IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 21, 2020, accepted November 2, 2020, date of publication November 18, 2020, date of current version December 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3038853

Dynamic System Approach for Improved $PM_{2.5}$ Prediction in Taiwan

LARRY LIN[1,2], CHIH-YUAN CHEN[3], HSIEN-YUEH YANG[1,2], ZHEZHUANG XU[4], (Member, IEEE), AND SHIH-HAU FANG[1,2], (Senior Member, IEEE)

[1]Department of Electrical Engineering, Yuan Ze University, Taoyuan 32003, Taiwan
[2]MOST Joint Research Center for AI Technology and All Vista Healthcare, Taipei 10617, Taiwan
[3]Department of Geography, Chinese Culture University, Taipei 111, Taiwan
[4]School of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108, China

Corresponding author: Shih-Hau Fang (shfang@saturn.yzu.edu.tw)

This work was supported by the Ministry of Science and Technology, Taiwan, under Grant MOST 109-2634-F-155-001. This work is also partially supported by NSF of China (61973085, 61673116).

ABSTRACT This study presents two dynamic models, namely recurrent neural network and long short-term memory (LSTM) models, for predicting $PM_{2.5}$ concentrations in Taiwan by using $PM_{2.5}$ time series obtained at air quality monitoring stations and weather information obtained at neighboring weather stations. The proposed models can efficiently predict $PM_{2.5}$ by incorporating a learned memory structure with a forgetting gate. To evaluate the predictive performance of the proposed models, large-scale databases established by Taiwan's Environmental Protection Administration, and Central Weather Bureau were used; these databases include hourly data from 77 air quality monitoring stations and 580 weather stations over a 1-year period. The results demonstrated that the proposed models outperformed three traditional machine learning methods (gradient boosting, support vector machine, and classification and regression tree models) by 27.12% and 33.69% on average in terms of the coefficient of determination and root mean square error, respectively. A geographical divergence analysis was conducted to compare predictive performance in different regions. The results revealed that the most significant improvement in predictive performance was achieved in central Taiwan. The seasonal and pollution effect on predictive performance were reduced by the LSTM and the source distribution of $PM_{2.5}$ emission in Taiwan was also analyzed.

INDEX TERMS Dynamic system, air quality, time series prediction, geographical monitoring stations.

I. INTRODUCTION

Air pollution is a major environmental concern and continues to pose a serious threat to health worldwide [1]. Long-term exposure to fine particulate matter, especially the particle with a diameter not larger than 2.5 μm ($PM_{2.5}$), has been linked to an increasing range of adverse health effects such as stroke, ischemic heart disease, lung cancer, respiratory infection, and chronic obstructive pulmonary disease [2], [3]. $PM_{2.5}$ has drawn considerable attention from citizens, scientists, and governments since it is a crucial indicator for air pollution early warning system [4]–[6]. As a result, the ability of predicting PM2.5 is considered as a critical part to help governments take countermeasures and improve environmental management [5], [7]. There are two major paradigms of methodology related to $PM_{2.5}$ prediction, deterministic and statistical methods. The statistical methods have shown better performance because the nonlinear and heterogeneous nature of processes in the formation and transportation of air pollution [8]. With the advance of technology and decreasing cost of sensors for collecting air quality data, data mining and machining learning methods become more and more important, such as time series analysis [9], [10], random forest [11], [12], principal component analysis [13], Kalman filters [14], support vector machines (SVMs) [15], [16], and artificial neural networks (ANNs) [5], [7], [17].

$PM_{2.5}$ predictions are challenging because the formation and transportation of $PM_{2.5}$ is strongly influenced by spatial and temporal variations at both micro- and macro-scales [18], [19]. These limitations of time and space also results in variations of predictive performance when applying different models in different countries [20]–[24]. For example, a study applied a gradient boosting model (GBM) to predict $PM_{2.5}$ in Taiwan; however, the model did not provide satisfactory

The associate editor coordinating the review of this manuscript and approving it for publication was Vlad Diaconita.

predictive performance in central and southern Taiwan because of the effect of industrial emissions in these regions [25]. Moreover, the trends of $PM_{2.5}$ time series are highly related to seasonal variations. Although the model considers features such as date attributes, it requires improvement to effectively capture the sequential pattern of measurements through a dynamic system.

To overcome the spatio-temporal heterogeneity of $PM_{2.5}$ predictions, several types of approaches have been proposed, which can be divided into two main streams: dynamic models and geographical divergence analysis. Dynamic models are designed to handle the sequential inputs, which is capable to capture the spatio-temporal evolution of $PM_{2.5}$ time series data [26]. Recurrent neural network (RNN) and long-short term memory (LSTM) model are the two major models which have been widely used in time series data prediction including $PM_{2.5}$ [27]–[29]. On the other hand, geographical divergence analysis focuses on identifying the spatial patterns, which can predict or explore the spatio-temporal heterogeneity of geographical events. Several geostatistical and machine learning models such as geographically weighted regression (GWR), mixed-effect model, geographically weighted gradient boosting machine (GW-GBM), and clustering analysis have been adopted to predict $PM_{2.5}$ concentrations [18], [30], [31].

This study focused on $PM_{2.5}$ concentration prediction in Taiwan along with a geographical divergence analysis. Traffic emissions and local meteorological conditions are major factors associated with the daily variability of $PM_{2.5}$ in the urban areas of western Taiwan [32]–[35]. The seasonal variability of $PM_{2.5}$ in Taiwan is strongly influenced by the southwesterly and northeasterly monsoonal flow [36]–[38]. Moreover, Taiwan is surrounded by the Pacific Ocean and its topography is dominated by the Central Mountain Range that runs from north to south. The complex terrain and various meteorological conditions engender large spatial and temporal variations in $PM_{2.5}$ concentration. The Environmental Protection Administration (EPA) of Taiwan established air quality monitoring stations in seven regions to record air quality.

The recent study integrated air quality data provided by the EPA and meteorological data provided by the Central Weather Bureau (CWB) to predict $PM_{2.5}$ [25]. Accordingly, to improve predictive performance [25], the present study presents two dynamic models, namely recurrent neural network (RNN) and long-short term memory (LSTM) models that can provide long-term memory of the correlations among time series items. An RNN is an artificial neural network that connects a current input and the state of the previous time step and subsequently outputs a prediction [39]. Because an RNN has memory of only the state of the previous time step, the proposed LSTM system with a forgetting gate was developed to prevent the RNN from storing only short-term states and forgetting long-term states [40]. To demonstrate the effectiveness of the proposed models, this study applied identical experimental data to those used in [25] to test the predictive performance of the proposed models; additionally, three traditional models (GBM, classification and regression tree model [CART], and support vector machine model [SVM]) were used for comparison.

The results revealed that the performance of the LSTM model was superior to that of the RNN model. Specifically, the average coefficient of determination (R^2), root mean square error (RMSE), normalized RMSE (NRMSE), and mean absolute percentage error (MAPE) of the LSTM model were 0.86, 4.46, 0.24, and 0.30, respectively, which were superior to those of the RNN model by 53.94%, 40.94%, 42.29%, and 49.29%, respectively.

The contributions of this study are summarized as follows:

1) We employed two dynamic models, RNN and LSTM models that can provide long-term memory of the correlations among time series items. Besides, three traditional machine learning models, CART, GBM, and SVM were applied for the predictive performance comparison. The predictive performance of the LSTM model was considerably superior to that of the RNN model and all three traditional machine learning models.

2) This study provides a geographical divergence analysis that was conducted to compare predictive performance in different seven air monitoring regions. To our best knowledge, it is the first study investigating the seasonal variation and domestic emission in the task of air quality prediction in Taiwan.

3) Although $PM_{2.5}$ is strongly influenced by seasons, the LSTM model can satisfactorily overcome the seasonal variation of $PM_{2.5}$ in Taiwan. Otherwise, compared with other regions in Taiwan, we determined that CT had the lowest predictive performance of traditional models resulting from highest domestic $PM_{2.5}$ emission. Nevertheless, the LSTM model clearly improved predictive performance in this region.

II. DATA COLLECTION

The EPA and CWB databases constitute the main sources of data used by researchers for air quality forecasting. The EPA database provides 264799 samples collected from 77 air stations for 2017. The data are typically collected in seven monitoring regions (Fig. 1): northern Taiwan (NT), the Chu-Miao (CM) area, central Taiwan (CT), the Yun-Chia-Nan (YCN) area, the Kao-Ping (KP) area, the Hua-Dong (HD) area, and Yilan (YI). The data include the following attributes: index, city, county, station name, date, detected items, and time in hours. Additionally, the following substances are monitored: $PM_{2.5}$, NO_2, PM_{10}, NO, NO_X, SO_2, CO, O_3, THC, NMHC, and CH_4. These substances are primarily released into the air through burning of fuel in vehicles and power plants.

