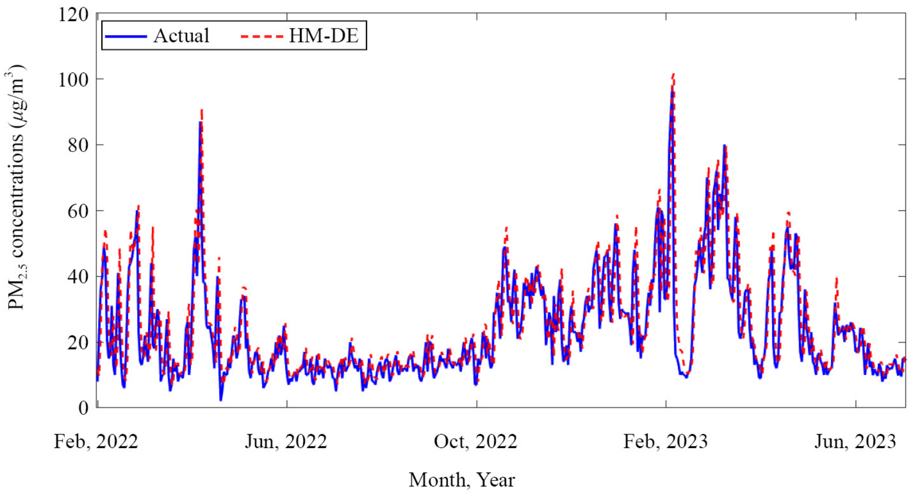


Figure 10. Predicted values of $PM_{2.5}$ pollution obtained from the HM-DE model and its actual values at station 27T

A comparison of the predicted $PM_{2.5}$ on the test set obtained from the suitable hybrid models and its observed $PM_{2.5}$ between February 2022 and June 2023 are illustrated in Figures 8 to 10.

In each figure, the blue lines represent the original $PM_{2.5}$, while the red dashed lines display the predicted $PM_{2.5}$ that was obtained from the most effective hybrid models on the test set between February 2022 and June 2023. According to the appropriate hybrid models, the predicted $PM_{2.5}$ concentration behaves similarly to the actual $PM_{2.5}$.

5. Conclusion and discussion

This work aims to enhance the weight optimization of a hybrid machine learning model on the daily average of $PM_{2.5}$ concentration in Thailand. Three stations, including 36T (Chiang Mai), 46T (Kon Kaen), and 27T (Samut Sakhon), located in the Northern, Northeastern, and Central parts of Thailand, are used in the case study of this work. We employ the ANN, the LSTM, and the CNN, which are all well-known machine learning models. The Sherpa algorithm is applied to determine the hyperparameters for each machine learning approach. After receiving the predicted $PM_{2.5}$ from each machine learning technique, the hybrid methods, including the linear and nonlinear models, are presented. The eight hybrid procedures include the simple average method (AVG), the variance no covariance method (VAR-NO-CORR), the discount mean square forecast error method (DMSFE), the simple average method with differential evolution method (AVG-DE), the geometric mean method (GM), the harmonic mean method (HM), geometric mean with differential evolution method (GM-DE), and the harmonic mean with differential evolution method (HM-DE) are investigated. The mean absolute error (MAE) and the median absolute error (MdAE) are the performance metrics employed to assess hybrid procedures. The experimental results confirm that the differential evolution weight optimization is superior to the others for linear and nonlinear hybrid techniques between ANN and LSTM models in most stations. The harmonic mean using the differential evolution approach is appropriate for stations 36T and 27T. In addition, the simple average with differential evolution method is considered adequate for predicting $PM_{2.5}$ pollution at station 46T.

According to the results presented in Table 5, the performance metrics at station 36T (Chiang Mai) are slightly greater than the other stations due to the high levels of $PM_{2.5}$ concentration. As a result, it experiences fluctuations, especially during the summer months. The data at stations 46T and 27T, which contain 436 and 388 missing values, respectively, illustrate the impact of missing values on errors. Focusing on the percentage improvement, there is no significant difference between the best hybrid model and the individual procedure at station 36T. This is due to insufficient samples, which may not be conducive to learning. Nevertheless, the percentage improvement of the remaining stations offers a notable enhancement compared to the individual techniques. The results of our study supported the findings of AH Mohamed [22] and Hu et al. [23], indicating that hybrid procedures are more accurate than the individual forecast method. The benefit of this work is that complex ML models are not required. Under our proposed technique, using only a simple ML method on parallel hybrid techniques can reduce the performance metrics. This study can be expanded for future work by investigating alternative imputation approaches to replace the incomplete data. In addition, other widely used machine learning models, such as the bidirectional long short-term memory (BiLSTM) and the support vector machine (SVM), can be utilized to predict the $PM_{2.5}$ concentration.

