sustainability

MDPI

Article

A New Combination Model for Air Pollutant Concentration Prediction: A Case Study of Xi'an, China

Fan Yang *, Guangqiu Huang * and Yanan Li

School of Management, Xi'an University of Architecture and Technology, Xi'an 710055, China; lyn@xauat.edu.cn
* Correspondence: yangfan@xauxt.edu.cn (F.Y.); gqhuang@xauat.edu.cn (G.H.)

Abstract: As energy demand continues to increase, the environmental pollution problem is becoming more severe. Governments and researchers have made great efforts to avoid and reduce air pollution. The prediction of $PM_{2.5}$, as an important index affecting air quality, has great significance. However, $PM_{2.5}$ concentration has a complex change process that makes its prediction challenging. By calculating both $PM_{2.5}$ concentration and that of other pollutants in the atmosphere and meteorological factors, it is evident that the variation in $PM_{2.5}$ concentration is influenced by multiple factors, and that relevant features also influence each other. To reduce the calculated loss, with full consideration given to the influencing factors, we used the maximum correlation and minimum redundancy (MRMR) algorithm to calculate the correlation and redundancy between features. In addition, it is known from the Brock–Dechert–Scheinman (BDS) statistical results that the change in $PM_{2.5}$ is nonlinear. Due to the outstanding performance of bidirectional long short-term memory (BiLSTM) neural networks in nonlinear prediction, we constructed an encoder–decoder model based on BiLSTM, named ED-BiLSTM, to predict the $PM_{2.5}$ concentration at monitoring stations. For areas without monitoring sites, due to the lack of historical data, the application of neural networks is limited. To obtain the pollutant concentration distribution in the study area, we divided the study area into a 1 km × 1 km grid and combined the ED-BiLSTM model via the use of the inverse distance weighting (IDW) algorithm to obtain the $PM_{2.5}$ concentration values in a region without monitoring stations. Finally, ArcGIS was used to visualize the results. The data for the case study were obtained from Xi'an. The results show that, compared with the standard long short-term memory (LSTM) model, the RMSE, MAE, and MAPE of our proposed model were reduced by 24.06%, 24.93%, and 22.9%, respectively. The proposed model has a low error for $PM_{2.5}$ prediction and can provide a theoretical basis for the formulation of environmental protection policies.

Keywords: air pollution; $PM_{2.5}$ prediction; correlation analysis; feature selection; bidirectional long short-term memory; encoder–decoder; inverse distance weighting; grid interpolation

check for updates

Citation: Yang, F.; Huang, G.; Li, Y. A New Combination Model for Air Pollutant Concentration Prediction: A Case Study of Xi'an, China. *Sustainability* **2023**, *15*, 9713. https://doi.org/10.3390/su15129713

Academic Editors: Linchuan Yang, Bao-Jie He and Junqing Tang

Received: 17 May 2023
Revised: 8 June 2023
Accepted: 14 June 2023
Published: 17 June 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Environment pollution is becoming increasingly severe as industrialization and urbanization accelerate, attracting the attention of governments and experts worldwide [1]. $PM_{2.5}$, an essential factor causing haze [2] and reducing atmospheric visibility [3], contributes much more to heavily polluted weather than other pollutants [4]. Many researchers have studied the relationship between $PM_{2.5}$ and health. One study shows that breathing air with excessive pollution levels for a long time can worsen cardiovascular and respiratory diseases [5]. Ashtari et al. [6] demonstrated a positive relationship between air quality and multiple sclerosis severity through an eight-year study in the Isfahan region of Iran. There are an estimated 1.65 million to 2.19 million early deaths every year associated with $PM_{2.5}$ [7]. In order to prevent air pollution and protect the environment, air pollution prediction is crucial [8]. An accurate $PM_{2.5}$ concentration forecast model can help the public to make sensible travel arrangements, lessen the risks to people's health, and guide governments and the public to make the right choices.

Obtaining accurate $PM_{2.5}$ concentration prediction values through the existing large quantity of meteorological and pollutant historical monitoring data is a crucial issue in air pollution research. The concentration of pollutants at a given point in time can affect future pollutant concentrations. The effects may last hours, days, or longer. At present, the time granularity in pollutant prediction is mainly assessed in hourly [9] and daily [10] terms, and some studies have examined pollutant emissions on a yearly scale [11,12]. The current pollutant prediction models are classified into mechanism models, statistical models and deep learning models. Mechanism models simulate the chemical and physical evolution of pollutant transport through atmospheric kinetic theory. The typical methods are community multiscale air quality (CMAQ) [13], the nested air quality prediction and modeling system (NAQPMS) [14], and the weather research and forecasting (WRF) model [15]. The advantage of these methods is that the modeling does not require historical data. Still, the complexity of model calculations and parameter settings requires a strong background in atmospheric science, limiting the practical application of the method. With the development of sensing technology, data acquisition has become easy. Statistical models are used for pollutant prediction. Statistical models are divided into traditional statistical models and machine learning models. For pollutant prediction, the traditional statistical models known as autoregressive integrated moving average (ARIMA) [16] and autoregressive moving average with extra input (ARMAX) [17] are frequently utilized. However, they are based on linear presumptions and have straightforward model structures that cannot be used to solve nonlinear issues. Machine learning techniques are also used to address nonlinear issues in pollutant prediction. Machine learning-based prediction methods include back propagation (BP) [18], extreme learning machine (ELM) [19], and support vector regression (SVR) [20]. Even though machine learning algorithms have produced positive results for pollutant prediction, $PM_{2.5}$ concentration sequences exhibit complex fluctuations over time and are influenced by a variety of factors. The prediction accuracy of statistical models is constrained because they cannot satisfy the requirements of multivariate nonlinear data prediction.

With the development of computer technology, deep learning has demonstrated outstanding nonlinear data processing capabilities. For example, computer vision [21,22], image classification [23], and natural language processing [24] have demonstrated the effectiveness of deep learning models. Because recurrent neural networks (RNNs) are good at dealing with the difficulties encountered when dealing with sequential data, they have been used in numerous models to predict pollution concentration [25]. Nevertheless, RNNs suffer from gradient disappearance and gradient explosion when predicting long time series. Long short-term memory (LSTM) [26], as a variant of an RNN, can solve this problem [27]. To demonstrate the viability of using LSTM models in short-term prediction, LSTM-based models have been compared with statistical models [28–30]. Both Wu et al. [31] and Wang et al. [32] predicted air pollution using LSTM; the difference was that Wang et al. [32] took into account the role of influencing factors through the chi-square test. As the LSTM can only encode data in one direction, Graves and Schmidhuber [33] proposed the bidirectional long short-term memory (BiLSTM) model, which contained forward LSTM unit and a backward LSTM unit. Experiments have shown that the BiLSTM is more accurate and stable in pollutant prediction than the single LSTM model, particularly for maximums and minimums. A single neural network cannot handle complex prediction tasks since the changes in pollutant concentration are complex and have spatiotemporal characteristics. So, more and more combination models are being developed. Bai et al. [34] considered the seasonal characteristics of $PM_{2.5}$ concentration variation and then obtained $PM_{2.5}$ predicted values using a stacked auto-encoder model. Pak et al. [35] extracted the spatial features of multidimensional time series using a convolutional neural network and then used LSTM to make predictions. Chang et al. [36] achieved $PM_{2.5}$ prediction by aggregating multiple LSTM units and considered the correlation of influencing factors in the model. A combination of auto-encoder and BiLSTM was proposed as a new model [37]. Furthermore, the encoder–decoder model based on LSTM was also applied to time series

prediction [38,39]. Because encoders and decoders based on LSTM can fully extract time series features, they are widely used in pollutant concentration prediction [40,41]. These models use an LSTM unit as an encoder to extract the features of the sequence data and then use another LSTM unit as a decoder to decode the encoded vector to obtain the predicted values. The existing studies show that BiLSTM performs better than LSTM in time series prediction and so we construct an encoder–decoder prediction model using BiLSTM.

In addition to prediction models, the presence of excessive features in $PM_{2.5}$ short-term prediction can cause information redundancy and lead to error accumulation [42]. Multiple-series data can lead to the overfitting of models [43]. Based on the above issues, we select features from the many pollutants and a large quantity of meteorological historical monitoring data. Feature selection improves prediction accuracy by selecting strongly correlated variables from multidimensional data. The methods commonly used for feature selection include mutual information [44,45], Pearson correlation coefficient [46,47], and Kendall correlation coefficient [48]. As part of their study, Bai et al. [34] calculated the Kendall correlation coefficient for $PM_{2.5}$ and meteorological variables. Zhang et al. [37] solved the same problem using the Pearson correlation coefficient, but the Pearson correlation coefficient is not a good measure of nonlinear problems [49]. To measure correlations between $PM_{2.5}$ and other pollutants, Wang et al. [2] used interval gray incidence analysis. However, the above study did not consider the redundancy between features. This study shows that adjacent monitoring sites can provide important information for the prediction model. To overcome the limitation, we use the maximum correlation minimum redundancy (MRMR) algorithm based on mutual information [50] to decide the model's input variables. Mutual information is also used to calculate the correlation of $PM_{2.5}$ between adjacent stations.

Although the existing methods can adequately extract time series features, they only consider the correlation between multidimensional feature variables and ignore the effect of feature redundancy on model accuracy. In addition, few studies have been conducted to predict pollutant concentrations without monitoring stations. In this paper, a new combination model for $PM_{2.5}$ concentration prediction is proposed based on the above analysis. This study contributes to the following areas:

(1) Based on the nonlinear characteristics of the $PM_{2.5}$ concentration series, this paper adopts the MRMR algorithm based on mutual information for feature selection, removing the effect of redundant features with full consideration of the influencing factors.
(2) To fully extract of spatial and temporal features, we construct a dual encoder–decoder prediction model based on BiLSTM for predicting $PM_{2.5}$ concentration named ED-BiLSTM. The proposed model accurately predicts $PM_{2.5}$ concentration.
(3) To obtain the $PM_{2.5}$ concentration distribution in Xi'an, China, we combine the ED-BiLSTM model with inverse distance weight (IDW) spatial interpolation to get the predicted $PM_{2.5}$ concentration values without monitoring stations. Finally, the overall distribution is obtained by ArcGIS visualization.

2. Study Area and Available Data

2.1. Study Data

The study area is located in Xi'an, Shaanxi Province, China. Xi'an is in the middle of the Guanzhong Basin, with the Loess Plateau in the north and the Qinling Mountains in the south. The area is long and narrow from east to west, with a low altitude in the east and high altitude in the west. Only the cold air from the northwest is beneficial to the diffusion of pollutants, while the contaminants entering the basin from the east are not easy to diffuse and remove. However, the northeast direction is the distribution area of heavy industries in China and is prone to the retention and accumulation of pollutants due to the influence of climate and topography. Therefore, the pollution problem is an urgent issue to be solved in Xi'an. $PM_{2.5}$ prediction can provide a theoretical basis for pollution control. We have collected pollutant data from Xi'an environmental monitoring station. The information on the monitoring station is shown in Table 1.

Table 1. The information on the monitoring stations.

Monitoring Station Coding	Monitoring Station Name	Longitude	Latitude
1462A	High-voltage switch factory	108.882	34.2749
1463A	Xinqing district	108.993	34.2629
1464A	Textile city	109.06	34.2572
1465A	Xiaozhai	108.94	34.2324
1466A	City stadium	108.954	34.2713
1467A	Xian high-tech zone	108.883	34.2303
1468A	Economic development zone	108.935	34.3474
1469A	Chang'an district	108.906	34.1546
1470A	Yanliang district	109.2	34.6575
1471A	Lintong district	109.2186	34.3731
1472A	Caotan	108.869	34.378
1473A	Qujiang Culture Industry Group	108.985	34.1978
1474A	Guangyuntan	109.043	34.3274

The study data include pollutant and meteorological data for each hour from 1 January 2017 to 31 December 2019. With 26,280 historical records, the data set contains 1180 missing values. The pollutant features are $PM_{2.5}$, PM_{10}, SO_2, NO_2, O_3, and CO, and the meteorological features are wind direction (WD), wind speed (WS), temperature (Tem), dewpoint temperature (TDew), pressure (Pre), and relative humidity (RH), totaling 12 features in total. A detailed description of the feature variables can be found in Table 2.

Table 2. The attributes of feature variables.

Categories	Variable	Unit	Min	Max	Mean	Std
Pollutant variables	$PM_{2.5}$	$\mu g/m^3$	1	795	64.26	62.98
	PM_{10}	$\mu g/m^3$	1	1967	118.30	95.7
	SO_2	$\mu g/m^3$	1	475	13.87	10.96
	NO_2	$\mu g/m^3$	1	293	53.16	29.9
	O_3	$\mu g/m^3$	1	426	51.48	52.39
	CO	mg/m^3	1	10.5	1.2	0.61
Meteorological variables	WD	°	0	360	148.34	104.55
	WS	m/s	0	12	2.36	1.34
	Tem	°C	−9.30	41.40	15.68	10.45
	TDew	°C	−24.20	25.60	6.64	10.67
	Pre	hPa	991.90	1048.80	1026.89	11.18
	RH	%	8.03	99.33	59.25	21.75

2.2. Data Processing

2.2.1. Missing Value Filling

Pollutant and meteorological data are time series data collected continuously by sensors. Some missing values inevitably exist due to instrument failure and other reasons. In order to ensure data integrity, missing values need to be filled. We use a linear function to fill the missing value in Equation (1). The missing values at the time x_t is replaced by $F(x_t)$.

$$F(x_t) = ax_t + b \tag{1}$$

where a and b represent the slope and intercept obtained by fitting the sample points of the characteristic variable x. The fitting value at the time x_t is expressed as $F(x_t)$.

2.2.2. Data Normalization

The multidimensional feature variables collected in this paper have different units and value ranges. We utilize the maximum and minimum normalization to eliminate the

dimensional difference between data to reduce the computational complexity and increase convergence speed. Equation (2) shows the calculation formula.

$$x_i = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{2}$$

$x_{\max}$ and $x_{\min}$, respectively, represent the maximum and minimum values of the original sample x, and x_i represents the normalized sample value.

3. Methods

We propose a new three-stage combination model for use in short-term $PM_{2.5}$ prediction. In the first part of the research, we select the temporal and spatial features related to $PM_{2.5}$ concentration through the MRMR algorithm. For the second part, we construct a dual encoder–decoder model based on BiLSTM to train and learn the spatiotemporal features and perform prediction. Finally, the study area is divided into 1 km × 1 km grids, and the IDW method is used to interpolate the grids to obtain $PM_{2.5}$ concentration values of all grids. Figure 1 presents a flow chart for the proposed model. This section describes the methods involved in the model in detail.

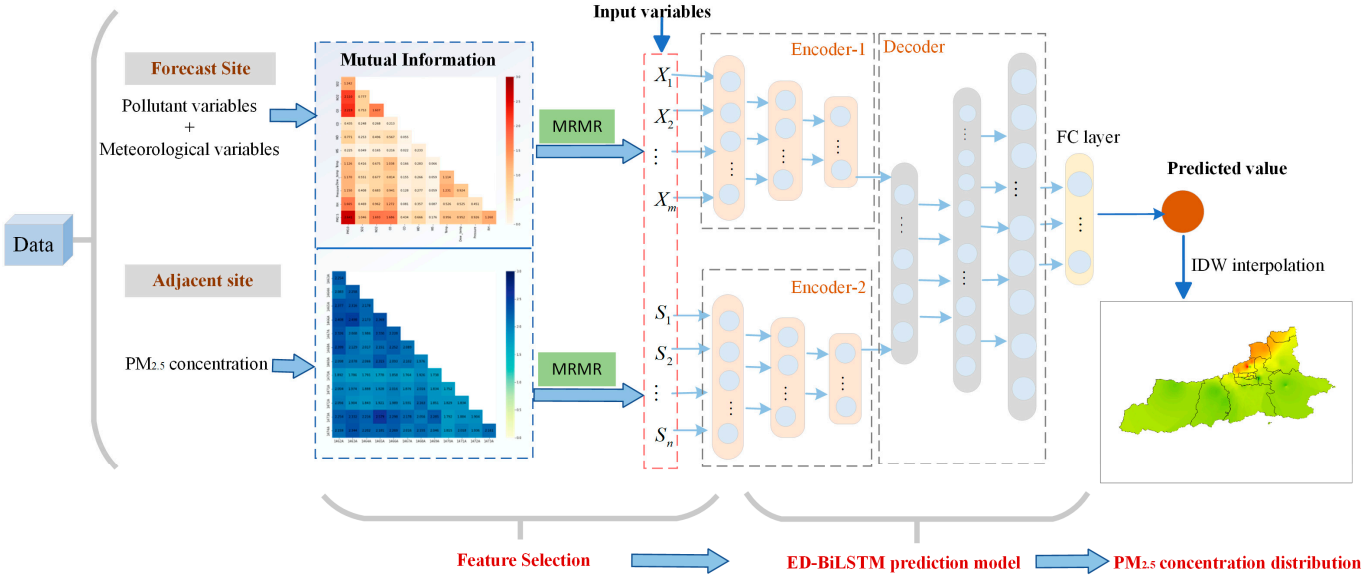

Figure 1. The flow chart of the proposed model.

3.1. MRMR Feature Selection Algorithm

$PM_{2.5}$ concentration is affected by other atmospheric pollutants, meteorological conditions [51], and $PM_{2.5}$ concentration at adjacent sites [52]. However, using multiple-feature variables may result in information redundancy and reduce the prediction model's accuracy [42]. To eliminate information redundancy between features, we use the MRMR feature selection algorithm based on mutual information to remove irrelevant and redundant features, reduce computational consumption, and improve prediction accuracy.

The MRMR algorithm considers the correlation and redundancy between features simultaneously. The correlation between the random variables x and y can be calculated using mutual information (MI) from Equation (3).

$$I(x,y) = \iint P(x,y) lg^{P(x,y)}_{P(x)P(y)} dxdy \tag{3}$$

where $P(x)$ and $P(y)$ are the probability distribution functions of x and y, $P(x,y)$ is the joint probability distribution function of x and y, and $I(x,y)$ is the mutual information of x and y.

Suppose that S denotes the set of feature variables x_i and $|S| = m$, and that m represents the number of features. The correlation between the characteristic variables and the target variable is calculated by Equation (4). The redundancy between the characteristic variables is calculated by Equation (5).

$$C(S, y) = max\left(\frac{1}{|S|}\sum_{x_i \in S} I(x_i; y)\right) \tag{4}$$

$$R(S) = min\left(\frac{1}{|S|^2}\sum_{x_i, x_j \epsilon S} I(x_i; x_j)\right) \tag{5}$$

where x_i is the value of the ith feature variable, y is the value of the target variable, $I(x_i; x_j)$ is the MI between the feature variables, $I(x_i; y)$ is the MI between the target variable and the feature variables, C is the correlation between the feature variable and the target variable, and R is the redundancy between the features.

The objective function for selecting the initial feature subset is shown in Equations (6) and (7). After determining the initial feature subset, an incremental search is performed using Equation (8). The steps for conducting an incremental search are as follows:

Step 1: The initial feature set determined by the MRMR feature selection algorithm is represented as S_{k-1}, which contains $k-1$ features. The remaining features are expressed as $S - S_{k-1}$. The feature variables contained in $S - S_{k-1}$ are denoted as $x_1, x_2, \ldots x_j$.

Step 2: Add x_1 to S_{k-1} to form feature set S'. Input S_{k-1} and S' into the prediction model to calculate the evaluation index of the model, respectively. If the prediction error corresponding to input S_{k-1} is small, the original data set is kept. If the error corresponding to input S' is small, the S' replaces S_{k-1} as the new data set.

Step 3: Add the remaining features in $S - S_{k-1}$ to the feature set with the minimum error in turn and perform Step 2 until the final feature set with the minimum error is obtained.

$$max\varnothing(C, R), \varnothing(C, R) = C - R \tag{6}$$

$$max\varnothing(C, R), \varnothing(C, R) = C/R \tag{7}$$

$$\max_{x_j \epsilon S - S_{k-1}}\left[I(x_j; y) - \frac{1}{k-1}\sum_{x_i \epsilon S_{k-1}} I(x_j; x_i)\right] \tag{8}$$

The final feature set fully preserves the correlation between the multidimensional characteristic variables and the prediction target, and achieves a data dimensionality reduction of the prediction model, which reduces the calculation loss.

3.2. ED-BiLSTM Prediction Model

The ED-BiLSTM model comprises a dual encoder and a decoder based on BiLSTM. An encoder encodes a temporal feature, and an encoder encodes a spatial feature. Then, the encoded vectors are aggregated and inputted to the decoder to obtain the ED-BiLSTM model. Finally, the $PM_{2.5}$ concentration prediction is realized via the test set.

In the deep neural network prediction model, the time series modeling and time step calculation for the characteristic variables are the basis of the realization of the prediction model. Therefore, before proposing the ED-BiLSTM prediction model, we first introduce the time series modeling and time step calculation methods.

3.2.1. Time Series Sample Modeling

To convert the data into a form that the computer can understand, we model the time series through a rolling window. An example of time series sample modeling by a rolling

window is shown in Figure 2. Suppose that the period contains 10 records X_{T1}, X_{T2}, ..., X_{T10}, where $X = \{x_1, x_2, \ldots, x_n\}$, and n is the number of feature dimensions. When Δt = 6, for Sample 1, $X_{T1}, X_{T2}, X_{T3}, X_{T4}, X_{T5}$, and X_{T6} are the features and X_{T7} is the label. For Sample 2, the features are $X_{T2}, X_{T3}, X_{T4}, X_{T5}, X_{T6}$ and X_{T7}; X_{T8} is the label, and all samples are modeled in the same way. In the prediction model, a small Δt will make the input information incomplete, and a larger one will increase the noise and the computational consumption [27]. We determined the Δt by calculating the autocorrelation coefficient (ACF) and the partial autocorrelation coefficient (PACF), as implemented in Section 3.2.2.

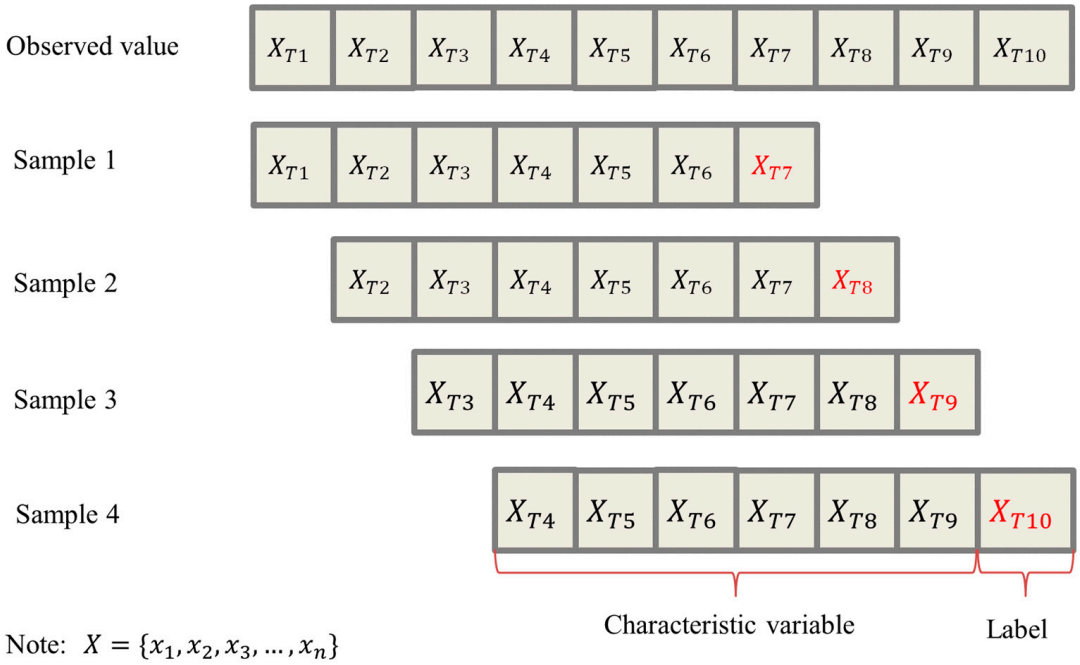

Figure 2. Modeling of time series samples. Black is the feature, red is the label.

3.2.2. Time Step Calculation

The number of stamps in each sample is determined by the size of time step Δt, which is called the lag effect. A small value of Δt does not ensure sufficient information input, while a large value will introduce irrelevant information and cause computational loss. Therefore, determining an appropriate Δt is crucial to designing the prediction model. We determined the Δt by analyzing ACF and PACF.

As a measure of correlation, the ACF describes the degree of correlation between values over time. For the time series $\{Y_t, t \in T\}$, the ACF between the values of $[Y_{t-\Delta t}, Y_t)$ and Y_t can be measured by Equation (9). A larger ACF value indicates a stronger correlation between $[Y_{t-\Delta t}, Y_t)$ and Y_t.

$$\rho_{\Delta t} = \frac{\sum_{t=\Delta t+1}^{T} (Y_t - \overline{Y})(Y_{t-\Delta t} - \overline{Y})}{\sum_{t=1}^{T} (Y_t - \overline{Y})^2} \tag{9}$$

where Y_t is the label, $Y_{t-\Delta t}$ is the sample with a Δt time lag, and $\overline{Y}$ is the average value of the sample.

Compared to ACF, PACF is more focused on the correlation between $Y_{t-\Delta t}$ and Y_t. The PACF of $Y_{t-\Delta t}$ and Y_t is the correlation coefficient between $Y_{t-\Delta t}$ and Y_t after removing the indirect effects of $[Y_{t-\Delta t+1}, Y_{t-\Delta t+2}, \ldots Y_{t-1}]$. The formula of PACF is shown in

Equations (10)–(12). When the value of PACF is closer to 0, it indicates that the correlation between $Y_{t-\Delta t}$ and Y_t is weaker.

$$\rho_{Y_t, Y_{t-\Delta t}|Y_{t-1},\cdots,Y_{t-\Delta t+1}} = \frac{E\left[\left(Y_t - \hat{E}Y_t\right)\left(Y_{t-\Delta t} - \hat{E}Y_{t-\Delta t}\right)\right]}{E\left[\left(Y_{t-\Delta t} - \hat{E}Y_{t-\Delta t}\right)^2\right]} \tag{10}$$

$$\hat{E}Y_t = E[Y_t|Y_{t-1}, \cdots, Y_{t-\Delta t+1}] \tag{11}$$

$$\hat{E}Y_{t-\Delta t} = E[Y_{t-\Delta t}|Y_{t-1}, \cdots, Y_{t-\Delta t+1}] \tag{12}$$

3.2.3. ED-BiLSTM Model

The BiLSTM [33] neural network structure contains a forward LSTM cell and a backward LSTM cell. The BiLSTM can fully mine time series information and has shown outstanding performance capacity in time series prediction tasks [27,29,53]. Figure 3 shows a comparison between LSTM and BiLSTM. We can see that BiLSTM performs not only forward, but also backward, extraction of features compared with LSTM. To extract bi-directional features from sequence data, we use the BiLSTM as the main component to construct the dual encoder and decoder model. To provide a deeper understanding of BiLSTM, we give the detailed calculation process of LSTM.

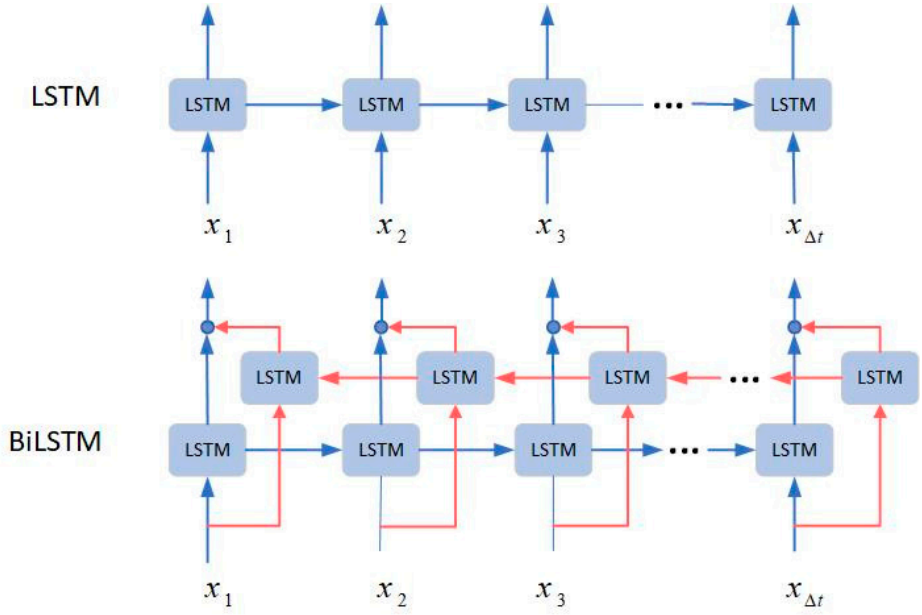

Figure 3. LSTM and BiLSTM. Blue indicates forward propagation and orange indicates backward propagation.

Figure 4 shows that each LSTM neural unit consists of an input gate, an output gate, and a forget gate. Data are received through the input gate, historical information is preserved through the forgetting gate, and information is output through the output gate. Equations (13)–(18) give the calculation formula.

$$f_t = \sigma\left(U_f h_{t-1} + W_f x_t + b_f\right) \tag{13}$$

$$i_t = \sigma(U_i h_{t-1} + W_i x_t + b_i) \tag{14}$$

$$\widetilde{c}_t = \tan \mathrm{h}(U_c h_{t-1} + W_c x_t + b_c) \tag{15}$$

$$c_t = f_t * c_{t-1} + i_t * \widetilde{c}_t \tag{16}$$

$$o_t = \sigma(U_O h_{t-1} + W_O x_t + b_o) \tag{17}$$

$$h_t = o_t * tanh(c_t) \tag{18}$$

where h_{t-1} is the output of the hidden at $t-1$, c_{t-1} is the output of the memory cell at $t-1$, and x_t is the current input vector. U_f, U_i, U_c, and U_O are the parameter matrices controlling the hidden state. W_f, W_i, W_c, and W_O are the parameter matrices controlling the input information, b_f, b_i, b_c, and b_o are the bias vectors, σ is the sigmoid function, tanh is the function tanh, and * is the Hadamard operation. f_t is the forget gating used to control the degree of forgetting of the previous state c_{t-1}, $\widetilde{c}_t$ is the memory of the current input information, i_t is memory gating, and c_t and i_t perform Hadamard operations to determine the storage of new information. The forgetting of past states and the updating of current input information together determine the c_t state vector. The output state h_t is obtained by the operation of o_t and c_t.

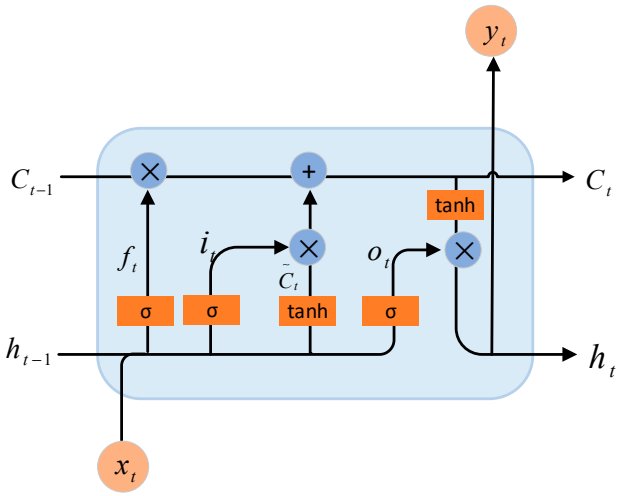

Figure 4. LSTM network structure.

The output of the BiLSTM neural network unit is a combination of forward LSTM hidden layer output $\overrightarrow{h}_t$ and backward LSTM hidden layer output $\overleftarrow{h}_t$. The calculation process is shown in Equation (19).

$$h_t = \alpha \overrightarrow{h}_t + \beta \overleftarrow{h}_t \tag{19}$$

where α and β are the output weights of the forward and backward LSTM hidden layers, respectively.

For multivariate time series, one hidden layer cannot express all the information. Adding hidden layers may give more accurate results. The general approach is to stack multiple hidden layers [54] and use the output of the previous hidden layer as the input of the next hidden layer. Since both temporal and spatial features are considered in our study, simply adding hidden layers cannot solve the complex data feature problem. Therefore, we propose the use of a dual encoder and decoder model (ED-BiLSTM) based on BiLSTM. Figure 5 shows the model structure, which consists of two parts: the encoder and the decoder. One of the encoders encodes the temporal features obtained from MRMR, and the other encoder extracts spatial features from adjacent sites. Then, the encoded vectors from both encoders are aggregated and fed into the decoder for prediction. The ED-BiLSTM

prediction model will be more beneficial for calculating the grid space interpolation without monitoring stations.

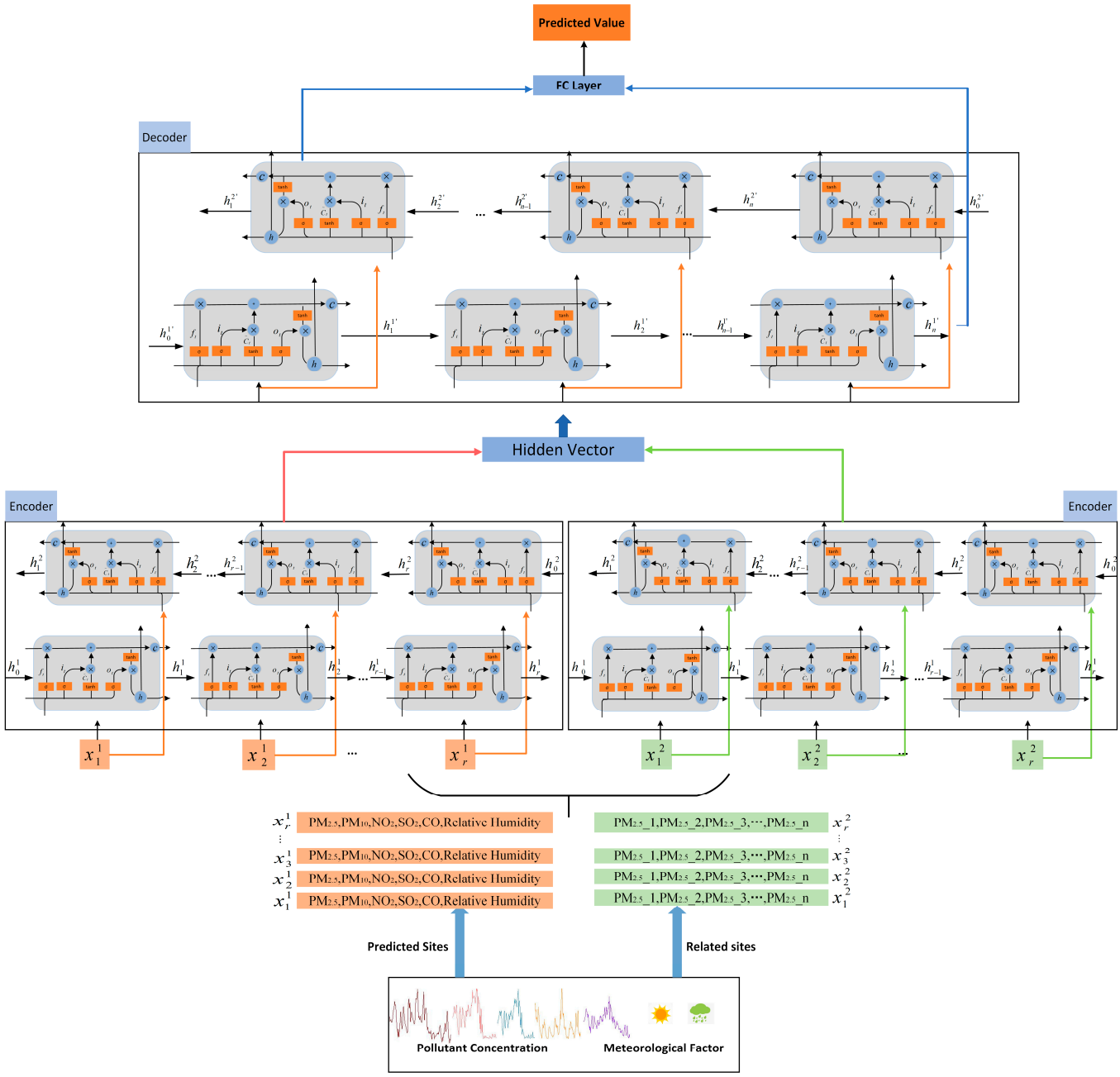

Figure 5. ED-BiLSTM model.

3.3. Inverse Distance Weight (IDW) Interpolation Method

With increasing popular concern over environmental pollution, predictions about air pollutants have made significant progress in recent years. However, for deep learning to be effective, large amounts of historical data are required for prediction, which leads most of the research about pollutant prediction to target monitoring stations [30,42]. There have also been a few studies into new sites with limited historical data [27]. However, for areas without monitoring sites, the absence of historical data has led to a lack of research in this area. In this study, the ED-BiLSTM prediction model was combined with the IDW algorithm to interpolate the 1 km × 1 km spatial grid and obtain each grid's predicted values. The method is divided into the following steps. Firstly, the $PM_{2.5}$ concentration of

all monitoring stations in the study area is predicted using the ED-BiLSTM model at the next moment. Then, the place is divided into grids of 1 km × 1 km; then, the coordinates of each grid center point are determined based on latitude and longitude. Finally, the IDW spatial interpolation algorithm obtains the predicted values for all grids.

IDW is a spatial interpolation algorithm based on points. For the interpolated grid with the center point coordinates (x, y), the predicted $PM_{2.5}$ value at this grid is calculated by Equation (20), where n is the number of monitoring stations, z_i is the predicted value at the monitoring station i and d_i indicate the distance between the interpolation grid (x, y), and the monitoring point (x_i, y_i) is calculated by Equation (21). k is the power exponent of distance. When defining higher values of k, the more adjacent stations have a more significant effect on the interpolation. By comparison, when defining smaller values of k, the more remote stations have more significant effects on interpolation. The core of the IDW algorithm is the selection of k. It has been shown that the search radius does not affect the results when k is 3 or more [55]. Based on the analysis of previous studies, we set k as 3.

$$z = \left[\sum_{i=1}^{n} \frac{z_i}{d_i^k}\right] / \left[\sum_{i=1}^{n} \frac{1}{d_i^k}\right] \tag{20}$$

$$d_i = \sqrt{(x - x_i)^2 + (y - y_i)^2} \tag{21}$$

3.4. Evaluation Index

To quantify the prediction model performance, we evaluated the ED-BiLSTM model and compared models using the root mean square error (RMSE), the mean absolute error (MAE), the mean absolute percentage error (MAPE), and R squared (R^2). Deviation between actual and predicted values is measured by RMSE and MAE, and MAPE can determine the percentage deviation. The smaller the RMSE, MAE, and MAPE values are, the more accurate the prediction will be. R^2 reflects the ability to fit data, and the value ranges from 0 to 1. The higher the R^2 is, the better the model fits the data will be. Extreme values do not affect MAE and MAPE, while RMSE uses the square of the error. This amplifies the prediction error and is more sensitive to anomalies. Therefore, we combine the evaluation metrics to evaluate the ED-BiLSTM model. The formulas are shown in Equations (22)– (25).

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{22}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{23}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{y_i} \tag{24}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y}_i)^2} \tag{25}$$

where y_i is the observed value of $PM_{2.5}$ concentration, $\hat{y}_i$ is the predicted value, $\overline{y}_i$ is the mean of the monitored value, and N denotes the number of observed samples.

4. Case Study

4.1. Feature Selection

$PM_{2.5}$ concentration is affected by other atmospheric pollutants, meteorological conditions, and the concentration of $PM_{2.5}$ at adjacent sites. However, in short-term predictions, the presence of multiple feature variables may result in information redundancy and reduce the prediction model's accuracy. To eliminate information redundancy caused by excessive

features, we calculated the MI of $PM_{2.5}$ using other pollutants, meteorological factors, and the MI on $PM_{2.5}$ from adjacent sites.

To choose the appropriate correlation analysis method, we first checked the linearity of the $PM_{2.5}$ series using Brock–Dechert–Scheinman (BDS) statistics [56]. The BDS test result is presented in Table 3 for when the dimension is set to 6. Here, the Z-statistic is higher than the critical value 1.96, and the p value is less than 0.05, indicating that the BDS statistic rejects the initial linearity hypothesis and proves that nonlinearity of $PM_{2.5}$ time series. Therefore, we used mutual information to calculate the correlation between variables. Then, we used the MRMR algorithm to remove redundant features and determine the optimal subset of features with which to implement the dimensionality reduction of the input data.

Table 3. The BDS results of the $PM_{2.5}$ concentration series at the 1463A site.

Dimension	Z-Statistic	p Value	95% Confidence Interval
2	258.908	0.000	[−1.96, 1.96]
3	273.814	0.000	[−1.96, 1.96]
4	291.613	0.000	[−1.96, 1.96]
5	318.066	0.000	[−1.96, 1.96]
6	354.348	0.000	[−1.96, 1.96]

Figure 6 displays the heat map of the mutual information values between $PM_{2.5}$ concentrations and influence factors at the 1463A site. The heat map's color is darker, suggesting a higher value for mutual information and a higher correlation between the variables. $PM_{2.5}$ concentration at time $t+1$ and other influence factors values at time t were used to calculate mutual information. Figure 6 illustrates that PM_{10}, NO_2, O_3, RH, SO_2, and $PM_{2.5}$ are highly correlated, and that the mutual information value is greater than 1. When k = 6, we determined the feature subsets according to the mutual information values as $PM_{2.5}$, PM_{10}, NO_2, SO_2, O_3, and RH. However, the subset of features calculated by the MRMR algorithm considering redundancy included $PM_{2.5}$, PM_{10}, NO_2, SO_2, CO, and RH. In Figure 6, we see that O_3 strongly correlates with PM_{10}, NO_2, and RH, indicating that the MRMR feature selection algorithm fully accounts for the redundancy between features.

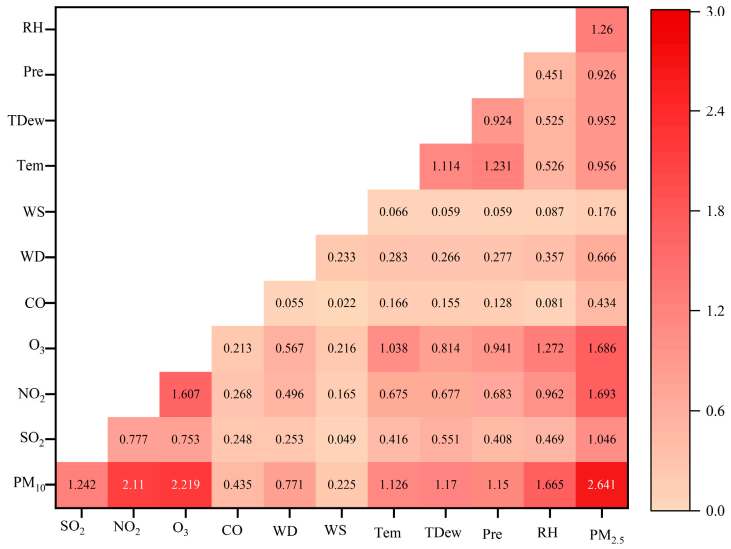

Figure 6. Mutual information of characteristic variables for site 1463A.

We conducted experiments by inserting a subset of features, determined by two feature selection algorithms based on mutual information and the MRMR algorithm, into the prediction model. Table 4 shows the values of evaluation metrics for different feature input prediction models. The MRMR algorithm obtained the optimal feature set. Then, we added

the remaining features to the subset of features using an incremental search algorithm. We found by comparing the model evaluation metrics that the features identified by the MRMR algorithm at $k = 6$ could be used to achieve the best prediction results. Therefore, the input feature subset of the prediction model contains $PM_{2.5}$, PM_{10}, NO_2, SO_2, RH, and CO. In addition, we calculated the mutual information value for $PM_{2.5}$ concentration between adjacent stations. Figure 7 shows a strong correlation between $PM_{2.5}$ concentrations at adjacent sites. Comparing the model before and after adding spatial features, we found that RMSE decreased by 12.2% and MAE decreased by 27.3% after adding spatial features. Based on the feature selection algorithm, we used $PM_{2.5}$, PM_{10}, NO_2, SO_2, RH, CO, and $PM_{2.5}$ from adjacent sites as inputs in the prediction model.

Table 4. $PM_{2.5}$ prediction results of different feature selection algorithms.

Feature Selection Method	Feature	RMSE	MAE	MAPE	R^2
Original Feature	$PM_{2.5}$, PM_{10}, NO_2, SO_2, RH, CO, O_3, Tem, TDew, Pre, WD, WS	7.660	5.696	0.132	0.976
MI	$PM_{2.5}$, PM_{10}, NO_2, SO_2, RH, O_3	7.080	4.397	0.126	0.982
MRMR	$PM_{2.5}$, PM_{10}, NO_2, SO_2, RH, CO	6.603	4.066	0.101	0.986
Incremental Search	$PM_{2.5}$, PM_{10}, NO_2, SO_2, RH, CO, O_3	6.686	4.075	0.103	0.985
	$PM_{2.5}$, PM_{10}, NO_2, SO_2, RH, CO, Tem	7.686	5.065	0.117	0.985
	$PM_{2.5}$, PM_{10}, NO_2, SO_2, RH, CO, TDew	7.853	5.264	0.122	0.984
	$PM_{2.5}$, PM_{10}, NO_2, SO_2, RH, CO, Pre	8.004	5.238	0.105	0.984
	$PM_{2.5}$, PM_{10}, NO_2, SO_2, RH, CO, WD	7.706	5.262	0.128	0.986
	$PM_{2.5}$, PM_{10}, NO_2, SO_2, RH, CO, WS	7.689	5.169	0.115	0.986

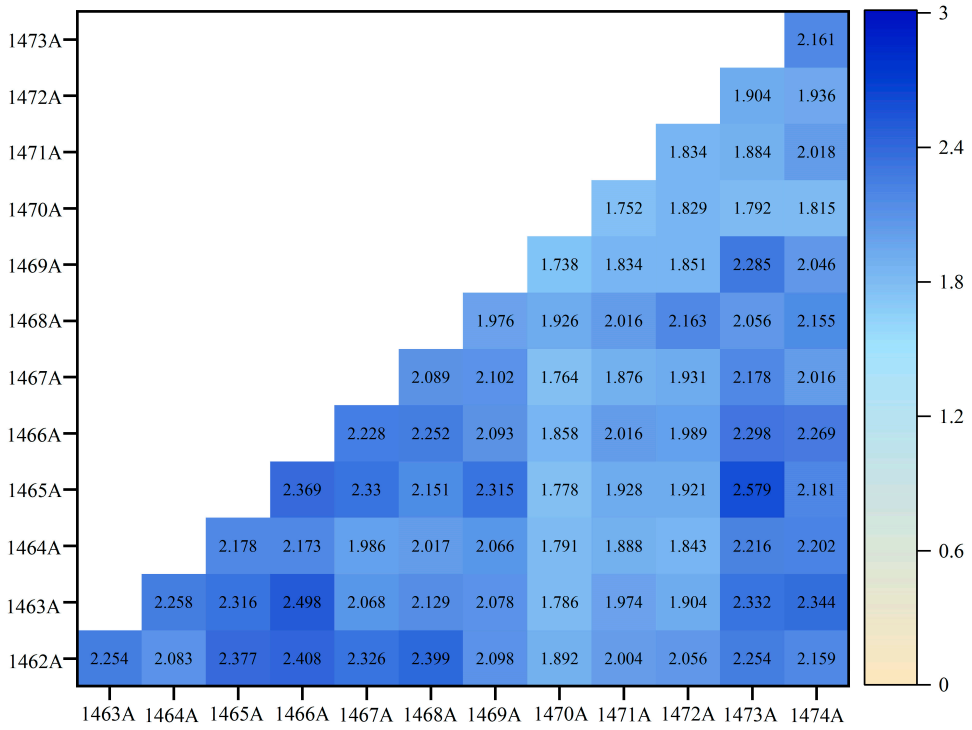

Figure 7. Mutual information between $PM_{2.5}$ concentration at sites.

4.2. Parameter Setting of ED-BiLSTM Model

4.2.1. Time Step Calculation

In the time series forecasting task, the time step Δt determines how many historical record values are input to the model, and a suitable value of Δt can increase the model's

forecasting performance. In this paper, we decide the value of Δt by calculating ACF and PACF.

Figure 8 shows the ACF calculation results of $PM_{2.5}$ concentration at different sites in Xi'an. We can see that the ACF presents a decreasing trend as the time lag lengthens. When Δt is less than 5, the ACF of most sites is greater than 0.9, and the data are highly correlated. When Δt is greater than 30, the correlation is reduced.

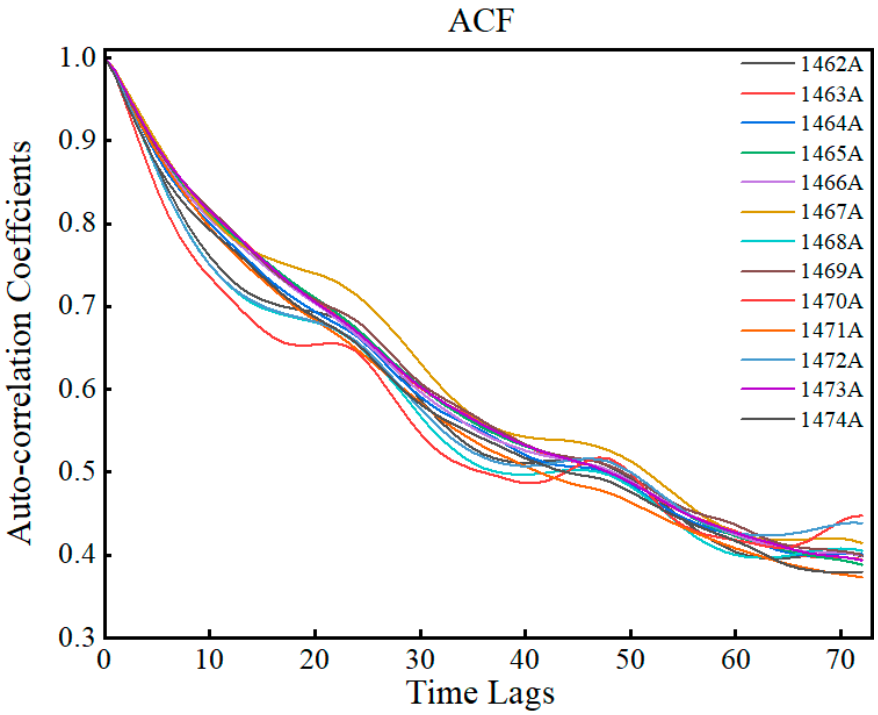

Figure 8. The ACF of different time lags for each site.

Compared to ACF, PACF is more focused on the correlation between $Y_{t-\Delta t}$ and Y_t. Figure 9 shows the PACF calculation results for $PM_{2.5}$ concentrations at site 1463A. The PACF tends to 0 when Δt = 4, and the ACF is 0.905 at this point, indicating a strong correlation between Y_t, Y_{t-1}, Y_{t-2}, Y_{t-3}, and Y_{t-4}. However, when Δt > 4, the PACF fluctuates around zero until Δt > 20, when the fluctuation becomes relatively flat.

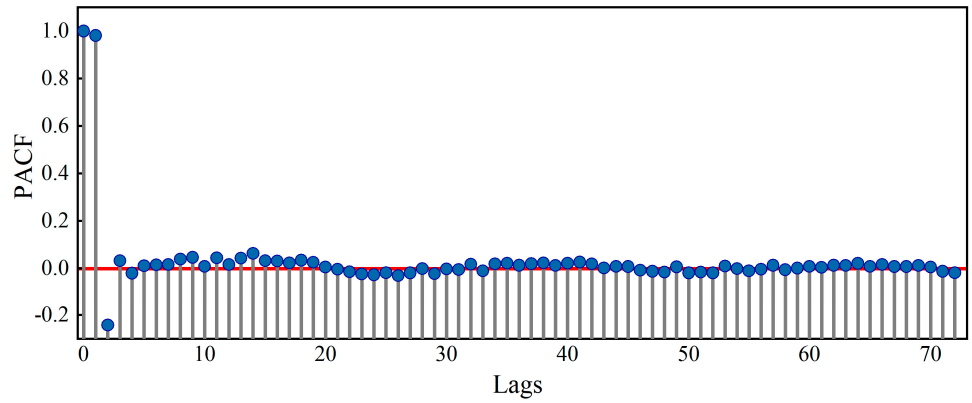

Figure 9. The PACF at the 1463A site.

To further determine Δt, we construct training samples with Δt in the range of [1,24] in order to train and make predictions using ED-BiLSTM. The model prediction results under different Δt settings are shown in Table 5. As seen in Table 5, when Δt is set to 4, the RMSE is 6.603, the MAE is 4.066, and the prediction model has the most minor error. To see

the effect of Δt more intuitively on the prediction performance, we plot the line graphs of RMSE and MAE for different time lags in Figure 10. There is a similar trend between the RMSE and MAE, and a regular pattern can establish that the prediction error is more minor when Δt is a multiple of 4. This is in agreement with the conclusion reached by performing PACF, and is also confirmed by the study of Huang et al. [9]. Through ACF and PACF calculations, we set Δt to 4, which means that when predicting Y_{t-1}, Y_{t-2}, Y_{t-3}, and Y_{t-4} are input into the prediction model.

Table 5. ED-BiLSTM model evaluation index with different time lags.

Δt	RMSE	MAE	MAPE	R^2	Δt	RMSE	MAE	MAPE	R^2
1	7.756	4.955	0.135	0.981	13	8.001	5.417	0.179	0.980
2	6.795	4.287	0.107	0.986	14	7.786	4.647	0.106	0.981
3	7.240	4.625	0.106	0.984	15	6.883	4.383	0.130	0.985
4	6.603	4.066	0.101	0.986	16	6.738	4.375	0.139	0.986
5	7.271	4.467	0.115	0.984	17	6.792	4.202	0.103	0.986
6	7.430	4.826	0.146	0.983	18	7.582	4.955	0.110	0.982
7	8.075	5.023	0.116	0.980	19	7.084	4.499	0.114	0.984
8	6.843	4.489	0.146	0.985	20	6.736	4.187	0.104	0.986
9	7.010	4.251	0.104	0.985	21	7.069	4.474	0.108	0.984
10	7.027	4.526	0.136	0.985	22	6.866	4.332	0.113	0.985
11	6.926	4.582	0.141	0.985	23	8.374	5.079	0.109	0.978
12	7.355	4.642	0.111	0.983	24	7.573	4.539	0.108	0.982

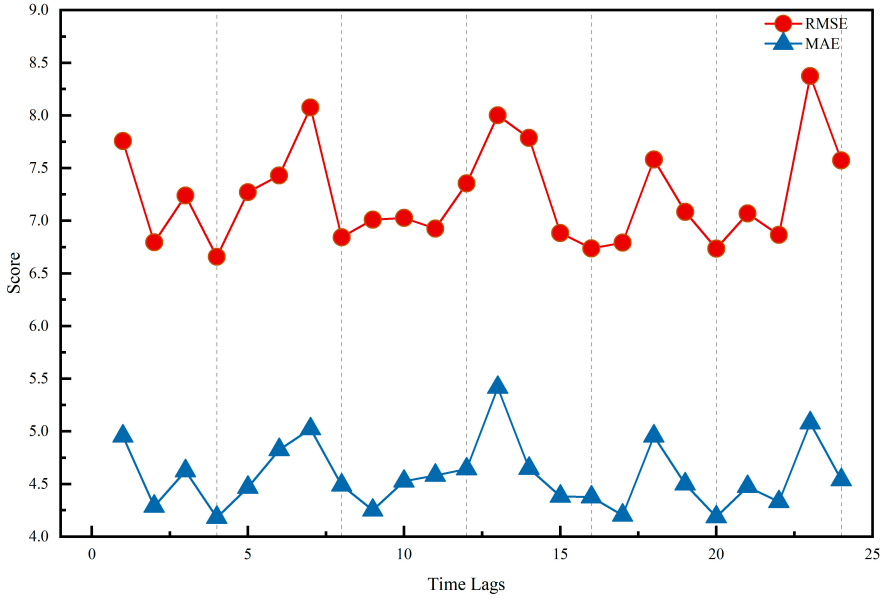

Figure 10. The RMSE and MAE with different time lags.

4.2.2. Experimental Settings

The experimental data in this paper, taken from the historical records of $PM_{2.5}$, PM_{10}, NO_2, SO_2, RH, CO, and adjacent sites' $PM_{2.5}$ concentrations from a previous 4 h period, are used to predict the $PM_{2.5}$ concentration for the next hour. For time series modeling, we individually model the time series features and spatial features. In model construction, the hidden layer of the encoder is the BiLSTM unit, and a concatenate layer is added after the encoder to fuse the feature vectors of the two encoders. Finally, the decoder sets a hidden layer. The prediction model with the lowest error on the training set is obtained by continuously adjusting of the parameters. Table 6 shows the parameters set for the ED-BiLSTM model.

Table 6. The parameter setting of the prediction model.

Parameter	Value
Training set	80% (21,020)
Validation set	10% (2628)
Test set	10% (2628)
Number of BiLSTM neurons	128
Epochs	200
Batch size	128
Optimizer	Adam
Loss	MAE
Activation function	ReLU
Learning rate	1×10^{-3}
Dropout	0.5

We model 26,280 historical records. Due to the time step being set to 4, we get 26,276 samples, including 21,020 training samples, 2628 validation samples, and 2628 test samples. In the model training, we input the samples into the model in batches; the batch size is 128. We set the epoch to 200 and the number of neurons to 128 in the BiLSTM. The optimizer is Adam, the loss function is a mean absolute error (MAE), and the learning rate is 1×10^{-3}. The activation function is set to ReLU. At the same time, the dropout parameter is set to 0.5 to prevent overfitting. The number of encoder and decoder layers is set to 1. As soon as the model training process is complete, a prediction is performed on the basis of the test set, and an evaluation index is calculated for the model.

4.3. Prediction Results

To verify the predictive performance of the proposed model, we predicted the $PM_{2.5}$ concentration at station 1463A for the next hour using the ED-BiLSTM and the comparative models. We constructed seven comparative models, including the machine learning model SVR, single LSTM, and single BiLSTM, as well as four combination models, namely, ED-LSTM, AE-BiLSTM, ED-BiLSTM-One, and ED-BiLSTM-IF. ED-LSTM means that the hidden layers of the encoder and decoder use LSTM neural network units. The combined model AE-BiLSTM denotes that the model uses an auto-encoder for the encoder and BiLSTM for the decoder. ED-BiLSTM-One indicates that an encoder with the hidden layer of BiLSTM encodes spatiotemporal features. ED-BiLSTM-IF indicates the selection of influencing factors, a practice which considers the correlation feature variables of the target site alone and disregards the spatial correlation of $PM_{2.5}$ concentrations at adjacent sites. All models were constructed in the Keras framework. To ensure the validity of the experimental results, all models were kept with the same parameter settings. Additionally, a 10-time repetition of each deep learning model was conducted to reduce randomness in neural network training, and Table 7 shows the experimental results.

Table 7. The prediction result of all models at the 1463A site.

Prediction Model	RMSE	MAE	MAPE	R^2
SVR	14.327	11.864	0.416	0.936
LSTM	8.695	5.416	0.131	0.985
BiLSTM	8.016	5.225	0.112	0.980
ED-LSTM	8.937	5.353	0.164	0.980
AE-BiLSTM	10.501	8.515	0.404	0.966
ED-BiLSTM-One	7.545	4.748	0.119	0.982
ED-BiLSTM-IF	7.521	5.593	0.114	0.982
ED-BiLSTM	6.603	4.066	0.101	0.986

Table 7 displayed the quantitative results of RMSE, MAE, MAPE, and R^2, comparing the ED-BiLSTM model with SVM, LSTM, BiLSTM, ED-LSTM, AE-BiLSTM, ED-BiLSTM-

One, and ED-BiLSTM-IF. The hyperparameter settings for all deep learning models are identical to those of the ED-BiLSTM model. The results in Table 7 show that the RMSE, MAE, and MAPE of the ED-BiLSTM model are lower than those of the comparative model, and that the R^2 is closer to 1. This indicates that the ED-BiLSTM model can better fit the $PM_{2.5}$ time series than the other machine learning and deep learning models. The average RMSE of the ED-BiLSTM model is 6.603, and the MAE is 4.066. Compared with the machine learning model SVM, the average RMSE is 53.91% lower, and the MAE is 65.73% lower. Compared with the single-LSTM and BiLSTM models, the average RMSE is 24.06% and 17.63% lower, and the average MAE is 24.93% and 22.18% lower. It proved that the ED-BiLSTM model is more suitable for $PM_{2.5}$ short-term prediction than the machine learning and single-neural network models. By comparing all models' predicted and true value curves, Figure 11 shows that the ED-BiLSTM model has the best fit.

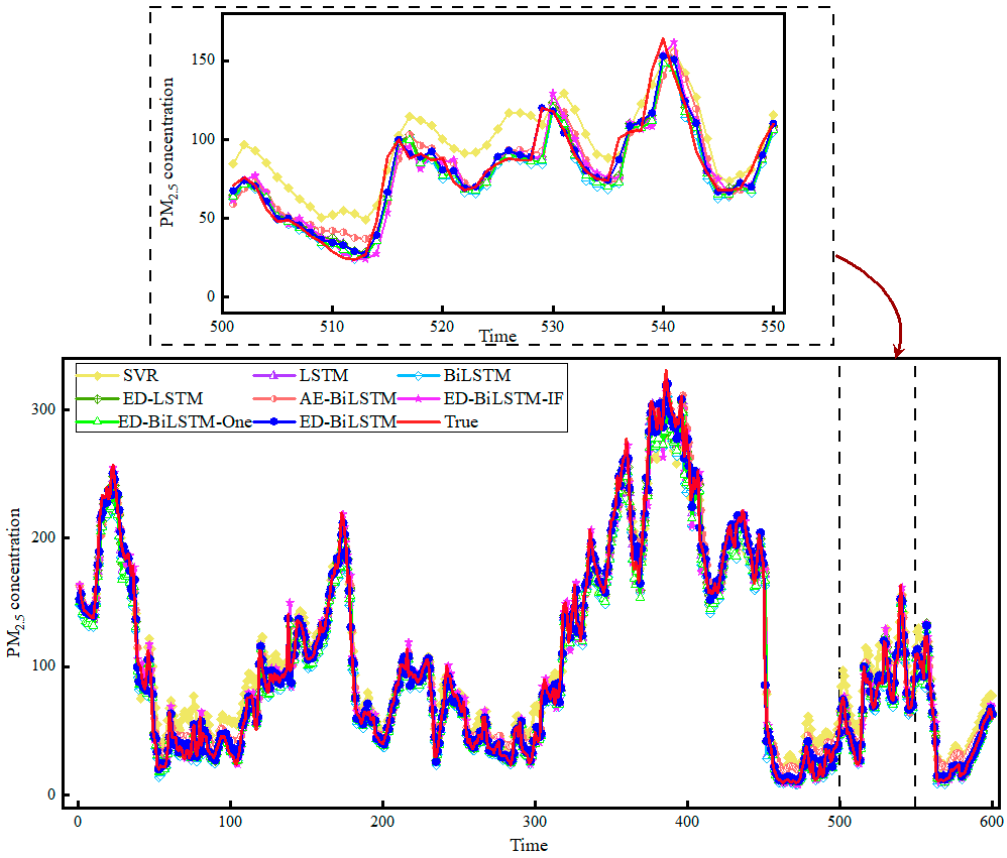

Figure 11. Results of all models and partial magnification.

In addition, we construct four hybrid models and compare the prediction results with our model. Table 7 shows the experimental results. The AE-BiLSTM model is provided by the work of Zhang et al. [37]. We construct the ED-LSTM, ED-BiLSTM-One, and ED-BiLSTM-IF models. There is a significant difference between these models due to the time series feature modeling and the hidden layer setting. The difference between ED-LSTM and ED-BiLSTM model is that LSTM replaces the neural network unit of the hidden layer. Compared with ED-LSTM, ED-BiLSTM is 26.12% lower in RMSE and 24.04% lower in MAE on average, indicating that using BiLSTM as the hidden layer unit of encoder and decoder can better extract the dependencies forward and backward from the sequence data. The encoder of AE-BiLSTM adopts an auto-encoder. Compared with AE-BiLSTM, the average RMSE of our model is 37.12% lower, and the MAE is 52.25% lower, indicating that BiLSTM has a better interpretation of sequence data than the auto-encoder. The ED-BiLSTM-One model uses one encoder to extract spatiotemporal features. Compared with ED-BiLSTM-One, our model has an average reduction of 12.49% in RMSE and 14.27%

in MAE, indicating that the use of independent encoders for spatial feature extraction of adjacent stations improves the model's prediction accuracy. The ED-BiLSTM-IF model ignores the mutual influence between adjacent stations, with an average reduction of 12.21% in RMSE and 27.3% in MAE for ED-BiLSTM compared with ED-BiLSTM-IF, which demonstrates the necessity of considering spatial features in pollutant prediction. The comprehensive comparison shows that ED-BiLSTM has the smallest error and the best fitting in $PM_{2.5}$ short-term prediction.

Figure 11 shows a graph of the prediction results for all models and a partial enlargement. As the prediction effect of each model is difficult to see from multiple curves directly, the 50 prediction points are partially enlarged to observe the prediction results of the different models more clearly. We can see from Figure 11 that the ED-BiLSTM model fits the predicted value curve better than the machine learning model, the simplex deep learning model, and the other combined models. Figure 12 plots a curve of predicted and actual values for the ED-BiLSTM model; the red line is the predicted value, and the blue line is the monitored value. The figure shows that the ED-BiLSTM model can track the fluctuations of $PM_{2.5}$ concentration well, especially the sudden changes and the maximum and minimum points. Figure 13 plots the scatter plot of predicted and actual values, with the horizontal and vertical coordinates representing the monitored and predicted, respectively. The black line symbolizes that y = x, and the scatter points are closely around y = x. Combining the prediction result plot in Figure 12 and the scatter plot in Figure 13, it is easy to see that the ED-BiLSTM fits the $PM_{2.5}$ concentration sequence best. The proposed model is significantly better than the comparative model and can be effectively applied to $PM_{2.5}$ concentration prediction.

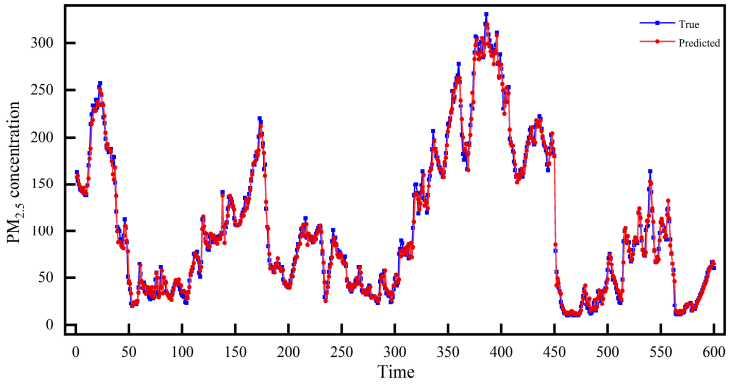

Figure 12. Result of the ED-BiLSTM prediction model.

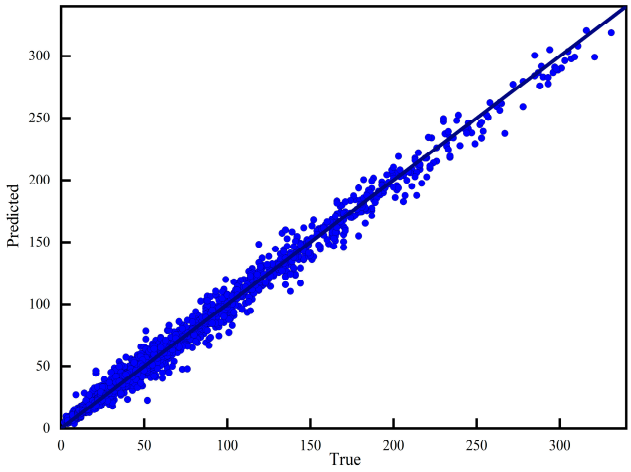

Figure 13. The scatter plot of predicted and true values on the test set.

4.4. Other Cases

To evaluate the robustness and stability of our model, we select 12 monitoring sites in Xi'an from 1 January 2017 to 31 December 2019 for $PM_{2.5}$ concentration prediction, and the information on the monitoring stations is shown in Table 1. The feature selection and prediction model construction process is the same as the one for the 1463A site. To simplify the description and avoid repetition, we directly give the prediction results of 12 monitoring stations. The evaluation indexes of each model are shown in Tables 8 and 9 for the 12 stations, and the prediction results differ among monitoring stations. The evaluation index of the proposed model's RMSE, MAE, and MAPE is smaller than that of other comparative models, and R^2 is the largest. The largest value of RMSE, MAE, and MAPE in the proposed model is 9.553, 5.786, and 0.128. The smallest R^2 is 0.974. Despite the different prediction results for each site, the proposed model's RMSE, MAE, and MAPE remain at a small value, and the R^2 is greater than 0.97. It appears that the proposed model can be used to estimate $PM_{2.5}$ concentrations.

Table 8. Prediction results of different monitoring stations, and the bolded represents the best value.

Model	Metric	1462A	1464A	1465A	1466A	1467A	1468A
SVR	RMSE	23.765	16.996	21.101	20.704	18.472	15.683
	MAE	21.415	14.814	18.093	15.501	14.315	12.639
	MAPE	0.644	0.508	0.582	0.480	0.420	0.433
	R^2	0.832	0.909	0.854	0.885	0.903	0.927
LSTM	RMSE	10.173	8.778	9.470	9.883	10.102	10.199
	MAE	6.756	5.415	6.217	4.873	6.918	7.269
	MAPE	0.261	0.101	0.145	0.098	0.115	0.114
	R^2	0.969	0.976	0.971	0.975	0.972	0.965
BiLSTM	RMSE	9.607	7.959	8.088	9.738	10.002	9.378
	MAE	5.741	4.781	5.001	4.848	6.718	7.010
	MAPE	0.126	0.110	0.125	0.093	0.112	0.125
	R^2	0.976	0.978	0.976	0.976	0.970	0.968
ED-LSTM	RMSE	9.092	7.924	8.531	9.594	10.105	9.113
	MAE	5.864	4.868	5.229	4.615	6.325	5.759
	MAPE	0.139	0.094	0.134	0.078	0.108	0.121
	R^2	0.975	0.980	0.976	0.976	0.972	0.977
AE-BiLSTM	RMSE	10.228	9.870	9.318	13.773	14.784	11.622
	MAE	6.999	6.953	6.013	6.103	10.361	7.665
	MAPE	0.256	0.160	0.131	0.125	0.264	0.131
	R^2	0.967	0.975	0.972	0.949	0.938	0.959
ED-BiLSTM-One	RMSE	8.427	7.871	8.075	8.938	9.864	9.061
	MAE	5.364	4.975	**4.919**	4.525	5.911	5.760
	MAPE	0.127	0.103	0.124	0.084	0.106	0.120
	R^2	**0.979**	0.980	0.978	0.979	0.975	0.975
ED-BiLSTM-IF	RMSE	8.916	8.319	9.110	9.001	10.556	9.108
	MAE	5.560	4.935	5.639	4.465	6.921	5.880
	MAPE	0.129	0.097	0.122	**0.074**	0.105	0.123
	R^2	0.976	0.978	0.973	0.978	0.974	0.978
ED-BiLSTM	RMSE	**7.393**	**6.609**	**7.997**	**8.023**	**9.553**	**8.001**
	MAE	**4.223**	**4.593**	4.938	**4.365**	**5.786**	**4.745**
	MAPE	**0.114**	**0.092**	**0.119**	0.076	**0.091**	**0.106**
	R^2	**0.979**	**0.982**	**0.979**	**0.983**	**0.978**	**0.985**

Table 9. Prediction results from monitoring stations 1469A-1474A, and the bolded represents the best value.

Model	Metric	1469A	1470A	1471A	1472A	1473A	1474A
SVR	RMSE	19.743	11.150	14.902	22.465	16.371	21.171
	MAE	17.325	8.570	11.695	19.500	13.489	18.849
	MAPE	0.553	0.191	0.312	0.842	0.531	0.768
	R^2	0.892	0.944	0.924	0.850	0.920	0.863
LSTM	RMSE	9.248	7.334	10.485	10.067	9.196	10.132
	MAE	6.060	4.913	6.352	5.849	5.386	6.002
	MAPE	0.114	0.128	0.118	0.121	0.131	0.151
	R^2	0.971	0.975	0.962	0.970	0.974	0.963
BiLSTM	RMSE	8.689	7.111	9.953	9.669	8.083	9.886
	MAE	5.287	4.725	5.802	5.558	4.543	6.132
	MAPE	0.107	0.115	0.117	0.129	0.128	0.143
	R^2	0.979	0.973	0.968	0.972	0.982	0.965
ED-LSTM	RMSE	9.227	7.263	10.018	9.801	8.048	10.050
	MAE	5.688	4.547	5.882	5.563	5.044	5.812
	MAPE	0.108	0.111	0.114	**0.117**	0.123	0.131
	R^2	0.976	0.974	0.966	0.971	0.981	0.969
AE-BiLSTM	RMSE	13.199	7.804	11.253	11.292	9.484	11.334
	MAE	10.676	6.328	6.973	7.800	6.910	7.210
	MAPE	0.122	0.117	0.116	0.326	0.286	0.209
	R^2	0.952	0.973	0.957	0.962	0.973	0.961
ED-BiLSTM-One	RMSE	8.300	5.518	9.557	8.694	7.656	9.259
	MAE	5.284	4.245	5.725	4.693	4.662	5.578
	MAPE	0.105	0.112	0.113	0.150	0.112	0.132
	R^2	0.981	0.986	0.969	0.972	0.983	0.973
ED-BiLSTM-IF	RMSE	8.645	5.807	9.787	8.959	8.127	9.453
	MAE	5.319	4.383	5.684	4.828	4.871	5.775
	MAPE	0.106	0.095	0.108	0.150	0.124	0.133
	R^2	0.979	0.985	0.967	0.971	0.980	0.973
ED-BiLSTM	RMSE	**8.095**	**4.872**	**8.124**	**8.625**	**7.037**	**9.228**
	MAE	**4.962**	**3.591**	**5.248**	**3.343**	**4.304**	**5.003**
	MAPE	**0.102**	**0.069**	**0.107**	0.124	**0.111**	**0.128**
	R^2	**0.982**	**0.989**	**0.979**	**0.981**	**0.985**	**0.974**

In addition, the prediction results are compared for all monitoring stations in Xi'an, and it is found that the proposed model has a minor prediction error for 1470A, while the worst result is obtained for 1467A. 1470A is Yanliang Industrial Park, with a high $PM_{2.5}$ pollutant emission concentration. In contrast, 1467A is in Xi'an High-Tech Industrial Park. A high-tech industrial park combines modern technology and natural ecology with a low concentration of pollutant emissions. It proves that the $PM_{2.5}$ prediction value of our model tends to be high, but that the error is still in the low range.

4.5. $PM_{2.5}$ Concentration Prediction on Areas without Stations

In Sections 4.3 and 4.4, we discuss the feasibility and prediction performance of the model. However, the information one can obtain from the $PM_{2.5}$ prediction results at the monitoring stations is limited. To have a more comprehensive grasp of $PM_{2.5}$ in the study area in the future time, a 1 km $\times$ 1 km resolution gridding is performed for Xi'an. Then, according to the prediction results of monitoring stations, the IDW algorithm is used for grid interpolation to obtain the $PM_{2.5}$ concentration prediction value for each grid. Finally, the interpolation results are visualized through ArcGIS for researchers and managers to get the distribution of $PM_{2.5}$ concentration in the whole region.

To see more clearly the variation process of $PM_{2.5}$ concentration in a whole day, Figure 14 visualizes the distribution of $PM_{2.5}$ concentration at two-hour intervals. We can see from the figure that the $PM_{2.5}$ concentration varies significantly at each moment of the day, and the distribution of $PM_{2.5}$ concentration in each region is imbalanced. To explain more clearly the reason for this phenomenon, the distribution of administrative divisions of Xi'an is given in Figure 15. Xi'an, the largest central city in northwest China, consists of 11 districts (Xincheng District, Beilin District, Lianhu District, Baqiao District, Weiyang District, Yanta District, Yanliang District, Lintong District, Chang'an District, Goaling District, and Huyi District) and 2 counties (Lantian County and Zhouzhi County), the main urban areas of which include Weiyang District, Baqiao District, Yanta District, Lianhu District, Xincheng District, and Beilin District.

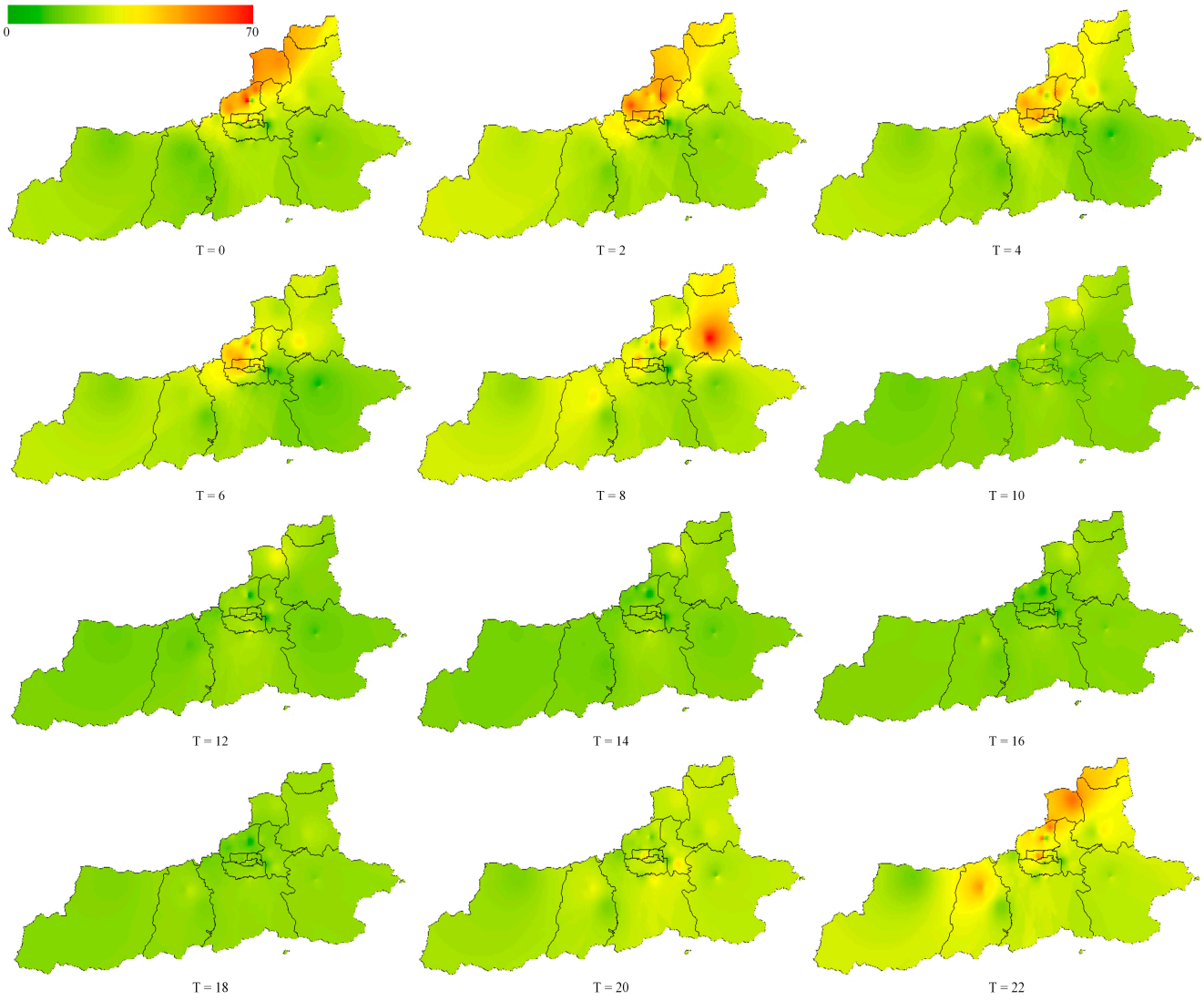

Figure 14. The distribution of $PM_{2.5}$ concentrations at two-hour intervals on 10 September 2019.

We combined Figures 14 and 15 to analyze the variation process of $PM_{2.5}$ concentration in Xi'an on 10 September 2019. From the perspective of the regional division, the $PM_{2.5}$ concentrations in Zhouzhi County, Huyi County, and Lantian County are low. In contrast, the $PM_{2.5}$ concentrations in the main urban area and Yanliang District are higher than in other districts and counties. Because of the large population and dense traffic in the central metropolitan area, Yanliang District has a dense industrial park. From the perspective of temporal evolution, $PM_{2.5}$ concentrations are higher at night than during the day. The

reason for this phenomenon is that the atmosphere easily forms a cold inversion layer at night in autumn and winter, making it impossible for air to undergo convection. The accumulation of $PM_{2.5}$ emissions during the day and the passage of high-emission trucks at night are also significant sources of $PM_{2.5}$. After a night of accumulation and motor vehicle emissions, the highest $PM_{2.5}$ concentration occurs at 8:00 am during the latter part of the morning peak. Then, as the ground temperature gradually rises, the inversion layer is broken, and the airflow rate accelerates. The lowest pollutant concentration of the day occurs from 14:00 to 16:00, which is also the best time for outdoor activities. The $PM_{2.5}$ concentration gradually increases after 18:00 as the ground-level temperature decreases and the evening peak time arrives.



Figure 15. The Map of Xi'an District.

In summary, the complex road traffic in the main urban area and the enormous number of $PM_{2.5}$ emission sources increase the difficulty of the model prediction. The worst prediction result was obtained at 1467A site in Section 4.4. 1467A site is in the Yanta district of the urban area, and so the interpolation analysis results are the same as the conclusions obtained from the model prediction results. Therefore, we can add more stations if we want to predict the $PM_{2.5}$ concentration in areas with complex emission sources. Secondly, we strongly recommend that the government invest more in efforts to deal with the problem of high $PM_{2.5}$ concentration in industrial parks. Finally, according to the distribution of pollutant concentrations in the entire day, it is recommended that residents should choose the mid-afternoon for outdoor activities when the $PM_{2.5}$ concentration is lowest.

5. Discussion

This paper proposes a new combination prediction model, combining the MRMR feature selection algorithm, the BiLSTM neural network, and IDW spatial interpolation algorithm to predict $PM_{2.5}$ concentrations in Xi'an. Firstly, we select the appropriate feature subset using the MRMR algorithm, and experiments show that the algorithm fully considers the correlation and redundancy among features, which helps to improve the accuracy of the prediction model. Secondly, we compare our model with the machine learning and single deep learning models. Due to our model considering the spatiotemporal correlation of multivariate time series, it shows strong advantages. Thirdly, the model based on BiLSTM units is compared with the model based on LSTM and autoencoder units, and

it is found that BiLSTM has a more robust feature extraction ability. Next, our model improves the predictive ability by considering the spatial correlation between adjacent sites. Finally, the experimental analysis finds that the dual encoder model outperforms a single encoder in extracting the spatiotemporal features of multiple time series. In our proposed comparative model prediction experiments, the ED- BiLSTM model has the lowest prediction error. In addition, we divided the Xi'an into 1 km × 1 km grids to predict $PM_{2.5}$ concentrations in each grid and explored the process of $PM_{2.5}$ concentration variation all day through ArcGIS visualization. The proposed model combines the feature selection algorithm, neural network, and interpolation algorithm, which solves the feature redundancy in $PM_{2.5}$ prediction and the difficulty of $PM_{2.5}$ prediction in regions without monitoring stations.

According to the analysis of $PM_{2.5}$ distribution results in Figure 14, it can be found that industrial pollution and traffic pollution are the primary sources of $PM_{2.5}$. The public can reduce $PM_{2.5}$ emissions by choosing more public transportation and reducing the use of motor vehicles. Regarding industrial emissions, industrial production processes should be optimized to reduce and purify industrial waste gas. In addition, the emission of $PM_{2.5}$ precursors, such as NOx, SO_2, and volatile organic compounds, should be controlled in the management of $PM_{2.5}$. When the level of $PM_{2.5}$ pollution in the atmosphere is high, it is reasonable for the public to wear masks when traveling, which can effectively protect the cardiovascular system. Effective source prevention and control can cut off pollution sources and solve pollution problems. As such, environmental protection departments should formulate more reasonable pollution prevention and control countermeasures to eliminate pollution at the source.

As our model predicts $PM_{2.5}$ concentration at the hourly level in the short term, and in the case analysis, the model's accuracy is verified by selecting stations only in Xian. However, this does not mean that the proposed model is geographically limited, and the model can be applied to $PM_{2.5}$ or other pollutant prediction in different cities. Due to the limitation of data sources, we only consider the effect of other pollutants and meteorological factors on $PM_{2.5}$ concentration. In addition, we only consider the effects of feature correlation and $PM_{2.5}$ concentration nonlinear features on the prediction results in this paper, ignoring the errors caused by the non-stationarity of $PM_{2.5}$ concentration series. In the future, we will analyze the non-stationarity of $PM_{2.5}$ to achieve a more accurate $PM_{2.5}$ concentration prediction. Besides, model parameter optimization by using optimization algorithms is also a future research direction.

6. Conclusions

To accurately predict the $PM_{2.5}$ concentration, we construct a new combination prediction model based on feature selection, BiLSTM, encoder–decoder, and spatial interpolation algorithms. The model introduces the following three parts. (1) Feature selection, whereby the correlation calculation of $PM_{2.5}$ concentration with other pollutants and meteorological factors is performed using the maximum correlation minimum redundancy algorithm based on mutual information. (2) Neural network model construction, whereby the temporal and spatial correlation features of $PM_{2.5}$ concentration are extracted using a dual encoder–decoder model. (3) Spatial interpolation, whereby the predicted value of $PM_{2.5}$ concentration without monitoring site is obtained by the IDW spatial interpolation algorithm, and ArcGIS obtains the distribution map of pollutant concentration in the study area. This paper takes the research data of Xi'an city as an example with which to verify the proposed model, and the main conclusions are as follows.

(1) Considering the correlation and redundancy between features can achieve data dimension reduction and increase prediction performance. The experimental results showed that the prediction models considering the influence factors RMSE, MAE and MAPE decreased by 12.2%, 27.3%, and 11.4%, respectively.
(2) The model based on dual encoder–decoder fully extracts the multi-dimensional time series features. The results showed that RMSE, MAE, and MAPE of ED-BiLSTM

were reduced by 17.63%, 22.18%, and 9.8%, respectively, compared with a single-BiLSTM model.

(3) IDW spatial interpolation results and ArcGIS visualization conclude that Weiyang, Gaoling, and Yangling districts are seriously polluted, and that pollution prevention and control in these areas should be strengthened. In addition, it is recommended that more air quality monitoring stations be installed in the main urban area of Xi'an, where air quality prediction is complicated due to the complex roads and high traffic flow during the daytime.

The main contribution of this paper is to propose a new method for predicting $PM_{2.5}$ concentration. The prediction results of the model can not only provide real-time air quality warnings and fill in the missing monitoring values caused by instrument failures, they can also capture the distribution of $PM_{2.5}$ concentrations in the study area. In addition, it should be noted that the innovation of this paper is the idea that other feature selection algorithms can be used in future research and that the encoder–decoder model structure can be adapted by other neural network models. The proposed model has a good generalization ability.

Author Contributions: Writing—original draft and methodology, F.Y.; writing—review and funding acquisition, G.H. writing—reviewing and editing, revision. Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: The article is supported by the National Natural Science Foundation of China (71874134). Key Project of Basic Natural Science Research Plan of Shaanxi Province (2019JZ-30).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this paper are publicly available at China Air Quality Online Monitoring and Analysis Platform (https://www.aqistudy.cn/historydata/, accessed on 1 May 2022), and China Meteorological Data Network (http://data.cma.cn/, accessed on 1 May 2022).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

CMAQ	community multiscale air quality
NAQPMS	nested air quality prediction and modeling system
WRF	weather research and forecasting
ARIMA	autoregressive integrated moving average
ARMAX	autoregressive moving average with extra input
BP	back propagation
SVR	support vector regression
ELM	extreme learning machine
RNN	recurrent neural network
LSTM	long- and short-term memory
BiLSTM	bidirectional long short-term memory
MRMR	maximum correlation minimum redundancy
IDW	inverse distance weight
MI	mutual information
ACF	autocorrelation coefficient
PACF	partial autocorrelation coefficient
RMSE	root-mean-square error
MAE	mean absolute error
MAPE	mean absolute percentage error
R^2	R squared
BDS	Brock–Dechert–Scheinman

References

1. Jiang, F.; Zhang, C.; Sun, S.; Sun, J. Forecasting hourly PM2.5 based on deep temporal convolutional neural network and decomposition method. *Appl. Soft Comput.* **2021**, *113*, 107988. [CrossRef]
2. Wang, Z.; Chen, L.; Ding, Z.; Chen, H. An enhanced interval PM2.5 concentration forecasting model based on BEMD and MLPI with influencing factors. *Atmos. Environ.* **2020**, *223*, 117200. [CrossRef]
3. Liu, F.; Tan, Q.; Jiang, X.; Yang, F.; Jiang, W. Effects of relative humidity and PM2.5 chemical compositions on visibility impairment in Chengdu, China. *J. Environ. Sci.* **2019**, *86*, 15–23. [CrossRef] [PubMed]
4. Kan, H.; Chen, R.; Tong, S. Ambient air pollution, climate change, and population health in China. *Environ. Int.* **2012**, *42*, 10–19. [CrossRef]
5. Weber, S.A.; Insaf, T.Z.; Hall, E.S.; Talbot, T.O.; Huff, A.K. Assessing the impact of fine particulate matter (PM2.5) on respiratory-cardiovascular chronic diseases in the New York City Metropolitan area using Hierarchical Bayesian Model estimates. *Environ. Res.* **2016**, *151*, 399–409. [CrossRef]
6. Ashtari, F.; Esmaeil, N.; Mansourian, M.; Poursafa, P.; Mirmosayyeb, O.; Barzegar, M.; Pourgheisari, H. An 8-year study of people with multiple sclerosis in Isfahan, Iran: Association between environmental air pollutants and severity of disease. *J. Neuroimmunol.* **2018**, *319*, 106–111. [CrossRef]
7. Li, J.; Liu, H.; Lv, Z.; Zhao, R.; Deng, F.; Wang, C.; Qin, A.; Yang, X. Estimation of PM2.5 mortality burden in China with new exposure estimation and local concentration-response function. *Environ. Pollut.* **2018**, *243*, 1710–1718. [CrossRef]
8. Zhang, B.; Zou, G.; Qin, D.; Lu, Y.; Jin, Y.; Wang, H. A novel Encoder-Decoder model based on read-first LSTM for air pollutant prediction. *Sci. Total Environ.* **2021**, *765*, 144507. [CrossRef]
9. Huang, G.; Li, X.; Zhang, B.; Ren, J. PM2.5 concentration forecasting at surface monitoring sites using GRU neural network based on empirical mode decomposition. *Sci. Total Environ.* **2021**, *768*, 144516. [CrossRef]
10. Sun, W.; Huang, C. A hybrid air pollutant concentration prediction model combining secondary decomposition and sequence reconstruction. *Environ. Pollut.* **2020**, *266*, 115216. [CrossRef]
11. Bildirici, M.; Ersin, Ö. Markov-switching vector autoregressive neural networks and sensitivity analysis of environment, economic growth and petrol prices. *Environ. Sci. Pollut. Res.* **2018**, *25*, 31630–31655. [CrossRef] [PubMed]
12. Bildirici, M.; Ersin, Ö.Ö. Economic growth and CO_2 emissions: An investigation with smooth transition autoregressive distributed lag models for the 1800–2014 period in the USA. *Environ. Sci. Pollut. Res.* **2018**, *25*, 200–219. [CrossRef] [PubMed]
13. Astitha, M.; Luo, H.; Rao, S.T.; Hogrefe, C.; Mathur, R.; Kumar, N. Dynamic evaluation of two decades of WRF-CMAQ ozone simulations over the contiguous United States. *Atmos. Environ.* **2017**, *164*, 102–116. [CrossRef]
14. Wang, Z.; Itahashi, S.; Uno, I.; Pan, X.; Osada, K.; Yamamoto, S.; Nishizawa, T.; Tamura, K.; Wang, Z. Modeling the long-range transport of particulate matters for January in East Asia using NAQPMS and CMAQ. *Aerosol Air Qual. Res.* **2017**, *17*, 3065–3078. [CrossRef]
15. Tie, X.; Geng, F.; Peng, L.; Gao, W.; Zhao, C. Measurement and modeling of O_3 variability in Shanghai, China: Application of the WRF-Chem model. *Atmos. Environ.* **2009**, *43*, 4289–4302. [CrossRef]
16. Zhang, L.; Lin, J.; Qiu, R.; Hu, X.; Zhang, H.; Chen, Q.; Tan, H.; Lin, D.; Wang, J. Trend analysis and forecast of PM2.5 in Fuzhou, China using the ARIMA model. *Ecol. Indic.* **2018**, *95*, 702–710. [CrossRef]
17. Yu, H.; Yuan, J.; Yu, X.; Zhang, L.; Chen, W. Tracking prediction model for PM2.5 hourly concentration based on ARMAX. *J. Tianjin Univ.* **2017**, *50*, 105–111. [CrossRef]
18. Dong, L.; Li, S.; Yang, J.; Shi, W.; Zhang, L. Investigating the performance of satellite-based models in estimating the surface PM2.5 over China. *Chemosphere* **2020**, *256*, 127051. [CrossRef]
19. Wang, D.; Wei, S.; Luo, H.; Yue, C.; Grunder, O. A novel hybrid model for air quality index forecasting based on two-phase decomposition technique and modified extreme learning machine. *Sci. Total Environ.* **2017**, *580*, 719–733. [CrossRef]
20. Zhou, Y.; Chang, F.J.; Chang, L.C.; Kao, I.F.; Wang, Y.S.; Kang, C.C. Multi-output support vector machine for regional multi-step-ahead PM2.5 forecasting. *Sci. Total Environ.* **2018**, *651*, 230–240. [CrossRef]
21. Voulodimos, A.; Doulamis, N.; Doulamis, A.; Protopapadakis, E. Deep learning for computer vision: A brief review. *Comput. Intell. Neurosci.* **2018**, *2018*, 7068349. [CrossRef] [PubMed]
22. Ioannidou, A.; Chatzilari, E.; Nikolopoulos, S.; Kompatsiaris, I. Deep learning advances in computer vision with 3d data: A survey. *ACM Comput. Surv. (CSUR)* **2017**, *50*, 1–38. [CrossRef]
23. Rawat, W.; Wang, Z. Deep convolutional neural networks for image classification: A comprehensive review. *Neural Comput.* **2017**, *29*, 2352–2449. [CrossRef] [PubMed]
24. Kasthuri, E.; Balaji, S. Natural language processing and deep learning chatbot using long short term memory algorithm. *Mater. Today Proc.* **2021**, *81*, 690–693. [CrossRef]
25. Biancofiore, F.; Busilacchio, M.; Verdecchia, M.; Tomassetti, B.; Aruffo, E.; Bianco, S.; Tommaso, S.D.; Colangeli, C.; Rosatelli, G.; Carlo, P.D. Recursive neural network model for analysis and forecast of PM10 and PM2.5. *Atmos. Pollut. Res.* **2017**, *8*, 652–659. [CrossRef]
26. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]

27. Ma, J.; Li, Z.; Cheng, J.C.P.; Ding, Y.; Lin, C.; Xu, Z. Air quality prediction at new stations using spatially transferred bi-directional long short-term memory network. *Sci. Total Environ.* **2020**, *705*, 135771. [CrossRef]
28. Zhao, J.; Deng, F.; Cai, Y.; Chen, J. Long short-term memory—Fully connected (LSTM-FC) neural network for PM2.5 concentration prediction. *Chemosphere* **2019**, *220*, 486–492. [CrossRef]
29. Zhang, L.; Liu, P.; Zhao, L.; Wang, G.; Zhang, W.; Liu, J. Air quality predictions with a semi-supervised bidirectional LSTM neural network. *Atmos. Pollut. Res.* **2021**, *12*, 328–339. [CrossRef]
30. Seng, D.; Zhang, Q.; Zhang, X.; Chen, G.; Chen, X. Spatiotemporal prediction of air quality based on LSTM neural network. *Alex. Eng. J.* **2021**, *60*, 2021–2032. [CrossRef]
31. Wu, X.; Liu, Z.; Yin, L.; Zheng, W.; Song, L.; Tian, J.; Yang, B.; Liu, S. A Haze Prediction Model in Chengdu Based on LSTM. *Atmosphere* **2021**, *12*, 1479. [CrossRef]
32. Wang, J.; Li, J.; Wang, X.; Wang, J.; Huang, M. Air quality prediction using CT-LSTM. *Neural Comput. Appl.* **2021**, *33*, 4779–4792. [CrossRef]
33. Graves, A.; Schmidhuber, J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Netw.* **2005**, *18*, 602–610. [CrossRef] [PubMed]
34. Bai, Y.; Li, Y.; Zeng, B.; Li, C.; Zhang, J. Hourly PM2. 5 concentration forecast using stacked autoencoder model with emphasis on seasonality. *J. Clean. Prod.* **2019**, *224*, 739–750. [CrossRef]
35. Pak, U.; Ma, J.; Ryu, U.; Ryom, K.; Juhyok, U.; Pak, K.; Pak, C. Deep learning-based PM2.5 prediction considering the spatiotemporal correlations: A case study of Beijing, China. *Sci. Total Environ.* **2020**, *699*, 133561. [CrossRef]
36. Chang, Y.-S.; Chiao, H.-T.; Abimannan, S.; Huang, Y.-P.; Tsai, Y.-T.; Lin, K.-M. An LSTM-based aggregated model for air pollution forecasting. *Atmos. Pollut. Res.* **2020**, *11*, 1451–1463. [CrossRef]
37. Zhang, B.; Zhang, H.; Zhao, G.; Lian, J. Constructing a PM2.5 concentration prediction model by combining auto-encoder with Bi-LSTM neural networks. *Environ. Model. Softw.* **2020**, *124*, 104600. [CrossRef]
38. Bui, T.-C.; Le, V.-D.; Cha, S.-K. A deep learning approach for air pollution forecasting in south korea using encoder-decoder networks & lstm. *arXiv* **2018**, arXiv:1804.07891. [CrossRef]
39. Lyu, P.; Chen, N.; Mao, S.; Li, M. LSTM based encoder-decoder for short-term predictions of gas concentration using multi-sensor fusion. *Process Saf. Environ. Prot.* **2020**, *137*, 93–105. [CrossRef]
40. Feng, H.; Zhang, X. A novel encoder-decoder model based on Autoformer for air quality index prediction. *PLoS ONE* **2023**, *18*, e0284293. [CrossRef]
41. Jamei, M.; Ali, M.; Malik, A.; Karbasi, M.; Rai, P.; Yaseen, Z.M. Development of a TVF-EMD-based multi-decomposition technique integrated with Encoder-Decoder-Bidirectional-LSTM for monthly rainfall forecasting. *J. Hydrol.* **2023**, *617*, 129105. [CrossRef]
42. Yang, H.; Wang, C.; Li, G. A new combination model using decomposition ensemble framework and error correction technique for forecasting hourly PM2.5 concentration. *J. Environ. Manag.* **2022**, *318*, 115498. [CrossRef] [PubMed]
43. Sun, W.; Sun, J. Daily PM2. 5 concentration prediction based on principal component analysis and LSSVM optimized by cuckoo search algorithm. *J. Environ. Manag.* **2017**, *188*, 144–152. [CrossRef] [PubMed]
44. Jain, N.; Murthy, C.A. A new estimate of mutual information based measure of dependence between two variables: Properties and fast implementation. *Int. J. Mach. Learn. Cybern.* **2016**, *7*, 857–875. [CrossRef]
45. Ryu, U.; Wang, J.; Kim, T.; Kwak, S.; Juhyok, U. Construction of traffic state vector using mutual information for short-term traffic flow prediction. *Transp. Res. Part C Emerg. Technol.* **2018**, *96*, 55–71. [CrossRef]
46. Jebli, I.; Belouadha, F.-Z.; Kabbaj, M.I.; Tilioua, A. Prediction of solar energy guided by pearson correlation using machine learning. *Energy* **2021**, *224*, 120109. [CrossRef]
47. Li, X.; Peng, L.; Yao, X.; Cui, S.; Hu, Y.; You, C.; Chi, T. Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. *Environ. Pollut.* **2017**, *231*, 997–1004. [CrossRef]
48. Gao, D.; Zhou, Y.; Wang, T.; Wang, Y. A method for predicting the remaining useful life of lithium-ion batteries based on particle filter using Kendall rank correlation coefficient. *Energies* **2020**, *13*, 4183. [CrossRef]
49. Schober, P.; Boer, C.; Schwarte, L.A. Correlation Coefficients: Appropriate Use and Interpretation. *Anesth. Analg.* **2018**, *126*, 1763–1768. [CrossRef]
50. Peng, H.; Long, F.; Ding, C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238. [CrossRef]
51. Huang, G.; Zhao, X.; Lu, Q. A new cross-domain prediction model of air pollutant concentration based on secure federated learning and optimized LSTM neural network. *Environ. Sci. Pollut. Res.* **2023**, *30*, 5103–5125. [CrossRef] [PubMed]
52. Liu, X.; Qin, M.; He, Y.; Mi, X.; Yu, C. A new multi-data-driven spatiotemporal PM2.5 forecasting model based on an ensemble graph reinforcement learning convolutional network. *Atmos. Pollut. Res.* **2021**, *12*, 101197. [CrossRef]
53. Ma, J.; Ding, Y.; Gan, V.J.L.; Lin, C.; Wan, Z. Spatiotemporal Prediction of PM2.5 Concentrations at Different Time Granularities Using IDW-BLSTM. *IEEE Access* **2019**, *7*, 107897–107907. [CrossRef]
54. Zhou, Y.; Chang, F.-J.; Chang, L.-C.; Kao, I.-F.; Wang, Y.-S. Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts. *J. Clean. Prod.* **2019**, *209*, 134–145. [CrossRef]

55. Harman, B.I.; Koseoglu, H.; Yigit, C.O. Performance evaluation of IDW, Kriging and multiquadric interpolation methods in producing noise mapping: A case study at the city of Isparta, Turkey. *Appl. Acoust.* **2016**, *112*, 147–157. [CrossRef]
56. Greenblatt, S.A. Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence. *Econ. J.* **1993**, *103*, 751–752. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.