To provide quality assurance, the EPA has strict guidelines for air quality monitoring. The EPA provides an annual report on quality assurance operations for the air quality monitoring system. The accuracy of the $PM_{2.5}$ data (2017) used in present study was 96.3%. The annual report for the period from 2001 to 2017 can be downloaded from the EPA website [41]. The air quality monitoring stations established by the

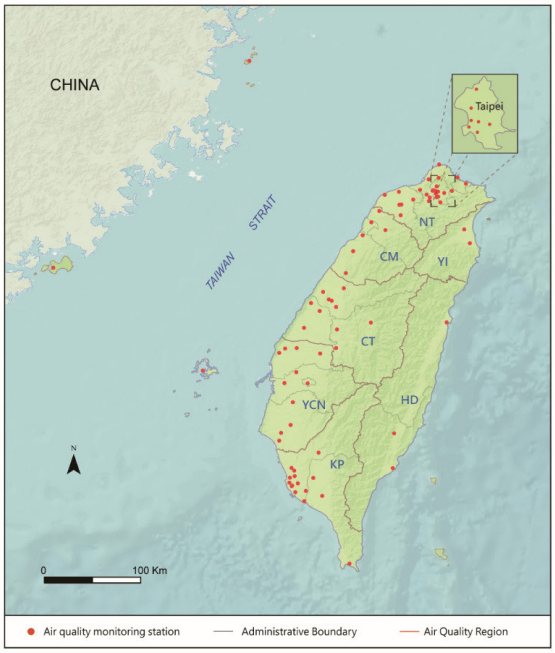

FIGURE 1. Air quality monitoring stations.

EPA can be divided into six categories. The first category comprises 60 general stations, most of which are located in urban and suburban areas. The second category contains five stations located close to industrial factories. The third category comprises five stations that are located close to main roads and have a relatively low sampling altitude. The fourth category contains four background stations (two of which simultaneously function as general stations) that are located away from pollution sources and provide baseline monitoring data. The fifth category comprises two national park stations (one of which serves as the general station). Finally, the sixth category contains two stations located in rural area.

Fig. 2 illustrates the monthly temporal variation of the mean $PM_{2.5}$ concentrations in the seven regions, where YI has missing data except for December. This figure indicates that the $PM_{2.5}$ concentrations reach a peak in most regions during spring (February, March, and April), drop dramatically during summer (May, June, and July), and start to go upwards during autumn (September, October, and November) and winter (November, December, and January).

The CWB has established 580 weather stations for monitoring and reporting weather conditions in Taiwan [42] (Fig. 3). The collected weather data includes the following attributes: longitude, latitude, station name, city, county, wind speed, wind direction, temperature, and pressure. The number and date of data sets from the CWB and EPA must be synchronized for model training. This study used fivefold cross validation to confirm the prediction results in the study [25]. The data were randomly split into five folds, four of which were used for training, with the remaining one used for testing.

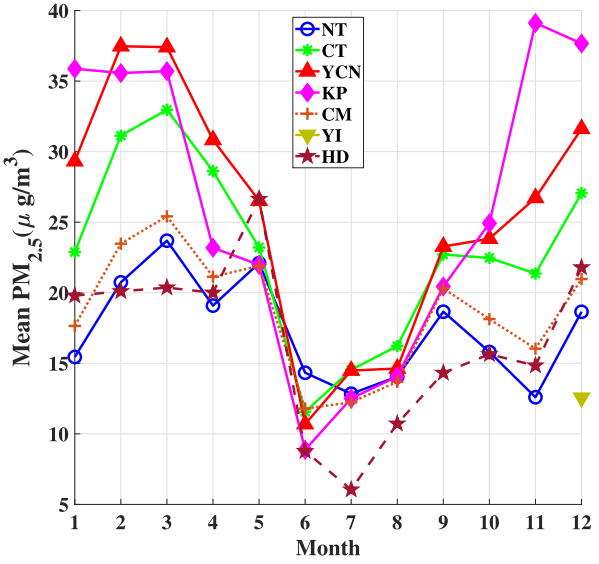

FIGURE 2. Monthly variations of the mean $PM_{2.5}$ concentrations in the seven air monitoring regions.

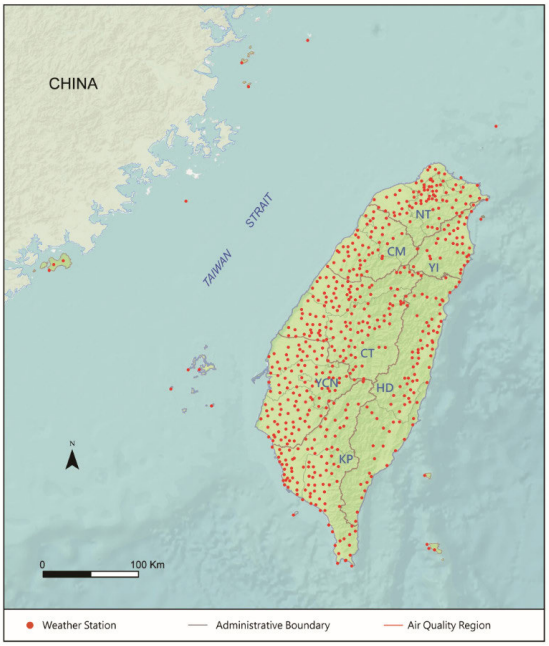

FIGURE 3. Weather stations.

III. METHODOLOGIES

We analyzed the predictive performance of the RNN and LSTM models. Next, we applied three traditional machine learning models, namely a CART model, GBM, and SVM model, for comparison.

A. TWO DYNAMIC MODELS

1) RNN MODEL

An RNN exhibits a temporal dynamic behavior in which an artificial neural network connects the current input and the state of the previous time step; the network then outputs a prediction. Furthermore, an RNN has an input layer, a hidden layer, and an output layer [39]. The input layer uses an input

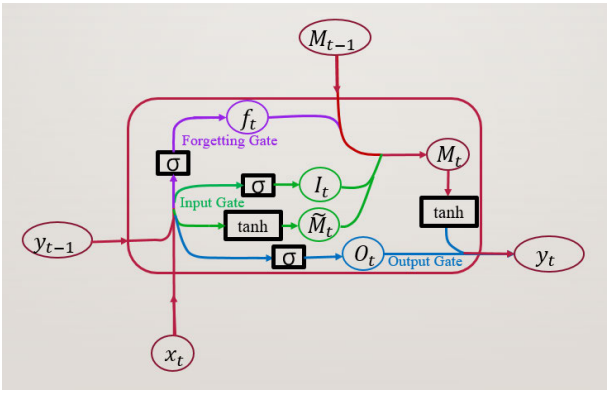

FIGURE 4. Flowchart of LSTM.

vector $\boldsymbol{x}$, and the output layer produces an output variable y. After an input value $\boldsymbol{x_t}$ and output value y_{t-1} are given, the hidden layer with the hidden state h_t applies a sigmoid function $f(z)$ to determine what percentage of $\boldsymbol{x_t}$ and y_{t-1} should be transmitted into an output value y_t.

$$h_t = V * x_t + U * h_{t-1}, \tag{1}$$
$$o_t = W * h_t, \tag{2}$$

where V, U, and W are weighting vectors and o_t is an intermediate result. Finally, the output value y_t is determined using the sigmoid function $f(z)$,

$$y_t = f(o_t). \tag{3}$$

In the RNN, the current output depends on the current input and previous output. The RNN remembers only short-term information (i.e., the state of the previous time step) and long-term information is forgotten with time. Accordingly, this study proposes the LSTM model to overcome this problem.

2) LSTM MODEL

In a $PM_{2.5}$ time series, neighboring samples are highly related to each other. Furthermore, $PM_{2.5}$ is strongly affected by meteorological factors for different seasons. Therefore, this study developed the LSTM model for prediction using $PM_{2.5}$ time series with long-term correlations. The LSTM model can prevent the RNN model from forgetting long-term states because the RNN model remembers only the state of the previous time step [40].

The LSTM model is a modified version of the RNN model. The LSTM model handles datasets comprising items with long-term correlations; by contrast, the RNN model cannot [43]. The LSTM architecture comprises an input layer, a hidden layer, and an output layer. Hence, the LSTM and RNN models have the same architecture, except for hidden layer. The input layer uses 81 features as the input vector $\boldsymbol{x}$ and the output layer produces an output variable y for $PM_{2.5}$ concentration prediction. The hidden layer's weighting vector and bias number are determined according to the data length.

