toxics

MDPI

Article

A Hybrid Wavelet-Based Deep Learning Model for Accurate Prediction of Daily Surface $PM_{2.5}$ Concentrations in Guangzhou City

Zhenfang He [1,2,*], Qingchun Guo [1,2,3], Zhaosheng Wang [4] and Xinzhou Li [3]

[1] School of Geography and Environment, Liaocheng University, Liaocheng 252000, China; guoqingchun@lcu.edu.cn
[2] Institute of Huanghe Studies, Liaocheng University, Liaocheng 252000, China
[3] State Key Laboratory of Loess and Quaternary Geology, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an 710061, China; lixz@ieecas.cn
[4] National Ecosystem Science Data Center, Key Laboratory of Ecosystem Network Observation and Modeling, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; wangzs@igsnrr.ac.cn
* Correspondence: hezhenfang@lcu.edu.cn

Abstract: Surface air pollution affects ecosystems and people's health. However, traditional models have low prediction accuracy. Therefore, a hybrid model for accurately predicting daily surface $PM_{2.5}$ concentrations was integrated with wavelet (W), convolutional neural network (CNN), bidirectional long short-term memory (BiLSTM), and bidirectional gated recurrent unit (BiGRU). The data for meteorological factors and air pollutants in Guangzhou City from 2014 to 2020 were utilized as inputs to the models. The W-CNN-BiGRU-BiLSTM hybrid model demonstrated strong performance during the predicting phase, achieving an R (correlation coefficient) of 0.9952, a root mean square error (RMSE) of 1.4935 $\mu g/m^3$, a mean absolute error (MAE) of 1.2091 $\mu g/m^3$, and a mean absolute percentage error (MAPE) of 7.3782%. Correspondingly, the accurate prediction of surface $PM_{2.5}$ concentrations is beneficial for air pollution control and urban planning.

Keywords: deep learning; ANN; LSTM; GRU; CNN; BiLSTM; BiGRU; $PM_{2.5}$; wavelet

check for updates

Academic Editor: Nikiforos Alygizakis

Received: 3 March 2025
Revised: 26 March 2025
Accepted: 27 March 2025
Published: 28 March 2025

Citation: He, Z.; Guo, Q.; Wang, Z.; Li, X. A Hybrid Wavelet-Based Deep Learning Model for Accurate Prediction of Daily Surface $PM_{2.5}$ Concentrations in Guangzhou City. *Toxics* **2025**, *13*, 254. https://doi.org/10.3390/toxics13040254

Copyright: © 2025 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

Surface air pollution affects the ecological environment, plant growth, climate change, food production, sustainable social development, and human health [1–9]. Nine out of ten people worldwide breathe polluted air, which causes probably seven million premature deaths each year [10]. Exposure to ambient fine particulate matter ($PM_{2.5}$) causes more than four million premature deaths worldwide each year [11]. Therefore, the accurate prediction of surface $PM_{2.5}$ is of great significance for human health, air quality management, urban planning, and government decision-making.

The highly nonlinear and complex relationship between meteorological variables and surface $PM_{2.5}$ concentrations is often encountered, which makes it impossible to map using conventional statistical regression models and nonlinear statistical models, including linear regression, generalized linear regression, land-use regression, geographically weighted regression, autoregressive moving averages, and several nonlinear statistical regression models [12–14]. With the development of artificial intelligence (AI), it has been widely applied in many fields [15–19]. Machine learning (ML)-based models like artificial neural networks (ANNs) that have become increasingly fashionable in air pollution prediction

provide an attractive alternative [20–22]. Wavelet-ANN (WANN) has the characteristics of nonlinearity, adaptability, and self-organization. It performs better than ANN in predicting $PM_{2.5}$ time series, but WNN still has the problems of slow convergence speed and low prediction accuracy [23,24].

In recent years, deep learning (DL) techniques have more parameters and deeper structures, which can extract low-level features from original data and improve forecast performance [25–27]. Thanks to multi-layer learning, DL can also exactly approximate highly complex nonlinear relations and demonstrate advantages over traditional ML [28,29]. More and more environment researchers have applied deep learning models to environment research [16,30–33]. The DL algorithm has acquired good outcomes in the analysis and forecast of $PM_{2.5}$, but there are still some challenges that need further in-depth research. CNN can extract features from time-series data, but it cannot uncover the problem of long dependencies in time-series data, and a single CNN model cannot fully capture the time-series information of historical data [34,35]. Therefore, RNN with strong modeling ability for time series can be introduced. The RNN has significantly improved the forecast accuracy of $PM_{2.5}$ concentrations in the Seoul metropolitan area [36]. Specifically, more complex model units have been developed in RNN using gating mechanisms. GRU, as a variant of RNNs, can not only process time-series information but also effectively solve the problem of model gradient vanishing. However, GRU only focuses on the forward sequence information of the time series and does not consider the correlation of the reverse sequence [37]. However, Bidirectional GRU (BiGRU) can be used to fully capture the long-term dependencies present in the time series [38]. The internal structure of LSTM allows the network to selectively retain or forget information, which helps to handle long-term dependencies in temporal data [39]. Compared with RNN and multivariate linear regression (MLR), the LSTM model has a higher performance in predicting particulate matter (PM) in South Korea [40,41]. However, for longer time series, LSTM may forget earlier information and, therefore, cannot learn all the content of the data [42,43]. Bidirectional LSTM (BiLSTM) can consider the global information of the data and avoid forgetting earlier content due to too long temporal data. Meanwhile, BiLe BiLSTM model results show good performance in predicting time series of daily $PM_{2.5}$ concentrations [44]. When considering robustness, any single artificial STM can effectively learn time-series data and solve the problem of long dependencies that CNN cannot handle [45]. The intelligence model has many limitations [46].

However, the combination model can compensate for the shortcomings of various models, capture the characteristics of daily $PM_{2.5}$ concentrations, and complete good forecast performance. Consequently, combination strategies are more suitable for forecasting air pollutants. A handful of studies have developed CNN-ANN, CNN-RNN, CNN-GRU, CNN-LSTM, CNN-BiGRU, and CNN-BiLSTM models [47–51]. The hybrid CNN-GRU method can learn the variability and complexity of time series, aiming to predict the concentration of $PM_{2.5}$. The 3D CNN-GRU model exhibited better performance than methods such as ARIMA, ANN, support vector regression machine (SVR), GRU, and LSTM [49]. The hybrid CNN-LSTM model outperformed the other models in predicting $PM_{2.5}$ concentration [52]. The hybrid CNN-GRU-LSTM method remarkably improves the performances of CNN, LSTM, and GRU methods in forecasting $PM_{2.5}$ concentrations in Dezhou City [53]. The performance of the 1D-CNN BiLSTM model for $PM_{2.5}$ forecasting is better than that of the BiLSTM model. One reason may be that 1D-CNN BiLSTM can use three different pooling sizes to obtain multi-scale temporal information [54]. The proposed CLSTM-BiGRU model integrates a CNN, an LSTM, and a BiGRU network. The CLSTM-BiGRU model demonstrates its superiority in air pollutant prediction and outperforms the baseline (CLSTM, DTR, BiGRU) methods [55]. The MTCAN method combines

the fast feature extraction ability of CNN and the temporal modeling features of RNN, improving $PM_{2.5}$ prediction accuracy. Moreover, the MTCAN method performance is compared with statistical methods such as SARIMA, ML methods such as SVR and ANN, and deep learning methods such as LSTM, CNN, BiLTM, GRU, and BiGRU. Specifically, the MTCAN method has the best performance among the benchmark methods [56]. A Learning Rate Schedule (LRS) is introduced in deep learning models. Compared with the RMSE of the other methods, LRS-BiLSTM-CNN has the highest prediction performance in forecasting $PM_{2.5}$ concentrations [57]. The findings revealed that the proposed model provided more accurate predictions. The GCN-LSTM-ResNet model significantly improved hourly $PM_{2.5}$ concentration prediction performance. It reduced MAE by about 10.6–20.0% and reduced RMSE by about 13.2–17.1% [58]. The SA–EMD–LSTM method for long-term $PM_{2.5}$ prediction is more accurate compared to other benchmark models [59]. CSBO-VMD-QRGRU-MGO-LSSVM is proposed to predict $PM_{2.5}$ concentration, and the model has the greatest prediction precision [60]. Wavelet transform is suitable for feature extraction and the noise removal of time-series signals and can obtain the high-frequency components of input time-series signals, making it very suitable for handling scenarios with nonlinear $PM_{2.5}$ fluctuations. Wavelet-based deep learning models can effectively overcome the problem of insufficient data over the training period and improve the accuracy of model forecasts [61]. To determine the optimal wavelet's layers in forecasting PM_{10}, a coupled WT-LSTM-SAE method is developed. The results show that the coupled method outperforms BiLSTM and LSTM [62]. The above results indicate that the hybrid models are more suitable for the combination of wavelet transform than the single models in predicting $PM_{2.5}$ concentrations.

However, these studies still have some shortcomings: (1) $PM_{2.5}$ concentration is prone to change over time and seasons, making it difficult for unoptimized prediction models and single prediction models to achieve high prediction accuracy. (2) $PM_{2.5}$ concentration is easily affected by related air pollutants and meteorological factors, which can interfere with prediction accuracy. Not considering these factors will result in poor prediction accuracy. (3) The selection of prediction methods is relatively simple, and most prediction models continue to use simple prediction methods, so there may be some limitations in the accuracy and stability of predictions. (4) The wavelet function affects the prediction results, but its selection is usually subjective, which leads to poor decomposition performance.

Based on the above analysis, there is an urgent need to develop a high-precision hybrid optimization prediction model. Here, our study aims to predict air pollution more accurately using wavelet-based deep learning models. We develop a multi-factor $PM_{2.5}$ optimization prediction model based on wavelet (W), CNN, BiGRU, and BiLSTM, named W-CNN-BiGRU-BiLSTM. Meanwhile, different factors are considered in the models, such as meteorological variables and air pollutants. The input predictors (parameters) are selected by R. By considering time dependence and frequency correlation, wavelet transform is applied to decompose the meteorological data and the air pollutant data into sub-level sequences. The selections of wavelet functions and parameters are accomplished through an optimization algorithm. Different machine learning models and wavelet-based hybrid models are used to forecast daily $PM_{2.5}$ concentrations in Guangzhou City. Among all the models under investigation, the W-CNN-BiGRU-BiLSTM model exhibits the best performance. In summary, this study constructed a wavelet-based deep learning prediction model, which can provide efficient and simplified models for air pollution prediction and help further explore the application of the networks in other fields.

2. Data and Methods

2.1. Data

Guangzhou City is the capital of Guangdong Province in China and a world-class city. The total area is about 7434.0 km^2, with a population of 18.7341 million people and an urbanization rate of 86.48% in 2022. The regional GDP of Guangzhou is about 288,390 billion yuan. Air quality data and meteorological data are collected in Figure 1 and Table 1. The models need to be validated, so it is necessary to partition the raw data. 80% of the data were selected as the training set, 10% as the validation set, and the remaining 10% were used as the test set. In other words, these are divided into three groups: training set (from 1 January 2014 to 30 June 2019), validation set (from 1 July 2019 to 31 March 2020), and testing set (from 1 April 2020 to 31 December 2020).

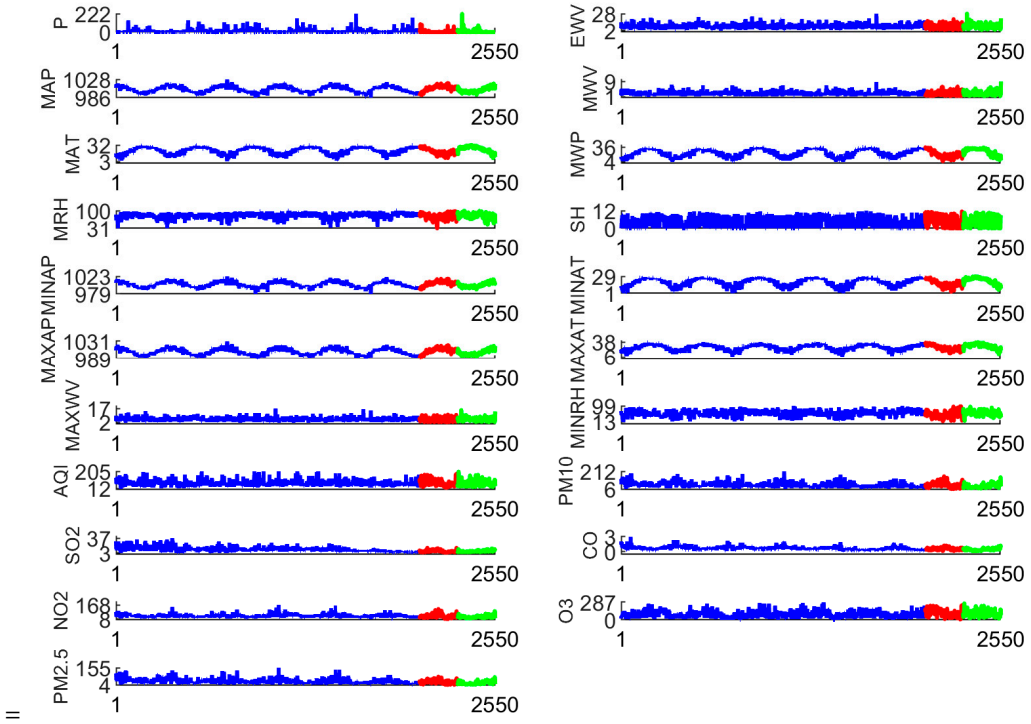


Figure 1. Meteorological and air pollution data of Guangzhou City. The blue line represents training data, the red line represents validation data, and the green line represents testing data.

Table 1. The Pearson correlation coefficients (R) between $PM_{2.5}$ (t + 1) and input predictors used for the model training.