Conflict of interest

The authors declare no competing financial interest.

References

[1] Mohammed MO, Song WW, Ma WL, Li WL, Ambuchi JJ, Thabit M, et al. Trends in indoor-outdoor $PM_{2.5}$ research: A systematic review of studies conducted during the last decade (2003-2013). *Atmospheric Pollution Research*. 2015; 6(5): 893-903. Available from: https://doi.org/10.5094/APR.2015.099.

[2] Lung SCC, Hien TT, Cambaliza MOL, Hlaing OMT, Oanh NTK, Latif MT, et al. Research priorities of applying low-cost $PM_{2.5}$ sensors in Southeast Asian countries. *International Journal of Environmental Research and Public Health*. 2022; 19(3): 1522. Available from: https://doi.org/10.3390/ijerph19031522.

[3] Thangavel P, Park D, Lee YC. Recent insights into particulate matter ($PM_{2.5}$)-mediated toxicity in humans: An overview. *International Journal of Environmental Research and Public Health*. 2022; 19(12): 7511. Available from: https://doi.org/10.3390/ijerph19127511.

[4] Zhang L, Lin J, Qiu R, Hu X, Zhang H, Chen Q, et al. Trend analysis and forecast of $PM_{2.5}$ in Fuzhou, China using the ARIMA model. *Ecological Indicators*. 2018; 95: 702-710. Available from: https://doi.org/10.1016/j.ecolind.2018.08.032.

[5] Bhatti UA, Yan Y, Zhou M, Ali S, Hussain A, Qingsong H, et al. Time series analysis and forecasting of air pollution particulate matter ($PM_{2.5}$): A SARIMA and factor analysis approach. *IEEE Access*. 2021; 9: 41019-41031. Available from: https://doi.org/10.1109/ACCESS.2021.3060744.

[6] Hansun S, Wicaksana A, Kristanda MB. Prediction of Jakarta City air quality index: Modified double exponential smoothing approaches. *International Journal of Innovative Computing, Information and Control*. 2021; 17(4): 1363-1371. Available from: https://doi.org/10.24507/ijicic.17.04.1363.

[7] Ratchagit M, Xu H. A two-delay combination model for stock price prediction. *Mathematics*. 2022; 10(19): 3447. Available from: https://doi.org/10.3390/math10193447.

[8] Bera B, Bhattacharjee S, Sengupta N, Saha S. $PM_{2.5}$ concentration prediction during COVID-19 lockdown over Kolkata metropolitan city, India using MLR and ANN models. *Environmental Challenges*. 2021; 4: 100155. Available from: https://doi.org/10.1016/j.envc.2021.100155.

[9] Feng R, Gao H, Luo K, Fan JR. Analysis and accurate prediction of ambient $PM_{2.5}$ in China using Multi-layer Perceptron. *Atmospheric Environment*. 2020; 232: 117534. Available from: https://doi.org/10.1016/j.atmosenv.2020.117534.

[10] Yang J, Yan R, Nong M, Liao J, Li F, Sun W. $PM_{2.5}$ concentration forecasting in Beijing through deep learning with different inputs, model structures and forecast time. *Atmospheric Pollution Research*. 2021; 12(9): 101168. Available from: https://doi.org/10.1016/j.apr.2021.101168.

[11] Zhang M, Wu D, Xue R. Hourly prediction of $PM_{2.5}$ concentration in Beijing based on Bi-LSTM neural network. *Multimedia Tools and Applications*. 2021; 80: 24455-24468. Available from: https://doi.org/10.1007/s11042-021-10852-w.

[12] Danesh Yazdi M, Kuang Z, Dimakopoulou K, Barratt B, Suel E, Amini H, et al. Predicting fine particulate matter ($PM_{2.5}$) in the greater London area: An ensemble approach using machine learning methods. *Remote Sensing*. 2020; 12(6): 914. Available from: https://doi.org/10.3390/rs12060914.

[13] Amnuaylojaroen T. Prediction of $PM_{2.5}$ in an urban area of northern Thailand using multivariate linear regression model. *Advances in Meteorology*. 2022; 2022: 1-9. Available from: https://doi.org/10.1155/2022/3190484.

[14] Wongrin W, Chaisee K, Suphawan K. Comparison of statistical and deep learning methods for forecasting $PM_{2.5}$ concentration in northern Thailand. *Polish Journal of Environmental Studies*. 2023; 32(2): 1419-1431. Available from: https://doi.org/10.15244/pjoes/157072.

[15] Saiohai J, Bualert S, Thongyen T, Duangmal K, Choomanee P, Szymanski WW. Statistical $PM_{2.5}$ prediction in an urban area using vertical meteorological factors. *Atmosphere*. 2023; 14(3): 589. Available from: https://doi.org/10.3390/atmos14030589.