Given an input value $\boldsymbol{x_t}$ and output value y_{t-1}, the hidden layer uses a forgetting gate, an input gate, and an output gate to determine what percentage of $\boldsymbol{x_t}$ and y_{t-1} should be transmitted into an output value y_t, as shown in Fig 4. First, the forgetting gate determines the percentage of $\boldsymbol{x_t}$ and y_{t-1} to be reserved by applying a sigmoid function σ, as expressed in Eq. (4):

$$f_t = \sigma(W_{fy} * y_{t-1} + W_{fx} * \boldsymbol{x_t} + b_f), \tag{4}$$

where W_{fy} and W_{fx} are the weighting vectors and b_f is the bias vector, respectively. The sigmoid function σ returns a value f_t in the range [0, 1]. Second, the input gate saves and modulates $\boldsymbol{x_t}$ and y_{t-1} by applying the sigmoid function and *tanh* activation function, respectively. The sigmoid function returns a percentage I_t, and a *tanh* function returns a temporary modulation parameter $\tilde{M}_t$, as expressed in Eq. (5) and Eq. (6):

$$I_t = \sigma(W_{Iy} * y_{t-1} + W_{Ix} * \boldsymbol{x_t} + b_I), \tag{5}$$
$$\tilde{M}_t = tanh(W_{My} * y_{t-1} + W_{Mx} * \boldsymbol{x_t} + b_M). \tag{6}$$

$\tilde{M}_t$ can add or subtract $\boldsymbol{x_t}$ and y_{t-1} because the *tanh* function returns a value in the range [−1, 1]. Subsequently, the modulation parameter M_t is determined as follows:

$$M_t = I_t * \tilde{M}_t + f_t * M_{t-1}. \tag{7}$$

M_{t-1} is a modulation parameter in a previous time stamp. Finally, the output gate evaluates a percentage O_t that is similar to f_t and I_t as follows:

$$O_t = \sigma(W_{Oy} * y_{t-1} + W_{Ox} * \boldsymbol{x_t} + b_O). \tag{8}$$

The output value y_t is then derived using O_t, M_t, and the *tanh* function:

$$y_t = O_t * tanh(M_t). \tag{9}$$

The procedure of the LSTM model is described in Algorithm 1.

B. THREE TRADITIONAL MACHINE LEARNING MODELS

1) CART

In the CART model proposed by Breiman and Friedman [44], a decision tree algorithm creates a series of decision rules for data classification. The rules are set according to training data attributes. Subsequently, the splitting nodes of the decision tree algorithm are determined. Maximum number of splitting note was set as 32 during model training. When a data attribute is numerical, the splitting nodes can be fitted using regression functions. Next, an optimized fitting function is applied to predict testing data.

2) GBM

The GBM combines fitting functions, loss functions, a decision tree, and gradient descent analysis [22]. The decision tree produces initial values for the fitting function through multiple regression and this process can handle numerous input variables, similar to those considered in the study. Errors between the observed datasets and output values are then calculated using a loss function. Frequently used loss functions

Algorithm 1 LSTM

Input:
0: $\boldsymbol{x_t}$: Input value at time t
0: y_{t-1} : Output value at time $t-1$
0: M_{t-1} : Modulation parameter at time $t-1$
0: T : Time length
Output:
0: y_t : Output value at time t
0: M_t : Modulation parameter at time t
0: **For** $t = 1$ to T
0: The forgetting gate:
0: $f_t = \sigma(W_{fy} * y_{t-1} + W_{fx} * \boldsymbol{x_t} + b_f)$
0: The input gate:
0: $I_t = \sigma(W_{Iy} * y_{t-1} + W_{Ix} * \boldsymbol{x_t} + b_I)$
0: $\tilde{M}_t = tanh(W_{My} * y_{t-1} + W_{Mx} * \boldsymbol{x_t} + b_M)$
0: $M_t = I_t * \tilde{M}_t + f_t * M_{t-1}$
0: The output gate:
0: $O_t = \sigma(W_{Oy} * y_{t-1} + W_{Ox} * \boldsymbol{x_t} + b_O)$
0: $y_t = O_t * tanh(M_t)$
0: **End**

include square error, absolute error, and negative binomial log-likelihood functions [45], [46]. Subsequently, gradient descent analysis is applied to determine the fitting function for which expected loss function value is minimized. The aforementioned procedure is repeated to derive an optimal fitting function. Maximum number of splitting note, learning rate, and epoch were set as 32, 0.0875, 500 during model training. The derived fitting function is thus applied to predict the $PM_{2.5}$ concentration.

3) SVM

An SVM is a supervised learning technique used for regression and classification [47]. An SVM model applies a kernel function that maps data points into a high-dimensional feature space. Frequently used kernel functions includes linear functions, nonlinear functions, polynomial functions, sigmoid functions, and radial basis functions (RBFs). The performance of SVM is associated with the selection of kernel functions. From polynomial functions, sigmoid functions, and RBFs, we choose RBFs as kernel functions of SVM resulting in the best forecast performance. The RBF selection process in this study entailed deriving a hyperplane with the maximum distance to the nearest data points. The selected RBF was then used to execute $PM_{2.5}$ prediction.

C. FEATURE EXTRACTION

We extracted features from the EPA and CWB data for model training. Eighty-one features were extracted for each air quality monitoring station and its four neighboring weather stations. Specifically, 21 features were identified for each air quality monitoring station and 15 features were identified for each of its four neighboring weather stations. Tables 1 and 2 list the details of the 81 features used to generate the $PM_{2.5}$ prediction models.

To predict air pollution on a given day, the pollution levels of the previous 2 days should also be considered because of the memory effect. In addition, because the flow of traffic strongly influences air quality, whether a given day is a regular day, weekend, or holiday should also be considered. An air quality monitoring station's weather conditions were represented by the mean of the pressure and temperature levels measured at nearby weather stations. Features in the training model also included the hour of the day, day of the week, and year to learn the trend and period of the temporal index.

On the basis of the features, we averaged the pressure, temperature, and wind speed levels at nearby weather stations to represent the weather conditions at a nearby air quality monitoring station. Because wind can blow from any direction, wind was split into the four cardinal directions for simplicity. In summary, 15 features were inferred from each weather station (Table 2). We used one air station and four neighboring weather stations to generate 81-dimensional feature vectors for model learning and $PM_{2.5}$ prediction. In addition, features' data were standardized to improve the convergence speed and predictive performance.

IV. PREDICTIVE PERFORMANCE

Experiments were conducted to determine the predictive performance of the proposed models, and R^2, RMSE, NRMSE, and MAPE served as performance metrics. R^2 is a measure of the proportion of the variance of observed values that is predictable in a multiple regression analysis. When the R^2 value is 1, the observed values are perfectly predicted. RMSE indicates the mean fluctuation between the observed and predicted values. NRMSE indicates the ratio of RMSE to mean observed values. Furthermore, NRMSE represents the percentage of prediction errors in observed values, and so does MAPE. Low RMSE, NRMSE, and MAPE values indicate high predictive performance.

A. HYPERPARAMETER OPTIMIZATION

We applied decision-tree-based models (CART model and GBM), neural-network-based (RNN and LSTM) models, and SVM. When a neural-network-based model is trained, parameter adjustment should be executed to identify optimal predictive performance. The LSTM and RNN architectures jointly comprise an input layer, a hidden layer, and an output layer. The size of the input layer and that of the output layer depend on the number of features. However, the size of the hidden layer (N) should be adjusted according to length of the data sample to avoid overfitting [48]. This is because the lengths of data samples are different in different air quality monitoring stations in Taiwan. Accordingly, we systematically varied the N value to determine the value associated with optimal predictive performance. Furthermore, the parameters used for model training included initial learning rates and maximum number of epochs, which were set as 0.005 and 250 in the experiment.

TABLE 1. Input features obtained from an air monitoring station.

Air Monitoring Station		
Day	**Features**	**Dimensions**
Predicting day	Holiday or not, weekend or not, Saturday or not, Sunday or not, concentration difference, extrapolation of concentration, average pressure, average temperature, hour of the day, day of the week, and which year	11
Day before the predicting day Days before the predicting day	Holiday or not, weekend or not, Saturday or not, and Sunday or not, Concentration of air pollutant	5×2

TABLE 2. Input features from a weather station. The features from four weather stations and one air station are required for model training.