Influence Factor	Abbreviation	R
Precipitation (t)	P (t)	−0.2021
Extreme wind velocity (t)	EWV (t)	−0.3554
Mean atmospheric pressure (t)	MAP (t)	0.4027
Mean wind velocity (t)	MWV (t)	−0.1720
Mean atmospheric temperature (t)	MAT (t)	−0.3696
Mean water pressure (t)	MWP (t)	−0.4440
Mean relative humidity (t)	MRH (t)	−0.3034
Sunshine hours (t)	SH (t)	0.1549
Minimum atmospheric pressure (t)	MINAP (t)	0.4024
Minimum atmospheric temperature (t)	MINAT (t)	−0.4373
Maximum atmospheric pressure (t)	MAXAP (t)	0.4084
Maximum atmospheric temperature (t)	MAXAT (t)	−0.2348
Maximum wind velocity (t)	MAXWV (t)	−0.3388
Minimum relative humidity (t)	MINRH (t)	−0.4134

Table 1. *Cont.*

Influence Factor	Abbreviation	R
AQI (t)	AQI (t)	0.4978
PM_{10} (t)	PM_{10} (t)	0.7203
SO_2 (t)	SO_2 (t)	0.5717
CO (t)	CO (t)	0.5119
NO_2 (t)	NO_2 (t)	0.6166
O_3 (t)	O_3 (t)	0.1748
$PM_{2.5}$ (t)	$PM_{2.5}$ (t)	0.7507
$PM_{2.5}$ (t − 1)	$PM_{2.5}$ (t − 1)	0.5577
$PM_{2.5}$ (t − 2)	$PM_{2.5}$ (t − 2)	0.4705
$PM_{2.5}$ (t − 3)	$PM_{2.5}$ (t − 3)	0.4306
$PM_{2.5}$ (t − 4)	$PM_{2.5}$ (t − 4)	0.3871
$PM_{2.5}$ (t − 5)	$PM_{2.5}$ (t − 5)	0.3511
$PM_{2.5}$ (t − 6)	$PM_{2.5}$ (t − 6)	0.3028

2.2. Wavelet Transformation (WT)

WT is a multi-resolution data analysis method for original signal processing to extract useful frequency information. Compared to Fourier transformation (FT), wavelet transformation (WT) has better time-frequency analysis capability. WT can reveal hidden, detailed information and realize multi-scale decomposition of the original data through the translation of the wavelet function [63,64]. Owing to the simplicity of discrete WT (DWT), DWT is a more common technique in many studies than continuous WT (CWT). DWT can improve the forecast accuracy and reduce overfitting [65]. Therefore, the DWT is used to attain features of the original signals. Decomposition algorithms can reduce the volatility of the original signals and improve the recognition ability of deep learning models for the original signals.

The DWT is used to analyze meteorological factors and air pollutants in this research. The DWT continuously decomposes raw data through high-pass and low-pass filters to obtain high-frequency components (detailed coefficients, CD) and low-frequency components (approximation coefficients, CA) [66]. The DWT of the original data $f(t)$ is calculated by

$$f(t) = \sum_{i=1}^{J} CD_i(t) + CA_J(t), \quad (1)$$

where J is the number of decomposition levels. DWT represents $f(t)$ in terms of the sum of subseries, consisting of high-frequency detail signals CD_1, CD_2, and D_J and a low-frequency approximation signal CA_J.

2.3. Artificial Neural Network (ANN)

An ANN simulates the biological brain, and it is used to predict $PM_{2.5}$ [15,67]. The ANN architecture consists of 3 layers, and each layer is composed of some artificial neurons (nodes) and an activation function (transfer function). Each node is contacted via weights and thresholds. The proposed ANN model for predicting $PM_{2.5}$ concentrations is showcased in Figure 2a.

The activation functions (transfer functions) of the ANN and deep learning models are usually logarithmic sigmoid transfer function (*logsig*, or sigmoid), hyperbolic tangent activation function (*tansig*, or tanh), linear transfer function (*purelin*), and rectified linear units (*relu*, or poslin), respectively. The activation functions are calculated by

$$logsig(b) = \frac{1}{1 + e^{-b}}, \quad (2)$$

$$tansig(b) = \frac{e^{b} - e^{-b}}{e^{b} + e^{-b}}, \tag{3}$$

$$purelin(b) = b, \tag{4}$$

$$relu(b) = \max(0, b), \tag{5}$$

where b represents the corresponding input variable.

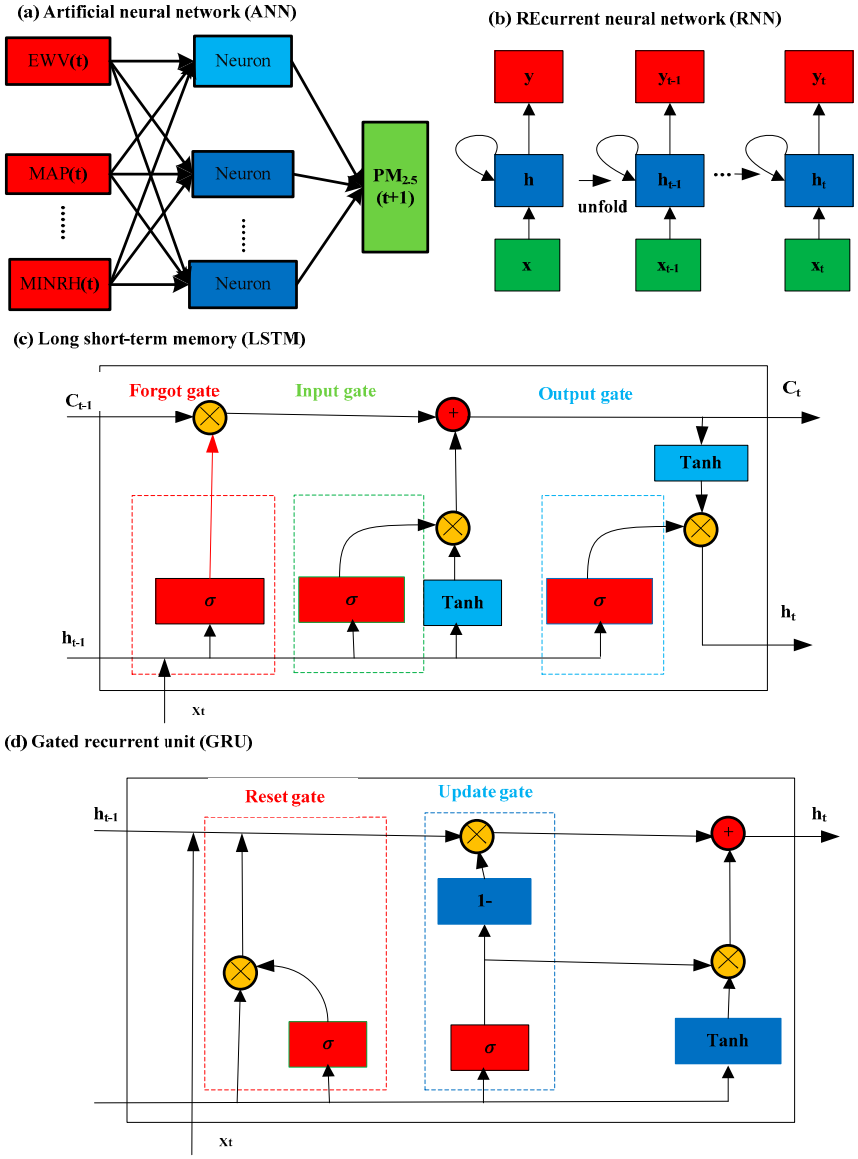


Figure 2. The architectures for predicting $PM_{2.5}$ concentrations.

2.4. Recurrent Neural Network (RNN)

An RNN is a variant of ANN, and it can easily capture the temporal dynamic behavior of input (Figure 2b). Nodes (neurons) in an RNN have a "recursive" property [68]. An RNN has memory ability and can capture temporal dependencies. Therefore, it is suitable for $PM_{2.5}$ time-series analysis.

2.5. Long Short-Term Memory (LSTM)

LSTM is an improved RNN, benefits from the advantages of the RNN technique, and utilizes the unique structure of gates to effectively solve the problem of gradient explosion and vanishing in the RNN. LSTM comprises several cyclic cells whose inputs include the input

features of the current time, the state of network cells at the previous time, and the output of the hidden layer [69]. The LSTM structure is showcased in Figure 2c. σ represents a sigmoid function, Ct − 1 and *Ct* represent cell states, and ht − 1 and ht represent hidden states.

2.6. Gated Recurrent Unit (GRU)

The GRU is a variant of LSTM that maintains the predicted performance of LSTM and effectively solves the problems of gradient explosion and vanishing in RNN while simplifying the structure. The GRU has easier training and simpler construction than LSTM, and it can improve the efficiency of model training. The GRU only has two gates: the reset gate and the update gate. Compared to LSTM, it reduces one less and also reduces matrix multiplication operations. The GRU has fewer parameters and lower computational complexity. Therefore, the GRU takes less time while the magnitude of input data is large [69]. The structure of the GRU is showcased in Figure 2d.

2.7. Bidirectional Long Short-Term Memory (BiLSTM)

BiLSTM is an evolution of LSTM based on Bidirectional RNN, an extension of RNN. BiLSTM consists of two LSTMs: forward and backward (Figure 3a). In the BiLSTM architecture, input information can be processed forward and backward. Information of t can use t − 1 and t + 1 information. Taken overall, BiLSTM is more effective and accurate than unidirectional LSTM [70].

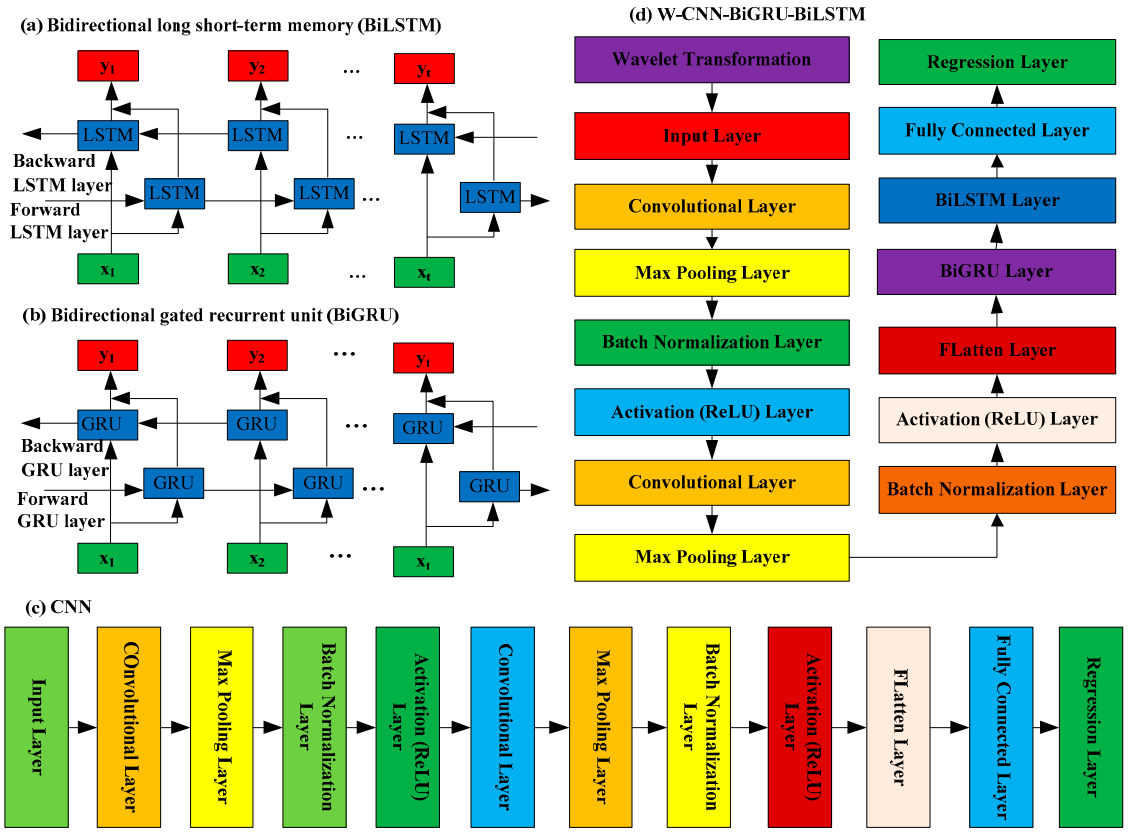


Figure 3. The network architectures of the proposed models.

2.8. Bidirectional Gated Recurrent Unit (BiGRU)

BiGRU consists of two unidirectional GRUs with opposite directions. BiGRU has an additional layer for hidden states [71]. The BiGRU architecture is shown in Figure 3b.

2.9. Convolutional Neural Network (CNN)

The CNN extracts map features from input data and reduces input data dimensionality, and the CNN is successfully applied in air pollution prediction (Figure 3c). The CNN includes convolution, pooling, fully connected, and regression layers [72]. The convolutional kernel, as the core component of a CNN, plays a crucial role in deep learning. The convolutional kernel performs convolution operations of input data through sliding windows to extract local features from the input data. The weights of the convolution kernel are shared. When using the same convolution kernel to perform convolution operations of different regions of the input data, the same weights are used. This parameter-sharing method greatly reduces the number of model parameters and improves the training efficiency of the CNN model. The pooling layer then extracts the most representative features from the obtained convolutional features, and it can reduce overfitting and dimensionality. The CNN is applied to $PM_{2.5}$ time-series data prediction, and the convolution kernel moves only in the temporal direction, so it can extract correlations between local variables.

2.10. Hybrid Models

The complete integrated framework for better forecasting of daily $PM_{2.5}$ concentrations is developed. Hybrid models include CNN-BiGRU, CNN-LSTM, CNN-GRU, CNN-BiLSTM, CNN-LSTM-GRU, CNN-GRU-LSTM, CNN-BiLSTM-BiGRU, CNN-BiGRU-BiLSTM, W-ANN, W-RNN, W-GRU, W-BiGRU, W-CNN, W-LSTM, W-BiLSTM, W-CNN-GRU, W-CNN-BiGRU, W-CNN-LSTM, W-CNN-BiLSTM, W-CNN-LSTM-GRU, W-CNN-GRU-LSTM, W-CNN-BiLSTM-BiGRU, and W-CNN-BiGRU-BiLSTM. By combining wavelet transform with ANN and deep learning, fifteen wavelet-based hybrid models are established. The W-CNN-BiGRU-BiLSTM model combines the advantages of the CNN model, the BiGRU model, the BiLSTM model, and the wavelet transformation technique. That is to say, it is a deep learning framework that combines the decomposition ability of wavelet transform, the powerful feature extraction ability of CNN, and the ability of time-series memory of BiGRU and BiLSTM for learning features related to $PM_{2.5}$ concentrations from input variables. The model structure of W-CNN-BiGRU-BiLSTM is shown in Figure 3d. Firstly, wavelets decompose the raw data to obtain a series of sub-components; then, the CNN extracts features from these sub-components; finally, the BiGRU and BiLSTM utilize these features for $PM_{2.5}$ time-series prediction.

The specific process is as follows:

(1) Feature selection: The correlation coefficient is used to discover the best input features that have the strongest relationship with $PM_{2.5}$ concentration.
(2) Data decomposition: Wavelet functions are used to decompose input variables into high-frequency and low-frequency components.
(3) Combination prediction: Predict $PM_{2.5}$ concentration using multiple deep learning models.
(4) Model evaluation: The prediction results of multiple models are evaluated using evaluation indices.

2.11. Normalization

The raw data were normalized between zero and one to acquire the minimal root mean square error (RMSE) values and the fast convergence of artificial intelligence models [48]. The formula is as follows:

$$b_j^{\Delta} = \frac{b_j - b_{min}}{b_{max} - b_{min}}, \quad (6)$$

where b_j^{Δ} expresses the normalized data, b_{max} is the maximum value of the raw sequence, and b_{min} represents the minimum value of the raw sequence.

After model simulation, the forecasted values should be reversely normalized, and the formula is calculated by

$$b_j = (b_{max} - b_{min})b_j^{\Delta} + b_{min} \quad (7)$$

2.12. Performance Criteria (Metrics)

In order to appraise the predicting performance of the artificial intelligence models, R, *RMSE*, MAPE, and MAE are defined using the following formulas [64]:

$$R = \frac{\sum (G_m - \overline{G})(K_m - \overline{K})}{\sqrt{\sum (G_m - \overline{G})^2 (K_m - \overline{K})^2}}, \quad (8)$$

$$MAPE = \frac{1}{I}\sum \frac{|G_m - K_m|}{G_m} \times 100, \quad (9)$$

$$RMSE = \sqrt{\frac{\sum (G_m - K_m)^2}{I}}, \quad (10)$$

$$MAE = \frac{1}{I}\sum |G_m - K_m|, \quad (11)$$

where G_m and K_m are, respectively, the values of observed and predicted $PM_{2.5}$ data; and I is the length of original data. $\overline{G}$ and $\overline{K}$ are, respectively, the average of the observed and forecasted $PM_{2.5}$ data.

MATLAB 2024 software is used to prepare the methods. Compared to methods that do not use wavelets, the computational burden of using wavelets increases threefold.

3. Results

3.1. Correlation Between Input Predictors and $PM_{2.5}$

It is crucial to conduct an association study between input features and pollutants for a better prediction model. The concentrations of $PM_{2.5}$ are influenced by various variables. Atmospheric temperature has a significant impact on the generation and diffusion of $PM_{2.5}$. Precipitation can affect the wet deposition and chemical reactions of chemicals in the atmosphere. Atmospheric pressure affects atmospheric stability and vertical mixing. Wind speed and direction are crucial for the transmission and diffusion of $PM_{2.5}$. Relative humidity affects the rate of chemical reactions and aerosol formation in the atmosphere. For instance, high wind speed helps to diffuse pollutant concentrations, high atmospheric pressure improves air quality, and high humidity deteriorates air quality [4,73,74]. Therefore, the characteristics of meteorological elements play an important role in air quality forecasting tasks [75]. AQI and other air pollutants also affect $PM_{2.5}$ [76]. In order to select the suitable predictors, the R between the predictors and $PM_{2.5}$ (t + 1) were calculated (Table 1). Here, t represents the current day, t − 1 represents the past day, and t + 1 represents the next day. The threshold of the absolute value of R is 0.15, and the significance level is 0.05. The order of the correlation coefficients of the elements in Table 1 is $PM_{2.5}$ (t), PM_{10} (t), NO_2 (t), SO_2 (t), $PM_{2.5}$ (t − 1), CO (t), AQI (t), $PM_{2.5}$ (t − 2), MWP (t), MINAT (t), $PM_{2.5}$ (t −3), MINRH (t), MAXAP (t), MAP (t), MINAP (t), $PM_{2.5}$ (t − 4), MAT (t), $PM_{2.5}$ (t − 5), EWV (t), MAXWV (t), MRH (t), $PM_{2.5}$ (t − 6), MAXAT (t), P (t), O_3 (t), MWV (t), and SH (t). $PM_{2.5}$ (t + 1) had the highest correlation with $PM_{2.5}$ (t), followed by PM_{10} (t), NO_2 (t), and SO_2 (t). Among meteorological elements, $PM_{2.5}$ (t + 1) was most correlated with mean water pressure (t), followed by minimum atmospheric temperature (t) and minimum relative humidity (t). $PM_{2.5}$ (t + 1) was positively correlated with mean atmospheric pressure (t), sunshine hours (t), minimum atmospheric pressure (t), and maximum atmospheric pressure (t), and

negatively correlated with other meteorological factors. Overall, this analysis validated the substantial impact of meteorological factors on $PM_{2.5}$ levels. Therefore, 27 predictors were selected as input features to the used models. $PM_{2.5}$ (t + 1) was used as the output variable of the used models.

3.2. Selection of Mother Wavelets

The important information of input predictors is extracted by wavelet transformation. The selection of appropriate wavelet functions and decomposition scales affects the prediction results of $PM_{2.5}$. Due to the decisive role of wavelet transformation in input signal feature extraction, it is necessary to select wavelet functions based on the required feature extraction. In general, input signals have different waveform characteristics at different scales. In addition, the spectra of input signals and noise are both nonlinear functions. Due to the unique characteristics of input signals and noise, they can be separated at different scales. Therefore, it is necessary to decompose input signals and noise separately at different scales. The various trends and details of input predictors are acquired by two-level wavelet decomposition. After two-level wavelet decomposition and reconstruction, the input predictors are decomposed into three sections. CA2 expresses the low-frequency information of the input predictors, and CD2 and CD1 express the high-frequency information of the input predictors. The variation characteristics of the input predictors are the key factors affecting mother wavelet selection [77]. The optimal wavelet function can be acquired through continuous iterative search and correlation coefficients [61]. The used mother wavelets are mainly symlets (sym), coiflets (coif), Daubechies (db), and biorthogonal wavelets (bior) [78]. Twenty-seven mother wavelets (wavelet functions) are used to decompose the raw data in order to select the appropriate wavelet function. The smaller the R between CA2, CD1, and CD2 of various wavelet functions, the better. From Table 2, it can be seen that the results of Bior1.1 were smallest, so Bior1.1 was chosen to decompose the original data. The wavelet function (Bior1.1) has good time-domain and frequency-domain characteristics, making it suitable for input signal feature extraction. The quantitative assessment revealed that the components of the input predictors were independent of one another. Figure 4 shows the decomposition results of the original signals using the wavelet function (Bior1.1). Each raw data point is decomposed into three components (approximative CA2, detailed CD1, and CD2). The decomposed components are more periodic and can better reflect the data characteristics of the original data. The raw data are decomposed to reduce their complexity. The fitting degree is significantly improved after wavelet decomposition.

Table 2. R between CD1, CA2, and CD2 for the wavelet bases.

Mother Wavelets	CA2 and CD1	CA2 and CD2	CD1 and CD2
db2	−0.0009	0.9643	0.0006
db3	0.0015	−0.0012	0.0017
db4	0.0005	−0.0005	0.0006
db5	−0.0012	0.0026	−0.001
db6	−0.0019	0.0022	−0.0014
db7	−0.0015	0.0015	0.0001
db8	0.0003	0.0008	0.0003
db9	0.0015	−0.0005	0.0017
db10	0.002	−0.0014	0.0022
sym2	0.0015	−0.0009	0.0006
sym3	0.0015	−0.0012	0.0017
sym4	−0.0014	0.0018	−0.0012
sym5	0.0001	0.0001	0.0006

Table 2. *Cont.*

Mother Wavelets	CA2 and CD1	CA2 and CD2	CD1 and CD2
sym6	−0.0014	0.0017	−0.0016
sym7	0.0015	−0.0006	0.0027
sym8	−0.0018	0.0019	−0.0009
coif1	−0.0016	0.002	−0.0012
coif2	−0.002	0.0021	0.9153
coif3	−0.0022	0.0019	−0.0009
coif4	−0.0021	0.0022	−0.0014
coif5	−0.0021	0.002	−0.0016
bior1.1	0	0	0
bior2.2	0.0029	0.0143	0.0169
bior3.3	0.0005	−0.061	0.0006
bior4.4	−0.0011	−0.0002	0.0098
bior5.5	0.0033	−0.0122	0.0002
bior6.8	−0.0016	0.0013	0.0024

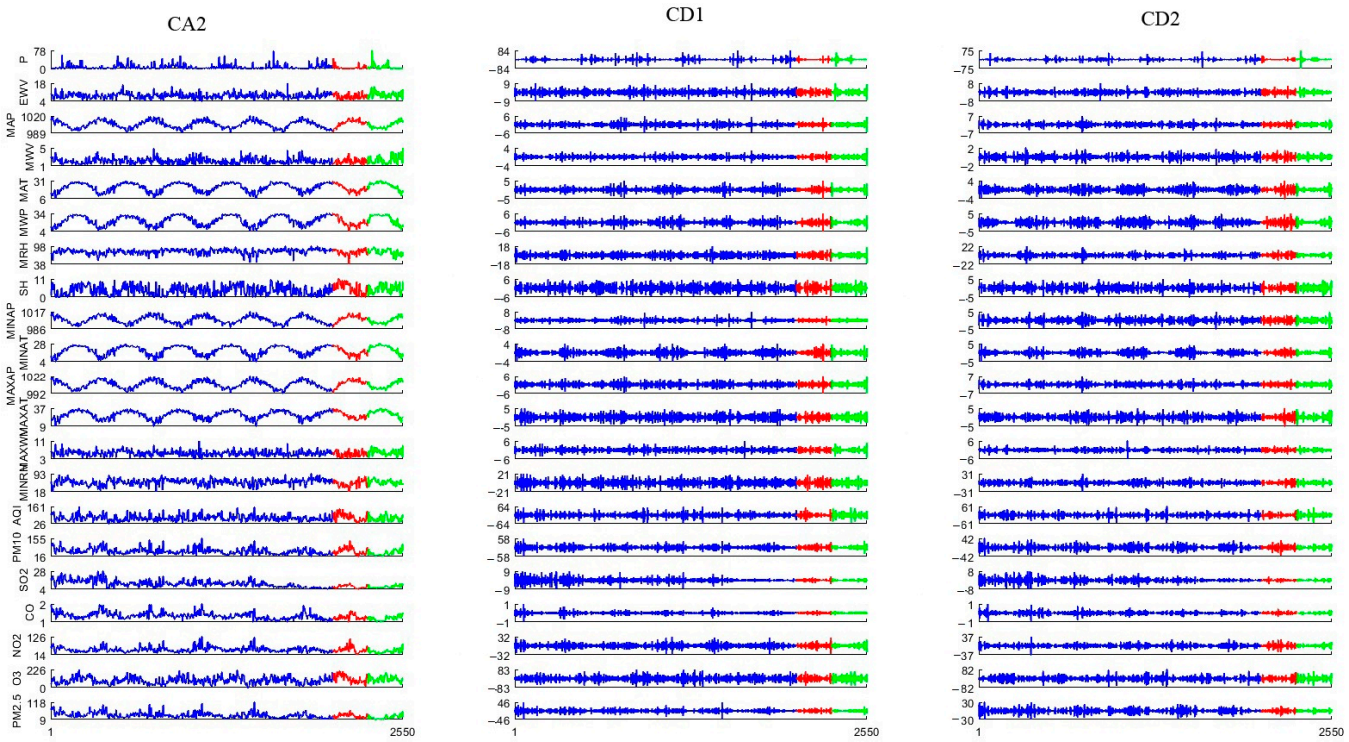


Figure 4. The result of wavelet decomposition. The blue line represents training data, the red line represents validation data, and the green line represents testing data.

3.3. Selection of the Hyperparameters in the Models

A series of hyperparameters of the proposed ANN and deep learning models are ablated, including suitable learning rate, appropriate batch size, and good epochs. A grid search is used to optimize the parameters. Through model training and validation, the hyperparameters of the models are selected in Table 3. The hidden layer of the ANN has 21 neurons. The units of hidden layers in deep learning models are 100. The activation functions of the ANN are logsig and purelin. The activation functions of deep learning models are tanh and sigmoid, the number of epochs is 100, the learning rate is 0.001, the kernel size of the CNN is 3×1, the batch size is 15, the Convolution Filters are 16 and 32, Max-pooling is 2×1, and the Adam optimizer is utilized in deep learning models. The networks are trained and the parameters are updated using the adaptive moment estimation (Adam) optimizer. The Adam optimizer is used to improve the gradient descent method, which is conducive to the convergence of deep learning models [79]. The

hyperparameters of the hybrid models are a combination of the parameters of the single models. The configuration of the W-CNN-BiGRU-BiLSTM model is as follows: the kernel size of the convolutional layer is 3 × 1, the Convolution Filters are 16 and 32, Max-pooling is 2 × 1, the units of hidden layers in BiGRU and BiLSTM are 100, the number of epochs is 100, the learning rate is 0.001, the batch size is 15, and the optimizer is Adam. Furthermore, the learning curve of the W-CNN-BiGRU-BiLSTM model is presented in Figure 5, where the MSE enables a comprehensive conclusion that the best model effectively learns.

Table 3. The hyperparameters of the single models.

Hyperparameters	ANN	RNN	GRU	BiGRU	LSTM	BiLSTM	CNN
Units in hidden layer	21	100	100	100	100	100	
Activation function	logsig-purelin			tanh-sigmoid			Relu
Learning rate	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Batch size	15	15	15	15	15	15	15
Epochs	100	100	100	100	100	100	100
Optimizer	Trainbr	Adam	Adam	Adam	Adam	Adam	Adam
Kernel size							3 × 1
Max-pooling							2 × 1
Convolution Filters							16–32

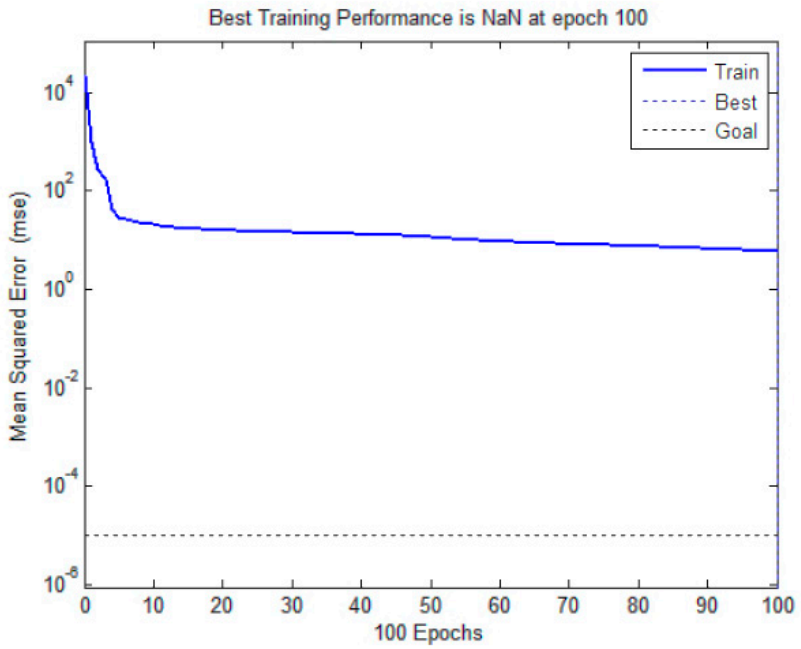


Figure 5. Learning curve convergence for the W-CNN-BiGRU-BiLSTM model.

3.4. Performance Comparison of the Various Models

The model performance indicators (R, RMSE, MAE, and MAPE) of ANN and deep learning-based forecasting techniques are given in Table 4. The performance of deep learning models was much better than that of the ANN. The hybrid models were significantly superior to the single deep learning predicting techniques for all periods. The hybrid models greatly improved the performance of the single models. This means that hybrid models were better able to capture the complex relationship between the predictors and $PM_{2.5}$. For example, the performance of the CNN-LSTM model was much better than that of single LSTM and CNN. Compared to other models, The CNN-BiGRU-BiLSTM model had the highest R value and the lowest MAE value. The CNN-BiGRU-BiLSTM outperformed other models by comparing all levels (R, MAPE, RMSE, and MAE) during the training, validation, and predicting periods. The RMSE values of CNN-BiGRU-BiLSTM are, respectively, 9.0623, 8.8398, and 7.6418 $\mu g/m^3$ during the training, validation, and predicting stages. Nevertheless, the corresponding RMSE values are, respectively, 10.7041, 11.6762, and 10.3658 $\mu g/m^3$ for CNN; 11.4148, 12.1985, and

11.3294 μg/m^3 for BiGRU; and 9.9673, 10.6220, and 9.5961 μg/m^3 for BiLSTM. Similarly, the MAPE values of CNN-BiGRU-BiLSTM are, respectively, 25.7337%, 27.4619%, and 26.2684% during the training, validation, and predicting periods. However, the corresponding MAPE values are, respectively, 29.8351%, 45.0852%, and 48.2678% for CNN; 30.3307%, 45.3682%, and 52.7421% for BiGRU; and 28.5275%, 37.6773%, and 39.7882% for BiLSTM. It is indicated that the employed CNN-BiGRU-BiLSTM could achieve the best simulation performance in all evaluation criteria (R, RMSE, MAE, and MAPE) when compared with other deep learning models.

Table 4. Performance comparison of various deep learning models for simulating $PM_{2.5}$ concentrations during the predicting phase.

Models	R	RMSE (μg/m^3)	MAE (μg/m^3)	MAPE (%)
ANN	0.6630	11.9688	9.9072	62.7491
RNN	0.6779	11.3927	9.7002	54.7036
GRU	0.6793	11.3511	8.9722	54.2364
BiGRU	0.6802	11.3294	8.8436	52.7421
CNN	0.7330	10.3658	8.2465	48.2678
LSTM	0.7458	10.0380	8.1713	45.6636
BiLSTM	0.7609	9.5961	7.6862	39.7882
CNN-GRU	0.7426	10.1655	8.1557	47.2628
CNN-BiGRU	0.7580	9.6270	7.7992	41.7416
CNN-LSTM	0.7810	8.5610	7.5092	38.9337
CNN-BiLSTM	0.7856	8.2600	6.6141	37.7028
CNN-LSTM-GRU	0.8005	7.8311	5.8929	31.9232
CNN-GRU-LSTM	0.8123	7.8235	5.7128	31.1673
CNN-BiLSTM-BiGRU	0.8183	7.8196	5.6382	29.0080
CNN-BiGRU-BiLSTM	0.8323	7.6418	5.5519	26.2684

Similarly, the model performance criteria of W-ANN and wavelet-based deep learning models are given in Table 5. The performance of wavelet-based deep learning models was much better than that of the W-ANN. The proposed W-CNN-BiGRU-BiLSTM model obtained the best simulation performance in all evaluation criteria (R, RMSE, MAE, and MAPE) during the training, validation, and predicting phases. The W-CNN-BiGRU-BiLSTM model had the highest R value and the lowest MAE, RMSE, and MAPE values. For example, the RMSE values of W-CNN-BiGRU-BiLSTM are, respectively, 1.9301, 1.4933, and 1.4935 μg/m^3 during the training, validation, and predicting stages. Nevertheless, the corresponding RMSE values are, respectively, 8.8393, 7.4215, and 7.5615 μg/m^3 for W-CNN; 8.9401, 8.4798, and 8.6062 μg/m^3 for W-BiGRU; and 7.3921, 7.0324, and 6.7973 μg/m^3 for W-BiLSTM. As for the MAPE, the MAPE values of W-CNN-BiGRU-BiLSTM are, respectively, 4.8861%, 5.2563%, and 7.3782% during the training, validation, and predicting stages. Nevertheless, the corresponding MAPE values are, respectively, 19.3726%, 19.2812%, and 28.6538% for W-CNN; 19.6657%, 22.5931%, and 31.4817% for W-BiGRU; and 18.3820%, 18.1092%, and 18.4883% for W-BiLSTM. More importantly, the W-CNN-BiGRU-BiLSTM achieved better prediction performance than W-CNN-LSTM-GRU, W-CNN-GRU-LSTM, and W-CNN-BiLSTM-BiGRU. In particular, the MAPE of the W-CNN-BiGRU-BiLSTM acquired a more accurate forecast value than the corresponding MAPE values among all the other fourteen wavelet-based deep learning forecasting models. Therefore, it has been proven that the proposed W-CNN-BiGRU-BiLSTM model can accomplish relatively stable and more accurate $PM_{2.5}$ prediction.

Table 5. Performance comparison of various wavelet-based deep learning models for simulating $PM_{2.5}$ concentrations during the predicting phase.

Models	R	RMSE (μg/m^3)	MAE (μg/m^3)	MAPE (%)
W-ANN	0.8188	11.5113	9.5490	58.4057
W-RNN	0.8718	8.8379	7.2191	44.6752
W-GRU	0.8890	8.7971	6.8600	34.1093
W-BiGRU	0.9029	8.6062	6.2471	31.4817
W-CNN	0.9161	7.5615	5.5796	28.6538
W-LSTM	0.9133	7.2930	5.0708	20.6160
W-BiLSTM	0.9223	6.7973	4.4111	18.4883
W-CNN-GRU	0.9122	7.3323	5.0871	20.6178
W-CNN-BiGRU	0.9212	7.2920	4.4431	18.7684
W-CNN-LSTM	0.9344	6.6869	3.6688	17.7805
W-CNN-BiLSTM	0.9404	4.6816	3.2193	17.5674
W-CNN-LSTM-GRU	0.9464	4.5310	3.1840	16.5858
W-CNN-GRU-LSTM	0.9489	4.4583	2.7451	14.4150
W-CNN-BiLSTM-BiGRU	0.9859	2.5590	2.0056	10.0762
W-CNN-BiGRU-BiLSTM	0.9952	1.4935	1.2091	7.3782

As shown in Tables 4 and 5, the prediction errors of wavelet-based deep learning models were lower than other deep learning models. By using secondary DWT, the sequence of input variables is decomposed into three subsequences with good stationarity for model training. This facilitates the extraction of mainstream components, thereby significantly improving prediction performance. Through a comprehensive comparison of thirty models, wavelet preprocessing can help improve the prediction of daily $PM_{2.5}$ concentrations. The W-LSTM method outperformed the existing LSTM method in predicting $PM_{2.5}$ concentrations. The W-CNN-BiGRU-BiLSTM model had the best performance by using various evaluation criteria. In addition, the W-ANN was superior to the ANN. Compared to the ANN, the RMSE, MAE, and MAPE of the hybrid W-CNN-BiGRU-BiLSTM prediction model decreased by 87.52%, 87.80%, and 88.24%, respectively. Compared with the LSTM prediction model, the RMSE of the hybrid W-CNN-BiGRU-BiLSTM model decreased by 71.39%, the MAE decreased by 70.27%, and the MAPE decreased by 61.01%. The performance of the W-CNN-BiGRU-BiLSTM model was much better than CNN-BiGRU-BiLSTM. Similarly, the prediction accuracy of the W-CNN-BiLSTM model was higher than CNN-BiLSTM. The prediction error of the W-CNN-GRU-LSTM model was much smaller than CNN-GRU-LSTM. The R values of the W-BiLSTM model were much higher than BiLSTM. Compared to the deep learning prediction models, the wavelet-based deep learning models could better predict nonlinear and non-stationary $PM_{2.5}$ concentrations. The consistency of these results demonstrated the effectiveness of the proposed wavelet-based hybrid prediction models in improving the prediction accuracy of $PM_{2.5}$ concentrations in Guangzhou City. The prediction results of all wavelet-based deep learning models are within the acceptable range. The experimental results show that wavelet transform can effectively improve the prediction accuracy of deep learning, further demonstrating the effectiveness of the combined W-CNN- BiGRU-BiLSTM model proposed in this paper.

The simulating results of $PM_{2.5}$ concentrations using the deep learning models are presented in Figures 6–8. The scatter plots of the ANN model exhibit high bias, with loosely distributed points that overestimate $PM_{2.5}$ values, demonstrating weak capability in capturing nonlinear temporal relationships (Figures 6a, 7a and 8a). Similarly, the RNN, BiGRU, LSTM, CNN, GRU, CNN-GRU, CNN-BiGRU, CNN-LSTM, and CNN-BiLSTM models also overestimate $PM_{2.5}$ values (Figures 6b–k and 8b–k). In contrast, the CNN-LSTM-GRU, CNN-GRU-LSTM, CNN-BiLSTM-BiGRU, and CNN-BiGRU-BiLSTM models

show scatter points closely aligned along the diagonal (Figures 6i–o, 7i–o and 8i–o). Among these, the CNN-BiGRU-BiLSTM model's scatter points are the closest to the diagonal. The simulating performance of the CNN-BiGRU-BiLSTM method was superior to other deep learning models during the training, validation, and prediction stages. Although all deep learning models could predict the long-term future trend of $PM_{2.5}$ concentrations, the proposed CNN-BiGRU-BiLSTM method could better fit the raw $PM_{2.5}$ concentration curve.

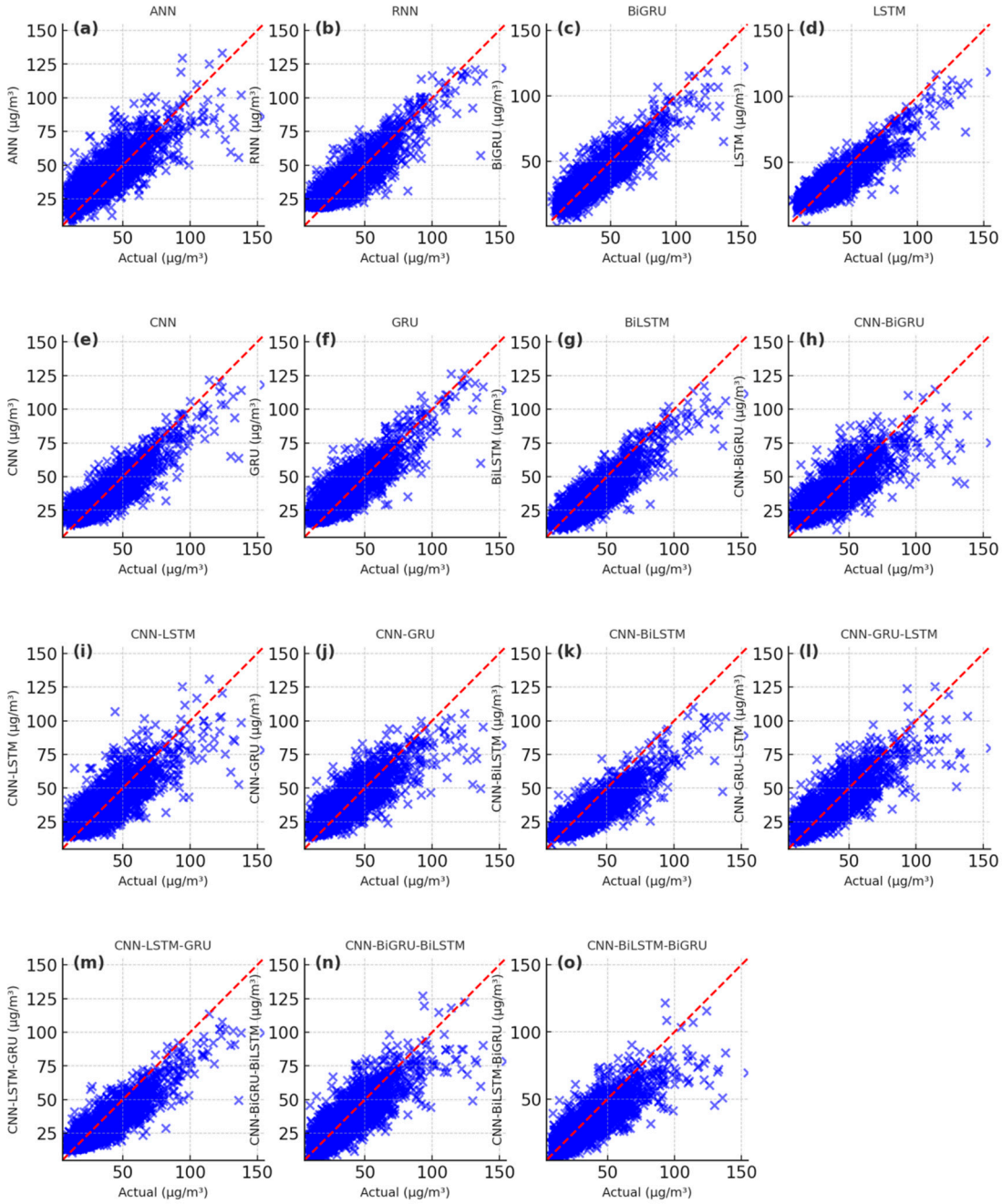


Figure 6. Scatter plot of the simulated $PM_{2.5}$ results during the training period with the deep learning models. (**a**) the simulated results of ANN, (**b**) the simulated results of RNN, (**c**) the simulated results of BiGRU, (**d**) the simulated results of LSTM, (**e**) the simulated results of CNN, (**f**) the simulated results of GRU, (**g**) the simulated results of BiLSTM, (**h**) the simulated results of CNN-BiGRU, (**i**) the simulated results of CNN-LSTM, (**j**) the simulated results of CNN-GRU, (**k**) the simulated results of CNN-BiLSTM, (**l**) the simulated results of CNN-GRU-LSTM, (**m**) the simulated results of CNN-LSTM-GRU, (**n**) the simulated results of CNN-BiGRU-BiLSTM, (**o**) the simulated results of CNN-BiLSTM-BiGRU.

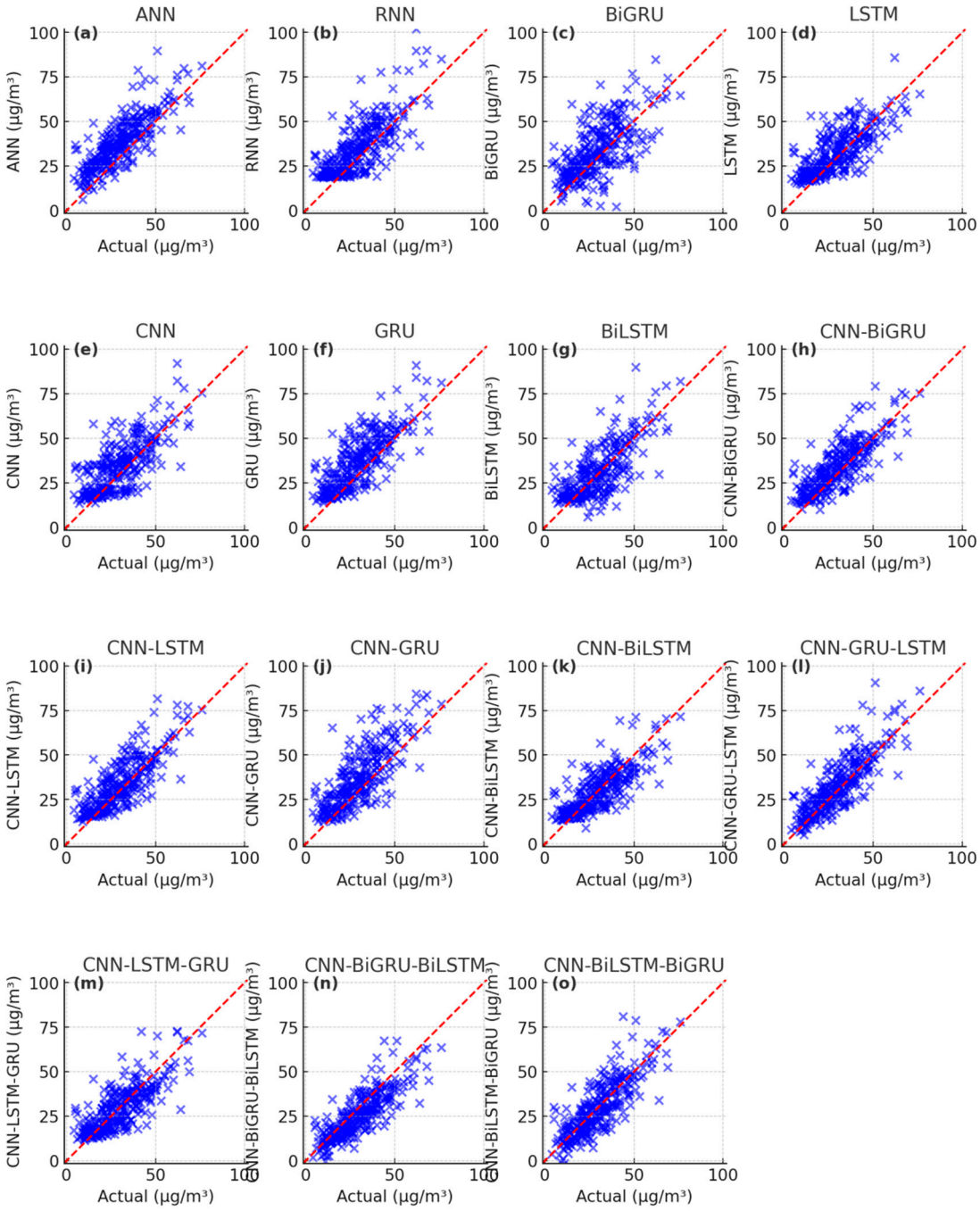


Figure 7. Scatter plot of the simulated $PM_{2.5}$ results during the validation period with the deep learning models. (**a**) the simulated results of ANN, (**b**) the simulated results of RNN, (**c**) the simulated results of BiGRU, (**d**) the simulated results of LSTM, (**e**) the simulated results of CNN, (**f**) the simulated results of GRU, (**g**) the simulated results of BiLSTM, (**h**) the simulated results of CNN-BiGRU, (**i**) the simulated results of CNN-LSTM, (**j**) the simulated results of CNN-GRU, (**k**) the simulated results of CNN-BiLSTM, (**l**) the simulated results of CNN-GRU-LSTM, (**m**) the simulated results of CNN-LSTM-GRU, (**n**) the simulated results of CNN-BiGRU-BiLSTM, (**o**) the simulated results of CNN-BiLSTM-BiGRU.

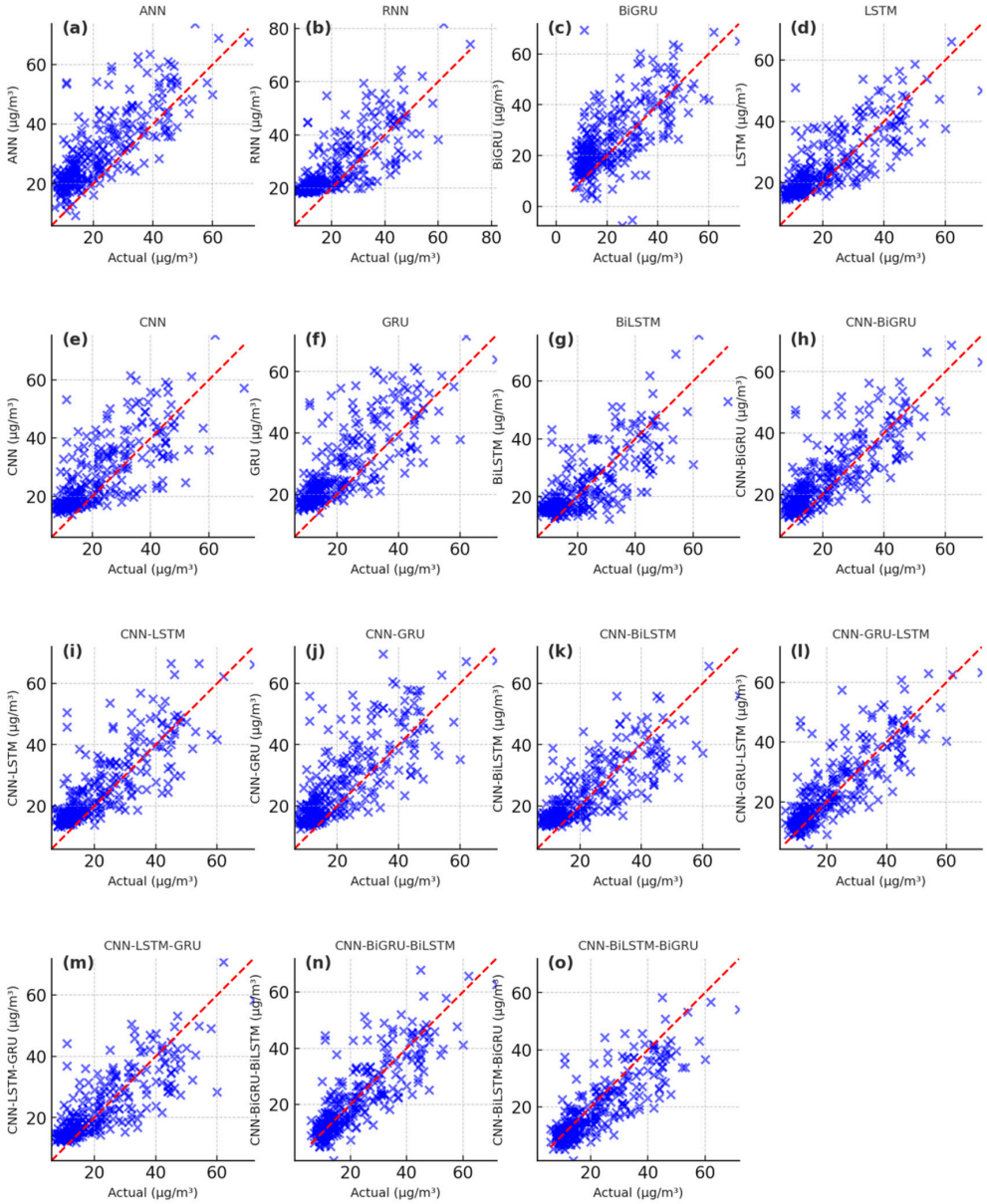


Figure 8. Scatter plot of the predicting $PM_{2.5}$ results during the predicting period with the deep learning models. (**a**) the predicting results of ANN, (**b**) the predicting results of RNN, (**c**) the predicting results of BiGRU, (**d**) the predicting results of LSTM, (**e**) the predicting results of CNN, (**f**) the predicting results of GRU, (**g**) the predicting results of BiLSTM, (**h**) the predicting results of CNN-BiGRU, (**i**) the predicting results of CNN-LSTM, (**j**) the predicting results of CNN-GRU, (**k**) the predicting results of CNN- BiLSTM, (**l**) the predicting results of CNN-GRU-LSTM, (**m**) the predicting results of CNN-LSTM-GRU, (**n**) the predicting results of CNN-BiGRU-BiLSTM, (**o**) the predicting results of CNN-BiLSTM-BiGRU.

The $PM_{2.5}$ fitting and testing results of wavelet-based deep learning models are showcased in Figures 9–11. The scatter plots of the W-ANN, W-RNN, and W-BiGRU models display bias and loosely distributed points (Figures 9a–c, 10a–c and 11a–c), while the W-CNN, W-LSTM, and W-GRU models also exhibit underestimation of peak values, particularly in long-term predictions (Figures 9d–f, 10d–f and 11d–f). The W-BiLSTM model's scatter plots fit the true values better than W-GRU but still show some lag (Figures 9g, 10g and 11g). The W-CNN-GRU, W-CNN-BiGRU, W-CNN-LSTM, and W-CNN-BiLSTM models demonstrate smoothed local fluctuations but suffer from underfitting in long-term trends (Figures 9h–k, 10h–k and 11h–k). On the other hand, the W-CNN-

LSTM-GRU, W-CNN-GRU-LSTM, W-CNN-BiLSTM-BiGRU, and W-CNN-BiGRU-BiLSTM models exhibit scatter points tightly clustered around the diagonal, with smaller errors in high-value regions (Figures 9i–o, 10i–o and 11i–o). The W-CNN-BiLSTM-BiGRU model is more stable than W-CNN-LSTM-GRU and more sensitive to sudden rises/drops in $PM_{2.5}$, especially in long-term predictions (Figures 9n, 10n and 11n). The W-CNN-BiGRU-BiLSTM model's scatter points are more symmetric and closer to the diagonal compared to W-CNN-BiLSTM-BiGRU, demonstrating the best overall performance, the most concentrated error distribution, and higher accuracy in predicting abrupt changes and sustained high $PM_{2.5}$ intervals (Figures 9o, 10o and 11o). In summary, by analyzing the distribution patterns, error concentration, and the ability to capture extreme $PM_{2.5}$ values in these scatter plots, we can intuitively assess the simulation and predictive capabilities of different models.

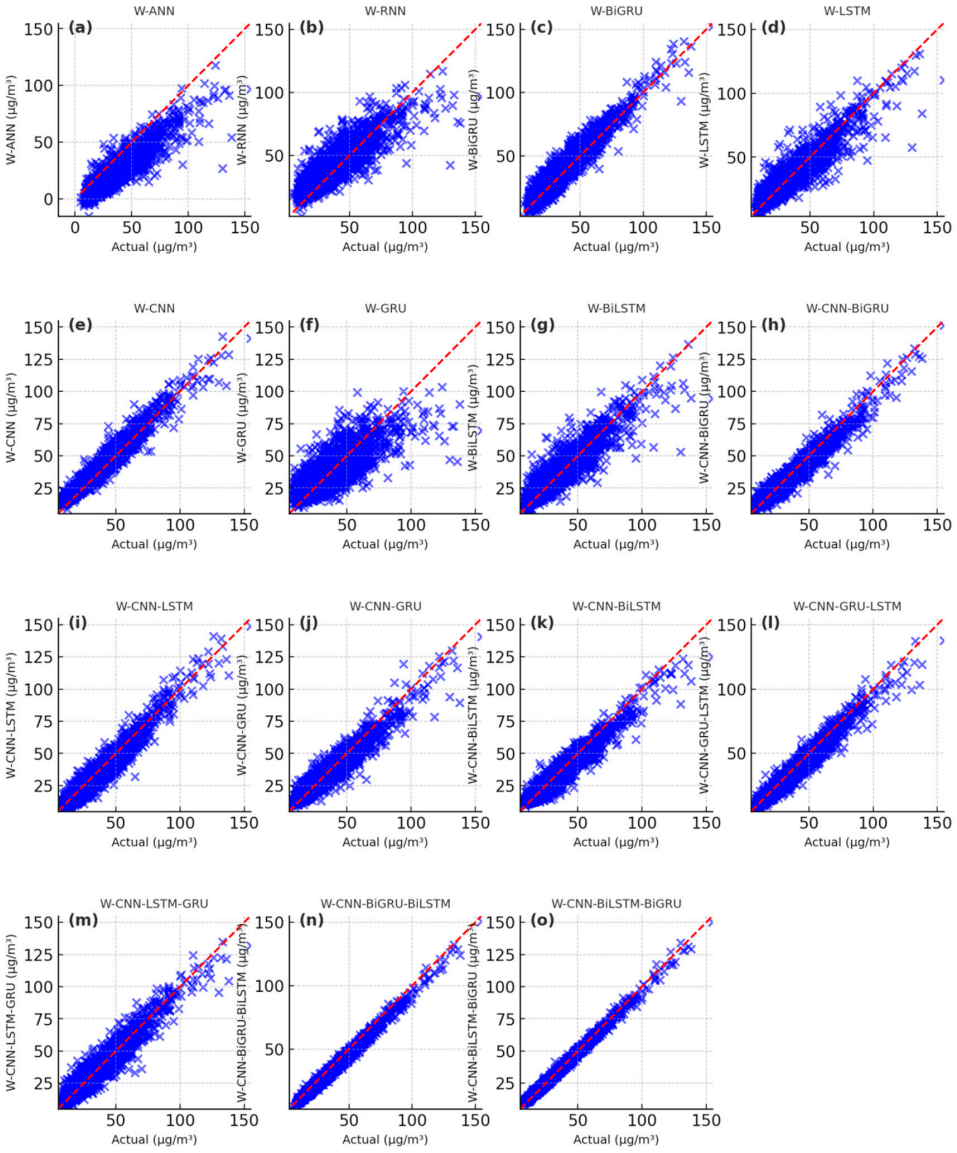


Figure 9. Scatter plot of the simulated $PM_{2.5}$ results during the training period with the wavelet-based deep learning models. (**a**) the simulated results of W-ANN, (**b**) the simulated results of W-RNN, (**c**) the simulated results of W-BiGRU, (**d**) the simulated results of W-LSTM, (**e**) the simulated results of W-CNN, (**f**) the simulated results of W-GRU, (**g**) the simulated results of W-BiLSTM, (**h**) the simulated results of W-CNN-BiGRU, (**i**) the simulated results of W-CNN-LSTM, (**j**) the simulated results of W-CNN-GRU, (**k**) the simulated results of W-CNN- BiLSTM, (**l**) the simulated results of W-CNN-GRU-LSTM, (**m**) the simulated results of W-CNN-LSTM-GRU, (**n**) the simulated results of W-CNN-BiGRU-BiLSTM, (**o**) the simulated results of W-CNN-BiLSTM-BiGRU.

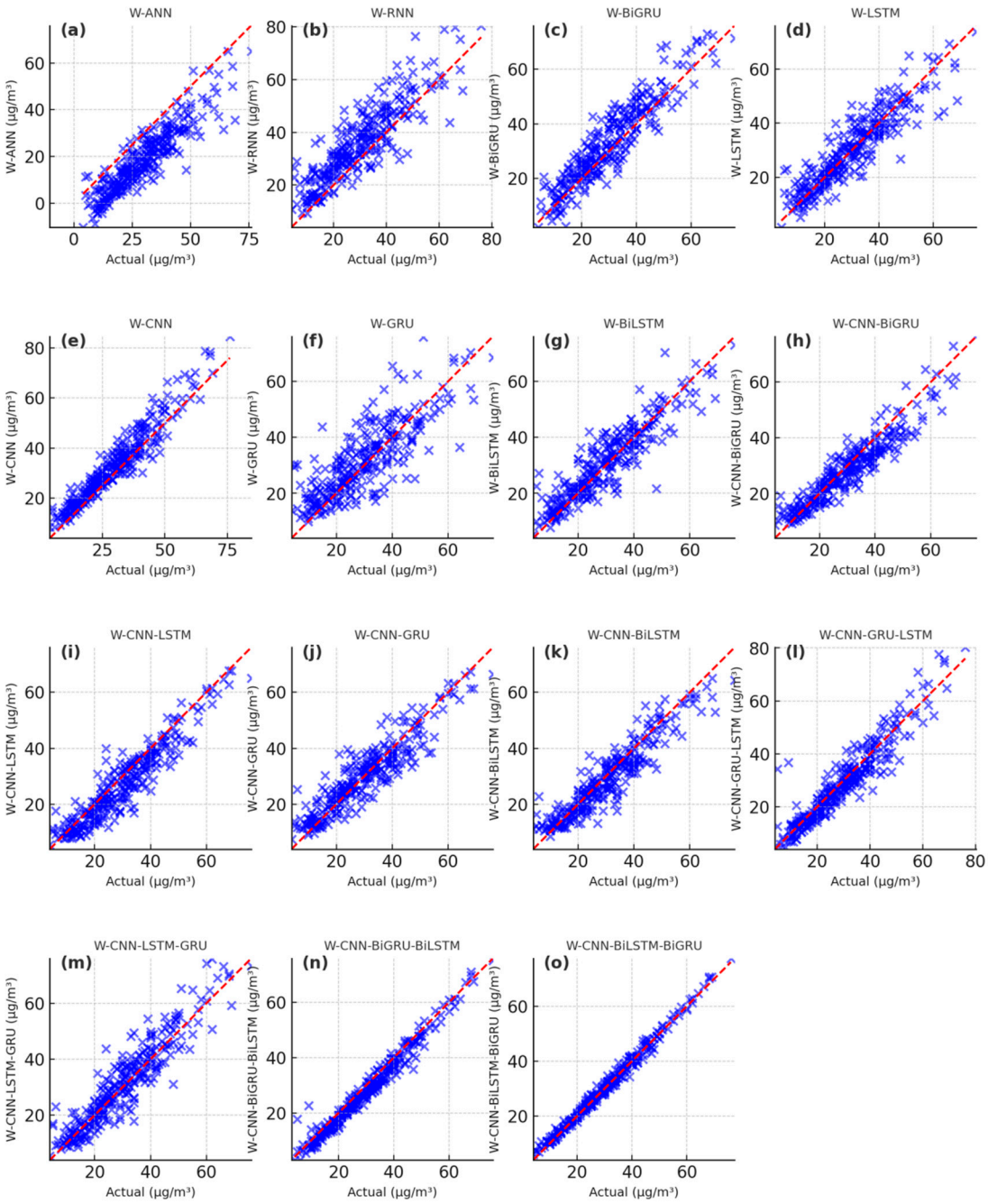


Figure 10. Scatter plot of the simulated $PM_{2.5}$ results during the validation period with the wavelet-based deep learning models. (**a**) the simulated results of W-ANN, (**b**) the simulated results of W-RNN, (**c**) the simulated results of W-BiGRU, (**d**) the simulated results of W-LSTM, (**e**) the simulated results of W-CNN, (**f**) the simulated results of W-GRU, (**g**) the simulated results of W-BiLSTM, (**h**) the simulated results of W-CNN-BiGRU, (**i**) the simulated results of W-CNN-LSTM, (**j**) the simulated results of W-CNN-GRU, (**k**) the simulated results of W-CNN-BiLSTM, (**l**) the simulated results of W-CNN-GRU-LSTM, (**m**) the simulated results of W-CNN-LSTM-GRU, (**n**) the simulated results of W-CNN-BiGRU-BiLSTM, (**o**) the simulated results of W-CNN-BiLSTM-BiGRU.

To our surprise, the hybrid models performed very well in reproducing dynamic features. The simulating performance of the W-CNN-BiGRU-BiLSTM method was superior to other deep learning models during the three stages. Moreover, the figures showcase clearly that the overall performance of W-CNN-BiLSTM-BiGRU and W-CNN-BiGRU-BiLSTM were superior to the W-CNN and W-BiGRU models during the training, validation, and prediction stages. The trends of the W-CNN-BiLSTM-BiGRU and W-CNN-BiGRU-

BiLSTM models had better agreement with observations than the W-CNN and W-BiGRU models. It is indicated that the W-CNN-BiGRU-BiLSTM model performed satisfyingly in the three stages. The trend of the W-CNN-BiGRU-BiLSTM prediction result was closely accordant with the observed $PM_{2.5}$ concentration data.

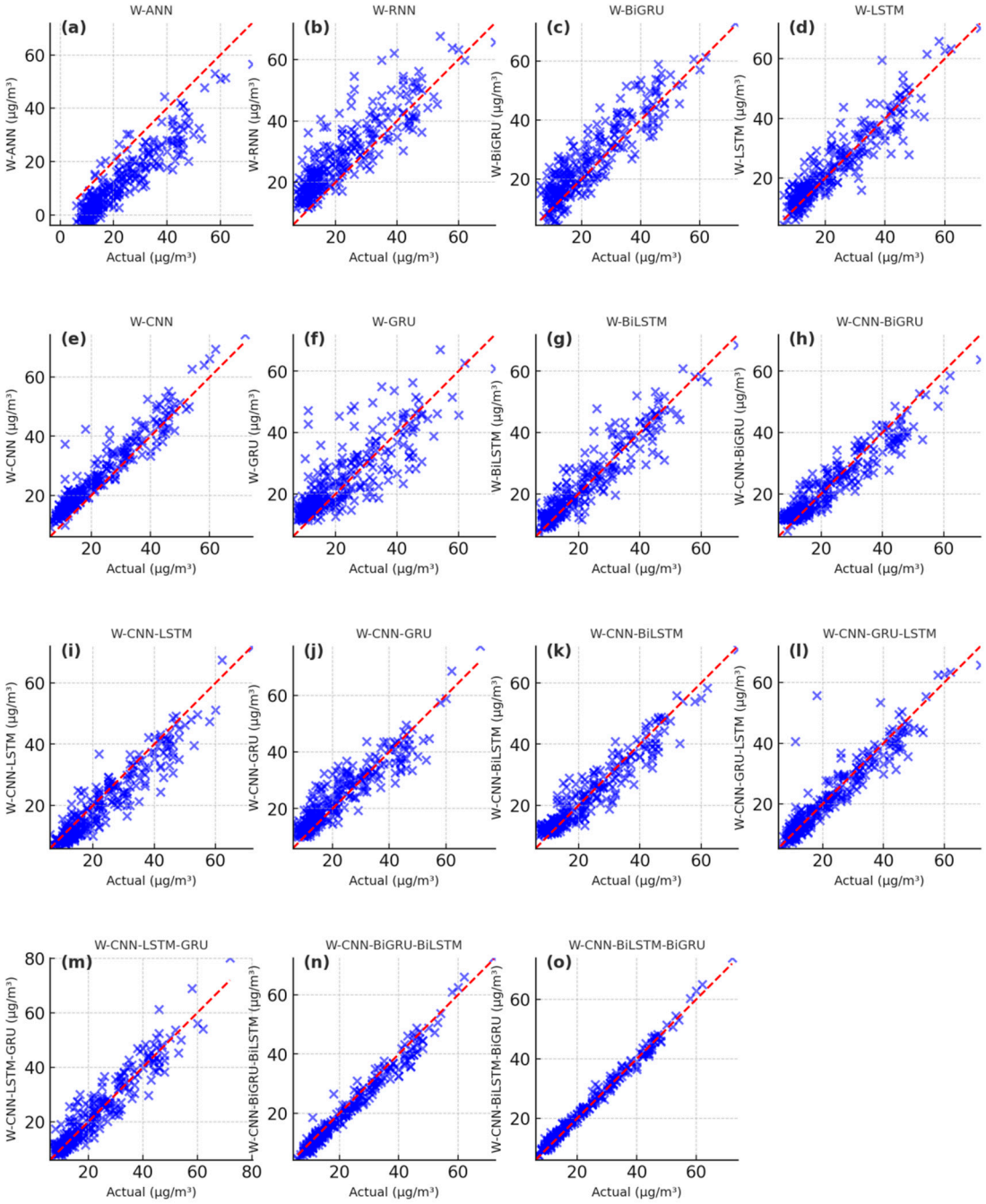


Figure 11. Scatter plot of the predicting $PM_{2.5}$ results during the predicting period with the wavelet-based deep learning models. (**a**) the predicting results of W-ANN, (**b**) the predicting results of W-RNN, (**c**) the predicting results of W-BiGRU, (**d**) the predicting results of W-LSTM, (**e**) the predicting results of W-CNN, (**f**) the predicting results of W-GRU, (**g**) the predicting results of W-BiLSTM, (**h**) the predicting results of W-CNN-BiGRU, (**i**) the predicting results of W-CNN-LSTM, (**j**) the predicting results of W-CNN-GRU, (**k**) the predicting results of W-CNN-BiLSTM, (**l**) the predicting results of W-CNN-GRU-LSTM, (**m**) the predicting results of W-CNN-LSTM-GRU, (**n**) the predicting results of W-CNN-BiGRU-BiLSTM, (**o**) the predicting results of W-CNN-BiLSTM-BiGRU.

4. Discussion

For accurately predicting $PM_{2.5}$ concentrations in Guangzhou City, different machine learning and wavelet-based deep learning models were proposed to forecast $PM_{2.5}$ in Guangzhou City. The performance of the methods was compared by RMSE, MAE, MAPE, and R. The W-CNN-BiGRU-BiLSTM model outperforms the other models. The W-CNN-BiGRU-BiLSTM exhibited significant advantages over other models in processing input factors. The W-CNN-BiGRU-BiLSTM method combines wavelet (W), CNN, BiLSTM, and BiGRU. Wavelet transformation is applied to decompose the meteorological and air pollutant data into sub-level sequences. The CNN algorithm is good at extracting local and spatial features. The BiLSTM and the BiGRU can better handle the bidirectional dependencies of sequence data while considering forward and backward contextual information, thereby improving the accuracy of $PM_{2.5}$ prediction. Correspondingly, simulation outcomes also indicated that the W-CNN-BiGRU-BiLSTM can significantly improve the predictive performance compared to other deep learning models. The wavelet transform technique plays an important role in enhancing the $PM_{2.5}$ prediction ability and accuracy of the deep learning (DL) prediction model. From the above results, the wavelet-based deep learning models outperform the deep learning models in predicting $PM_{2.5}$ concentrations.

Wavelet transforms are powerful tools for signal processing, allowing for multi-resolution analysis. In the context of neural networks, wavelets can be used for preprocessing or feature extraction, enabling the capture of both frequency and time-domain characteristics in input signals. This preprocessing step can significantly enhance the performance of deep learning models when applied to $PM_{2.5}$ time-series data or sequences. LSTM, GRU, and their bidirectional counterparts excel in $PM_{2.5}$ sequence-based tasks due to their ability to capture long-term dependencies.

Hybrid models combine CNNs with other neural networks to leverage the strengths of both architectures. For instance, CNN-LSTM combines CNN's feature extraction capabilities with LSTM's ability to capture temporal dependencies, providing a powerful tool for $PM_{2.5}$ time-series forecasting. Other hybrid models, such as CNN-BiLSTM or CNN-BiGRU, enhance the capability to process sequential data by considering both past and future contexts while also extracting rich features through convolutional layers. These more complex hybrid architectures further combine multiple models to capture various levels of features and temporal dependencies. For example, CNN-LSTM-GRU could be used in scenarios requiring both short-term and long-term memory with feature extraction. Hybrid models improve performance further by combining the feature extraction capabilities of CNN with the sequential learning power of GRU and LSTM. CNN-LSTM-GRU provides flexibility in handling various types of input data, making it suitable for $PM_{2.5}$ time-series forecasting. The intricate architectures like CNN-BiLSTM-BiGRU offer multi-layered mechanisms to process data from both directions (BiLSTM, BiGRU) while extracting features through CNN layers, optimizing performance on $PM_{2.5}$ prediction tasks with complex patterns and dependencies.

Accurate $PM_{2.5}$ prediction is crucial for providing critical information for policy decisions and public health. By providing reliable $PM_{2.5}$ predictions, wavelet-based deep learning models can help stakeholders develop targeted strategies to improve air quality and safeguard public health. In summary, this study demonstrates the effectiveness of wavelet deep learning models in $PM_{2.5}$ prediction and provides valuable insights into optimization strategies for improving model performance. By addressing key hyperparameters, we can improve our understanding of $PM_{2.5}$ prediction. Based on the predicted results, reference data can be provided for urban planning, land uses, and spatial structures to minimize $PM_{2.5}$ emissions.

Although our models have achieved good predictive performance, there are several noteworthy limitations of the wavelet-based deep learning models for $PM_{2.5}$ prediction

in Guangzhou City. Firstly, the prediction accuracy of the prediction models is closely related to input parameters, such as meteorological factors and other air pollutants. Due to the dependence of model development on the specific meteorological conditions of Guangzhou City, the model may not be able to capture the complexity of all environmental factors that affect $PM_{2.5}$ concentrations. In addition, model development is based on daily average data. The lack of real-time hourly meteorological data and pollution emission data in Guangzhou City may affect the accuracy of the proposed model [80]. Thirdly, the lack of external pollution inputs from other regions may lead to the neglect of potential sources of pollution [81]. Additionally, we do not consider the impact of different seasons of El Niño and El Niño on $PM_{2.5}$ concentrations [82]. Fourthly, optimizing all the hyperparameters of each model can be a time-consuming process, so we only considered the common hyperparameters. However, the wavelet-based deep learning model will be widely applied. By incorporating supplementary input variables, such as real-time meteorological data and hourly data from pollution sources and other regions, the accuracy and reliability of the wavelet-based deep learning model can be significantly improved. These data can be obtained through meteorological stations and air quality monitoring stations, which will greatly improve the accuracy and comprehensiveness of the hybrid model. Finally, integrating wavelet-based deep learning models with chemical transport models could improve the accuracy of $PM_{2.5}$ prediction.

5. Conclusions

Wavelet-based deep learning methods for predicting $PM_{2.5}$ concentrations in Guangzhou from 2014 to 2020 were developed and compared. The input variables were derived from the correlation coefficients between $PM_{2.5}$ and predictors, including meteorological factors and air pollutants. Second-level discrete wavelet transformation (DWT) was used to preprocess the raw data and decompose it into three different frequency subsequences. Fifteen wavelet-based hybrid models were compared with fifteen machine learning models using the same historical data and a validation mechanism. When multiple frequency subsequences were utilized as inputs, the prediction accuracy of $PM_{2.5}$ concentrations was significantly improved. Further, the findings showed the potential of the wavelet-based hybrid models in forecasting $PM_{2.5}$ concentrations. The W-CNN-BiGRU-BiLSTM model demonstrated the highest predictive accuracy for $PM_{2.5}$ levels. This proposed W-CNN-BiGRU-BiLSTM method successfully outperformed the benchmark deep learning methods (CNN, BiGRU, and BiLSTM). For the 1-day-ahead forecasting, the proposed W-CNN-BiGRU-BiLSTM model showed an RMSE advantage of 85.59%, 86.82%, and 84.44%, respectively, over the CNN, BiGRU, and BiLSTM models in Guangzhou. The W-CNN-BiGRU-BiLSTM model provided more accurate results. The W-CNN-BiGRU-BiLSTM model approached the W-CNN-BiLSTM-BiGRU model in some individual coefficients, but the relative error distribution indicated that the W-CNN-BiGRU-BiLSTM model had more stable performance. Also, the performance of deep learning models is much better than that of the ANN. Additionally, the prediction accuracy of hybrid models is higher than that of the single models. Moreover, the prediction errors of wavelet-based hybrid models are lower than for other hybrid models. The proposed hybrid W-CNN-BiGRU-BiLSTM model is a promising and practical method for $PM_{2.5}$ prediction. It should be emphasized that the detailed information about the deep learning models and parameter setting rules for the wavelet transformation technique in this research can offer a useful reference for future study. Overall, air pollution prediction models can contribute to sustainable development and to improving people's health, and they can provide information for government agencies to formulate air pollution prevention and control strategies.

In the future, we will apply this method to other cities to verify its effectiveness. Meanwhile, a multi-lead time predicting model of daily $PM_{2.5}$ concentrations will be

studied and developed. $PM_{2.5}$ has certain differences in space and time, but it is feasible to use meteorological elements and pollutants as the input factors of the model in prediction and analysis. The addition of input parameters as predictors can improve the predicting model performance. Other factors need to be considered, including building height, terrain, traffic data, pollutant emissions, vegetation, transportation, land use, population, GDP, etc. New hybrid models will be utilized, such as wavelet transform and transformer, graph convolutional neural (GCN), transfer learning, generative adversarial network (GAN), etc.

Author Contributions: Conceptualization, Z.H. and Q.G.; methodology, Z.H. and Q.G.; software, Z.H. and Q.G.; validation, Z.H. and Q.G.; formal analysis, Z.H., Q.G., Z.W. and X.L.; investigation, Z.H. and Q.G.; resources, Z.H. and Q.G.; data curation, Z.H. and Q.G.; writing—original draft, Z.H. and Q.G.; writing—review and editing, Z.H. and Q.G.; visualization, Z.H. and Q.G.; supervision, Z.H. and Q.G.; project administration, Z.H. and Q.G.; funding acquisition, Z.H. and Q.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Shandong Provincial Natural Science Foundation (ZR2023MD075), the State Key Laboratory of Loess and Quaternary Geology Foundation (SKLLQG2419), LAC/CMA (2023B02), and the Shandong Province Higher Educational Humanities and Social Science Program (J18RA196).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors upon request.

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

References

1. Crippa, M.; Solazzo, E.; Guizzardi, D.; Van Dingenen, R.; Leip, A. Air pollutant emissions from global food systems are responsible for environmental impacts, crop losses and mortality. *Nat. Food* **2022**, *3*, 942–956. [CrossRef]
2. Tripathi, D.P.; Nema, A.K. Seasonal variation of biochemical parameters and air pollution tolerance index (APTI) of selected plant species in Delhi city, and detailed meta-analysis from Indian metropolitan cities. *Atmos. Environ.* **2023**, *309*, 119862. [CrossRef]
3. Abadi, A.R.S.; Hamzeh, N.H.; Kaskaoutis, D.G.; Opp, C.; Kazemi, A.F. Long-Term Spatio-Temporal Analysis, Distribution, and Trends of Dust Events over Iran. *Atmosphere* **2025**, *16*, 334. [CrossRef]
4. Guo, Q.; He, Z.; Wang, Z. Change in Air Quality during 2014–2021 in Jinan City in China and Its Influencing Factors. *Toxics* **2023**, *11*, 210. [CrossRef] [PubMed]
5. Liu, J.; Wang, P.; Shang, L.; Ye, F.; Liu, L.; He, Z. Adverse Associations of Long-Term Exposure to $PM_{2.5}$ and Its Components with Platelet Traits among Subway Shift-Workers without Air Purifier Use. *Toxics* **2024**, *12*, 529. [CrossRef]
6. Botto, L.; Bulbarelli, A.; Lonati, E.; Cazzaniga, E.; Palestini, P. Correlation between Exposure to UFP and ACE/ACE2 Pathway: Looking for Possible Involvement in COVID-19 Pandemic. *Toxics* **2024**, *12*, 560. [CrossRef]
7. Sun, L.; Wei, P.; Westerdahl, D.; Xue, J.; Ning, Z. Evaluating Indoor Air Quality in Schools: Is the Indoor Environment a Haven during High Pollution Episodes? *Toxics* **2024**, *12*, 564. [CrossRef]
8. Zhang, Y.; Fan, L.; Wang, S.; Luo, H. Short-Term Interaction Effects of $PM_{2.5}$ and O_3 on Daily Mortality: A Time-Series Study of Multiple Cities in China. *Toxics* **2024**, *12*, 578. [CrossRef]
9. Yang, D.; Li, M.; Geng, X.; Feng, Z. Sources and Specified Health Risks of 12 $PM_{2.5}$-Bound Metals in a Typical Air-Polluted City in Northern China during the 13th Five-Year Plan. *Toxics* **2024**, *12*, 581. [CrossRef]
10. Fowler, D.; Pyle, J.A.; Sutton, M.A.; Williams, M.L. Global Air Quality, past present and future: An introduction. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2020**, *378*, 20190323. [CrossRef]
11. Nansai, K.; Tohno, S.; Chatani, S.; Kanemoto, K.; Kagawa, S.; Kondo, Y.; Takayanagi, W.; Lenzen, M. Consumption in the G20 nations causes particulate air pollution resulting in two million premature deaths annually. *Nat. Commun.* **2021**, *12*, 6286. [CrossRef] [PubMed]
12. Tao, H.; Jawad, A.H.; Shather, A.H.; Al-Khafaji, Z.; Rashid, T.A.; Ali, M.; Al-Ansari, N.; Marhoon, H.A.; Shahid, S.; Yaseen, Z.M. Machine learning algorithms for high-resolution prediction of spatiotemporal distribution of air pollution from meteorological and soil parameters. *Environ. Int.* **2023**, *175*, 107931. [CrossRef] [PubMed]

13. Song, Y.; Zhang, C.; Jin, X.; Zhao, X.; Huang, W.; Sun, X.; Yang, Z.; Wang, S. Spatial prediction of $PM_{2.5}$ concentration using hyper-parameter optimization XGBoost model in China. *Environ. Technol. Innov.* **2023**, *32*, 103272. [CrossRef]
14. Kim, H.; Hong, S. Relationship between Land-Use Type and Daily Concentration and Variability of PM_{10} in Metropolitan Cities: Evidence from South Korea. *Land* **2022**, *11*, 23. [CrossRef]
15. Adnan, R.M.; Mo, W.; Kisi, O.; Heddam, S.; Al-Janabi, A.M.S.; Zounemat-Kermani, M. Harnessing Deep Learning and Snow Cover Data for Enhanced Runoff Prediction in Snow-Dominated Watersheds. *Atmosphere* **2024**, *15*, 1407. [CrossRef]
16. Guo, Q.; He, Z.; Wang, Z. Prediction of monthly average and extreme atmospheric temperatures in Zhengzhou based on artificial neural network and deep learning models. *Front. For. Glob. Chang.* **2023**, *6*, 1249300. [CrossRef]
17. Roy, B.; Malviya, L.; Kumar, R.; Mal, S.; Kumar, A.; Bhowmik, T.; Hu, J.W. Hybrid Deep Learning Approach for Stress Detection Using Decomposed EEG Signals. *Diagnostics* **2023**, *13*, 1936. [CrossRef]
18. He, D.; Dong, Z.; Zhu, B. An effective global biochar application strategy for reducing global cropland nitrogen emissions without compromising crop yield: Findings from a global meta-analysis and density functional theory calculation. *Environ. Technol. Innov.* **2024**, *33*, 103546. [CrossRef]
19. Poelzl, M.; Kern, R.; Kecorius, S.; Lovrić, M. Exploration of transfer learning techniques for the prediction of PM_{10}. *Sci. Rep.* **2025**, *15*, 2919. [CrossRef]
20. Guo, Q.; He, Z.; Wang, Z. Prediction of Hourly $PM_{2.5}$ and PM_{10} Concentrations in Chongqing City in China Based on Artificial Neural Network. *Aerosol Air Qual. Res.* **2023**, *23*, 220448. [CrossRef]
21. Pak, A.; Rad, A.K.; Nematollahi, M.J.; Mahmoudi, M. Application of the Lasso regularisation technique in mitigating overfitting in air quality prediction models. *Sci. Rep.* **2025**, *15*, 547. [CrossRef]
22. Rahimi, A. Short-term prediction of NO2 and NOxconcentrations using multilayer perceptron neural network: A case study of Tabriz, Iran. *Ecol. Process.* **2017**, *6*, 4. [CrossRef]
23. Guo, Q.; He, Z.; Li, S.; Li, X.; Meng, J.; Hou, Z.; Liu, J.; Chen, Y. Air Pollution Forecasting Using Artificial and Wavelet Neural Networks with Meteorological Conditions. *Aerosol Air Qual. Res.* **2020**, *20*, 1429–1439. [CrossRef]
24. Wang, P.; Zhang, G.; Chen, F.; He, Y. A hybrid-wavelet model applied for forecasting $PM_{2.5}$ concentrations in Taiyuan city, China. *Atmos. Pollut. Res.* **2019**, *10*, 1884–1894. [CrossRef]
25. Shakya, D.; Deshpande, V.; Goyal, M.K.; Agarwal, M. $PM_{2.5}$ air pollution prediction through deep learning using meteorological, vehicular, and emission data: A case study of New Delhi, India. *J. Clean. Prod.* **2023**, *427*, 139278. [CrossRef]
26. Singh, S.; Suthar, G. Machine learning and deep learning approaches for $PM_{2.5}$ prediction: A study on urban air quality in Jaipur, India. *Earth Sci. Inform.* **2024**, *18*, 97. [CrossRef]
27. Gonçalves, R.; Ribeiro, V.M. Convolutional attention with roll padding: Classifying $PM_{2.5}$ concentration levels in the city of Beijing. *Energy* **2024**, *289*, 130045. [CrossRef]
28. Wang, J.; Wang, P.; Tian, H.; Tansey, K.; Liu, J.; Quan, W. A deep learning framework combining CNN and GRU for improving wheat yield estimates using time series remotely sensed multi-variables. *Comput. Electron. Agric.* **2023**, *206*, 107705. [CrossRef]
29. Ahn, J.M.; Kim, J.; Kim, H.; Kim, K. Harmful Cyanobacterial Blooms forecasting based on improved CNN-Transformer and Temporal Fusion Transformer. *Environ. Technol. Innov.* **2023**, *32*, 103314. [CrossRef]
30. Dai, H.; Liu, Y.; Wang, J.; Ren, J.; Gao, Y.; Dong, Z.; Zhao, B. Large-scale spatiotemporal deep learning predicting urban residential indoor $PM_{2.5}$ concentration. *Environ. Int.* **2023**, *182*, 108343. [CrossRef]
31. Kim, Y.-b.; Park, S.-B.; Lee, S.; Park, Y.-K. Comparison of $PM_{2.5}$ prediction performance of the three deep learning models: A case study of Seoul, Daejeon, and Busan. *J. Ind. Eng. Chem.* **2023**, *120*, 159–169. [CrossRef]
32. Guo, Q.; He, Z.; Wang, Z. Monthly climate prediction using deep convolutional neural network and long short-term memory. *Sci. Rep.* **2024**, *14*, 17748. [CrossRef]
33. Guo, Q.; He, Z.; Wang, Z.; Qiao, S.; Zhu, J.; Chen, J. A Performance Comparison Study on Climate Prediction in Weifang City Using Different Deep Learning Models. *Water* **2024**, *16*, 2870. [CrossRef]
34. Sayeed, A.; Choi, Y.; Pouyaei, A.; Lops, Y.; Jung, J.; Salman, A.K. CNN-based model for the spatial imputation (CMSI version 1.0) of in-situ ozone and $PM_{2.5}$ measurements. *Atmos. Environ.* **2022**, *289*, 119348. [CrossRef]
35. Xia, S.; Zhang, R.; Zhang, L.; Wang, T.; Wang, W. Multi-dimensional distribution prediction of $PM_{2.5}$ concentration in urban residential areas based on CNN. *Build. Environ.* **2025**, *267*, 112167. [CrossRef]
36. Kim, D.; Ho, C.-H.; Park, I.; Kim, J.; Chang, L.-S.; Choi, M.-H. Untangling the contribution of input parameters to an artificial intelligence $PM_{2.5}$ forecast model using the layer-wise relevance propagation method. *Atmos. Environ.* **2022**, *276*, 119034. [CrossRef]
37. Huang, H.; Qian, C. Modeling $PM_{2.5}$ forecast using a self-weighted ensemble GRU network: Method optimization and evaluation. *Ecol. Indic.* **2023**, *156*, 111138. [CrossRef]
38. Zhang, Y.; Li, C.; Duan, H.; Yan, K.; Wang, J.; Wang, W. Deep learning based data-driven model for detecting time-delay water quality indicators of wastewater treatment plant influent. *Chem. Eng. J.* **2023**, *467*, 143483. [CrossRef]
39. Guo, Q.; He, Z.; Wang, Z. Assessing the effectiveness of long short-term memory and artificial neural network in predicting daily ozone concentrations in Liaocheng City. *Sci. Rep.* **2025**, *15*, 6798. [CrossRef]

40. Park, S.-Y.; Woo, S.-H.; Lim, C. Predicting PM_{10} and $PM_{2.5}$ concentration in container ports: A deep learning approach. *Transp. Res. Part D Transp. Environ.* **2023**, *115*, 103601. [CrossRef]
41. Kushwah, V.; Agrawal, P. Hybrid model for air quality prediction based on LSTM with random search and Bayesian optimization techniques. *Earth Sci. Inform.* **2024**, *18*, 32. [CrossRef]
42. Bedi, S.; Katiyar, A.; Krishnan, N.M.A.; Kota, S.H. Utilizing LSTM models to predict $PM_{2.5}$ levels during critical episodes in Delhi, the world's most polluted capital city. *Urban Clim.* **2024**, *53*, 101835. [CrossRef]
43. Lin, M.-D.; Liu, P.-Y.; Huang, C.-W.; Lin, Y.-H. The application of strategy based on LSTM for the short-term prediction of $PM_{2.5}$ in city. *Sci. Total Environ.* **2024**, *906*, 167892. [CrossRef]
44. Miao, L.; Tang, S.; Ren, Y.; Kwan, M.-P.; Zhang, K. Estimation of daily ground-level $PM_{2.5}$ concentrations over the Pearl River Delta using 1 km resolution MODIS AOD based on multi-feature BiLSTM. *Atmos. Environ.* **2022**, *290*, 119362. [CrossRef]
45. Teng, M.; Li, S.; Xing, J.; Song, G.; Yang, J.; Dong, J.; Zeng, X.; Qin, Y. 24-Hour prediction of $PM_{2.5}$ concentrations by combining empirical mode decomposition and bidirectional long short-term memory neural network. *Sci. Total Environ.* **2022**, *821*, 153276. [CrossRef]
46. Yin, S.; Liu, H.; Duan, Z. Hourly $PM_{2.5}$ concentration multi-step forecasting method based on extreme learning machine, boosting algorithm and error correction model. *Digit. Signal Process.* **2021**, *118*, 103221. [CrossRef]
47. Kumar, S.; Kumar, V. Multi-view Stacked CNN-BiLSTM (MvS CNN-BiLSTM) for urban $PM_{2.5}$ concentration prediction of India's polluted cities. *J. Clean. Prod.* **2024**, *444*, 141259. [CrossRef]
48. Zhang, J.; Li, S. Air quality index forecast in Beijing based on CNN-LSTM multi-model. *Chemosphere* **2022**, *308*, 136180. [CrossRef]
49. Faraji, M.; Nadi, S.; Ghaffarpasand, O.; Homayoni, S.; Downey, K. An integrated 3D CNN-GRU deep learning method for short-term prediction of $PM_{2.5}$ concentration in urban environment. *Sci. Total Environ.* **2022**, *834*, 155324. [CrossRef]
50. Yan, R.; Liao, J.; Yang, J.; Sun, W.; Nong, M.; Li, F. Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering. *Expert Syst. Appl.* **2021**, *169*, 114513. [CrossRef]
51. Kow, P.-Y.; Wang, Y.-S.; Zhou, Y.; Kao, I.F.; Issermann, M.; Chang, L.-C.; Chang, F.-J. Seamless integration of convolutional and back-propagation neural networks for regional multi-step-ahead $PM_{2.5}$ forecasting. *J. Clean. Prod.* **2020**, *261*, 121285. [CrossRef]
52. Cho, E.; Yoon, H.; Cho, Y. Evaluation of the impact of intensive $PM_{2.5}$ reduction policy in Seoul, South Korea using machine learning. *Urban Clim.* **2024**, *53*, 101778. [CrossRef]
53. He, Z.; Guo, Q. Comparative Analysis of Multiple Deep Learning Models for Forecasting Monthly Ambient $PM_{2.5}$ Concentrations: A Case Study in Dezhou City, China. *Atmosphere* **2024**, *15*, 1432. [CrossRef]
54. Zhu, M.; Xie, J. Investigation of nearby monitoring station for hourly $PM_{2.5}$ forecasting using parallel multi-input 1D-CNN-biLSTM. *Expert Syst. Appl.* **2023**, *211*, 118707. [CrossRef]
55. Ahmed, A.A.M.; Jui, S.J.J.; Sharma, E.; Ahmed, M.H.; Raj, N.; Bose, A. An advanced deep learning predictive model for air quality index forecasting with remote satellite-derived hydro-climatological variables. *Sci. Total Environ.* **2024**, *906*, 167234. [CrossRef]
56. Samal, K.K.R.; Babu, K.S.; Das, S.K. Multi-directional temporal convolutional artificial neural network for $PM_{2.5}$ forecasting with missing values: A deep learning approach. *Urban Clim.* **2021**, *36*, 100800. [CrossRef]
57. Wu, S.; Li, H. Prediction of $PM_{2.5}$ concentration in urban agglomeration of China by hybrid network model. *J. Clean. Prod.* **2022**, *374*, 133968. [CrossRef]
58. Shen, J.; Liu, Q.; Feng, X. Hourly $PM_{2.5}$ concentration prediction for dry bulk port clusters considering spatiotemporal correlation: A novel deep learning blending ensemble model. *J. Environ. Manag.* **2024**, *370*, 122703. [CrossRef]
59. Da, X.; Ye, D.; Shen, Y.; Cheng, P.; Yao, J.; Wang, D. A Novel Hybrid Method for Multi-Step Short-Term 70 m Wind Speed Prediction Based on Modal Reconstruction and STL-VMD-BiLSTM. *Atmosphere* **2024**, *15*, 1014. [CrossRef]
60. Yang, H.; Wang, W.; Li, G. Multi-factor $PM_{2.5}$ concentration optimization prediction model based on decomposition and integration. *Urban Clim.* **2024**, *55*, 101916. [CrossRef]
61. Zheng, Q.; Tian, X.; Yu, Z.; Jiang, N.; Elhanashi, A.; Saponara, S.; Yu, R. Application of wavelet-packet transform driven deep learning method in $PM_{2.5}$ concentration prediction: A case study of Qingdao, China. *Sustain. Cities Soc.* **2023**, *92*, 104486. [CrossRef]
62. Qiao, W.; Wang, Y.; Zhang, J.; Tian, W.; Tian, Y.; Yang, Q. An innovative coupled model in view of wavelet transform for predicting short-term PM_{10} concentration. *J. Environ. Manag.* **2021**, *289*, 112438. [CrossRef]
63. Ebrahimi-Khusfi, Z.; Taghizadeh-Mehrjardi, R.; Kazemi, M.; Nafarzadegan, A.R. Predicting the ground-level pollutants concentrations and identifying the influencing factors using machine learning, wavelet transformation, and remote sensing techniques. *Atmos. Pollut. Res.* **2021**, *12*, 101064. [CrossRef]
64. Guo, Q.; He, Z.; Wang, Z. Predicting of Daily $PM_{2.5}$ Concentration Employing Wavelet Artificial Neural Networks Based on Meteorological Elements in Shanghai, China. *Toxics* **2023**, *11*, 51. [CrossRef]
65. Liu, M.; Lu, Y.; Long, S.; Bai, J.; Lian, W. An attention-based CNN-BiLSTM hybrid neural network enhanced with features of discrete wavelet transformation for fetal acidosis classification. *Expert Syst. Appl.* **2021**, *186*, 115714. [CrossRef]
66. Hu, X.; Huang, G.; Zhou, F.; Qiu, Y.; Gou, X.; Chen, C. Pressure response using wavelet analysis in the process of hydraulic fracturing: Numerical simulation and field case. *J. Pet. Sci. Eng.* **2022**, *217*, 110837. [CrossRef]

67. Jida, S.N.; Hetet, J.-F.; Chesse, P.; Guadie, A. Roadside vehicle particulate matter concentration estimation using artificial neural network model in Addis Ababa, Ethiopia. *J. Environ. Sci.* **2021**, *101*, 428–439. [CrossRef]
68. Chandrasekar, A.; Zhang, S.; Mhaskar, P. A Hybrid Hubspace-RNN based approach for Modelling of Non-Linear Batch Processes. *Chem. Eng. Sci.* **2023**, *281*, 119118. [CrossRef]
69. Zhao, L.; Li, Z.; Qu, L.; Zhang, J.; Teng, B. A hybrid VMD-LSTM/GRU model to predict non-stationary and irregular waves on the east coast of China. *Ocean Eng.* **2023**, *276*, 114136. [CrossRef]
70. Ying, H.; Deng, C.; Xu, Z.; Huang, H.; Deng, W.; Yang, Q. Short-term prediction of wind power based on phase space reconstruction and BiLSTM. *Energy Rep.* **2023**, *9*, 474–482. [CrossRef]
71. Rezaei, M.; Mohammadifar, A.; Gholami, H.; Mina, M.; Riksen, M.J.P.M.; Ritsema, C. Mapping of the wind erodible fraction of soil by bidirectional gated recurrent unit (BiGRU) and bidirectional recurrent neural network (BiRNN) deep learning models. *CATENA* **2023**, *223*, 106953. [CrossRef]
72. Pan, S.; Yang, B.; Wang, S.; Guo, Z.; Wang, L.; Liu, J.; Wu, S. Oil well production prediction based on CNN-LSTM model with self-attention mechanism. *Energy* **2023**, *284*, 128701. [CrossRef]
73. Deng, C.; Qin, C.; Li, Z.; Li, K. Spatiotemporal variations of $PM_{2.5}$ pollution and its dynamic relationships with meteorological conditions in Beijing-Tianjin-Hebei region. *Chemosphere* **2022**, *301*, 134640. [CrossRef]
74. Guo, Q.; Wang, Z.; He, Z.; Li, X.; Meng, J.; Hou, Z.; Yang, J. Changes in Air Quality from the COVID to the Post-COVID Era in the Beijing-Tianjin-Tangshan Region in China. *Aerosol Air Qual. Res.* **2021**, *21*, 210270. [CrossRef]
75. Samal, K.K.R.; Panda, A.K.; Babu, K.S.; Das, S.K. An improved pollution forecasting model with meteorological impact using multiple imputation and fine-tuning approach. *Sustain. Cities Soc.* **2021**, *70*, 102923. [CrossRef]
76. Guo, Q.; He, Z.; Wang, Z. The Characteristics of Air Quality Changes in Hohhot City in China and their Relationship with Meteorological and Socio-economic Factors. *Aerosol Air Qual. Res.* **2024**, *24*, 230274. [CrossRef]
77. Sang, Y.F. A review on the applications of wavelet transform in hydrology time series analysis. *Atmos. Res.* **2013**, *122*, 8–15.
78. Moreno-Barón, L.; Cartas, R.; Merkoçi, A.; Alegret, S.; del Valle, M.; Leija, L.; Hernandez, P.R.; Muñoz, R. Application of the wavelet transform coupled with artificial neural networks for quantification purposes in a voltammetric electronic tongue. *Sens. Actuators B Chem.* **2006**, *113*, 487–499. [CrossRef]
79. Hou, X.; Ge, F.; Chen, D.; Shen, L.; Zou, F. Temporal distribution-based prediction strategy for dynamic multi-objective optimization assisted by GRU neural network. *Inf. Sci.* **2023**, *649*, 119627. [CrossRef]
80. Gao, Z.; Do, K.; Li, Z.; Jiang, X.; Maji, K.J.; Ivey, C.E.; Russell, A.G. Predicting $PM_{2.5}$ levels and exceedance days using machine learning methods. *Atmos. Environ.* **2024**, *323*, 120396. [CrossRef]
81. Liu, H.; Wang, Q.; Wei, P.; Zhang, Q.; Qu, Y.; Zhang, Y.; Tian, J.; Xu, H.; Zhang, N.; Shen, Z.; et al. The impacts of regional transport on anthropogenic source contributions of $PM_{2.5}$ in a basin city, China. *Sci. Total Environ.* **2024**, *917*, 170038. [CrossRef]
82. Wang, X.; Qin, J.; Zhong, S.; Yang, Y.; