

# CombineDeepNet: A Deep Network for Multistep Prediction of Near-Surface PM<sub>2.5</sub> Concentration

Prasanjit Dey , Soumyabrata Dev , *Member, IEEE*, and Bianca Schoen Phelan , *Member, IEEE*

**Abstract**—PM<sub>2.5</sub> is a type of air pollutant that can cause respiratory and cardiovascular problems. Precise PM<sub>2.5</sub> ( $\mu\text{g}/\text{m}^3$ ) concentration prediction may help reduce health concerns and provide early warnings. To better understand air pollution, a number of approaches have been presented for predicting PM<sub>2.5</sub> concentrations. Previous research used deep learning models for hourly predictions of air pollutants due to their success in pattern recognition, however, these models were unsuitable for multisite, long-term predictions, particularly in regard to the correlation between pollutants and meteorological data. This article proposes the combine deep network (CombineDeepNet), which combines multiple deep networks, including a bidirectional long short-term memory, bidirectional gated recurrent units, and a shallow model represented by fully connected layers, to create a hybrid forecasting system. It can effectively capture the complex relationships between air pollutants and various influencing factors to predict hourly PM<sub>2.5</sub> concentrations in multiple monitoring sites based in China. The best root mean square error achieved was 22.0  $\mu\text{g}/\text{m}^3$  (long-term) and 6.2  $\mu\text{g}/\text{m}^3$  (short-term), with mean absolute error values of 3.4  $\mu\text{g}/\text{m}^3$  (long-term) and 2.2  $\mu\text{g}/\text{m}^3$  (short-term). In addition, the correlation coefficient ( $R^2$ ) reached 0.96 (long-term) and 0.83 (short-term) across six monitoring sites. These results demonstrate that CombineDeepNet enhances prediction accuracy compared with popular deep learning methods. Therefore, CombineDeepNet proves to be an important framework for predicting PM<sub>2.5</sub> concentration.

**Index Terms**—Air pollution, bidirectional gated recurrent units (BiGRU), bidirectional long short-term memory (BiLSTM), prediction.

## I. INTRODUCTION

**N**EAR — surface air pollutants have created various adverse health effects globally [1]. Their impact continues to grow

Manuscript received 4 July 2023; revised 2 October 2023; accepted 28 October 2023. Date of publication 16 November 2023; date of current version 4 December 2023. This work was supported in part by the Science Foundation Ireland Centre for Research Training in Digitally-Enhanced Reality (d-real) under Grant 18/CRT/6224, in part by the Science Foundation Ireland under Grant 13/RC/2106\_P2 at the ADAPT SFI Research Centre, University College Dublin, and in part by the ADAPT SFI Research Centre for AI-Driven Digital Content Technology, Science Foundation Ireland, through the SFI Research Centres Programme. (Corresponding author: Prasanjit Dey.)

Prasanjit Dey is with the School of Computer Science, Technological University Dublin, D07 EWW4 Dublin, Ireland (e-mail: d22124678@mytudublin.ie).

Soumyabrata Dev is with the ADAPT SFI Research Centre, School of Computer Science, University College Dublin, Dublin 4 Belfield, Ireland (e-mail: soumyabrata.dev@ucd.ie).

Bianca Schoen Phelan is with the School of Computer Science, Technological University Dublin, D07 EWW4 Dublin, Ireland (e-mail: bianca.schoenphelan@tudublin.ie).

This article has supplementary downloadable material available at <https://doi.org/10.1109/JSTARS.2023.3333269>, provided by the authors.

Digital Object Identifier 10.1109/JSTARS.2023.3333269

across the globe. Economic expansion is one of the key reasons for the increasing concentration of pollutants. According to the research in [2], the number of fatalities in China caused by PM<sub>2.5</sub> increased by 0.39 million (23%) between 2002 and 2007, emphasizing the need for precisely predicting PM<sub>2.5</sub> concentration. PM<sub>2.5</sub> is considered the principal air pollutant that may have an impact on our quality of life, our health, and reduce air visibility [3]. The researchers have shown the relationship between PM<sub>2.5</sub> and respiratory system disorders [4], [5].

The first step in resolving the PM<sub>2.5</sub> problem is to increase forecast accuracy, which has been a crucial issue. It is important to note that accurately predicting PM<sub>2.5</sub> levels can be challenging, as there are many factors that can influence air quality [6]. To enhance the accuracy of PM<sub>2.5</sub> predictions, it is often necessary to incorporate a diverse range of data [7] and utilize techniques such as cross validation and regularization. These techniques play a crucial role in optimizing the model's performance by preventing overfitting, improving generalization, and ensuring robustness and accuracy. There are two types of prediction techniques available today: deterministic methods and statistical approaches. To predict the PM<sub>2.5</sub> concentration, deterministic approaches need complex prior knowledge [8]. Statistical methods involve analyzing PM<sub>2.5</sub> concentrations over a period of time in order to identify trends and patterns. Based on historical data, a time-series method like the autoregressive integrated moving average (ARIMA) model may be used to forecast future PM<sub>2.5</sub> concentrations [9]. However, the effectiveness of these statistical techniques may be limited due to the requirement of adherence to certain statistical assumptions. These assumptions play an important part in ensuring the validity and effectiveness of the techniques. Some common statistical assumptions include normality, independence, homoscedasticity, linearity, and no multicollinearity [10]. It is essential to carefully consider these assumptions when applying statistical techniques since violations can result in misleading results and adversely impact the reliability of the analysis.

When dealing with multidimensional time series data, machine learning (ML) methods often perform better. Particularly for the PM<sub>2.5</sub> concentration tasks, several models have been used, including decision tree [11], random forest [12], [13], support vector regression [14], support vector machine (SVM) [15], multilayer perceptron [16], artificial neural network (ANN) with density-based spatial clustering of applications with noise [17], extreme learning machines [18], echo state networks [18], hidden Markov model [19], XGBoost [20], and Light GBM [21], [22].

TABLE I

SUMMARY COMPARISON OF RECENTLY DEVELOPED DL MODELS IN TERMS OF THE YEAR OF DEVELOPMENT, THREE COMPARISON METRICS, PREDICTION OF PM<sub>2.5</sub> IN MULTIPLE SITES, CORRELATION OF PM<sub>2.5</sub> WITH DIFFERENT PARAMETERS, AND THEIR SHORT-TERM AND LONG-TERM PREDICTION CAPABILITIES

Model	Year	RMSE	MAE	R <sup>2</sup>	PM <sub>2.5</sub> prediction in multi site	Correlate between PM <sub>2.5</sub> and Meteorological parameters	Correlate among PM <sub>2.5</sub> , other pollutants, and Meteorological parameters	Short-term prediction	Long-term prediction
Autoencoder Bi-LSTM [33]	2020	7.93	–	–	–	✓	–	✓	✓
Bi-LSTM [34]	2021	17.2	14.1	0.95	–	✓	–	✓	✓
3D CNN-GRU [35]	2022	15.2	12.0	0.84	✓	✓	–	✓	–
STA-ResCNN [36]	2023	15.3	10.1	–	✓	✓	✓	✓	–
CNN-LSTM [37]	2023	21.8	15.6	–	–	✓	–	✓	✓

Recently, deep learning (DL) models have gained popularity and have demonstrated impressive results in various applications, such as pattern recognition [23], [24], remote sensing images [25], and built-up land expansion [26]. Advanced versions of recurrent neural networks (RNNs) or combinations of convolutional neural networks (CNNs) are frequently used for PM<sub>2.5</sub> prediction. These include long short-term memory (LSTM) [27], [28], gated recurrent units (GRU) [29], bidirectional LSTM (BiLSTM) [30], CNN-LSTM [31], and CNN-BiLSTM [32]. In this study, we propose the CombineDeepNet model using three DL models: BiLSTM, BiGRU, and fully connected (FC) for predicting near-surface PM<sub>2.5</sub> concentration.

#### A. Related Works

In recent decades, researchers have predominantly utilized ML models for PM<sub>2.5</sub> concentration prediction. Bhatti et al. [38] suggested a semisupervised learning strategy based on factor analysis and the seasonal autoregressive integrated moving average model. In this study, the authors correlate PM<sub>2.5</sub> with other pollutants to analyze the overall air quality of Lahore, Pakistan. Zhou et al. [39] created a hybrid framework that combines multi-output SVM with multitask learning to enhance long-term forecasting of PM<sub>2.5</sub> concentration in Taipei City, Taiwan. In this work, the authors also developed a single-output SVM benchmark technique for comparing the implemented model. Masood and Ahmad [40] used ANN and SVM models to forecast PM<sub>2.5</sub> level in Delhi. They trained their models using various meteorological and pollutant parameters over a two-year period (2016–2018).

Recently, numerous DL models have become more popular in time-series prediction applications. Zhang et al. [33] presented a DL system for forecasting PM<sub>2.5</sub> concentrations using an auto-encoder and a BiLSTM. In this work, the association between PM<sub>2.5</sub> concentration and other climatic factors was employed to produce a more accurate result. Similarly, Zhang et al. [34] developed a BiLSTM model for forecasting daily PM<sub>2.5</sub> concentrations in Beijing. Faraji et al. [35] proposed a combined 3-D CNN-GRU model for the spatial-temporal forecasting of PM<sub>2.5</sub> concentrations in urban areas. In order to train the prediction model, this study gathered PM<sub>2.5</sub> concentrations from several air quality monitoring sites around Tehran, Iran, from 2016

to 2019. In their proposal, Zhang et al. [36] combined CNN with spatial-temporal attention and residual learning to create a hybrid DL network for short-term PM<sub>2.5</sub> concentration prediction. Natsagdorj et al. [37] implemented two DL frameworks using Bayesian optimized LSTM and CNN-LSTM to forecast daily PM<sub>2.5</sub> levels in Ulaanbaatar, Mongolia. They collected hourly Himawari8 aerosol optical depth, PM<sub>2.5</sub> concentration, and meteorological data to train their prediction models.

Chen et al. [41] proposed a physical model for estimating PM<sub>2.5</sub> levels in the mideastern China. In their work, the authors utilized data from the AERONET official website. Wu et al. [42] developed a framework that utilizes satellite and station observations to estimate hourly PM<sub>2.5</sub> levels.

A summary comparison of recently developed DL models is presented in Table I. The table provides an overview of these models in terms of the year of development, three comparison metrics, prediction of PM<sub>2.5</sub> in multiple sites, correlation of PM<sub>2.5</sub> with different parameters, and their short-term and long-term prediction capabilities. Furthermore, certain studies have employed a combination of adjacent station parameters and target station parameters to effectively forecast PM<sub>2.5</sub> levels at the desired location. Zhao et al. [43] proposed an LSTM-FC framework to forecast PM<sub>2.5</sub> concentration for a given air quality monitoring station using 48 h of past air pollutant, meteorological, and weather forecasting data. However, this approach combines spatial information from surrounding monitoring stations located in different geographical regions, which may lead to overfitting issues and an increase in computational complexity. While combining information from surrounding monitoring stations can enhance prediction accuracy, there is a risk of overfitting. Overfitting can occur when the model captures noise or idiosyncratic patterns specific to the training dataset, resulting in reduced generalization ability. When neighboring station data is excessively relied upon, the model may become overly sensitive to local variations, compromising its ability to generalize to new locations or conditions. It is crucial to strike a balance between incorporating relevant information and avoiding overfitting to ensure robust and reliable predictions. Chang et al. [44] presented an aggregated LSTM to predict PM<sub>2.5</sub> level for 1–8 h using local air quality data from nearby industrial stations. Finally, Yeo et al. [45] combined CNN and GRU with a group of nearby stations to forecast PM<sub>2.5</sub> levels efficiently.

## B. Contributions of the Article

The previously described models are unable to extract deep features from  $PM_{2.5}$  datasets. Therefore, in this research, we propose the CombineDeepNet model for predicting  $PM_{2.5}$  levels at multiple sites. The CombineDeepNet is a combination architecture of three different networks, namely, BiLSTM, BiGRU, and an FC layer. The BiLSTM layer extracts preliminary features, while the BiGRU layer extracts deeper features from the air quality data. The FC layer predicts the  $PM_{2.5}$  level using these deep features. Deep features refer to the high-level abstract representations of  $PM_{2.5}$  data extracted from the BiGRU module using the lower level representation of the raw  $PM_{2.5}$  data from the BiLSTM module. These deep features capture important patterns and relationships in the data that are not easily discernible from a simple analysis of the raw data. The absence of these deep features can diminish the precision of  $PM_{2.5}$  level prediction. We evaluate the performance of this model using three key metrics: root mean square error (RMSE), mean absolute error (MAE), and correlation coefficient ( $R^2$ ). Furthermore, we explore the relationships between  $PM_{2.5}$  and meteorological factors, as well as the associations between  $PM_{2.5}$ , other pollutants, and meteorological data for both short-term and long-term forecasts. Our observations reveal that the CombineDeepNet model possesses all the necessary attributes for accurately predicting  $PM_{2.5}$  concentrations. The main contributions of this article can be outlined as follows.

- 1) The BiLSTM module is used to solve vanishing gradient problems and to integrate the relationship between pollutant and meteorological data in the initial feature investigation. The vanishing gradient problem refers to the difficulty of RNNs propagating information over long sequences due to the diminishing gradients during backpropagation. This issue can hamper the network's capability to capture long-term dependencies within the data. BiLSTM addresses this issue by employing gates. A standard LSTM consists of three types of gates: the forget gate, the input gate, and the output gate. The forget gate decides whether to keep or discard information from the previous time step. The input gate decides which new information should be added to the cell state, and the output gate decides which information gets transferred to the next step. In a BiLSTM, there are separate sets of gates for the forward LSTM and the backward LSTM. When we compute gradients during backpropagation in a BiLSTM, it involves gradients from both directions. If one direction experiences vanishing gradients for a particular gate, the other direction might have nonvanishing gradients, which can help prevent complete information loss. Therefore, the bidirectional nature of a BiLSTM makes it less likely for both directions to simultaneously suffer from severe vanishing gradients. Similar to the forget gate, the input and output gates in a BiLSTM also have bidirectional counterparts. This bidirectional gating allows information to flow from both directions, potentially reducing the impact of vanishing gradients.

BiLSTM can also effectively integrate the relationship between pollutant and meteorological data by considering temporal dependencies and contextual information. By inputting both the pollutant and meteorological data into the BiLSTM model, the network can learn the complex relationships and patterns between them. The sequential nature of BiLSTM allows it to capture dependencies over time, enabling the model to understand how meteorological factors affect pollutant levels at different time points.

- 2) The bidirectional gated recurrent units (BiGRU) module not only mitigates the vanishing gradient problem but also extracts deeper features from the preliminary features. In a standard gated recurrent unit (GRU), there are reset and update gates that regulate the information flow within the network. In a BiGRU, there are two sets of these gates: one for the forward GRU and one for the backward GRU. When computing gradients during backpropagation in a GRU, the gradients are influenced by both directions. If one direction suffers from vanishing gradients for a particular gate, the other direction may exhibit a nonvanishing gradient for that gate, helping to prevent complete information loss. This bidirectional aspect can make it less likely for both directions to simultaneously encounter severe vanishing gradients.

“Preliminary features” refer to the initial set of features that act as the initial point of reference for the BiGRU module. These preliminary features are derived from the output of the BiLSTM modules. The selection of preliminary features is typically based on the hidden patterns found in the historical pollution data. These features capture relevant information and characteristics from the data that are deemed important for further analysis. Subsequently, the BiGRU module processes these preliminary features and extracts deep features from them. Deep features represent higher level representations of the data, capturing more complex patterns and relationships.

- 3) The FC layer is utilized to create an accurate long-term forecast of  $PM_{2.5}$  concentrations across six monitoring locations. We are forecasting the next 180 h of lead time. We have measured the accuracy in terms of RMSE, MAE, and  $R^2$ . The FC layer is a type of ANN layer where each neuron is linked to all neurons in the preceding layer. It is commonly used for mapping high-level features extracted from the preceding layers to the desired output. Here, the FC layer takes the extracted features from the preceding layers as input and performs computations to generate predictions of  $PM_{2.5}$  concentrations.

By utilizing the FC layer, the model can capture complex relationships and patterns in the input data to create accurate long-term forecasts. The layer learns to combine the extracted features from earlier layers and map them to the desired output, which is the prediction of  $PM_{2.5}$  concentrations at various monitoring locations. In addition, the experimental outcomes demonstrate the superior performance of our proposed model compared with several other well-known DL models.

TABLE II  
LATITUDE AND LONGITUDE DETAILS OF SIX MONITORING STATIONS

Site Name	Site type	Latitude (°N)	Longitude (°E)
Aoti Zhongxin	Urban	39.98	116.40
Dongsi	Urban	39.93	116.42
Shunyicheng	Suburban	40.13	116.66
Tiantan	Urban	39.89	116.41
Haidan Wanliu	Urban	39.99	116.29
Wanshou Xigong	Urban	39.88	116.35

4) In the spirit of transparent research, we make available the code used to produce the findings of this work on the <https://github.com/Prasanjit-Dey/CombineDeepNet>.

In this work, the name “CombineDeepNet” signifies the combining aspect of the model. The model incorporates three modules: BiLSTM, BiGRU, and FC layers. The BiLSTM module takes the pollutants and meteorological data as input and extracts initial features. The BiGRU module then takes these initial features as input and further extracts deep features. Finally, the FC layers utilize the deep features to make predictions. The combining aspect of our framework lies in the integration of these three architectural components for long-term prediction. By combining BiLSTM, BiGRU, and FC layers, we leverage the capabilities of these modules to enhance predictive performance. Therefore, we named our framework “CombineDeepNet” to reflect its combining nature.

The rest of this article is organized as follows. Section II includes the study area and dataset analysis. The proposed method is described in Section III. It includes the CombineDeepNet framework. Section IV incorporates the results of the proposed model and five other popular DL models. Section V analyzes the discussion of the CombineDeepNet model and the other five models. Finally, Section VI concludes this article.

## II. STUDY AREA AND DATASET ANALYSIS

### A. Study Area

The six urban and suburban monitoring stations of China, namely, Aoti Zhongxin, Dongsi, Shunyicheng, Tiantan, Haidan Wanliu, and Wanshou Xigong, are selected as study areas. Table II shows the detailed description of six monitoring stations along with latitude and longitude details. Similarly, Fig. 1 depicts the physical location of the research region. These stations are in highly urbanized and industrialized regions. We have chosen the monitoring stations in these areas as our main study focus. We observe that there is still significant room for improvement in overall air quality in these chosen areas when compared with international standards. As a consequence, precise prediction of PM<sub>2.5</sub> concentrations in these locations is critical. While our study primarily focuses on providing accurate predictions of air quality pollutants, such as PM<sub>2.5</sub>, the information obtained from these predictions can be valuable in informing decision-making processes related to air pollution mitigation and control. In addition, accurate predictions enable the identification of critical periods or locations where pollution levels are likely to exceed the limits, allowing for proactive measures and interventions.



Fig. 1. Distribution of PM<sub>2.5</sub> pollution monitoring station in China. Here, there are six monitoring stations across urban and suburban areas.

### B. Dataset Analysis

1) *Data Description*: In this research, our dataset contains past pollution levels and meteorological parameters collected from six monitoring stations (Aoti Zhongxin, Dongsi, Shunyicheng, Tiantan, Haidan Wanliu, and Wanshou Xigong) in China between March 1, 2013 and February 28, 2017. It contains 35 064 h of data for each monitoring site. The dataset contains 12 pollutant and meteorological variables, such as PM<sub>2.5</sub>, PM<sub>10</sub>, SO<sub>2</sub>, NO<sub>2</sub>, CO, O<sub>3</sub>, temperature, air pressure, humidity, rain, wind direction, and wind speed. The inclusion of other pollutants and meteorological variables alongside PM<sub>2.5</sub> serves a specific purpose. By incorporating these additional variables, we aim to leverage the synergistic relationships and interactions among different factors affecting air quality. This approach allows the DL model to extract efficient hidden patterns and improve the overall prediction accuracy. Table III describes the statistical information of maximal values, minimal values, mean values, and standard deviation for six monitoring stations. We observe that high mean and standard deviation values for pollutant concentrations like PM<sub>2.5</sub>, PM<sub>10</sub>, SO<sub>2</sub>, NO<sub>2</sub>, CO, and O<sub>3</sub> over six monitoring stations can indicate significant fluctuations in air quality over time, and there are multiple sources of pollutants in the area.

2) *Correlation of PM<sub>2.5</sub> With Other Parameters*: We calculated the correlation coefficient between pollutants and meteorological parameters for six monitoring stations: (a) Aoti Zhongxin, (b) Dongsi, (c) Shunyicheng, (d) Tiantan, (e) Haidan Wanliu, and (f) Wanshou Xigong, as shown in Fig. 2. It shows that PM<sub>2.5</sub> has the highest correlation with other parameters, specifically PM<sub>2.5</sub> has the strongest correlation with PM<sub>10</sub> for all six monitoring sites. Furthermore, Table IV presents the grading table for the Spearman correlation coefficient ( $\rho$ ) [46]. The Spearman correlation coefficient is employed to assess both the magnitude and direction of the monotonic relationship between

TABLE III  
STATISTICAL INFORMATION OF SIX MONITORING STATIONS

Location		PM <sub>2.5</sub>	PM <sub>10</sub>	SO <sub>2</sub>	NO <sub>2</sub>	CO	O <sub>3</sub>	TEMP	PRES	DEWP	RAIN	WSPM
Zhongxin	Min	3.000	2.000	0.285	2.000	100.0	0.214	-16.80	985.9	-35.30	0.000	0.000
	Max	898.0	984.0	341.0	290.0	10000.0	423.0	40.50	1042.0	28.50	72.50	11.20
	Mean	83.16	110.7	17.57	59.29	1267.0	55.18	13.58	1011.8	3.120	0.067	1.708
	Std	82.29	95.37	22.81	37.00	1242.1	57.58	11.40	10.40	13.69	0.909	1.204
Dongsi	Min	3.000	2.000	0.285	2.000	100.0	0.642	-16.80	987.1	-35.30	0.000	0.000
	Max	737.0	955.0	300.0	258.0	10000.0	1071.0	41.10	1042.0	28.80	46.40	10.50
	Mean	86.31	110.5	18.52	52.19	1363.2	57.52	13.66	1012.5	2.445	0.063	1.860
	Std	86.33	98.18	22.97	34.47	1201.2	58.67	11.46	10.26	13.81	0.786	1.280
Shunyi	Min	2.000	2.000	0.285	2.000	100.0	0.214	-16.80	988.0	-36.00	0.000	0.000
	Max	941.0	999.0	239.0	258.0	10000.0	351.7	40.60	1042.8	27.50	37.30	12.80
	Mean	80.02	99.69	13.45	44.55	1203.9	53.86	13.37	1013.0	2.450	0.061	1.807
	Std	82.09	91.81	19.55	31.61	1171.4	54.88	11.48	10.17	13.73	0.761	1.287
Tiantan	Min	3.000	2.000	0.571	2.000	100.0	0.428	-16.80	987.1	-35.30	0.000	0.000
	Max	821.0	988.0	273.0	241.0	10000.0	674.0	41.10	1042.0	28.80	46.40	10.50
	Mean	82.25	106.65	14.47	53.23	1308.2	56.46	13.66	1012.5	2.445	0.063	1.860
	Std	81.13	90.43	20.50	31.99	1184.4	60.84	11.46	10.26	13.81	0.786	1.280
Wanliu	Min	2.000	2.000	0.285	1.642	100.0	0.214	-15.80	985.9	-34.90	0.000	0.000
	Max	957.0	951.0	282.0	264.0	10000.0	364.0	40.50	1040.3	28.50	72.50	11.20
	Mean	83.51	110.6	18.41	65.97	1331.7	46.92	13.42	1011.1	3.263	0.068	1.501
	Std	82.17	93.61	22.74	38.24	1267.9	54.45	11.34	10.35	13.67	0.896	1.104
Wanshou	Min	3.000	2.000	0.285	2.000	100.0	0.214	-16.80	985.1	-35.30	0.000	0.000
	Max	999.0	961.0	411.0	251.0	9800.0	358.0	40.60	1042.0	28.50	46.40	13.20
	Mean	85.35	112.7	17.36	55.64	1376.9	55.97	13.78	1011.5	2.705	0.064	1.745
	Std	86.62	98.70	24.27	36.09	1238.8	57.16	11.38	10.57	13.70	0.796	1.206

TABLE IV  
GRADING TABLE OF SPEARMAN CORRELATION COEFFICIENT ( $\rho$ ) [46]

Spearman ( $\rho$ )	Strength of correlation
0.9–1.0	Very strong
0.7–0.9	Strong
0.5–0.7	Moderate
0.3–0.5	Weak
0.0–0.3	Negligible

variables. According to the table, a strong correlation exists between PM<sub>2.5</sub> and other parameters, especially between PM<sub>2.5</sub> and PM<sub>10</sub>, as the correlation coefficient between them is at least 0.87 for all monitoring sites. This strong correlation implies that there is a hidden pattern between the parameters, which can be useful during the model training process to improve prediction accuracy.

In addition, PM<sub>2.5</sub> is a well-established air pollutant that has been consistently associated with numerous adverse health effects [1]. Consequently, our objective was to estimate the concentration of PM<sub>2.5</sub> by leveraging meteorological parameters and historical pollution levels, as these factors exhibit strong correlations with PM<sub>2.5</sub> levels. It is important to reiterate that the primary focus of this article is to predict PM<sub>2.5</sub> due to its significant impact. Hence, PM<sub>2.5</sub> has been chosen as the target variable in this study.

3) *Distribution Characteristic of Data*: We identified six monitoring sites: Aoti Zhongxin, Dongsi, Shunyi, Tiantan, Haidan Wanliu, and Wanshou Xigong as the study targets. These

monitoring locations were selected to investigate the properties of air pollution dispersion and meteorological data. Fig. 3 shows the numerical variations in PM<sub>2.5</sub> concentrations over various time periods across six monitoring sites. The deep blue line shows the 30-day exponential moving average of PM<sub>2.5</sub> concentrations. The reason for using a 30-day exponential moving average is to smooth out short-term fluctuations in the data, while still capturing long-term trends. This approach can provide a clearer representation of the overall pattern of PM<sub>2.5</sub> levels over time, particularly in situations where the data may be noisy or subject to short-term variability. The choice of 30 days may have also been based on practical considerations, such as the frequency of data collection and the time scale of relevant environmental and health effects associated with PM<sub>2.5</sub> exposure. In addition, it is a commonly used time frame for calculating moving averages in various fields, including finance and economics [47], [48]. It shows that the changes in PM<sub>2.5</sub> concentration have been pretty consistent over time. This suggests that there may be a hidden correlation in the historical data.

After statistical analysis, an average of 63.20% of the PM<sub>2.5</sub> levels at the six monitoring sites were determined to be greater than the World Health Organization's (WHO) first interim threshold of 35  $\mu\text{g}/\text{m}^3$ . Exposure to this pollutant can have adverse health effects on both vulnerable and healthy individuals [1]. The health effects may include respiratory and cardiovascular problems, among others, and can range in severity depending on the level and duration of exposure. Moreover, exposure to high levels of this pollutant can be harmful to the entire population, not just vulnerable individuals. The average

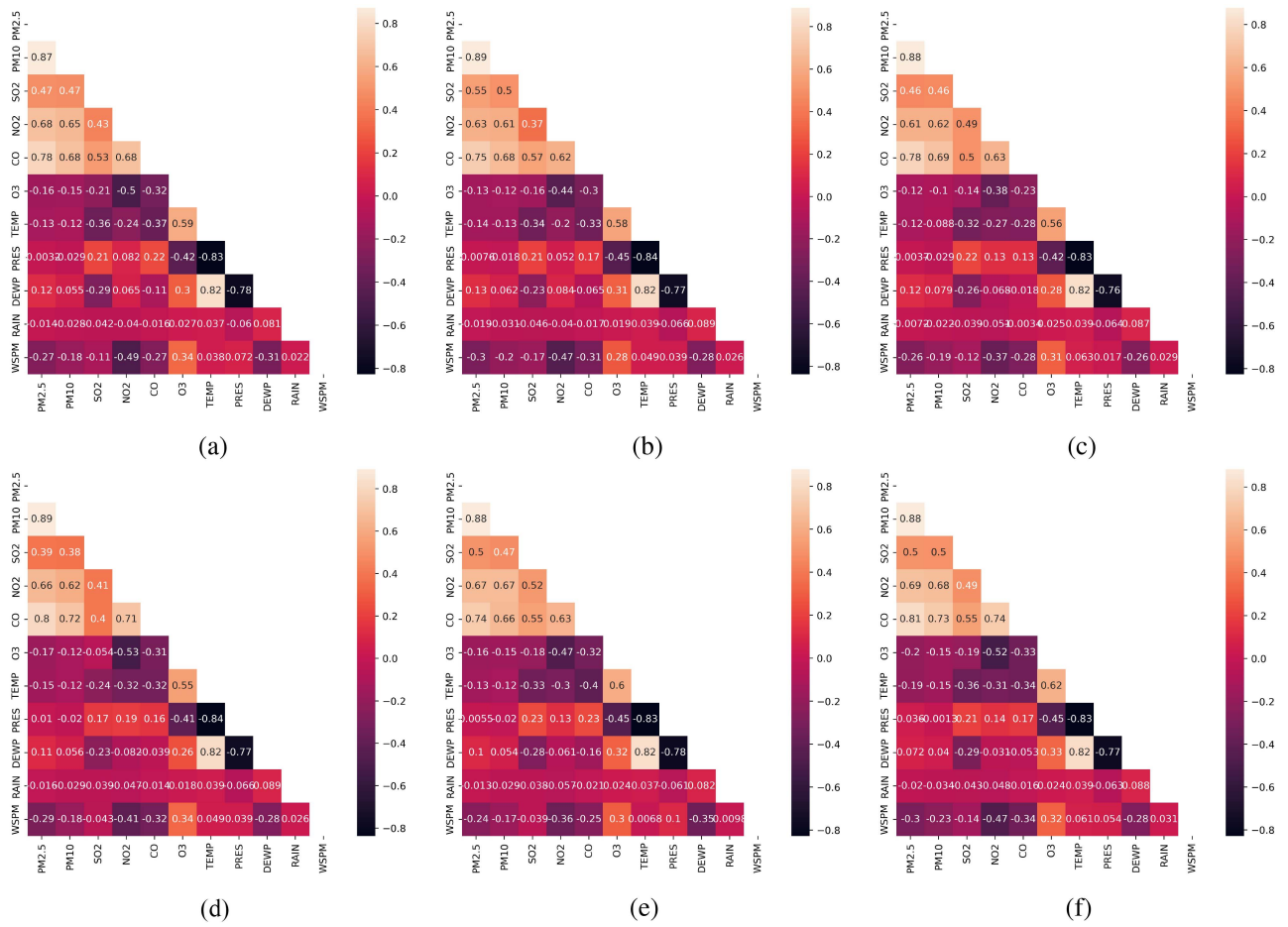


Fig. 2. Correlation coefficient between pollutants and meteorological parameters for six monitoring stations: (a) Aoti Zhongxin, (b) Dongsi, (c) Shunyicheng, (d) Tiantan, (e) Haidan Wanliu, and (f) Wanshou Xigong.

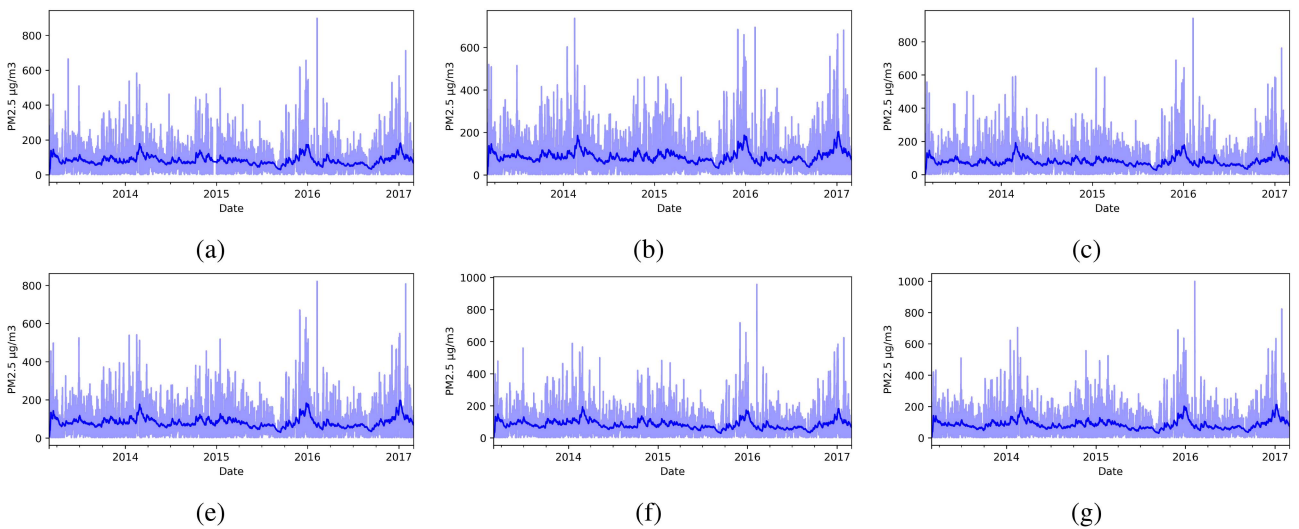


Fig. 3. Distribution of PM<sub>2.5</sub> concentrations across six monitoring stations: (a) Aoti Zhongxin, (b) Dongsi, (c) Shunyicheng, (d) Tiantan, (e) Haidan Wanliu, and (f) Wanshou Xigong. The deep blue line depicts the 30-d exponential moving average of PM<sub>2.5</sub> concentrations, indicating a consistent pattern.

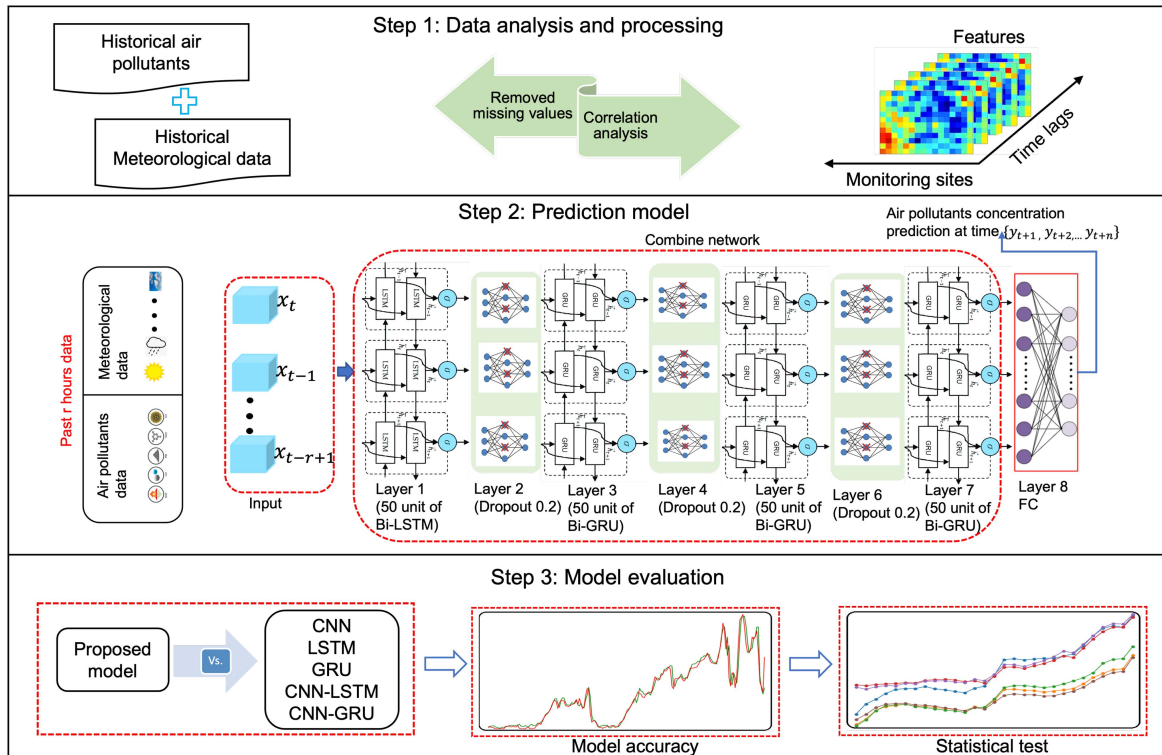


Fig. 4. Frameworks of the CombineDeepNet for  $PM_{2.5}$  concentration prediction.  $x_{t-i}$  is the pollutants' concentration and meteorological data input in the model for each timestamp.

$PM_{2.5}$  concentration of 40.02% was found to be greater than  $75 \mu\text{g}/\text{m}^3$ , which has a direct impact on physical health and human life. Therefore, to predict  $PM_{2.5}$  concentration accurately and prevent its impact on human health, we considered the hidden relationship of the historical data [49], [50]. In our experiment, we used 35 064 h of data from each of the monitoring sites. However, we removed some inconsistent data, such as missing data and data with known errors or anomalies. These data points were excluded from the analysis to ensure the reliability and accuracy of our models. Among the remaining data, we used 26 300 h to train the models and 8 760 h to test the models.

### III. METHOD

#### A. Overview of Proposed Framework

Fig. 4 depicts the proposed CombineDeepNet framework. The main structure is made up of the three stages listed below.

1) *Data Analysis and Preprocessing*: First, some of the data from the six monitoring stations could be lost because of equipment failure, routine maintenance, bad transmission, or other things that cannot be controlled. However, data loss significantly affects the performance of the prediction models. Therefore, replacing the empty or “nan” values is critical to ensuring the model's performance [51]. Using “nan” values in data during DL model training can pose three main challenges as follows:

i) *Data incompatibility*: Many DL frameworks and libraries expect numeric data as input. The “nan” is a non-numeric value, which can lead to compatibility issues when

training DL models. It may result in errors or inconsistencies during computations, affecting the training process.

ii) *Inconsistent data shape*: DL models typically require consistent data shapes for proper training. If “nan” values are present in the dataset, they can introduce inconsistencies in the data shape. This can cause difficulties in batching, vectorization, or applying certain operations, leading to training errors or unexpected behavior.

iii) *Gradient computation issues*: DL models rely on gradient computations during backpropagation for updating model parameters. The “nan” values can disrupt these computations because gradients for nonnumeric values are undefined. This can introduce errors or inaccuracies during gradient descent optimization, hindering the training process.

To address these challenges, some common strategies for handling empty or “nan” values have been employed in previous research, such as: 1) using the mean value of the previous and next values for the current “nan” value, 2) k-nearest neighbor, etc. In this research, we replace a missing value with the existing data's next value. Next, the Pearson correlation coefficient is applied to the six monitoring stations to assess the degree of correlation between pollutants and meteorological data, which can be utilized for geographical analysis [52]. Fig. 2 displays the correlation coefficients between variables for the six monitoring stations. In addition, Table IV presents the grading tables for the Spearman correlation coefficient, which assesses the strength of the correlation between variables. Both the figures and tables indicate a strong correlation between  $PM_{2.5}$  and the

other variables across all six monitoring sites. The presence of a strong correlation suggests the existence of hidden patterns, which can be leveraged during the model training process to enhance prediction accuracy. Furthermore, PM<sub>2.5</sub> is a widely recognized air pollutant that has been associated with numerous negative health impacts [1]. Therefore, we aimed to determine the PM<sub>2.5</sub> concentration based on meteorological parameters and past pollution levels, as these factors are closely related to PM<sub>2.5</sub> levels. Therefore, in this study, PM<sub>2.5</sub> is chosen as the target variable. Then, we normalize the dataset using the min-max technique, which is described as follows:

$$x_t = \frac{x_t - \min(x_t)}{\max(x_t) - \min(x_t)}. \quad (1)$$

In this case,  $x_t$  represents a single dataset from six monitoring stations, and  $\min(\cdot)$  and  $\max(\cdot)$  represent the minimal and maximal values of the  $x_t$ .

2) *Prediction Model*: CombineDeepNet is a powerful DL framework that combines the BiLSTM, BiGRU, and FC layers to address the PM<sub>2.5</sub> prediction problem. The motivation behind CombineDeepNet was to leverage the capabilities of both BiLSTM and BiGRU modules for extracting features from the input data. The BiLSTM module is employed to capture temporal dependencies and patterns in pollutants and meteorological data, generating initial features. These initial features are subsequently fed into the BiGRU module, which further extracts deeper features by leveraging its ability to model long-term dependencies. The FC layers in CombineDeepNet utilize these deep features to predict PM<sub>2.5</sub> concentration.

To enhance accuracy and understand the relationships between various variables, CombineDeepNet combines multiple pollutants and meteorological factors to generate more accurate predictions and identify meaningful relationships between the different variables. The training method employs a combination of input data and target results. In this scenario, the input data comprises air pollutants and meteorological parameters collected from six monitoring stations over the past  $r$  h. Here,  $r$  is equal to 26 300 h, and these data are fed into the hybrid models as  $x = x_{t-i}, \dots, x_{t-r-1}$ , where  $x_{t-i} \in \mathbb{R}^m$  (with  $m$  denoting the number of pollutants and meteorological parameters) for each station. The objective of this task is to forecast the PM<sub>2.5</sub> level for each monitoring station, specifically for  $n$  h into the future. This forecast is represented as  $y = y_{t+1}, \dots, y_{t+j}, \dots, y_{t+n}$ , where  $y_{t+j} \in \mathbb{R}$  represents the prediction outcome. In this context,  $n$  corresponds to 180 h.

The highest concentration of PM<sub>2.5</sub> is an indicator of air pollution and its potential adverse effects on human health [1]. In our research, we specifically focused on estimating and predicting the concentration of PM<sub>2.5</sub>, recognizing its importance in understanding and addressing air quality concerns. The training method for this framework is designed using a three-module architecture. In the first module, there is a 50-unit BiLSTM layer that is used to look for early features in the training data. The BiLSTM layer is a variant of the RNN capable of processing sequential data in both forward and backward directions. This makes it useful for finding patterns in time series data. The

second module consists of a BiGRU layer with 50 units, which further extracts deep features.

While BiLSTMs are well-regarded for their ability to capture long-range dependencies in sequential data, they tend to impose a higher computational and memory demand compared with BiGRUs [53]. BiGRUs, being a simpler variant, offer computational efficiency. In our experiment, we tackled the challenge of making accurate long-term predictions of PM<sub>2.5</sub> concentration within a complex and dynamic domain. Given this context, it was imperative to select a computationally efficient DL architecture. To strike a balance between model complexity and predictive accuracy, we initially integrated a single layer of BiLSTM into our model for primary feature extraction. Our analysis revealed that this single layer of BiLSTM was insufficient to capture the features necessary for accurate predictions. Adding more BiLSTM layers would have escalated computational complexity without commensurate gains in accuracy. Therefore, in our proposed architecture, we incorporated BiGRU layers to extract deeper features. This hybrid approach allowed us to harness the strengths of BiLSTM for primary feature extraction, whereas benefiting from the computational efficiency of BiGRU for capturing intricate, long-range dependencies. In our study, we conducted experiments with various numbers of units to identify the optimal configuration. After a thorough evaluation, we determined that utilizing 50 units in both the BiLSTM and BiGRU layers provided the optimal solution in terms of accuracy and computational efficiency.

To regularize the training process, dropout layers are added after each of the BiLSTM and BiGRU layers. Dropout is a strategy for preventing overfitting by randomly removing a particular proportion of neurons during training. It has proven to be a valuable technique for regularization in DL, helping networks generalize better and avoid overfitting. Researchers have found that dropout can lead to improved performance and generalization on a wide range of tasks [54], [55]. The third module is a FC layer that takes the output from the BiGRU layers and completes the PM<sub>2.5</sub> prediction as  $y = y_{t+1}, \dots, y_{t+j}, \dots, y_{t+n}$ . The FC layer is a sort of neural network layer that utilizes the characteristics retrieved by the preceding layers to produce predictions.

3) *Model Evaluation*: The last step of the proposed CombineDeepNet framework is model evaluation. The CombineDeepNet model, along with other popular DL models, is evaluated by comparing their predicted values to actual values using various metrics. These metrics are then compared to determine whether the proposed model is better at predicting PM<sub>2.5</sub>.

Overall, the combined neural network model considers both air pollutants and meteorological parameters, employing a three-module architecture to extract initial and deep features and make predictions. This approach was developed primarily to forecast PM<sub>2.5</sub> concentrations, which may have a substantial influence on air quality and human health.

## B. CombineDeepNet

Initially, air pollutants and meteorological parameters from six individual monitoring sites are fed into both the forward and



backward directions of an LSTM model in chronological order using the time-series notation  $x = x_t, \dots, x_{t-i}, \dots, x_{t-r+1}$  for preliminary feature extraction. Each forward and backward LSTM gate, such as the forget gate ( $f_t$ ), input gate ( $i_t$ ), output gate ( $O_t$ ), candidate cell state ( $\tilde{C}_t$ ), and hidden cell state ( $C_t$ ), extracts features from the training data. Here, the forget gate determines the extent to which information from the cell state is retained by the current unit, whereas the input gate decides how much information from the current input is saved in the cell state. Finally, the output gate controls the output state of the memory cell. The equations governing each gate are as follows:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (2)$$

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (3)$$

$$O_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (4)$$

where  $\sigma$  denotes the sigmoid activation function, the weight for the corresponding gate ( $x$ ) neurons is represented by  $W_x$ ,  $h_{t-1}$  denotes the output of the previous LSTM unit (at timestamp  $t-1$ ),  $x_t$  denotes input at the current period, and  $b_x$  denotes the biases for the corresponding gates ( $x$ ).

The following are the equations for the candidate cell state, hidden cell state, and final output:

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (5)$$

$$C_t = f_t * c_{t-1} + i_t * \tilde{C}_t \quad (6)$$

$$h_t = O_t * \tanh(C_t) \quad (7)$$

where  $\tanh$  represents the activation function.

Finally, forward and backward LSTM combines the features to create preliminary features in time series order  $O = \{O_t, \dots, O_{t-i}, \dots, O_{t-r+1}\}$ . Therefore, BiLSTM employed two hidden layers, including forward hidden layers ( $h_t^{\rightarrow}$ ) and backward hidden layers ( $h_t^{\leftarrow}$ ). Finally, the preliminary features are generated by combining  $h_t^{\rightarrow}$  and  $h_t^{\leftarrow}$ . The key equations of the BiLSTM network are shown as follows:

$$h_t^{\rightarrow} = \tanh\left(W_h^{\rightarrow} x_t + W_h^{\rightarrow} h_{t-1}^{\rightarrow} + b_h^{\rightarrow}\right) \quad (8)$$

$$h_t^{\leftarrow} = \tanh\left(W_h^{\leftarrow} x_t + W_h^{\leftarrow} h_{t-1}^{\leftarrow} + b_h^{\leftarrow}\right) \quad (9)$$

$$O_t = W^{\rightarrow} h_t^{\rightarrow} + W^{\leftarrow} h_t^{\leftarrow} + b_o \quad (10)$$

where  $O$  denotes the output features,  $w$  denotes the weight matrix,  $x$  represents the training data, and  $b$  represents the bias.

Subsequently, the output features  $O = \{O_t, \dots, O_{t-i}, \dots, O_{t-r+1}\}$  were utilized as input to the forward and backward GRU for further feature extraction. The update and reset gates within each unit of the forward and backward GRU were then employed to extract features. The update gate enables the model to evaluate the extent of historical information (from prior time steps) that should be conveyed to future steps. The reset gate enables the model to determine the amount of historical information that should be disregarded. The key equations of the update gates ( $z_t$ ), reset gates ( $r_t$ ), candidate cell state ( $\tilde{h}_t$ ), and hidden cell states ( $h_t$ ) are shown

as follows:

$$z_t = \sigma(W_z[h_{t-1}, O_t] + b_z) \quad (11)$$

$$r_t = \sigma(W_r[h_{t-1}, O_t] + b_r) \quad (12)$$

$$\tilde{h}_t = \tanh(W_h O_t + (r_t \odot h_{t-1}) * W_h + b_h) \quad (13)$$

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \quad (14)$$

where  $W_z$  and  $W_r$  represent the weight parameters,  $b_z$  and  $b_r$  denote the bias parameters, and  $\odot$  represents the Hadamard (elementwise) product operator.

Next, the forward and backward GRU merge the features and create the deep features in a temporal sequence  $y = y_t, \dots, y_{t-i}, \dots, y_{t-r+1}$ , as expressed in the equation representing the BiGRU network.

$$h_t^{\rightarrow} = \tanh\left(W_h^{\rightarrow} O_t + W_h^{\rightarrow} h_{t-1}^{\rightarrow} + b_h^{\rightarrow}\right) \quad (15)$$

$$h_t^{\leftarrow} = \tanh\left(W_h^{\leftarrow} O_t + W_h^{\leftarrow} h_{t-1}^{\leftarrow} + b_h^{\leftarrow}\right) \quad (16)$$

$$y_t = W^{\rightarrow} h_t^{\rightarrow} + W^{\leftarrow} h_t^{\leftarrow} + b_y. \quad (17)$$

Lastly, the FC layers use the extracted features to forecast the  $PM_{2.5}$  concentration. The approach presented is outlined in Algorithm 1.

### C. Evaluation Metrics

In this study, the CombineDeepNet model was compared against other prediction models using the same dataset. The forecasting accuracy of the CombineDeepNet model was evaluated using the RMSE, MAE, and  $R^2$  metrics. The following equation represents these metrics:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2} \quad (18)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - y_i| \quad (19)$$

$$R^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (20)$$

where,  $x_i$  and  $y_i$  are the test and forecast results of  $PM_{2.5}$  level, respectively,  $n$  signifies the quantity of data points within the test dataset,  $i$  represents a collection of test data, and  $\bar{x}$  and  $\bar{y}$  is the mean value of  $n$  observed sample. For RMSE and MAE metrics, lower values indicate higher forecasting accuracy. Similarly, for  $R^2$  metric, a higher value indicates higher forecasting accuracy.

## IV. RESULTS

A set of tests was carried out to assess the efficacy of the proposed model. The test dataset from six monitoring stations was employed in the CombineDeepNet model to observe the forecasting accuracy. The same dataset was then applied to five popular DL models (CNN, LSTM, GRU, CNN-LSTM, and CNN-GRU), and their prediction accuracy was recorded. Lastly, the CombineDeepNet model's prediction results were compared with those of the five popular DL models, confirming the model's

**Algorithm 1:** Pseudo Code for Proposed Method.

**Data:** Train and Test data of twelve variable ( $PM_{2.5}$ ,  $PM_{10}$ ,  $SO_2$ ,  $NO_2$ ,  $CO$ ,  $O_3$ , Temperature, Air pressure, Humidity, Rain, Wind direction, Wind speed),  $N$  : Number of monitoring stations,  $C$  : Number of Epoch

**Result:** Hourly predicted PM<sub>2.5</sub> Concentration

```

for  $n = 1 : N$  do
  1: Replace the missing value with the next value of
  the current data;
  2: Apply correlation coefficient ( $R^2$ ) to assess the
  degree of correlation between pollutants and
  meteorological data. Calculate  $R^2$  values for each
  pollutant with respect to PM2.5 concentrations.
  Select PM2.5 as the target variable for prediction
  based on the highest  $R^2$  value, as it is of
  paramount importance in air quality assessment
  and public health considerations;
  3: Apply min-max normalization to normalize the
  data;
  for  $c = 1 : C$  do
    4: Train data fed into the 50 unit BiLSTM
    layers to extract preliminary features;
    5: Apply dropout layers to maintain the
    regularization;
    6: Preliminary features fed into the three layers
    BiGRU (50 unit each) to extract the deep
    features;
    7: Apply dropout to maintain the regularization;
    8: Apply FC layer to calculate the predicted
    result;
    9: Calculate the  $MSE$  to test the model in
    each step;
    if  $MSE\ value \leq\ previous\ MSE\ value$  then
      10: Update the weight values of neurons;
      Goto Step 4;
    end
    else
      11: Stop the model training process;
      Break;
    end
  end
  12: Predict the hourly generated result;
  13: Calculate the  $RMSE$ ,  $MAE$ , and  $R^2$  to find
  out the model's performance
end

```

**end**

efficacy. The Supplementary Material includes the results from comparisons with other models.

#### A. Parameter Setting

In this article, we employed 26 300 h (75%) of data for training the models and 8760 h (25%) of data for model testing. A number of experiments were used to identify the hyperparameter configuration for this research, which resulted in the selection of the

TABLE V  
SPECIFIC HYPERPARAMETER SETTINGS ACCORDING TO THE BEST MODEL'S PERFORMANCE

Parameter	Models					
	CNN	LSTM	GRU	CNN-LSTM	CNN-GRU	Proposed
Filter Size	64	–	–	64	64	–
Kernel Size	2	–	–	5	5	–
Max Pooling	2	–	–	2	2	–
No of unit	50	50	50	50	50	50
Batch Size	64	64	64	64	64	64
Epoch	100	100	100	100	100	100
Window Size	4	4	4	4	4	4
Learning	$10^{-3}$	$10^{-3}$	$10^{-3}$	$10^{-3}$	$10^{-3}$	$10^{-3}$
Dropout	–	0.2	0.2	0.2	0.2	0.2
Optimizer	Adam	Adam	Adam	Adam	Adam	Adam
Loss	MSE	MSE	MSE	MSE	MSE	MSE

ideal hyperparameter set. To assess the prediction effectiveness of the algorithm, a validation set was utilized to calculate the mean squared error (MSE) after each epoch. Then, we selected the optimal model based on the validation set's error. We used the following methodology for the model training time: we set the number of epochs to 100 and tested the trained model against the validation set after each epoch. We changed and saved the model parameters if the forecasting model's MSE on the validation set improved. We ended the training after making several adjustments to the parameters and conducting various experiments, when the forecasting model's performance on the validation set was at its best. Lastly, in this experiment, we employed dropout to prevent the issue of model overfitting.

Also, we carefully addressed the selection of the time lag parameter for prediction tasks. We have addressed the concept of time lag differently. Instead of specifying a fixed time lag, we use a window size of 4 during the training phase. This means that when making predictions, we considered the previous 4 data points ( $t, t + 1, t + 2, t + 3$ ) as the "context" or "memory" for the model. In other words, we allowed the network to capture dependencies within this 4-h window. Furthermore, our predictions cover a time range of 0 to 180 h, which means that the model is designed to make forecasts for a relatively long-term horizon. The information within the 4-h window contributes to the model's understanding of temporal patterns and trends, allowing it to make predictions over this extended time frame.

The general hyperparameters used in this experiment are included in Table V. The layer-by-layer parameters are incorporated into the Supplementary Material. For further reference, we have included the code in the GitHub repository.<sup>1</sup>

#### B. Loss Function of the Proposed Model

Fig. 5 shows the training loss versus validation loss during training of the CombineDeepNet model over six monitoring stations, namely, (a) Aoti Zhongxin, (b) Dongsi, (c) Shunyicheng, (d) Tiantan, (e) Haidan Wanliu, and (f) Wanshou Xigong. The blue line in Fig. 5 shows the train loss, and the orange line shows the validation loss. The loss function is an important component

<sup>1</sup>[Online]. Available: <https://github.com/Prasanjit-Dey/CombineDeepNet>

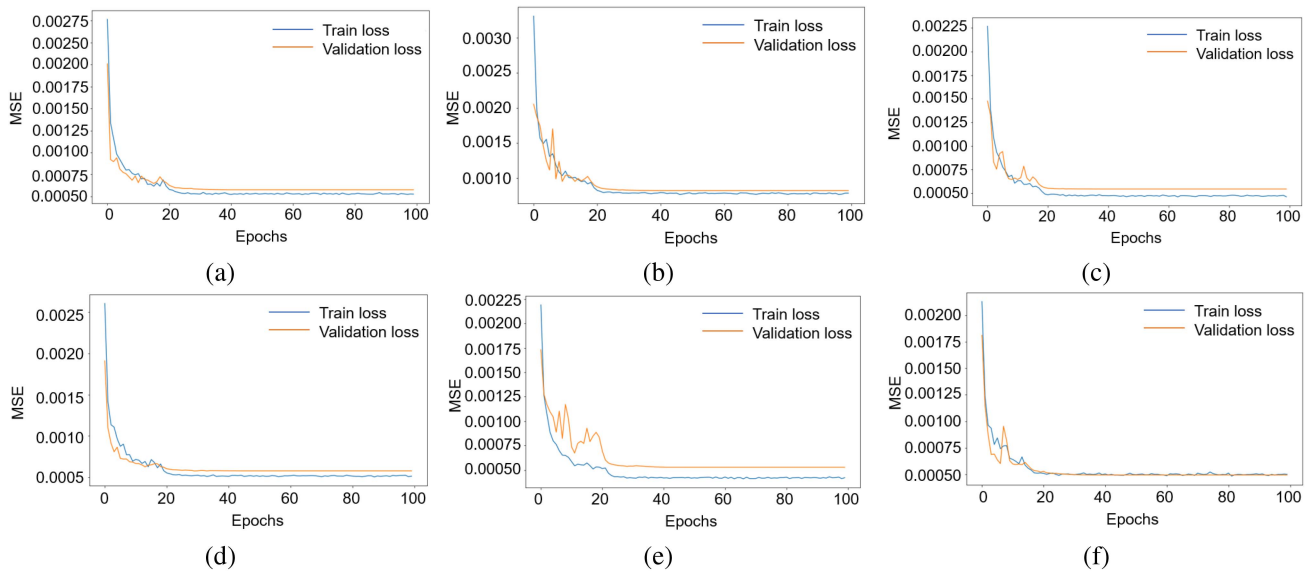


Fig. 5. Train loss versus validation loss of the proposed CombineDeepNet model over six monitoring stations: (a) Aoti Zhongxin, (b) Dongsi, (c) Shunyicheng, (d) Tiantan, (e) Haidan Wanliu, and (f) Wanshou Xigong.

of a DL model since it is used to evaluate the efficacy of a trained model on a training sample. The loss function's purpose is to minimize the disparity between the anticipated and real output. The performance of the model improves as the loss function value decreases. However, when a model is trained too closely on the input dataset, it can lead to an overfitting problem. It occurs when a model learns the noise in the data instead of the hidden patterns, which results in poor generalization performance on new data. To overcome overfitting, a loss strategy is used to remove the model's overfitting flaws. This involves adjusting the model's parameters to achieve a balance between minimizing the training loss function and maximizing the validation loss function. During the model training process, the training loss function is calculated during each iteration or epoch to monitor the model's performance on the training data. The validation loss function is then measured after each iteration or epoch to evaluate the model's generalization performance on new data. If the validation loss function starts increasing whereas the training loss function continues to decrease, it is a sign of overfitting, and the model needs to be adjusted to improve its generalization ability.

### C. Multistep Prediction of $PM_{2.5}$ Concentration Over Six Monitoring Stations

This section summarizes the findings and results of the study. Our main focus in this article is on the long-term, multistep forecasting of  $PM_{2.5}$  levels over six monitoring stations. In this context, Figs. 6 and 7 play a crucial role in illustrating how different models, namely, (a) CNN, (b) LSTM, (c) GRU, (d) CNN-LSTM, (e) CNN-GRU, and (f) the proposed model, can be generalized to the same test data over Aoti Zhongxin, and Wanshou Xigong monitoring stations. The  $x$ -axis of Figs. 6 and 7 corresponds to a time span of 180 h, indicating that the initial 180 consecutive lead times were chosen from the test data to

assess the performance of the forecasting models. In contrast, the  $y$ -axis depicts both the observed test data and the predicted  $PM_{2.5}$  concentration. Specifically, the red curve represents the observed test data, whereas the green curve denotes the predicted values.

Figs. 6 and 7 demonstrate an interesting comparison of the proposed model's efficiency with five popular DL models: (a) CNN, (b) LSTM, (c) GRU, (d) CNN-LSTM, and (e) CNN-GRU. The results show that the CombineDeepNet model excels in terms of generalization for longer lead times compared with the other models. Specifically, the green curve, which represents the forecasting value of the proposed model, closely aligns with the actual test data represented by the red curve, demonstrating a high degree of accuracy. The comparison presented in the figures highlights the significance of the CombineDeepNet model's ability to generalize better than other popular DL models. Generalization plays a pivotal role in evaluating prediction models as it assesses a model's ability to effectively handle new, previously unobserved data. A model that generalizes well can provide more accurate and reliable predictions, which makes it more suited for forecasting  $PM_{2.5}$  concentrations.

### D. Comparison With Statistical Result for Long-Term $PM_{2.5}$ Concentration Prediction

To further confirm the validation of the CombineDeepNet model by incorporating three statistical analysis techniques: RMSE, MAE, and  $R^2$ . To analyze the CombineDeepNet model's performance, the prediction task was divided into multistep lead times ranging from 40–180 h. In addition, five other popular DL models, namely, CNN, LSTM, GRU, CNN-LSTM, and CNN-GRU, were also considered for comparison.

Fig. 8 presented a comparison of the CombineDeepNet model's effectiveness in terms of RMSE with other models for the prediction of  $PM_{2.5}$  concentration over six monitoring

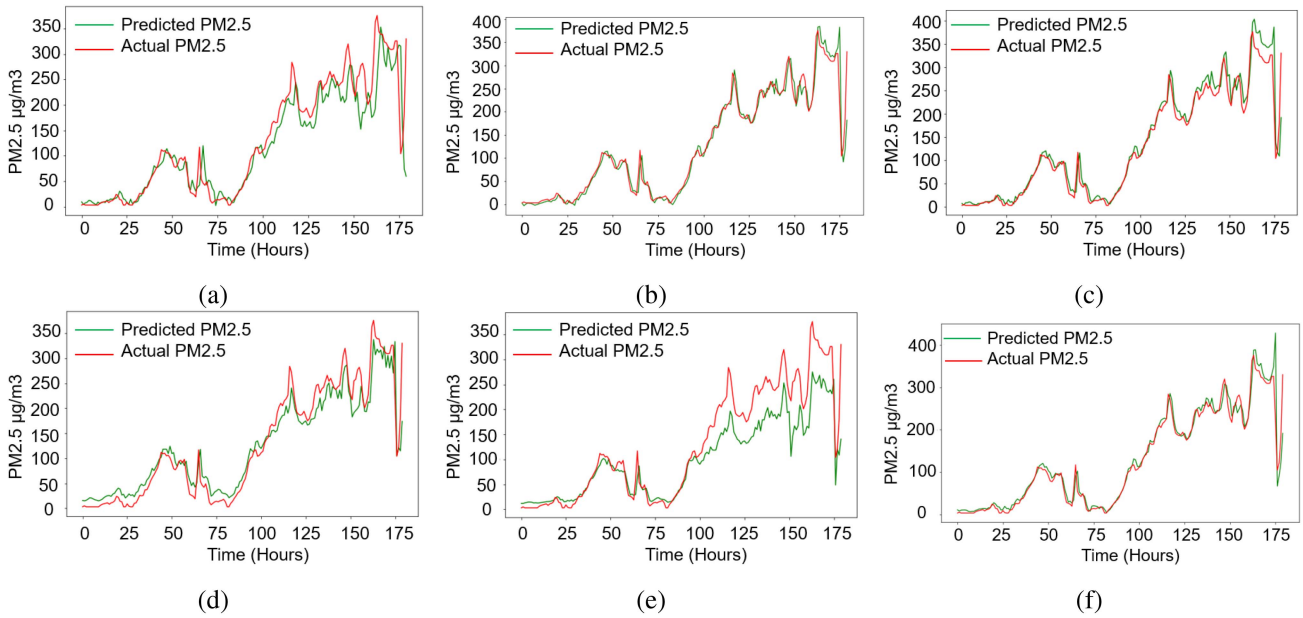


Fig. 6. Plot of actual PM<sub>2.5</sub> concentration versus predicted PM<sub>2.5</sub> concentration in the Aoti Zhongxin monitoring station for the 0 to 180 h lead times over six models: (a) CNN, (b) LSTM, (c) GRU, (d) CNN-LSTM, (e) CNN-GRU, and (f) the proposed model. It represents how well the model generalizes new data.

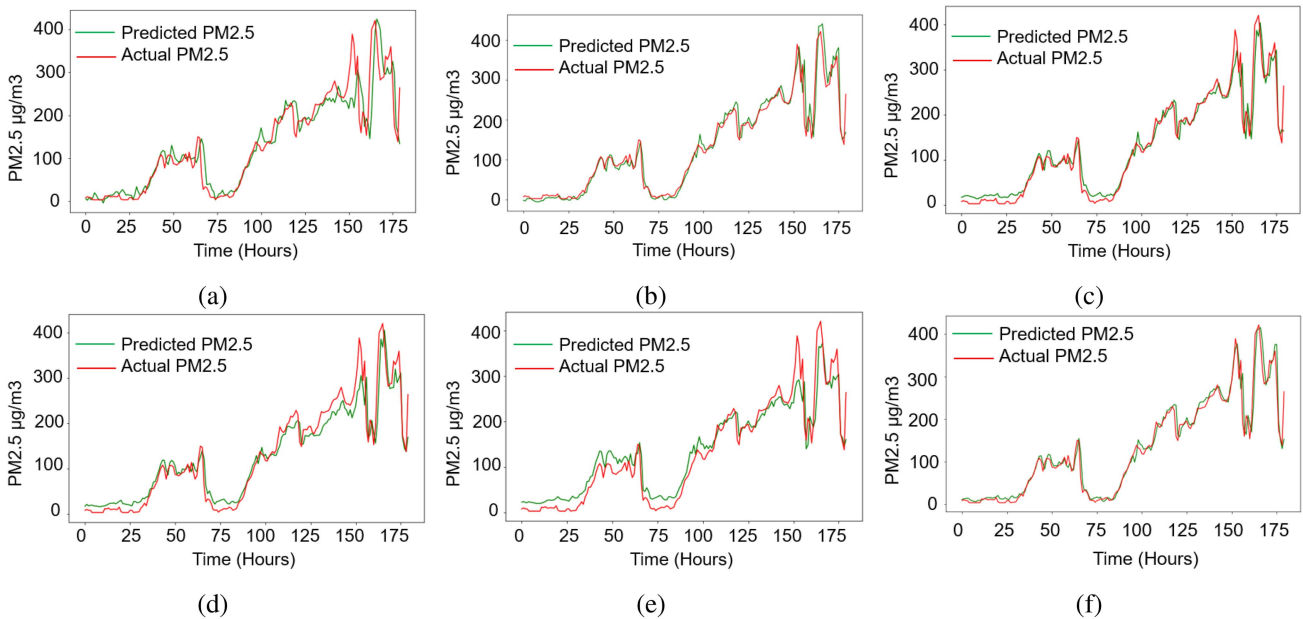


Fig. 7. Plot of actual PM<sub>2.5</sub> concentration versus predicted PM<sub>2.5</sub> concentration in the Wanshou Xigong monitoring station for the 0 to 180 h lead times over six models: (a) CNN, (b) LSTM, (c) GRU, (d) CNN-LSTM, (e) CNN-GRU, and (f) the proposed model. It represents how well the model generalizes new data.

stations: (a) Aoti Zhongxin, (b) Dongsu, (c) Shunyicheng, (d) Tiantan, (e) Haidan Wanliu, and (f) Wanshou Xigong. The findings demonstrated that the CombineDeepNet model outperformed the other five models in terms of RMSE for all six monitoring stations during long-term lead time, indicating its superiority in predicting PM<sub>2.5</sub> concentration.

Similarly, Fig. 9 presented the MAE values of the proposed model and the other five models over the same six monitoring stations. The findings demonstrated that the CombineDeepNet

model had lower MAE values than the other five models, indicating its better performance in accurately predicting PM<sub>2.5</sub> concentrations.

To further assess the effectiveness of CombineDeepNet, Table VI provides a numerical analysis comparing the CombineDeepNet model with five popular DL models, emphasizing RMSE for forecasting PM<sub>2.5</sub> concentration across the same six monitoring stations. The table highlighted that the CombineDeepNet model had significantly lower RMSE values than

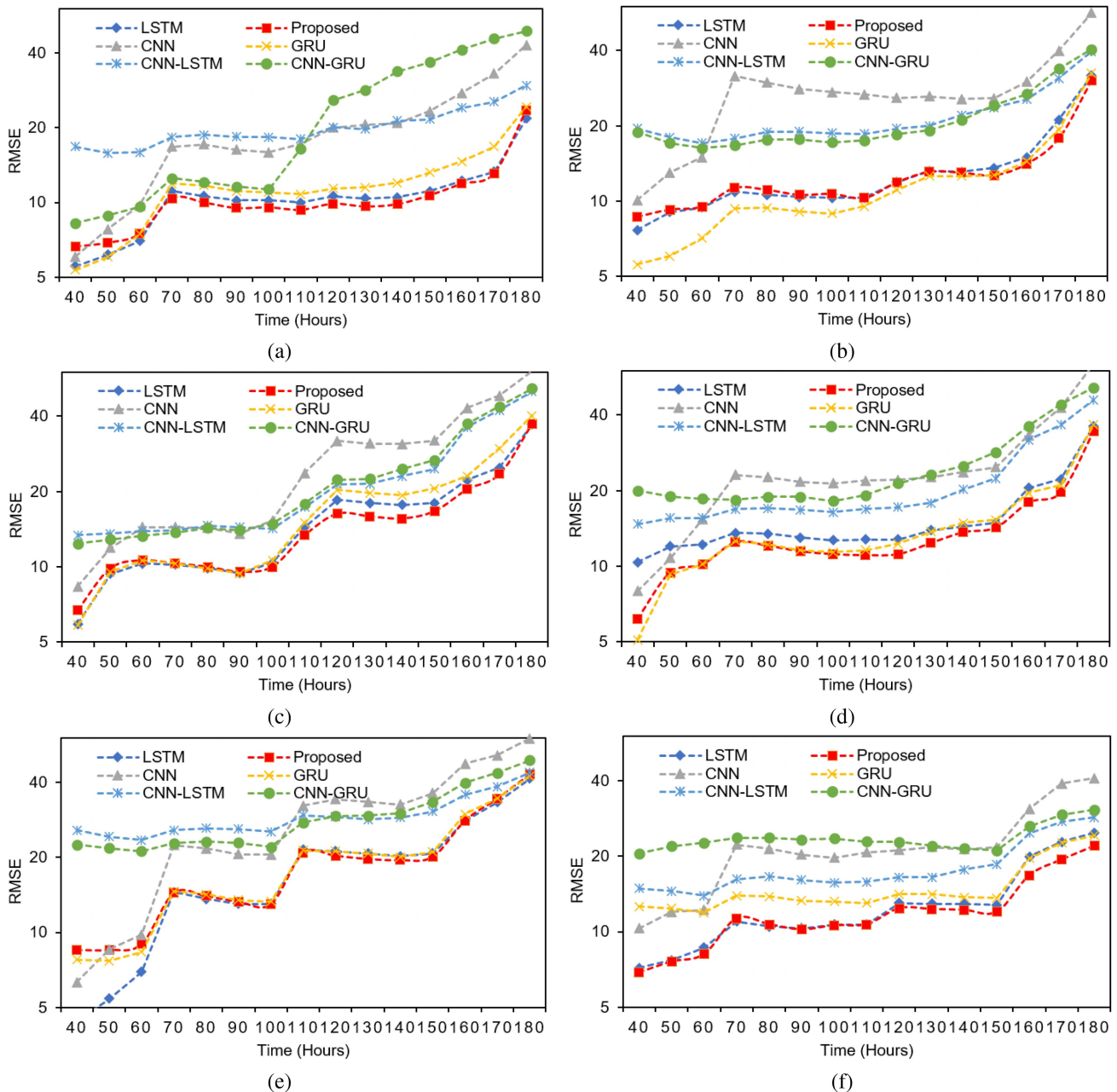


Fig. 8. Comparison of the performance of the proposed model with other models in terms of RMSE for the prediction of  $PM_{2.5}$  concentration over six monitoring stations: (a) Aoti Zhongxin, (b) Dongsi, (c) Shunyicheng, (d) Tiantan, (e) Haidian Wanliu, and (f) Wanshou Xigong.

the other five models during long-term lead time, thus, reinforcing the findings of the graphical analysis.

Similarly, Tables VII and VIII presented the numerical values of MAE and  $R^2$  for the CombineDeepNet model and the other five models over the same six monitoring stations. The results revealed that, in comparison with the other models, the CombineDeepNet model exhibited lower MAE values and higher  $R^2$  values, providing further evidence for the superiority of the proposed model in accurately forecasting  $PM_{2.5}$  concentration.

Thus, the statistical results demonstrate that the CombineDeepNet model achieves promising performance in accurately forecasting  $PM_{2.5}$  concentration levels when compared with other popular DL models.

#### E. Comparison Between the Top Three Models: CombineDeepNet, LSTM, and GRU

In this section, we further compare the prediction accuracy of the top three models: our proposed CombineDeepNet, LSTM, and GRU, using box plots. Fig. 10(a)–(f) illustrates the box plots of the RMSE values obtained for CombineDeepNet, LSTM, and GRU models across six monitoring stations: Aoti Zhongxin, Dongsi, Shunyicheng, Tiantan, Haidian Wanliu, and Wanshou Xigong. These plots demonstrate a consistent trend of RMSE values for CombineDeepNet, whereas for LSTM and GRU models, RMSE scores are more widely spread across most of the stations. In addition, the average RMSE values for

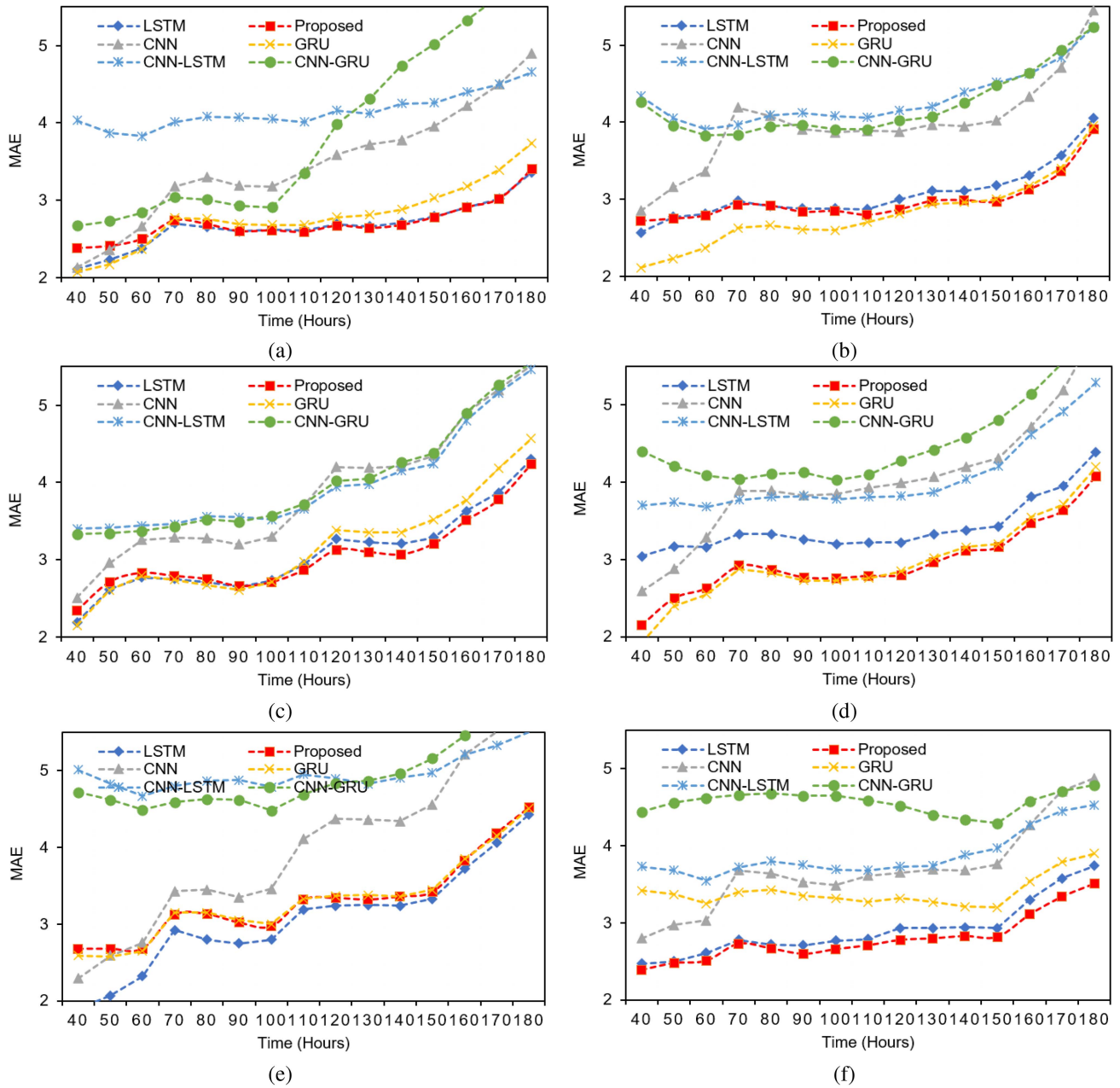


Fig. 9. Comparison of the performance of the proposed model with other models in terms of MAE for the prediction of PM<sub>2.5</sub> concentration over six monitoring stations: (a) Aoti Zhongxin, (b) Dongsi, (c) Shunyicheng, (d) Tiantan, (e) Haidan Wanliu, and (f) Wanshou Xigong.

CombineDeepNet are consistently lower than those for LSTM and GRU, highlighting the superiority of the CombineDeepNet model. Therefore, this model can be effectively applied in complex environments for long-term PM<sub>2.5</sub> concentration prediction.

#### F. Comparison With Fitting Curve

Further, the CombineDeepNet model's effectiveness was evaluated based on the degree of fit between the actual and predicted PM<sub>2.5</sub> concentrations. Fig. 11 presents the scatter plot that depicts the degree of fit between the actual and forecasting PM<sub>2.5</sub>

concentrations on the test set for the six monitoring stations over 0–180 h. The y-axis indicates the forecasted PM<sub>2.5</sub> level, whereas the x-axis indicates the actual PM<sub>2.5</sub> concentration. In Fig. 11, the red line represents the  $y = \hat{y}$  function, whereas the green dots represent the degree of variation between the actual and forecasted PM<sub>2.5</sub> concentration. We observed that the proposed model's forecasted PM<sub>2.5</sub> concentrations are generally consistent with the actual PM<sub>2.5</sub> concentrations.

Furthermore, we conducted a dispersion comparison analysis for the PM<sub>2.5</sub> concentrations between 0 and 400  $\mu\text{g}/\text{m}^3$ . The results demonstrate that the CombineDeepNet model's performance was best within this range. Overall, the results indicate

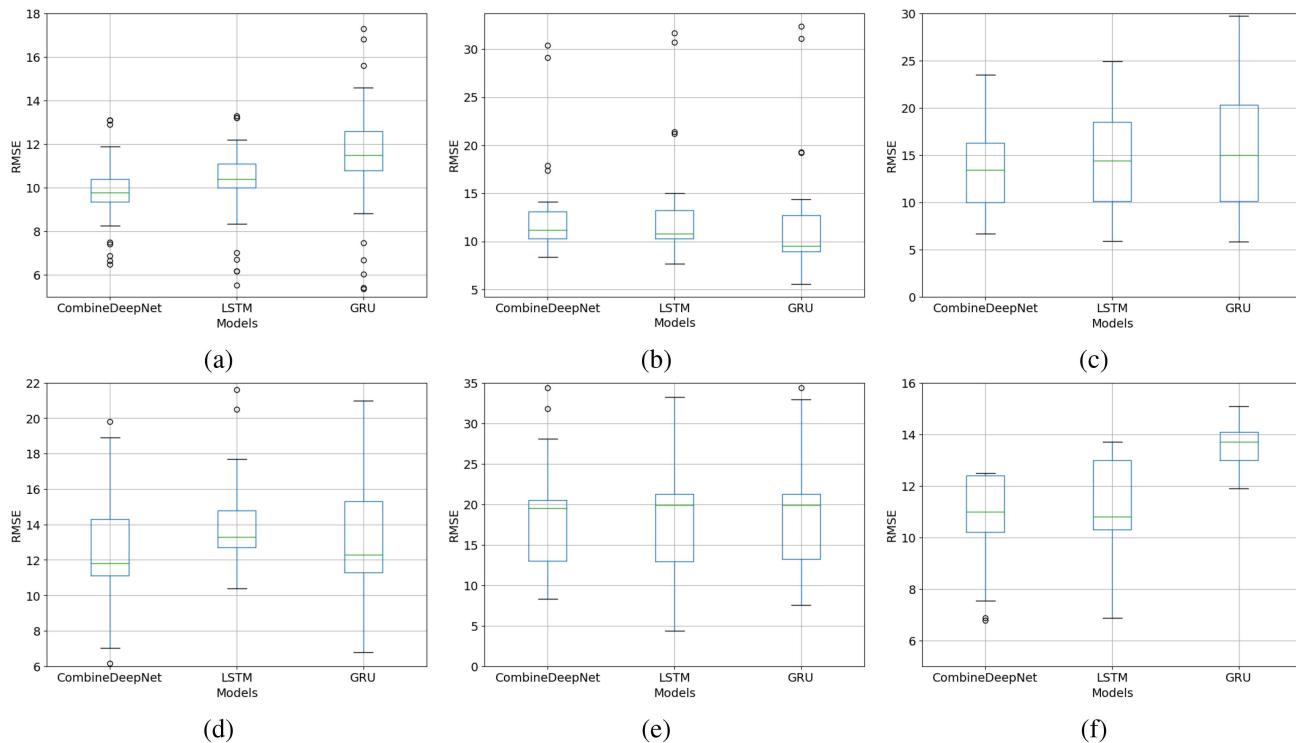


Fig. 10. Box plot of RMSE values for CombineDeepNet, LSTM, and GRU across six monitoring stations: (a) Aoti Zhongxin, (b) Dongsi, (c) Shunyicheng, (d) Tiantan, (e) Haidian Wanliu, and (f) Wanshou Xigong.

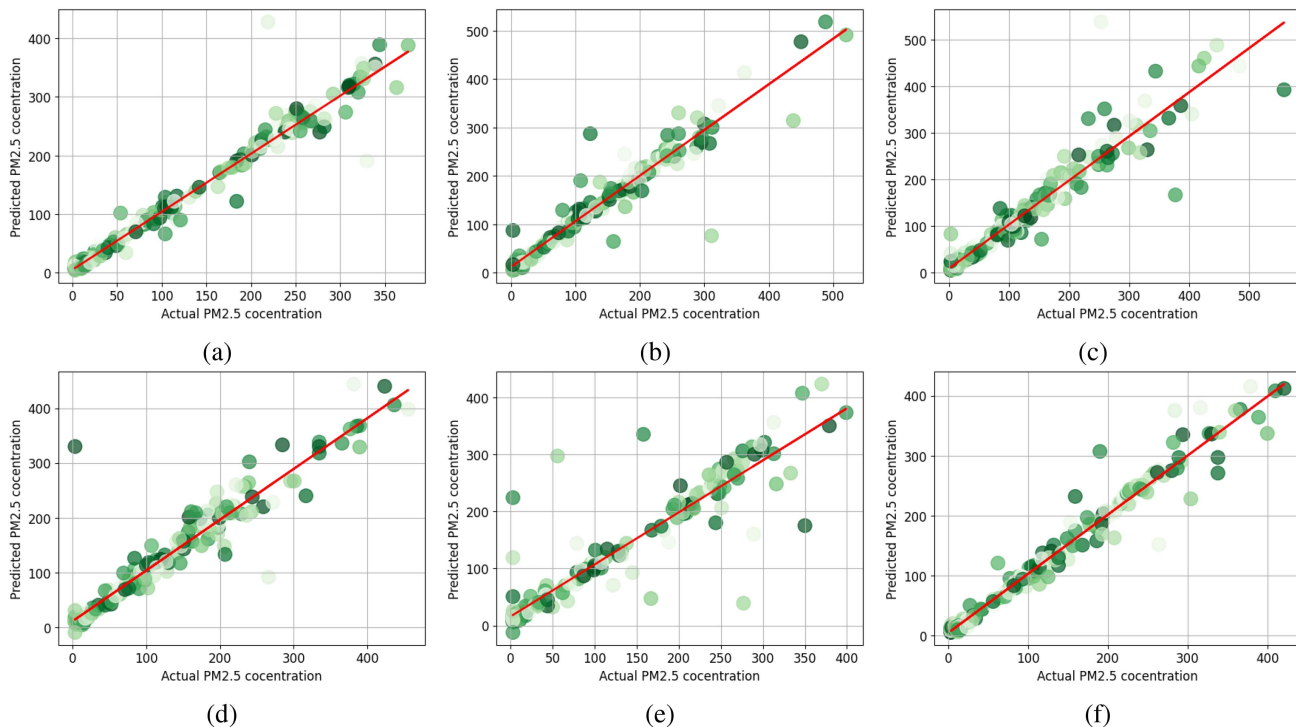


Fig. 11. Degree of fit between the actual and predicted values of the proposed model on the test set in the [0–180 h] task for the six monitoring stations: (a) Aoti Zhongxin, (b) Dongsi, (c) Shunyicheng, (d) Tiantan, (e) Haidian Wanliu, and (f) Wanshou Xigong.

TABLE VI

NUMERICAL EVALUATION OF THE COMBINEDEEPNET MODEL ALONGSIDE POPULAR DL MODELS REGARDING RMSE FOR FORECASTING PM<sub>2.5</sub> CONCENTRATION ACROSS SIX MONITORING STATIONS: (A) AOTI ZHONGXIN, (B) DONGSI, (C) SHUNYICHENG, (D) Tiantan, (E) Haidan Wanliu, and (F) Wanshou Xigong

Methods	40h	50h	60h	70h	80h	90h	100h	110h	120h	130h	140h	150h	160h	170h	180h	
<b>Zhongxin</b>	CNN	6.1	7.8	9.7	16.8	17.1	16.3	15.9	17.2	20.0	20.6	20.9	23.3	27.6	33.0	43.0
	LSTM	5.5	6.2	7.0	11.1	10.6	10.2	10.2	10.0	10.6	10.4	10.5	11.1	12.2	13.3	21.8
	GRU	5.4	6.0	7.5	11.9	11.7	11.1	11.0	10.8	11.4	11.5	12.0	13.2	14.6	16.8	24.3
	CNN-LSTM	16.8	15.8	15.9	18.3	18.7	18.4	18.3	18.0	20.1	19.7	21.3	21.6	24.0	25.4	29.6
	CNN-GRU	8.3	8.9	9.6	12.5	12.1	11.6	11.3	16.5	25.8	28.4	33.7	36.7	41.2	45.7	49.0
	Proposed	6.7	6.9	7.5	10.4	10.0	9.5	9.6	9.4	9.9	9.6	9.9	10.7	11.9	13.1	23.5
<b>Dongsi</b>	CNN	10.1	13.0	15.0	31.6	29.8	28.1	27.3	26.7	25.9	26.2	25.6	25.8	30.0	39.9	56.6
	LSTM	7.7	9.0	9.4	10.9	10.6	10.4	10.3	10.3	11.9	13.2	13.1	13.6	15.0	21.2	31.7
	GRU	5.6	6.0	7.1	9.3	9.4	9.1	8.9	9.5	11.1	12.6	12.6	12.7	14.4	19.3	32.4
	CNN-LSTM	19.5	18.0	17.1	17.8	18.9	19.0	18.7	18.6	19.5	20.0	22.1	23.7	25.5	31.0	39.7
	CNN-GRU	18.9	17.1	16.3	16.7	17.6	17.7	17.2	17.5	18.5	19.2	21.1	24.2	26.8	33.8	40.5
	Proposed	8.7	9.3	9.5	11.3	11.1	10.6	10.7	10.3	11.9	13.1	13.0	12.7	14.1	17.9	30.4
<b>Shunyicheng</b>	CNN	8.3	11.9	14.4	14.4	14.3	13.6	15.3	23.7	31.9	31.1	31.0	32.0	43.2	48.3	60.8
	LSTM	5.9	9.4	10.3	10.2	9.9	9.4	10.4	14.4	18.5	18.0	17.7	18.0	22.2	24.9	37.1
	GRU	5.8	9.5	10.6	10.3	9.8	9.4	10.6	15.0	20.3	19.7	19.3	20.6	23.0	29.7	40.2
	CNN-LSTM	13.4	13.6	13.9	14.0	14.6	14.4	14.2	17.3	21.4	21.5	23.0	24.6	36.2	42.3	50.2
	CNN-GRU	12.4	12.9	13.3	13.7	14.3	14.0	14.9	17.9	22.3	22.5	24.6	26.7	37.3	43.7	51.7
	Proposed	6.7	9.8	10.6	10.3	10.0	9.5	10.0	13.4	16.3	15.9	15.6	16.7	20.4	23.5	37.3
<b>Tiantan</b>	CNN	8.0	10.8	15.4	23.1	22.6	21.7	21.4	21.9	22.1	22.6	23.7	24.8	33.2	42.8	64.0
	LSTM	10.4	12.0	12.2	13.6	13.5	13.0	12.7	12.8	12.8	13.9	14.4	14.9	20.5	22.2	36.1
	GRU	5.1	9.3	10.2	12.6	12.2	11.6	11.4	11.6	12.3	13.7	14.9	15.3	19.6	21.0	36.7
	CNN-LSTM	14.7	15.6	15.6	16.9	17.0	16.8	16.4	16.9	17.2	17.9	20.2	22.4	31.9	36.6	45.9
	CNN-GRU	20.0	19.0	18.6	18.4	18.9	18.9	18.2	19.2	21.4	23.2	25.1	28.5	36.1	44.0	51.4
	Proposed	6.2	9.5	10.2	12.5	12.1	11.5	11.2	11.1	11.2	12.4	13.7	14.3	18.0	19.8	34.6
<b>Wanliu</b>	CNN	6.3	8.6	9.8	22.4	21.7	20.6	20.5	32.2	34.2	33.4	32.5	36.1	47.4	51.3	59.9
	LSTM	4.4	5.4	7.0	14.5	13.6	13.0	13.0	21.5	21.2	20.7	20.2	20.7	28.1	33.2	41.4
	GRU	7.8	7.7	8.3	14.6	14.2	13.5	13.3	21.3	21.2	20.7	20.1	21.0	29.8	34.4	42.4
	CNN-LSTM	25.7	24.2	23.4	25.7	26.1	26.0	25.3	29.5	29.1	28.4	28.9	30.5	35.6	38.4	43.7
	CNN-GRU	22.5	21.8	21.2	22.8	23.1	22.9	22.0	27.6	29.3	29.4	30.0	33.4	39.7	43.6	49.0
	Proposed	8.5	8.5	9.0	14.4	14.0	13.3	13.0	20.8	20.3	19.7	19.5	20.1	28.1	34.4	43.0
<b>Wanshou</b>	CNN	10.3	12.0	12.2	22.2	21.4	20.3	19.7	20.7	21.2	21.8	21.3	21.7	30.9	38.9	40.9
	LSTM	7.2	7.7	8.6	11.0	10.5	10.3	10.7	10.7	13.0	12.9	12.9	12.8	19.9	22.8	24.8
	GRU	12.6	12.4	11.9	13.9	13.8	13.3	13.2	13.0	14.1	14.1	13.7	13.6	19.6	22.6	24.2
	CNN-LSTM	14.9	14.5	13.9	16.2	16.6	16.1	15.7	15.8	16.5	16.5	17.7	18.6	24.7	27.5	28.6
	CNN-GRU	20.5	21.9	22.6	23.7	23.7	23.3	23.5	22.9	22.8	22.0	21.5	21.1	26.3	29.3	30.6
	Proposed	6.9	7.6	8.1	11.3	10.7	10.2	10.6	10.7	12.4	12.3	12.2	12.0	16.8	19.4	22.0

that the CombineDeepNet model can efficiently forecast PM<sub>2.5</sub> concentrations. The model's performance was consistent across the six monitoring stations, suggesting that it can be applied in different geographical locations. Further testing and validation on other datasets will be required to confirm the generalizability of the proposed model.

Therefore, our study presents a promising approach for predicting PM<sub>2.5</sub> concentrations, which might be valuable for air quality monitoring and management.

## V. DISCUSSION

Making long-term predictions of PM<sub>2.5</sub> is a difficult task, as it requires highly correlated historical pollutant and meteorological data. Moreover, it is difficult to establish the duration of the input sequence, which makes accurate prediction much more difficult. However, in this article, we propose CombineDeepNet, which can easily achieve long-term prediction for PM<sub>2.5</sub> concentration. The CombineDeepNet model achieves accurate long-term forecasting of PM<sub>2.5</sub> concentrations by combining various types of data, including historical pollutant and meteorological data, as well as other environmental data. This allows the model to capture complex relationships between different environmental variables and accurately predict PM<sub>2.5</sub> concentrations over a long period of time.

Table VI shows that the proposed model's RMSE values are closely related to the LSTM and GRU models, and outperform the CNN, CNN-LSTM, and CNN-GRU models for long-term lead time for all six monitoring stations. The RMSE values of the proposed model ranged from 6.70 to 23.50  $\mu\text{g}/\text{m}^3$  for Aoti Zhongxin, 8.7 to 30.4  $\mu\text{g}/\text{m}^3$  for Dongsi, 6.70 to 37.3  $\mu\text{g}/\text{m}^3$  for Shunyicheng, 6.20 to 34.6  $\mu\text{g}/\text{m}^3$  for Tiantan, 8.50 to 43.0  $\mu\text{g}/\text{m}^3$  for Haidan Wanliu, and 6.90 to 22.0  $\mu\text{g}/\text{m}^3$  for Wanshou Xigong. The RMSE values of the other five models ranged from 5.40 to 49.0  $\mu\text{g}/\text{m}^3$  for Aoti Zhongxin, 5.60 to 56.6  $\mu\text{g}/\text{m}^3$  for Dongsi, 5.80 to 60.8  $\mu\text{g}/\text{m}^3$  for Shunyicheng, 5.10 to 64.0  $\mu\text{g}/\text{m}^3$  for Tiantan, 4.40 to 59.9  $\mu\text{g}/\text{m}^3$  for Haidan Wanliu, and 7.20 to 40.9  $\mu\text{g}/\text{m}^3$  for Wanshou Xigong.

Similarly, Tables VII and VIII demonstrate that the CombineDeepNet models' MAE and R<sup>2</sup> values are closely related to the LSTM and GRU models, as well as outperform the CNN, CNN-LSTM, and CNN-GRU models for the long-term lead time for all six monitoring stations. The MAE values of the proposed model ranged from 2.40 to 3.40  $\mu\text{g}/\text{m}^3$  for Aoti Zhongxin, 2.70 to 3.90  $\mu\text{g}/\text{m}^3$  for Dongsi, 2.30 to 4.20  $\mu\text{g}/\text{m}^3$  for Shunyicheng, 2.20 to 4.10  $\mu\text{g}/\text{m}^3$  for Tiantan, 2.70 to 4.50  $\mu\text{g}/\text{m}^3$  for Haidan Wanliu, and 2.4 to 3.50  $\mu\text{g}/\text{m}^3$  for Wanshou Xigong. Conversely, the MAE values of the other five models ranged from 2.10 to 5.80  $\mu\text{g}/\text{m}^3$  for Aoti Zhongxin, 2.10 to 5.40  $\mu\text{g}/\text{m}^3$  for Dongsi, 2.10 to 5.50  $\mu\text{g}/\text{m}^3$  for Shunyicheng, 1.90 to 6.00  $\mu\text{g}/\text{m}^3$  for



TABLE VII

NUMERICAL EVALUATION OF THE COMBINEDEEPNET MODEL ALONGSIDE POPULAR DL MODELS REGARDING MAE FOR FORECASTING  $PM_{2.5}$  CONCENTRATION ACROSS SIX MONITORING STATIONS: (A) AOTI ZHONGXIN, (B) DONGSI, (C) SHUNYICHENG, (D) Tiantan, (E) Haidan Wanliu, AND (F) Wanshou Xigong

Methods	40h	50h	60h	70h	80h	90h	100h	110h	120h	130h	140h	150h	160h	170h	180h	
Zhongxin	CNN	2.1	2.4	2.7	3.2	3.3	3.2	3.2	3.4	3.6	3.7	3.8	4.0	4.2	4.5	4.9
	LSTM	2.1	2.2	2.4	2.7	2.6	2.6	2.6	2.6	2.7	2.7	2.7	2.8	2.9	3.0	3.4
	GRU	2.1	2.2	2.4	2.8	2.8	2.7	2.7	2.7	2.8	2.8	2.9	3.0	3.2	3.4	3.7
	CNN-LSTM	4.0	3.9	3.8	4.0	4.1	4.1	4.1	4.0	4.2	4.1	4.3	4.3	4.4	4.5	4.7
	CNN-GRU	2.7	2.7	2.8	3.0	3.0	2.9	2.9	3.3	4.0	4.3	4.7	5.0	5.3	5.6	5.8
	Proposed	2.4	2.4	2.5	2.7	2.7	2.6	2.6	2.6	2.7	2.6	2.7	2.8	2.9	3.0	3.4
Dongsi	CNN	2.9	3.2	3.4	4.2	4.1	3.9	3.9	3.9	3.9	4.0	3.9	4.0	4.3	4.7	5.4
	LSTM	2.6	2.8	2.8	3.0	2.9	2.9	2.9	2.9	3.0	3.1	3.1	3.2	3.3	3.6	4.1
	GRU	2.1	2.2	2.4	2.6	2.7	2.6	2.6	2.7	2.8	2.9	3.0	3.0	3.2	3.4	4.0
	CNN-LSTM	4.3	4.1	3.9	4.0	4.1	4.1	4.1	4.1	4.2	4.2	4.4	4.5	4.6	4.8	5.2
	CNN-GRU	4.3	4.0	3.8	3.8	3.9	4.0	3.9	3.9	4.0	4.1	4.3	4.5	4.6	4.9	5.2
	Proposed	2.7	2.7	2.8	2.9	2.9	2.8	2.9	2.8	2.9	3.0	3.0	3.0	3.1	3.4	3.9
Shunyicheng	CNN	2.5	3.0	3.3	3.3	3.3	3.2	3.3	3.7	4.2	4.2	4.2	4.4	4.9	5.2	5.5
	LSTM	2.2	2.6	2.8	2.7	2.7	2.6	2.7	2.9	3.3	3.2	3.2	3.3	3.6	3.9	4.3
	GRU	2.1	2.6	2.8	2.7	2.7	2.6	2.7	3.0	3.4	3.4	3.3	3.5	3.8	4.2	4.6
	CNN-LSTM	3.4	3.4	3.4	3.5	3.6	3.5	3.5	3.7	4.0	4.0	4.1	4.2	4.8	5.2	5.5
	CNN-GRU	3.3	3.3	3.4	3.4	3.5	3.5	3.6	3.7	4.0	4.0	4.3	4.4	4.9	5.3	5.5
	Proposed	2.3	2.7	2.8	2.8	2.7	2.7	2.7	2.9	3.1	3.1	3.1	3.2	3.5	3.8	4.2
Tiantan	CNN	2.6	2.9	3.3	3.9	3.9	3.8	3.9	3.9	4.0	4.1	4.2	4.3	4.7	5.2	6.0
	LSTM	3.0	3.2	3.2	3.3	3.3	3.3	3.2	3.2	3.2	3.3	3.4	3.4	3.8	3.9	4.4
	GRU	1.9	2.4	2.6	2.9	2.8	2.7	2.7	2.8	2.9	3.0	3.2	3.2	3.5	3.7	4.2
	CNN-LSTM	3.7	3.7	3.7	3.8	3.8	3.8	3.8	3.8	3.8	3.9	4.0	4.2	4.6	4.9	5.3
	CNN-GRU	4.4	4.2	4.1	4.0	4.1	4.1	4.0	4.1	4.3	4.4	4.6	4.8	5.1	5.6	5.9
	Proposed	2.2	2.5	2.6	2.9	2.9	2.8	2.8	2.8	2.8	3.0	3.1	3.2	3.5	3.6	4.1
Wanliu	CNN	2.3	2.6	2.8	3.4	3.5	3.4	3.5	4.1	4.4	4.4	4.3	4.6	5.2	5.5	5.9
	LSTM	1.9	2.1	2.3	2.9	2.8	2.7	2.8	3.2	3.2	3.3	3.2	3.3	3.7	4.1	4.4
	GRU	2.6	2.6	2.7	3.1	3.2	3.0	3.0	3.3	3.4	3.4	3.4	3.5	3.9	4.1	4.5
	CNN-LSTM	5.0	4.8	4.7	4.8	4.9	4.9	4.8	5.0	4.9	4.8	4.9	5.0	5.2	5.3	5.5
	CNN-GRU	4.7	4.6	4.5	4.6	4.6	4.6	4.5	4.7	4.8	4.9	5.0	5.2	5.5	5.7	6.0
	Proposed	2.7	2.7	2.7	3.1	3.1	3.0	3.0	3.3	3.3	3.3	3.4	3.4	3.8	4.2	4.5
Wanshou	CNN	2.8	3.0	3.0	3.7	3.6	3.5	3.5	3.6	3.7	3.7	3.7	3.8	4.3	4.7	4.9
	LSTM	2.5	2.5	2.6	2.8	2.7	2.7	2.8	2.8	2.9	2.9	2.9	2.9	3.3	3.6	3.7
	GRU	3.4	3.4	3.2	3.4	3.4	3.4	3.3	3.3	3.3	3.3	3.2	3.2	3.5	3.8	3.9
	CNN-LSTM	3.7	3.7	3.6	3.7	3.8	3.7	3.7	3.7	3.7	3.7	3.9	4.0	4.3	4.4	4.5
	CNN-GRU	4.4	4.6	4.6	4.7	4.7	4.6	4.7	4.6	4.5	4.4	4.3	4.3	4.6	4.7	4.8
	Proposed	2.4	2.5	2.5	2.7	2.7	2.6	2.7	2.7	2.8	2.8	2.8	2.8	3.1	3.3	3.5

Tiantan, 1.90 to 6.00  $\mu\text{g}/\text{m}^3$  for Haidan Wanliu, and 2.50 to 4.90  $\mu\text{g}/\text{m}^3$  for Wanshou Xigong. The  $R^2$  values of the proposed model ranged from 0.78 to 0.99 for Aoti Zhongxin, 0.68 to 0.98 for Dongsi, 0.66 to 0.96 for Shunyicheng, 0.83 to 0.97 for Tiantan, 0.60 to 0.95 for Haidan Wanliu, and 0.74 to 0.98 for Wanshou Xigong. Whereas the  $R^2$  values of the other five models ranged from 0.23 to 0.99 for Aoti Zhongxin, 0.46 to 0.98 for Dongsi, 0.22 to 0.95 for Shunyicheng, 0.08 to 0.96 for Tiantan, 0.24 to 0.97 for Haidan Wanliu, and 0.53 to 0.98 for Wanshou Xigong.

The experimental findings indicate that the CombineDeepNet is very accurate at forecasting  $PM_{2.5}$  concentrations for long-term scenarios at numerous target sites. In particular, Fig. 8 depicts the RMSE values of the CombineDeepNet model and other five models at six monitoring stations, showing that the CombineDeepNet model had the lowest RMSE values or similar RMSE values in comparison to the other models. This finding implies that the CombineDeepNet model performs better than other models in terms of long-term forecast accuracy for  $PM_{2.5}$  concentrations. In addition, the results presented in Fig. 9 show that the proposed model had the lowest MAE values for all six monitoring stations in the long-term scenarios. This finding suggests that the CombineDeepNet model is highly effective

in accurately predicting  $PM_{2.5}$  concentrations over an extended period.

The optimal values for the RMSE, MAE, and  $R^2$  for the proposed CombineDeepNet model were also reported. The optimal RMSE value was 22.0  $\mu\text{g}/\text{m}^3$ , the optimal MAE value was 3.4  $\mu\text{g}/\text{m}^3$ , and the optimal  $R^2$  value was 0.96 for the long-term lead time. These values indicate that the CombineDeepNet model can efficiently forecast  $PM_{2.5}$  concentrations for an extended period with minimal errors. The experimental findings demonstrate that the CombineDeepNet model is quite good at forecasting  $PM_{2.5}$  concentrations in long-term situations across various target sites. The suggested model seems to be able to accurately forecast  $PM_{2.5}$  concentrations with minimum errors over a long period of time, as shown by the model's superior accuracy compared with previous models.

In this research, we describe the findings of our investigation, in which we selected a consecutive 180-h test dataset and displayed it in Figs. 6 and 7 for Aoti Zhongxin and Wanshou Xigong monitoring stations. Our primary goal was to evaluate the fitting ability of the CombineDeepNet model and verify our supposition that it performs better on the new dataset. The figures also demonstrate that, in the long-term scenario, the CombineDeepNet model fits the new data better than the CNN,

TABLE VIII

NUMERICAL EVALUATION OF THE COMBINEDEEPNET MODEL ALONGSIDE POPULAR DL MODELS REGARDING R<sup>2</sup> FOR FORECASTING PM<sub>2.5</sub> CONCENTRATION ACROSS SIX MONITORING STATIONS: (A) AOTI ZHONGXIN, (B) DONGSI, (C) SHUNYICHENG, (D) TIAN TAN, (E) HAIDAN WANLIU, AND (F) WANSHOU XIGONG

Methods	40h	50h	60h	70h	80h	90h	100h	110h	120h	130h	140h	150h	160h	170h	180h	
<b>Zhongxin</b>	CNN	0.81	0.94	0.93	0.77	0.74	0.74	0.79	0.83	0.88	0.89	0.92	0.91	0.89	0.87	0.79
	LSTM	0.84	0.97	0.97	0.92	0.92	0.92	0.93	0.96	0.98	0.98	0.98	0.99	0.98	0.98	0.96
	GRU	0.86	0.97	0.96	0.90	0.90	0.90	0.92	0.95	0.97	0.98	0.98	0.98	0.98	0.98	0.96
	CNN-LSTM	0.23	0.77	0.79	0.71	0.67	0.65	0.73	0.82	0.87	0.89	0.90	0.92	0.90	0.91	0.89
	CNN-GRU	0.48	0.91	0.90	0.82	0.82	0.82	0.86	0.78	0.69	0.66	0.63	0.66	0.59	0.60	0.58
	Proposed	0.78	0.96	0.96	0.92	0.92	0.92	0.94	0.96	0.98	0.98	0.99	0.99	0.98	0.98	0.96
<b>Dongsi</b>	CNN	0.63	0.86	0.85	0.69	0.70	0.71	0.74	0.80	0.86	0.87	0.90	0.91	0.89	0.84	0.68
	LSTM	0.76	0.95	0.95	0.94	0.94	0.94	0.95	0.96	0.97	0.97	0.97	0.98	0.97	0.96	0.91
	GRU	0.86	0.98	0.97	0.96	0.95	0.95	0.96	0.97	0.97	0.97	0.97	0.98	0.97	0.97	0.91
	CNN-LSTM	0.46	0.75	0.80	0.79	0.75	0.72	0.77	0.83	0.88	0.88	0.88	0.88	0.88	0.87	0.80
	CNN-GRU	0.51	0.75	0.81	0.81	0.78	0.76	0.79	0.84	0.89	0.89	0.89	0.87	0.86	0.84	0.77
	Proposed	0.68	0.95	0.95	0.94	0.94	0.94	0.95	0.96	0.97	0.97	0.97	0.98	0.98	0.97	0.92
<b>Shunyicheng</b>	CNN	0.48	0.89	0.87	0.85	0.84	0.85	0.84	0.73	0.65	0.69	0.74	0.80	0.66	0.72	0.56
	LSTM	0.74	0.93	0.93	0.92	0.92	0.93	0.92	0.88	0.88	0.90	0.92	0.94	0.93	0.95	0.89
	GRU	0.78	0.94	0.94	0.93	0.93	0.94	0.93	0.89	0.87	0.89	0.92	0.93	0.93	0.94	0.89
	CNN-LSTM	0.22	0.80	0.84	0.81	0.77	0.77	0.80	0.76	0.77	0.78	0.79	0.82	0.66	0.70	0.61
	CNN-GRU	0.23	0.83	0.86	0.83	0.80	0.80	0.82	0.78	0.76	0.77	0.76	0.78	0.64	0.67	0.58
	Proposed	0.66	0.92	0.92	0.92	0.92	0.92	0.92	0.89	0.91	0.92	0.94	0.95	0.95	0.96	0.90
<b>Tiantan</b>	CNN	0.73	0.91	0.86	0.74	0.75	0.76	0.78	0.83	0.87	0.88	0.90	0.90	0.83	0.77	0.51
	LSTM	0.49	0.87	0.89	0.89	0.89	0.89	0.90	0.93	0.95	0.95	0.96	0.96	0.94	0.95	0.87
	GRU	0.89	0.93	0.93	0.92	0.92	0.92	0.93	0.95	0.96	0.96	0.96	0.96	0.95	0.96	0.88
	CNN-LSTM	0.16	0.77	0.80	0.82	0.80	0.79	0.82	0.86	0.89	0.89	0.88	0.87	0.78	0.79	0.68
	CNN-GRU	0.08	0.55	0.65	0.72	0.69	0.66	0.71	0.75	0.76	0.74	0.75	0.72	0.64	0.61	0.47
	Proposed	0.83	0.93	0.93	0.92	0.92	0.92	0.93	0.95	0.97	0.96	0.97	0.97	0.96	0.96	0.89
<b>Wanliu</b>	CNN	0.81	0.94	0.94	0.68	0.67	0.68	0.72	0.35	0.65	0.75	0.83	0.83	0.71	0.71	0.66
	LSTM	0.90	0.97	0.97	0.85	0.86	0.86	0.89	0.75	0.89	0.92	0.94	0.95	0.92	0.90	0.85
	GRU	0.63	0.94	0.95	0.84	0.83	0.84	0.88	0.73	0.88	0.91	0.94	0.94	0.91	0.88	0.83
	CNN-LSTM	0.35	0.35	0.53	0.42	0.35	0.31	0.49	0.43	0.71	0.78	0.82	0.83	0.80	0.80	0.75
	CNN-GRU	0.24	0.53	0.65	0.57	0.52	0.49	0.60	0.45	0.65	0.73	0.78	0.77	0.71	0.67	0.60
	Proposed	0.60	0.93	0.94	0.85	0.85	0.85	0.89	0.76	0.90	0.93	0.95	0.95	0.92	0.90	0.84
<b>Wanshou</b>	CNN	0.53	0.89	0.91	0.75	0.76	0.76	0.80	0.83	0.89	0.90	0.92	0.93	0.87	0.84	0.84
	LSTM	0.77	0.96	0.96	0.94	0.94	0.94	0.95	0.96	0.96	0.97	0.98	0.98	0.96	0.96	0.96
	GRU	0.80	0.86	0.89	0.88	0.87	0.87	0.89	0.92	0.94	0.95	0.96	0.97	0.95	0.95	0.94
	CNN-LSTM	0.73	0.82	0.86	0.83	0.81	0.81	0.85	0.87	0.91	0.92	0.93	0.93	0.90	0.91	0.91
	CNN-GRU	0.54	0.68	0.72	0.72	0.70	0.69	0.73	0.79	0.86	0.88	0.91	0.92	0.89	0.89	0.89
	Proposed	0.74	0.95	0.96	0.93	0.93	0.93	0.94	0.95	0.96	0.97	0.98	0.98	0.97	0.97	0.96

LSTM, GRU, CNN-LSTM, and CNN-GRU models. Figs. 6 and 7 depict the variations in prediction accuracy between CNN-LSTM, CNN-GRU, and models such as CombineDeepNet, CNN, LSTM, and GRU concerning short-term and long-term lead times in air quality data. These differences can be attributed to the architectural characteristics and capabilities of these models. CNN-LSTM and CNN-GRU are hybrid models that combine CNN layers for spatial feature extraction with recurrent layers (LSTM or GRU) for temporal sequence modeling. While they have the capacity to capture both short-term and long-term dependencies, the interplay between these layers can sometimes lead to a bias toward short-term patterns. In certain cases, the convolutional layers may overemphasize short-term fluctuations in the data, causing the models to focus excessively on recent data. This can lead to suboptimal performance for short-term predictions compared with models that are specifically designed for short-term tasks.

In addition, Fig. 11 displays the fitting trend of the test dataset. Furthermore, we conducted a dispersion comparison analysis for PM<sub>2.5</sub> concentrations between 0 and 400  $\mu\text{g}/\text{m}^3$ . Our results demonstrate that the CombineDeepNet model's performance was best within this range. Overall, the results indicate that the CombineDeepNet model can efficiently forecast PM<sub>2.5</sub> concentrations. Furthermore, the model's performance was consistent across the six monitoring stations, suggesting its applicability in

different geographical locations. However, further testing and validation on other datasets will be necessary to confirm the generalizability of the CombineDeepNet model.

In light of the experimental findings, we believe that the suggested CombineDeepNet can better forecast long-term PM<sub>2.5</sub> concentrations. The proposed model outperformed CNN, LSTM, GRU, CNN-LSTM, and CNN-GRU in terms of RMSE, MAE, and R<sup>2</sup> values for all six monitoring stations. The lower values of RMSE and MAE, as well as the higher values of R<sup>2</sup> indicate that the proposed model has a better fit to the observed data.

## VI. CONCLUSION

The CombineDeepNet model, proposed based on a hybrid architecture and data correlation principles, represents an effective approach for forecasting PM<sub>2.5</sub> concentration across several monitoring stations. The CombineDeepNet model consists of three modules that tract the preliminary and deep features of pollutants and meteorological data from six monitoring stations and combine the features to predict the PM<sub>2.5</sub> level. The advantages of the CombineDeepNet model include better extraction of deep features and better correlation of pollutants and meteorological data, which can be applied to multiple sites for multistep long-term forecasting of PM<sub>2.5</sub> concentration.

The experimental findings show that by completely extracting the correlated information between pollutants and meteorological data, the CombineDeepNet model outperforms conventional approaches like CNN, LSTM, GRU, CNN-LSTM, and CNN-GRU. Furthermore, the CombineDeepNet model effectively addresses the challenge of long-term dependencies, which play a crucial role in predicting pollutant hazards. Overall, the CombineDeepNet model provides a promising network for forecasting PM<sub>2.5</sub> concentration in multiple monitoring stations. The capability of the proposed model to extract profound features and establish correlations between pollutant and meteorological data, coupled with its ability to solve long-term dependency problems, makes it a valuable tool for predicting PM<sub>2.5</sub> concentration and warning against pollutant hazards.

One limitation of this article is that the information from neighboring locations is not linked to the intended target area. In future work, the correlation between nearby location data and the target location will be incorporated into the input features to improve the long-term prediction of PM<sub>2.5</sub> concentration.

## REFERENCES

- [1] S. S. Azhdari et al., "Associations of combined short-term exposures to ambient PM<sub>2.5</sub> air pollution and noise annoyance on mental health disorders: A panel study of healthy college students in Tehran," *Air Qual. Atmos. Health*, vol. 15, no. 8, pp. 1497–1505, 2022.
- [2] G. Geng et al., "Drivers of PM<sub>2.5</sub> air pollution deaths in China 2002–2017," *Nature Geosci.*, vol. 14, no. 9, pp. 645–650, 2021.
- [3] S. Li, G. Xie, J. Ren, L. Guo, Y. Yang, and X. Xu, "Urban PM<sub>2.5</sub> concentration prediction via attention-based CNN-LSTM," *Appl. Sci.*, vol. 10, no. 6, 2020, Art. no. 1953.
- [4] L.-J. Chen et al., "An open framework for participatory PM<sub>2.5</sub> monitoring in smart cities," *IEEE Access*, vol. 5, pp. 14 441–14 454, 2017.
- [5] L. Qibin et al., "The impact of PM<sub>2.5</sub> on lung function in adults with asthma," *Int. J. Tuberculosis Lung Dis.*, vol. 24, no. 6, pp. 570–576, 2020.
- [6] B. Alparslan, M. Jain, J. Wu, and S. Dev, "Analyzing air pollutant concentrations in New Delhi, India," in *Proc. Photon. Electromagn. Res. Symp.*, 2021, pp. 1191–1197.
- [7] N. Danesi, M. Jain, Y. H. Lee, and S. Dev, "Monitoring atmospheric pollutants from ground-based observations," in *Proc. IEEE USNC-URSI Radio Sci. Meeting (Joint AP-S Symp.)*, 2021, pp. 98–99.
- [8] H. Liu, Z. Duan, and C. Chen, "A hybrid framework for forecasting PM<sub>2.5</sub> concentrations using multi-step deterministic and probabilistic strategy," *Air Qual. Atmos. Health*, vol. 12, no. 7, pp. 785–795, 2019.
- [9] S. Mahajan, L.-J. Chen, and T.-C. Tsai, "An empirical study of PM<sub>2.5</sub> forecasting using neural network," in *Proc. IEEE SmartWorld Ubiquitous Intell. Comput. Adv. Trusted Comput. Scalable Comput. Commun. Cloud Big Data Comput. Internet People Smart City Innov.*, 2017, pp. 1–7.
- [10] Y. Hu and L. Plonsky, "Statistical assumptions in L2 research: A systematic review," *Second Lang. Res.*, vol. 37, no. 1, pp. 171–184, 2021.
- [11] K. Harishkumar et al., "Forecasting air pollution particulate matter (PM<sub>2.5</sub>) using machine learning regression models," *Procedia Comput. Sci.*, vol. 171, pp. 2057–2066, 2020.
- [12] K. Huang et al., "Predicting monthly high-resolution PM<sub>2.5</sub> concentrations with random forest model in the north China plain," *Environ. Pollut.*, vol. 242, pp. 675–683, 2018.
- [13] B. Guo et al., "Estimating PM<sub>2.5</sub> concentrations via random forest method using satellite, auxiliary, and ground-level station dataset at multiple temporal scales across China in 2017," *Sci. Total Environ.*, vol. 778, 2021, Art. no. 146288.
- [14] W. Yang, M. Deng, F. Xu, and H. Wang, "Prediction of hourly PM<sub>2.5</sub> using a space-time support vector regression model," *Atmospheric Environ.*, vol. 181, pp. 12–19, 2018.
- [15] R. Chuentawat and Y. Kan-ngan, "The comparison of PM<sub>2.5</sub> forecasting methods in the form of multivariate and univariate time series based on support vector machine and genetic algorithm," in *Proc. 15th Int. Conf. Elect. Eng./Electron. Comput. Telecommun. Inf. Technol.*, 2018, pp. 572–575.
- [16] R. Feng, H. Gao, K. Luo, and J.-R. Fan, "Analysis and accurate prediction of ambient PM<sub>2.5</sub> in China using multi-layer perceptron," *Atmospheric Environ.*, vol. 232, 2020, Art. no. 117534.
- [17] J.-H. Chang and C.-Y. Tseng, "Analysis of correlation between secondary PM<sub>2.5</sub> and factory pollution sources by using ANN and the correlation coefficient," *IEEE Access*, vol. 5, pp. 22 812–22 822, 2017.
- [18] G. Polezer et al., "Assessing the impact of PM<sub>2.5</sub> on respiratory disease using artificial neural networks," *Environ. Pollut.*, vol. 235, pp. 394–403, 2018.
- [19] M. Xu and Y.-X. Wang, "Quantifying PM<sub>2.5</sub> concentrations from multi-weather sensors using hidden Markov models," *IEEE Sensors J.*, vol. 16, no. 1, pp. 22–23, Jan. 2016.
- [20] L. Zhang, Y. Ji, T. Liu, and J. Li, "PM<sub>2.5</sub> prediction based on XGBoost," in *Proc. 7th Int. Conf. Inf. Sci. Control Eng.*, 2020, pp. 1011–1014.
- [21] Y. Su, "Prediction of air quality based on gradient boosting machine method," in *Proc. Int. Conf. Big Data Informatization Educ.*, 2020, pp. 395–397.
- [22] A. Banga, R. Ahuja, and S. C. Sharma, "Performance analysis of regression algorithms and feature selection techniques to predict PM<sub>2.5</sub> in smart cities," *Int. J. Syst. Assurance Eng. Manage.*, vol. 14, pp. 1–14, 2021.
- [23] M. Jain, S. Manandhar, Y. H. Lee, S. Winkler, and S. Dev, "Forecasting precipitable water vapor using LSTMs," in *Proc. IEEE USNC-CNC-URSI North Amer. Radio Sci. Meeting (Joint AP-S Symp.)*, 2020, pp. 147–148.
- [24] Z. Y. Ho, M. Jain, and S. Dev, "Multivariate convolutional LSTMs for relative humidity forecasting," in *Proc. Photon. Electromagn. Res. Symp.*, 2021, pp. 2317–2323.
- [25] M. M. Sheikholeslami, S. Nadi, A. A. Naeini, and P. Ghamisi, "An efficient deep unsupervised superresolution model for remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1937–1945, 2020.
- [26] H. Shojaei, S. Nadi, H. S.-Moghadam, A. Tayyebi, and J. V. Genderen, "An efficient built-up land expansion model using a modified u-net," *Int. J. Digit. Earth*, vol. 15, no. 1, pp. 148–163, 2022.
- [27] Y.-T. Tsai, Y.-R. Zeng, and Y.-S. Chang, "Air pollution forecasting using RNN with LSTM," in *Proc. IEEE 16th Int. Conf. Dependable Auton. Sec. Comput., 16th Int. Conf. Pervasive Intell. Comput., 4th Int. Conf. Big Data Intell. Comput. Cyber Sci. Technol. Congr.*, 2018, pp. 1074–1079.
- [28] L. Shi, H. Zhang, X. Xu, M. Han, and P. Zuo, "A balanced social LSTM for PM<sub>2.5</sub> concentration prediction based on local spatiotemporal correlation," *Chemosphere*, vol. 291, 2022, Art. no. 133124.
- [29] G. Huang, X. Li, B. Zhang, and J. Ren, "PM<sub>2.5</sub> concentration forecasting at surface monitoring sites using GRU neural network based on empirical mode decomposition," *Sci. Total Environ.*, vol. 768, 2021, Art. no. 144516.
- [30] Z. Zhang, X. Ma, and K. Yan, "A deep learning model for PM<sub>2.5</sub> concentration prediction," in *Proc. IEEE Int. Conf. Dependable Auton. Sec. Comput., Int. Conf. Pervasive Intell. Comput., Int. Conf. Cloud Big Data Comput., Int. Conf. Cyber Sci. Technol. Congr.*, 2021, pp. 428–433.
- [31] T. Li, M. Hua, and X. Wu, "A hybrid CNN-LSTM model for forecasting particulate matter (PM<sub>2.5</sub>)," *IEEE Access*, vol. 8, pp. 26 933–26 940, 2020.
- [32] Y. Chen et al., "Research on an attention-based hybrid CNN and BiLSTM model for air pollutant concentration prediction," in *Proc. 6th Int. Conf. Comput. Intell. Appl.*, 2021, pp. 79–84.
- [33] B. Zhang, H. Zhang, G. Zhao, and J. Lian, "Constructing a PM<sub>2.5</sub> concentration prediction model by combining auto-encoder with Bi-LSTM neural networks," *Environ. Modelling Softw.*, vol. 124, 2020, Art. no. 104600.
- [34] M. Zhang, D. Wu, and R. Xue, "Hourly prediction of PM<sub>2.5</sub> concentration in Beijing based on Bi-LSTM neural network," *Multimedia Tools Appl.*, vol. 80, no. 16, pp. 24 455–24 468, 2021.
- [35] M. Faraji, S. Nadi, O. Ghaffarparasad, S. Homayoni, and K. Downey, "An integrated 3D CNN-GRU deep learning method for short-term prediction of PM<sub>2.5</sub> concentration in urban environment," *Sci. Total Environ.*, vol. 834, 2022, Art. no. 155324.
- [36] K. Zhang, X. Yang, H. Cao, J. Thé, Z. Tan, and H. Yu, "Multi-step forecast of PM<sub>2.5</sub> and PM<sub>10</sub> concentrations using convolutional neural network integrated with spatial-temporal attention and residual learning," *Environ. Int.*, vol. 171, 2023, Art. no. 107691.
- [37] N. Natsagdorj et al., "Prediction of PM<sub>2.5</sub> concentration in Ulaanbaatar with deep learning models," *Urban Climate*, vol. 47, 2023, Art. no. 101357.
- [38] U. A. Bhatti et al., "Time series analysis and forecasting of air pollution particulate matter (PM<sub>2.5</sub>): An SARIMA and factor analysis approach," *IEEE Access*, vol. 9, pp. 41 019–41 031, 2021.
- [39] 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<sub>2.5</sub> forecasting," *Sci. Total Environ.*, vol. 651, pp. 230–240, 2019.

- [40] A. Masood and K. Ahmad, "A model for particulate matter (PM<sub>2.5</sub>) prediction for delhi based on machine learning approaches," *Procedia Comput. Sci.*, vol. 167, pp. 2101–2110, 2020.
- [41] G. Chen, J. Guang, Y. Xue, Y. Li, Y. Che, and S. Gong, "A physically based PM<sub>2.5</sub> estimation method using AERONET data in Beijing area," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 6, pp. 1957–1965, Jun. 2018.
- [42] J. Wu, T. Li, C. Zhang, Q. Cheng, and H. Shen, "Hourly PM<sub>2.5</sub> concentration monitoring with spatiotemporal continuity by the fusion of satellite and station observations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 8019–8032, 2021.
- [43] J. Zhao, F. Deng, Y. Cai, and J. Chen, "Long short-term memory-fully connected (LSTM-FC) neural network for PM<sub>2.5</sub> concentration prediction," *Chemosphere*, vol. 220, pp. 486–492, 2019.
- [44] Y.-S. Chang, H.-T. Chiao, S. Abimannan, Y.-P. Huang, Y.-T. Tsai, and K.-M. Lin, "An LSTM-based aggregated model for air pollution forecasting," *Atmospheric Pollut. Res.*, vol. 11, no. 8, pp. 1451–1463, 2020.
- [45] I. Yeo, Y. Choi, Y. Lops, and A. Sayeed, "Efficient PM<sub>2.5</sub> forecasting using geographical correlation based on integrated deep learning algorithms," *Neural Comput. Appl.*, vol. 33, no. 22, pp. 15 073–15 089, 2021.
- [46] Z. Yan, S. Wang, D. Ma, B. Liu, H. Lin, and S. Li, "Meteorological factors affecting pan evaporation in the Haihe River Basin, China," *Water*, vol. 11, no. 2, 2019, Art. no. 317.
- [47] O. B. Sezer and A. M. Ozbayoglu, "Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach," *Appl. Soft Comput.*, vol. 70, pp. 525–538, 2018.
- [48] M. A. Monteiro, B. Grové, and N. Matthews, "Developing a moving average crossover strategy as an alternative hedging strategy for the South Africa maize market," *Agriculture*, vol. 12, no. 8, p. 2022, Art. no. 1227.
- [49] N. R. Martins and G. C. Da Graca, "Impact of PM<sub>2.5</sub> in indoor urban environments: A review," *Sustain. Cities Soc.*, vol. 42, pp. 259–275, 2018.
- [50] M. Zhixiang, C. Cai, M. Xiangwei, L. Wei, and Z. Chuazhen, "Short-term effects of different PM<sub>2.5</sub> thresholds on daily all-cause mortality in Jinan, China," 2021.
- [51] J. Ma and J. C. Cheng, "Estimation of the building energy use intensity in the urban scale by integrating GIS and Big Data technology," *Appl. Energy*, vol. 183, pp. 182–192, 2016.
- [52] L. Zhang, J. Na, J. Zhu, Z. Shi, C. Zou, and L. Yang, "Spatiotemporal causal convolutional network for forecasting hourly PM<sub>2.5</sub> concentrations in Beijing, China," *Comput. Geosci.*, vol. 155, 2021, Art. no. 104869.
- [53] R. Cahuantzi, X. Chen, and S. Güttel, "A comparison of LSTM and GRU networks for learning symbolic sequences," in *Proc. Sci. Inf. Conf.*, 2023, pp. 771–785.
- [54] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [55] Z. Liu, Z. Xu, J. Jin, Z. Shen, and T. Darrell, "Dropout reduces underfitting," 2023, *arXiv:2303.01500*.



**Prasanjit Dey** received the M.Tech. degree in information technology from the Maulana Abul Kalam Azad University of Technology, Kolkata, India, in 2016. He is working toward the Ph.D. degree in short-term air quality prediction using deep neural networks with the School of Computer Science, Technological University Dublin, Dublin, Ireland.

After that, he was a Project Associate at CSIR-Central Institute of Mining and Fuel Research, Dhanbad, India. His research project is on monitoring and short-term forecasting of atmospheric air pollutants using deep neural networks. His research interests include predicting time-series data using deep learning and machine learning algorithms.



**Soumyabrata Dev** (Member, IEEE) received the B.Tech. degree (*summa cum laude*) in electronics and communication from the National Institute of Technology Silchar, Silchar, India, in 2010, and the Ph.D. degree in electrical and electronics engineering from Nanyang Technological University (NTU), Singapore, in 2017.

He is currently an Assistant Professor with the School of Computer Science, University College Dublin, Dublin, Ireland, and an SFI Funded Investigator with ADAPT SFI Research Centre, Dublin.

In 2015, he was a Visiting Student with Audiovisual Communication Laboratory (LCAV), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. He has authored over 150 publications in leading journals and conferences. His research interests include remote sensing, image processing, machine learning, and deep learning.



**Bianca Schoen-Phelan** (Member, IEEE) received the B.Sc. degree in computer science from the University of Applied Sciences, Darmstadt Germany, in 2002, and the Ph.D. degree in computer science from University College Dublin, Dublin, Ireland, in 2007.

She is currently a Senior Lecturer and Researcher with the School of Computer Science, TU Dublin City Campus, Dublin. She has previous industry experience as well as strong research experience on commercially focused and applied research projects. Her core areas of expertise are deep learning analysis

and VR simulations of geospatial data, with a special focus on environmental analytics topics. Her research interests are focused on the areas of geospatial data, which includes airborne LiDAR data, satellite data, and crowd-sourced data that contains geospatial components.