Weather Station		
Day	**Features**	**Dimensions**
Predicting day Day before the predicting day Two days before the predicting day	pressure, temperature, wind speed, eastern wind speed, northern wind speed	5×3

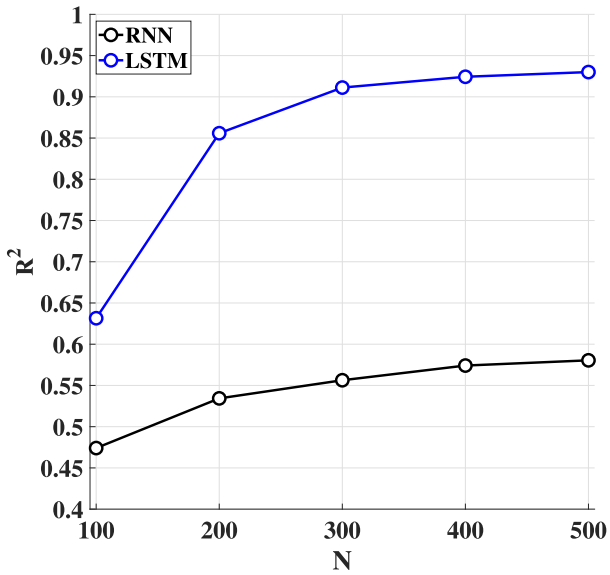

FIGURE 5. R^2 versus N.

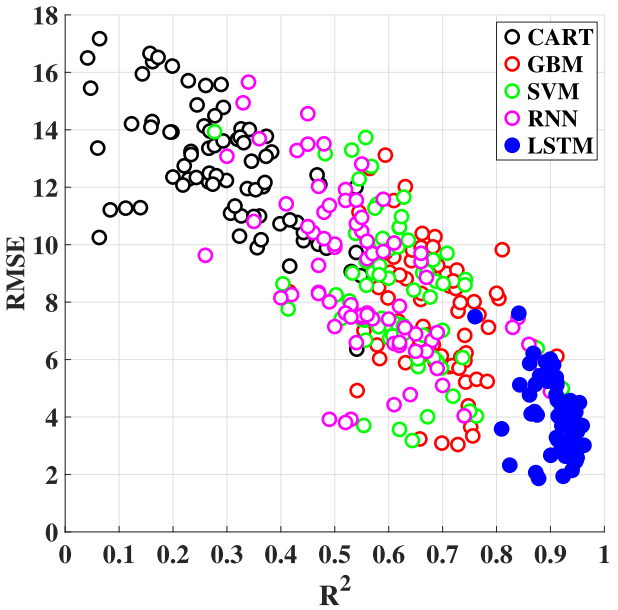

FIGURE 6. Comparison of $PM_{2.5}$ predictive performance of 77 air stations among the RNN model, GBM, SVM model, CART model, and LSTM model.

As displayed in Fig. 5, the R^2 values of the RNN and LSTM models varied with the N values. When the N value was 300, the R^2 value of the RNN model was less than 0.6, whereas that of the LSTM model was higher than 0.9. Furthermore, when the N value was higher than 300, the R^2 values of the RNN and LSTM models improves only slightly (less than 0.05). Furthermore, training the RNN and LSTM models required long calculation times when the N value was higher than 300. Therefore, we set N to 300 when comparing the predictive performance of the RNN and LSTM models.

B. COMPARISON OF MODEL PERFORMANCE

Fig. 6 illustrates the model performance comparison results. The black, purple, red, and green empty circles represent the predictive performance of the CART model, RNN model, SVM model, and GBM, respectively, and the blue solid circle represents the predictive performance of the LSTM model. The R^2 value of the LSTM model was 0.7-1.0, and those of the other models were 0-1.0 (Fig. 6). Furthermore, the RMSE of the LSTM model was 1-8, and those of the other models were in the range of 2-18. In summary, the predictive performance of the LSTM model was superior to that of the other models. However, Taiwan is characterized by complex terrains and meteorological conditions. Therefore, an in-depth comparison of predictive performance for different regions must be executed. Accordingly, we assessed predictive performance with respect to R^2, RMSE, NRMSE, and MAPE in seven regions, as presented in the next subsection.

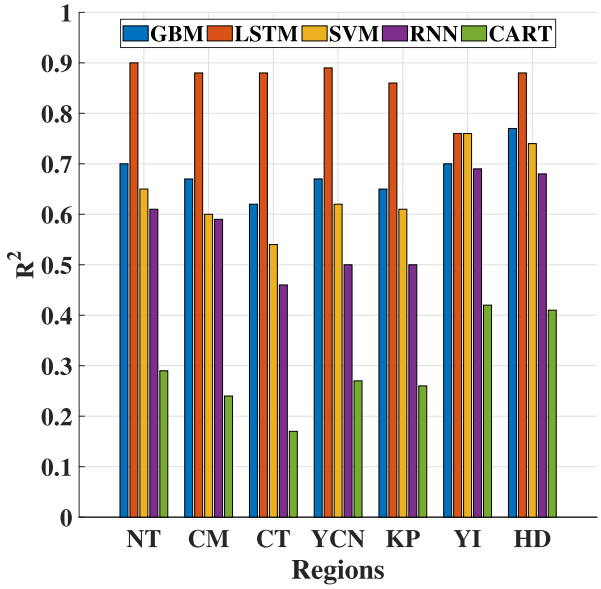

FIGURE 7. Comparison of the R^2 in the seven regions among the five models.

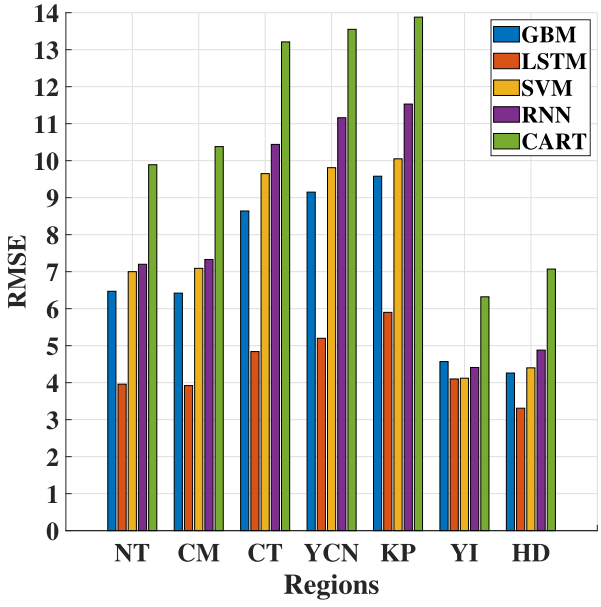

FIGURE 8. Comparison of the RMSE in the seven regions among the five models.

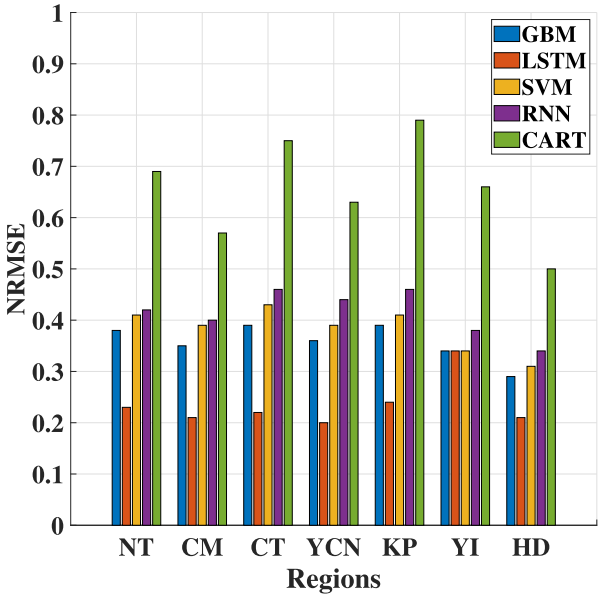

FIGURE 9. Comparison of the NRMSE in the seven regions among the five models.

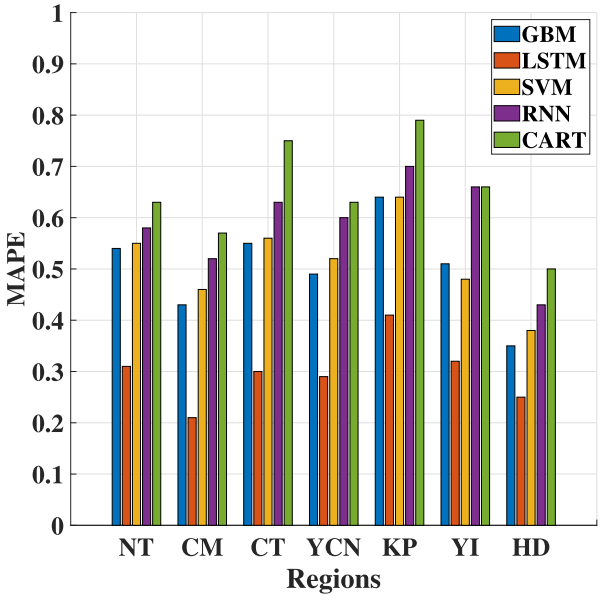

FIGURE 10. Comparison of the MAPE in the seven regions among the five models.

C. MODEL PERFORMANCE COMPARISON WITH RESPECT TO GEOGRAPHICAL DIVERGENCE

1) R^2

Fig. 7 illustrates the R^2 values of the five models when used to execute predictions in the seven regions. Among the models, the LSTM model had the highest R^2 value. Thus, the LSTM model exhibited superior predictive performance compared with the other models. Furthermore, the R^2 values of all models, except for the LSTM model, were lowest in CT.

2) RMSE

Fig. 8 displays the RMSE values of the five models in the seven regions. The CART model had the highest RMSE values. However, the LSTM model had the lowest RMSE values, indicating that the difference between the observed and predicted values was lowest for this model.

3) NRMSE AND MAPE

Figs. 9 and 10 present the NRMSE and MAPE values of the five models in the seven regions. The LSTM model exhibited the lowest NRMSE and MAPE values. This thus implies that the percentage of prediction errors in the observed values was lowest for this model. The aforementioned values are listed in Table 4.

Figs. 7, 8, and 9 reveal that in the YI region, the R^2, RMSE, and NRMSE values of the models were similar, except for

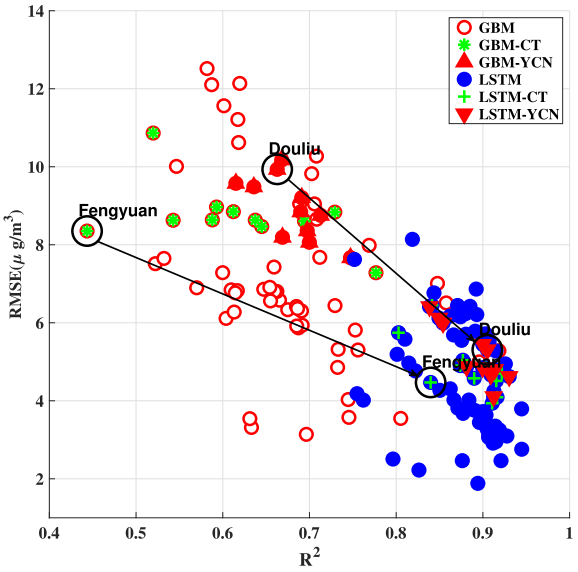

FIGURE 11. Predictive performance of the LSTM model and GBM in CT and YCN. Specifically, Douliu station in YCN and Fengyuan station in CT are two examples for model comparison.

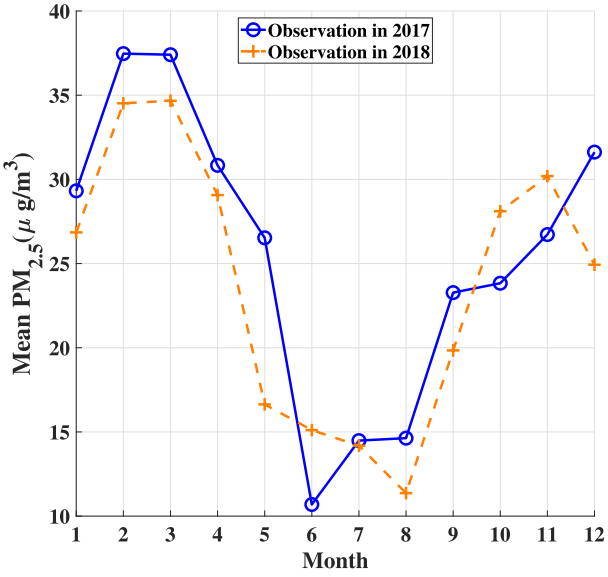

FIGURE 12. Monthly variations of the mean $PM_{2.5}$ concentration in YCN area.

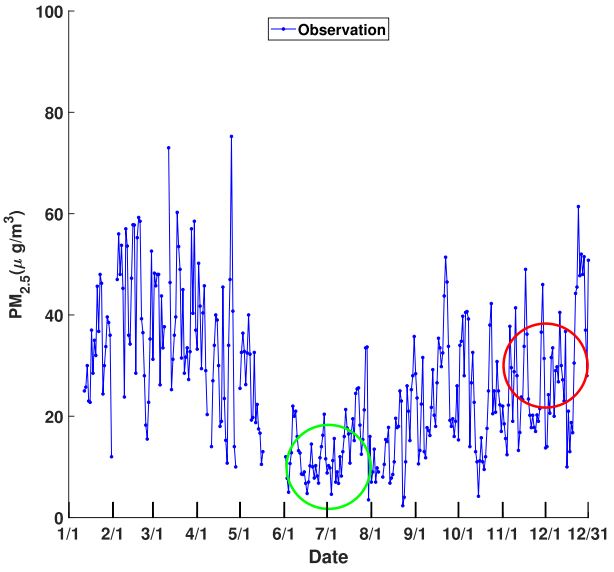

FIGURE 13. Annual $PM_{2.5}$ time series for Douliu station.

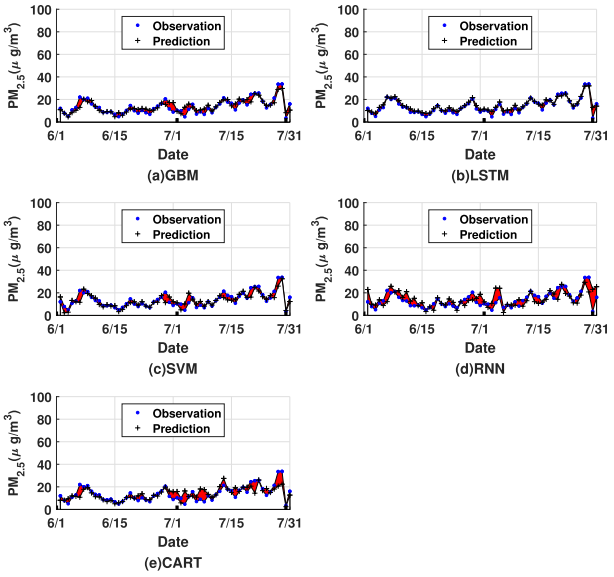

FIGURE 14. Difference between the observed and predicted $PM_{2.5}$ time series among five models for Douliu station in summer.

those of the CART model. Because the YI area had missing data for the observation period, except for December, the length of data limits the predictive performance of the models, especially that of the LSTM model. We discovered that the proposed LSTM model exhibited adequate performance in CT and KP, the most challenging regions for the other models in terms of $PM_{2.5}$ prediction.

D. EFFECTS OF SEASONS AND POLLUTION ON MODEL PERFORMANCE

1) EFFECT OF SEASONS

In Taiwan, $PM_{2.5}$ varies seasonally because the southwesterly monsoonal flow in summer and northeasterly monsoonal flow in winter strongly affect the $PM_{2.5}$ concentration [37], [38]. Less than 80% of observed values could be accurately predicted by the models, except for the LSTM model (Fig. 7). Because the LSTM model can adequately handle long-term correlations in time series of $PM_{2.5}$, it could accurately predict up to 90% of the observed values in all regions, except for YI. In addition, the RMSE, NRMSE, and MAPE values of the LSTM model were lower than those of the other models (Figs. 8-10).

$PM_{2.5}$ time series data (Fig. 11) obtained at the Douliu station in YCN were used for model comparison. To illustrate the effect of seasonal variations, Fig. 12 shows the similar monthly mean $PM_{2.5}$ concentration variations for YCN area in 2017 and 2018, respectively. Besides, Fig. 13 presents the annual $PM_{2.5}$ concentration variations for Douliu station. The green circle indicates relatively low $PM_{2.5}$ concentrations in

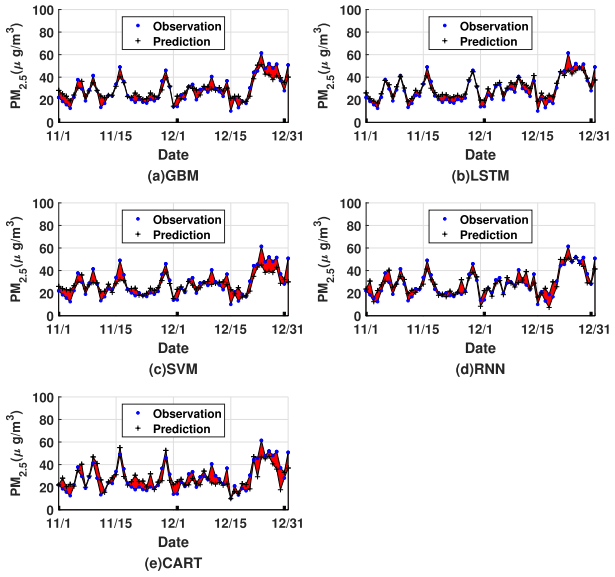

FIGURE 15. Difference between the observed and predicted $PM_{2.5}$ time series among five models for Douliu station in winter.

FIGURE 16. Box plot of residuals between observed and predicted $PM_{2.5}$ time series in summer and winter for Douliu station among the five models.

summer, and the red circle indicates relatively high $PM_{2.5}$ concentrations in winter. Furthermore, Figs. 14 and 15 show difference between the observed and predicted values among five models in summer and winter, respectively. Figs. 16(a) and 16(b) display a box plot of residuals between observed and predicted values in summer and winter for 2 months. The LSTM model had the lowest difference between the predicted and observed values.

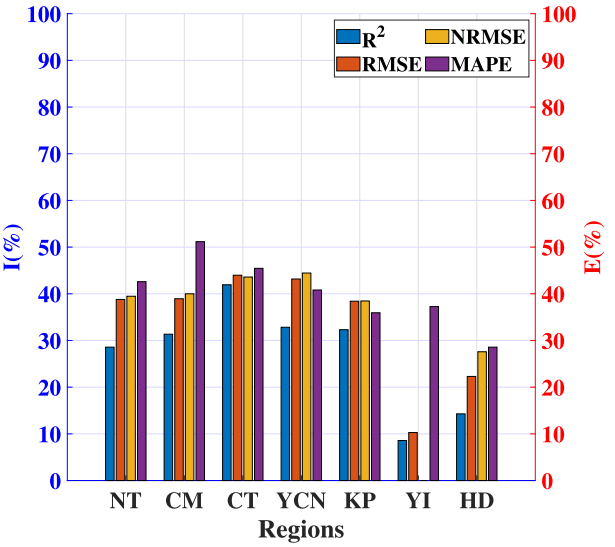

FIGURE 17. Improvement of the R^2 and error reduction rates of the RMSE, NRMSE, and MAPE from the GBM to the LSTM model in the seven regions.

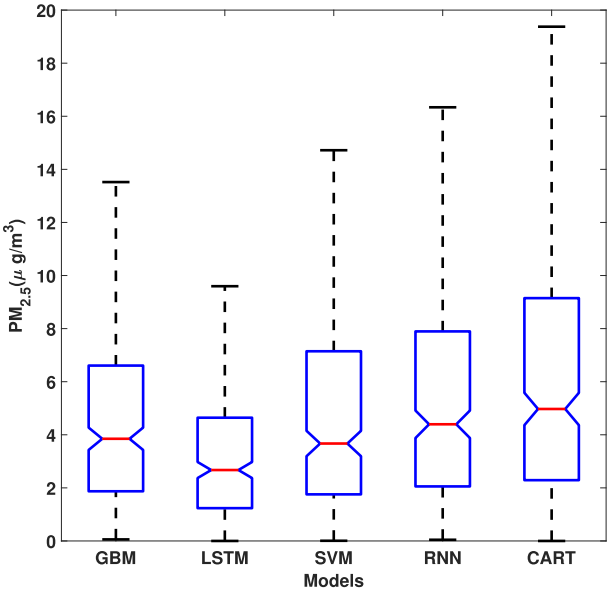

FIGURE 18. Box plots of absolute difference between the annual observed and predicted $PM_{2.5}$ time series for Fengyuan station among the five models.

2) EFFECT OF POLLUTION

Air pollution has deteriorated in CT, thus becoming a major public health concern. The EPA data indicated that CT had the highest percentage of domestic $PM_{2.5}$ emission (up to 21.46%), mainly contributed by industry, cars, and road dust (Table 3) [49]. Despite the complex mechanisms underlying this ominous pollution problem, we discovered that the LSTM model could adequately predict $PM_{2.5}$ concentrations in this region. The LSTM model exhibited the favorable R^2, RMSE, NRMSE, and MAPE values.

The LSTM model exhibited the highest R^2 in CT (Fig. 7). The improvement I could be derived as follows: $100 * (R^2_{LSTM} - R^2_{GBM})/R^2_{GBM}$. Fig. 17 illustrates improvement

TABLE 3. Domestic $PM_{2.5}$ Emission of Taiwan in 2017.

$PM_{2.5}$					
Regions	Industry	Cars	Road Dust	Total	Percentage
	Metric Tons Per Year				
NT	2559.33	**4240.69**	3764.21	10564.23	19.83%
CM	1164.28	1780.47	1582.87	4527.62	8.50%
CT	3536.21	3724.25	**4173.48**	11433.93	**21.46%**
YCN	2166.81	3365.67	3410.71	8943.19	16.78%
KP	**4863.23**	2920.55	2920.55	10459.82	19.63%
HD	1249.25	3241.54	3241.54	4991.90	9.37%
YI	970.63	968.33	968.33	2366.36	4.44%

TABLE 4. Model comparison.

Model	IndicatorRegion	NT	CM	CT	YCN	KP	YI	HD
LSTM	R^2	0.90±0.031	0.88±0.046	0.88±0.039	0.89±0.030	0.86±0.045	0.76±0.005	0.88±0.030
	RMSE	3.96±1.072	3.92±0.759	4.84±0.719	5.20±0.722	5.90±1.334	4.10±0.116	3.31±1.000
	NRMSE	0.23±0.037	0.21±0.018	0.22±0.035	0.20±0.026	0.24±0.051	0.34±0.091	0.21±0.043
	MAPE	0.31±0.141	0.21±0.083	0.30±0.113	0.29±0.100	0.41±0.223	0.32±0.230	0.25±0.141
RNN	R^2	0.61±0.128	0.59±0.079	0.46±0.113	0.50±0.069	0.50±0.106	0.69±0.074	0.68±0.149
	RMSE	7.20±0.937	7.33±0.389	10.44±1.164	11.16±1.358	11.53±3.121	4.41±0.493	4.88±0.142
	NRMSE	0.42±0.088	0.40±0.051	0.46±0.068	0.44±0.053	0.46±0.083	0.38±0.156	0.34±0.092
	MAPE	0.58±0.282	0.52±0.206	0.63±0.147	0.60±0.179	0.70±0.259	0.66±0.572	0.43±0.168
GBM	R^2	0.70±0.092	0.67±0.088	0.62±0.088	0.67±0.031	0.65±0.072	0.70±0.003	0.77±0.046
	RMSE	6.47±0.969	6.42±0.682	8.64±0.764	9.15±0.826	9.58±2.525	4.57±0.020	4.26±0.815
	NRMSE	0.38±0.071	0.35±0.046	0.39±0.054	0.36±0.030	0.39±0.078	0.34±0.093	0.29±0.024
	MAPE	0.54±0.212	0.43±0.182	0.55±0.173	0.49±0.154	0.64±0.319	0.51±0.378	0.35±0.117
SVM	R^2	0.65±0.121	0.60±0.077	0.54±0.109	0.62±0.041	0.61±0.076	0.76±0.009	0.74±0.065
	RMSE	7.00±1.700	7.09±0.511	9.65±1.508	9.81±1.020	10.05±2.563	4.12±0.118	4.40±0.718
	NRMSE	0.41±0.097	0.39±0.049	0.43±0.068	0.39±0.038	0.41±0.068	0.34±0.092	0.31±0.035
	MAPE	0.55±0.224	0.46±0.169	0.56±0.162	0.52±0.169	0.64±0.304	0.48±0.338	0.38±0.123
CART	R^2	0.29±0.243	0.24±0.078	0.17±0.126	0.27±0.114	0.26±0.155	0.42±0.146	0.41±0.123
	RMSE	9.89±1.629	10.38±0.774	13.21±1.308	13.55±0.827	13.88±3.627	6.32±0.728	7.07±1.577
	NRMSE	0.58±0.121	0.54±0.079	0.59±0.088	0.54±0.046	0.58±0.109	0.51±0.189	0.47±0.022
	MAPE	0.69±0.259	0.57±0.217	0.75±0.270	0.63±0.179	0.79±0.332	0.66±0.489	0.50±0.123

in the seven regions; the highest improvement in R^2 was observed between the LSTM model and GBM in CT.

Error reduction rates E with respect to RMSE, NRMSE, and MAPE can be defined as follows: $100 * |(\beta_{LSTM} - \beta_{GBM})|/\beta_{GBM}$, where β represents RMSE, NRMSE, or MAPE. Significant error reduction rates were also observed between the LSTM model and GBM (Fig. 17).

Consider, for example, model performance for the Fengyuan station in CT (Fig. 11). Box plots of the absolute difference between the observed and predicted values for a year are provided in Fig. 18. The lowest difference between the predicted and observed values was observed for the LSTM model.

On the basis of preceding results, we can conclude that the effects of seasonal variations and pollution on the LSTM model are low, which improves predictive performance.

V. CONCLUSION

This study proposes two dynamic models, namly RNN and LSTM models, for predicting $PM_{2.5}$ concentrations at individual air quality monitoring stations in Taiwan. The study investigated the divergence between predicted values and measured values in seven regional air quality monitoring regions. In Taiwan, the predictive performance of the LSTM model significantly outperformed the three traditional machine learning models. By the visualizations and the box plots of predictive performance in summer and winter, the LSTM model was robust in minimizing the residuals between the observed and predicted values under conditions of seasonal variations in $PM_{2.5}$ time series.

In 2017, the $PM_{2.5}$ emission percentage was largest in CT, which renders $PM_{2.5}$ prediction difficult. Nevertheless, regarding predictive performance in CT, the R^2, RMSE, NRMSE, and MAPE of the LSTM model were superior to those of the other traditional machine learning models by 41.93%, 43.98%, 43.59%, and 45.45% on average, respectively. In summary, the effects of seasons and pollution on predictive performance in Taiwan were apparently reduced by the LSTM model.

REFERENCES

[1] World Health Organization, Occupational and Environmental Health Team, "Who air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide: Global update 2005: Summary of risk assessment," World Health Organization, Geneva, Switzerland, Tech. Rep. WHO/SDE/PHE/OEH/06.02, 2006.

[2] J. S. Apte, J. D. Marshall, A. J. Cohen, and M. Brauer, "Addressing global mortality from ambient $PM_{2.5}$," *Environ. Sci. Technol.*, vol. 49, no. 13, pp. 8057–8066, Jul. 2015.
[3] L. Conibear, E. W. Butt, C. Knote, S. R. Arnold, and D. V. Spracklen, "Residential energy use emissions dominate health impacts from exposure to ambient particulate matter in india," *Nature Commun.*, vol. 9, no. 1, p. 617, Dec. 2018.
[4] W. Sun, H. Zhang, A. Palazoglu, A. Singh, W. Zhang, and S. Liu, "Prediction of 24-hour-average $PM_{2.5}$ concentrations using a hidden Markov model with different emission distributions in Northern California," *Sci. Total Environ.*, vol. 443, pp. 93–103, Jan. 2013.
[5] Q. Zhou, H. Jiang, J. Wang, and J. Zhou, "A hybrid model for $PM_{2.5}$ forecasting based on ensemble empirical mode decomposition and a general regression neural network," *Sci. Total Environ.*, vol. 496, pp. 264–274, Oct. 2014.
[6] Y. Xu, W. Yang, and J. Wang, "Air quality early-warning system for cities in China," *Atmos. Environ.*, vol. 148, pp. 239–257, Jan. 2017.
[7] M. Arhami, N. Kamali, and M. M. Rajabi, "Predicting hourly air pollutant levels using artificial neural networks coupled with uncertainty analysis by Monte Carlo simulations," *Environ. Sci. Pollut. Res.*, vol. 20, no. 7, pp. 4777–4789, Jul. 2013.
[8] U. Pak, J. Ma, U. Ryu, K. Ryom, U. Juhyok, K. Pak, and C. Pak, "Deep learning-based $PM_{2.5}$ prediction considering the spatiotemporal correlations: A case study of beijing, China," *Sci. Total Environ.*, vol. 699, Jan. 2020, Art. no. 133561.
[9] X. Qiu, Y. Ren, P. N. Suganthan, and G. A. J. Amaratunga, "Empirical mode decomposition based ensemble deep learning for load demand time series forecasting," *Appl. Soft Comput.*, vol. 54, pp. 246–255, May 2017.
[10] J. Li, X. Li, and K. Wang, "Atmospheric $PM_{2.5}$Concentration prediction based on time series and interactive multiple model approach," *Adv. Meteorol.*, vol. 2019, pp. 1–11, Oct. 2019.
[11] X. Hu, J. H. Belle, X. Meng, A. Wildani, L. A. Waller, M. J. Strickland, and Y. Liu, "Estimating $PM_{2.5}$ concentrations in the conterminous united states using the random forest approach," *Environ. Sci. Technol.*, vol. 51, no. 12, pp. 6936–6944, 2017.
[12] M. Stafoggia, T. Bellander, S. Bucci, M. Davoli, K. de Hoogh, F. de' Donato, C. Gariazzo, A. Lyapustin, P. Michelozzi, M. Renzi, M. Scortichini, A. Shtein, G. Viegi, I. Kloog, and J. Schwartz, "Estimation of daily PM_{10} and $PM_{2.5}$ concentrations in Italy, 2013–2015, using a spatiotemporal land-use random-forest model," *Environ. Int.*, vol. 124, pp. 170–179, Mar. 2019.
[13] A. Azid, H. Juahir, M. E. Toriman, M. K. A. Kamarudin, A. S. M. Saudi, C. N. C. Hasnam, N. A. A. Aziz, F. Azaman, M. T. Latif, S. F. M. Zainuddin, M. R. Osman, and M. Yamin, "Prediction of the level of air pollution using principal component analysis and artificial neural network techniques: A case study in Malaysia," *Water, Air, Soil Pollut.*, vol. 225, no. 8, p. 2063, Aug. 2014.
[14] I. Djalalova, L. D. Monache, and J. Wilczak, "$PM_{2.5}$ analog forecast and Kalman filter post-processing for the community multiscale air quality (CMAQ) model," *Atmos. Environ.*, vol. 108, pp. 76–87, May 2015.
[15] P. Wang, H. Zhang, Z. Qin, and G. Zhang, "A novel hybrid-garch model based on ARIMA and SVM for $PM_{2.5}$ concentrations forecasting," *Atmos. Pollut. Res.*, vol. 8, no. 5, pp. 850–860, Sep. 2017.
[16] Y. Zhou, F.-J. Chang, L.-C. Chang, I.-F. Kao, Y.-S. Wang, and C.-C. Kang, "Multi-output support vector machine for regional multi-step-ahead $PM_{2.5}$ forecasting," *Sci. Total Environ.*, vol. 651, pp. 230–240, Feb. 2019.
[17] P. Pérez, A. Trier, and J. Reyes, "Prediction of $PM_{2.5}$ concentrations several hours in advance using neural networks in santiago, chile," *Atmos. Environ.*, vol. 34, no. 8, pp. 1189–1196, Jan. 2000.
[18] H.-J. Chu, B. Huang, and C.-Y. Lin, "Modeling the spatio-temporal heterogeneity in the PM_{10}-$PM_{2.5}$ relationship," *Atmos. Environ.*, vol. 102, pp. 176–182, Feb. 2015.
[19] S. Mandal, K. K. Madhipatla, S. Guttikunda, I. Kloog, D. Prabhakaran, and J. D. Schwartz, "Ensemble averaging based assessment of spatiotemporal variations in ambient $PM_{2.5}$ concentrations over Delhi, India, during 2010–2016," *Atmos. Environ.*, vol. 224, Mar. 2020, Art. no. 117309.
[20] S. Deleawe, J. Kusznir, B. Lamb, and D. Cook, "Predicting air quality in smart environments," *J. Ambient Intell. Smart environments*, vol. 2, pp. 145–152, Jan. 2010.
[21] G. K. Kang, U. the Department of Computer EngineeringSan Jose State University, J. Z. Gao, S. Chiao, S. Lu, and G. Xie, "Air quality prediction: Big data and machine learning approaches," *Int. J. Environ. Sci. Develop.*, vol. 9, no. 1, pp. 8–16, 2018.
[22] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.
[23] E. Kalapanidas and N. Avouris, "Applying machine learning techniques in air quality prediction," in *Proc. ACAI*. Rio Achaia, Greece: Univ. of Patras, Department of Electrical and Computer Engineering, 1999, pp. 58–64.
[24] H. Zhao, J. Zhang, K. Wang, Z. Bai, and A. Liu, "A GA-ANN model for air quality predicting," in *Proc. Int. Comput. Symp. (ICS)*, Dec. 2010, pp. 693–699.
[25] M. Lee, L. Lin, C.-Y. Chen, Y. Tsao, T.-H. Yao, M.-H. Fei, and S.-H. Fang, "Forecasting air quality in taiwan by using machine learning," *Sci. Rep.*, vol. 10, no. 1, p. 4153, Dec. 2020.
[26] X. Wu, Y. Wang, S. He, and Z. Wu, "$PM_{2.5}$/PM_{10} ratio prediction based on a long short-term memory neural network in wuhan, china," *Geosci. Model Develop.*, vol. 13, no. 3, pp. 1499–1511, 2020.
[27] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transp. Res. C, Emerg. Technol.*, vol. 54, pp. 187–197, May 2015.
[28] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social LSTM: Human trajectory prediction in crowded spaces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 961–971.
[29] X. Li, L. Peng, X. Yao, S. Cui, Y. Hu, C. You, and T. Chi, "Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation," *Environ. Pollut.*, vol. 231, pp. 997–1004, Dec. 2017.
[30] Z. Ma, X. Hu, A. M. Sayer, R. Levy, Q. Zhang, Y. Xue, S. Tong, J. Bi, L. Huang, and Y. Liu, "Satellite-based spatiotemporal trends in $PM_{2.5}$ concentrations: China, 2004–2013," *Environ. Health Perspect.*, vol. 124, no. 2, pp. 184–192, Feb. 2016.
[31] Y. Zhan, Y. Luo, X. Deng, H. Chen, M. L. Grieneisen, X. Shen, L. Zhu, and M. Zhang, "Spatiotemporal prediction of continuous daily $PM_{2.5}$ concentrations across China using a spatially explicit machine learning algorithm," *Atmos. Environ.*, vol. 155, pp. 129–139, Apr. 2017.
[32] K.-S. Chen, C. F. Lin, and Y.-M. Chou, "Determination of source contributions to ambient $PM_{2.5}$in Kaohsiung, Taiwan, using a receptor model," *J. Air Waste Manage. Assoc.*, vol. 51, no. 4, pp. 489–498, Apr. 2001.
[33] Y.-H. Cheng and Y.-S. Li, "Influences of traffic emissions and meteorological conditions on ambient PM_{10} and $PM_{2.5}$ levels at a highway toll station," *Aerosol Air Qual. Res.*, vol. 10, no. 5, pp. 456–462, 2010.
[34] C.-H. Hsu and F.-Y. Cheng, "Classification of weather patterns to study the influence of meteorological characteristics on $PM_{2.5}$ concentrations in Yunlin County, Taiwan," *Atmos. Environ.*, vol. 144, pp. 397–408, Nov. 2016.
[35] H.-Y. Lu, S.-L. Lin, J. K. Mwangi, L.-C. Wang, and H.-Y. Lin, "Characteristics and source apportionment of atmospheric $PM_{2.5}$ at a coastal city in Southern Taiwan," *Aerosol Air Qual. Res.*, vol. 16, no. 4, pp. 1022–1034, 2016.
[36] F.-Y. Cheng and C.-H. Hsu, "Long-term variations in PM2.5 concentrations under changing meteorological conditions in Taiwan," *Sci. Rep.*, vol. 9, no. 1, p. 6635, Dec. 2019.
[37] K.-L. Yang, "Spatial and seasonal variation of PM10 mass concentrations in Taiwan," *Atmos. Environ.*, vol. 36, no. 21, pp. 3403–3411, Jul. 2002.
[38] M.-T. Chuang, J. S. Fu, C. J. Jang, C.-C. Chan, P.-C. Ni, and C.-T. Lee, "Simulation of long-range transport aerosols from the asian continent to taiwan by a southward asian high-pressure system," *Sci. Total Environ.*, vol. 406, nos. 1–2, pp. 168–179, Nov. 2008.
[39] T. Mikolov and G. Zweig, "Context dependent recurrent neural network language model," in *Proc. IEEE Spoken Lang. Technol. Workshop (SLT)*, Dec. 2012, pp. 234–239.
[40] F. A. Gers, N. N. Schraudolph, and J. Schmidhuber, "Learning precise timing with lstm recurrent networks," *J. Mach. Learn. Res.*, vol. 3, pp. 115–143, Mar. 2003.
[41] *Environmental Protection Administration in Taiwan. Quality Assurance of Air Quality Monitoring*. Retrieved From. Accessed: Aug. 1, 2018. [Online]. Available: https://taqm.epa.gov.tw/taqm/en/b0801.aspx
[42] Central Weather Bureau in Taiwan. *Data Sets of Weather Station [Data Files] Available Via*. Accessed: Aug. 1, 2019. [Online]. Available: http://farmer.iyard.org/cwb/cwb.htm
[43] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, Oct. 2000.
[44] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. London, U.K.: Chapman & Hall, 1984, p. 368.
[45] I. B. C. Matheson, "A critical comparison of least absolute deviation fitting (robust) and least squares fitting: The importance of error distributions," *Comput. Chem.*, vol. 14, no. 1, pp. 49–57, Jan. 1990.
[46] G. C. White and R. E. Bennetts, "Analysis of frequency count data using the negative binomial distribution," *Ecology*, vol. 77, no. 8, pp. 2549–2557, Dec. 1996.

[47] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[48] C.-F. Lai, W.-C. Chien, L. T. Yang, and W. Qiang, "LSTM and edge computing for big data feature recognition of industrial electrical equipment," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2469–2477, Apr. 2019.
[49] *Environmental Protection Administration in Taiwan. Taiwan Emission Data System*. Retrieved From. Accessed: Aug. 1, 2018. [Online]. Available: https://air.epa.gov.tw/EnvTopics/AirQuality_7.aspx

LARRY LIN received the Ph.D. degree from the Physical Department, Chung-Hsing University, Taiwan, in 2013. In 2015, he was a Postdoctoral Researcher with the Institute of Biomedical Sciences, Sinica. In 2016, he was a Research Assistant with the Department of Biomedical Engineering, I-Shou University. In 2017, he was a Postdoctoral Researcher with the Department of Optoelectronic Engineering, National Taipei University of Technology. He is currently a Research Assistant with the Department of Electrical Engineering, Yuan Ze University, and also with the MOST Joint Research Center for AI Technology and All Vista Healthcare.

CHIH-YUAN CHEN is currently an Assistant Professor with the Department of Geography, Chinese Culture University. His research interests include knowledge discovery framework for sptio-temporal data with a focus on machine learning, neural networks, and geo-visualization. He is also interested in theoretical theme of cyberinfrastructure and geographic information science. His current research interest includes spatio-temporal mining, spatial mobility, and mobile device.

HSIEN-YUEH YANG was born in 1998. He graduated from the Department of Electrical Engineering, Yuan Ze University (YZU), in 2020. His research interest includes air pollution prediction by using LSTM model and machine learning models.

ZHEZHUANG XU (Member, IEEE) received the B.Eng. degree in automation from Xiamen University, Xiamen, China, in 2005, the M.Eng. degree in measuring and testing technologies and instruments from Jiangsu University, Zhejiang, China, in 2008, and the Ph.D. degree in control science and engineering from Shanghai Jiao Tong University, Shanghai, China, in 2012. He joined the School of Electrical Engineering and Automation, Fuzhou University, Fuzhou, China, in 2012, where he is currently an Associate Professor. He is also the Director of the Public Platform of Industrial Big Data Application, Fujian Provincial Commission of Economy and Information Technology. He has authored or coauthored more than 30 refereed international journal and conference papers. His research interests include wireless communication and big data analysis in the industrial Internet of Things.

SHIH-HAU FANG (Senior Member, IEEE) received the B.S. degree from National Chiao Tung University, in 1999, and the M.S. and Ph.D. degrees from National Taiwan University, Taiwan, in 2001 and 2009, respectively, all in communication engineering.

From 2001 to 2007, he was a Software Architect with Chung-Hwa Telecom Ltd. He joined Yuan Ze University (YZU), in 2009. He is currently a Full Professor with the Department of Electrical Engineering, YZU, and also with the MOST Joint Research Center for AI Technology and All Vista Healthcare, Taiwan. He is also a Technical Advisor to HyXen and PTCom Technology Company Ltd. His research interests include artificial intelligence, mobile computing, machine learning, and signal processing. He has received several awards for his research work, including the Young Scholar Research Award from YZU, in 2012, the Project for Excellent Junior Research Investigators from MOST, in 2013, the Outstanding Young Electrical Engineer Award from the Chinese Institute of Electrical Engineering, in 2017, the Outstanding Research Award from YZU, in 2018, and the Best Synergy Award from the Far Eastern Group, in 2018. His team received the Third Place of IEEE Multimedia Big Data (BigMM) HTC Challenge in 2016 and the Third Place of Indoor Positioning and Indoor Navigation (IPIN) in 2017. He serves as an YZU President's Special Assistant. He is also an Associate Editor of *IEICE Transactions on Information and Systems*.