[16] Wanishsakpong W, Atsawachanakan S, Panityakul T. Application of deep belief network in weather modeling: $PM_{2.5}$ concentration in Thailand. *International Journal of Analysis and Applications*. 2024; 22: 30. Available from: https://doi.org/10.28924/2291-8639-22-2024-30.

[17] Gunasekar S, Kumar G, Agbulu GP. Air quality predictions in urban areas using hybrid ARIMA and metaheuristic LSTM. *Computing and Systems in Science and Engineering*. 2022; 43(3): 101761. Available from: https://doi.org/10.1016/j.apr.2023.101761.

[18] Shahriar SA, Kayes I, Hasan K, Hasan M, Islam R, Awang NR, et al. Potential of ARIMA-ANN, ARIMA-SVM, DT and CatBoost for atmospheric $PM_{2.5}$ forecasting in Bangladesh. *Atmosphere*. 2021; 12(1): 100. Available from: https://doi.org/10.3390/atmos12010100.

[19] Seabe PL, Moutsinga CRB, Pindza E. Forecasting cryptocurrency prices using LSTM, GRU, and bi-directional LSTM: A deep learning approach. *Fractals*. 2023; 7(2): 203. Available from: https://doi.org/10.3390/fractalfract7020203.

[20] Bates JM, Granger CW. The combination of forecasts. *Journal of the Operational Research Society*. 1969; 20: 451-468. Available from: https://doi.org/10.1057/jors.1969.103.

[21] Petropoulos F, Svetunkov I. A simple combination of univariate models. *International Journal of Forecasting*. 2020; 36(1): 110-115. Available from: https://doi.org/10.1016/j.ijforecast.2019.01.006.

[22] Mohamed RA. Enhancing forecast accuracy using combination methods for the hierarchical time series approach. *PLoS One*. 2023; 18(7): e0287897. Available from: https://doi.org/10.1371/journal.pone.0287897.

[23] Hu YC, Wu G, Jiang P. Tourism demand forecasting using nonadditive forecast combinations. *Journal of Hospitality and Tourism Research*. 2023; 47(5): 775-799. Available from: https://doi.org/10.1177/1096348021104785.

[24] Li L, Han C, Yao S, Ning L. Variable weights combination MIDAS model based on ELM for natural gas price forecasting. *IEEE Access*. 2022; 10: 52075-52093. Available from: https://doi.org/10.1109/ACCESS.2022.3174814.

[25] Sarker IH. Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*. 2021; 2(3): 160. Available from: https://doi.org/10.1007/s42979-021-00592-x.

[26] Jamous R, ALRahhal H, El-Darieby M. A new ANN-particle swarm optimization with center of gravity (ANN-PSOCOG) prediction model for the stock market under the effect of COVID-19. *Scientific Programming*. 2021; 2021(1): 1-17. Available from: https://doi.org/10.1155/2021/6656150.

[27] Kianfar N, Mesgari MS, Mollalo A, Kaveh M. Spatio-temporal modeling of COVID-19 prevalence and mortality using artificial neural network algorithms. *Spatial and Spatio-Temporal Epidemiology*. 2022; 40: 100471. Available from: https://doi.org/10.1016/j.sste.2021.100471.

[28] Kaur K, Kaur P. Improving BERT model for requirements classification by bidirectional LSTM-CNN deep model. *Computers and Electrical Engineering*. 2023; 108: 108699. Available from: https://doi.org/10.1016/j.compeleceng.2023.108699.

[29] Cang S. A comparative analysis of three types of tourism demand forecasting models: Individual, linear combination and nonlinear combination. *International Journal of Tourism Research*. 2014; 16(6): 596-607. Available from: https://doi.org/10.1002/jtr.1953.

[30] Storn R, Price K. Differential evolution-a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*. 1997; 11: 341-359. Available from: https://doi.org/10.1023/A:1008202821328.

[31] Pollution Control Department. *Daily Average $PM_{2.5}$*. Pollution Control Department; 2024. Available from: http://air4thai.pcd.go.th/webV3/#/History.

[32] Theodosiou M. Forecasting monthly and quarterly time series using STL decomposition. *International Journal of Forecasting*. 2011; 27(4): 1178-1195. Available from: https://doi.org/10.1016/j.ijforecast.2010.11.002.

[33] Zhao Y, Ma Y, Liu Z, Qin P, Wang W, Jiao H, et al. Evaluation of temperature modifying effect of $PM_{2.5}$ on respiratory mortality in high latitude monsoon area. *Theoretical and Applied Climatology*. 2024; 155: 6167-6175. Available from: https://doi.org/10.1007/s00704-024-05002-9.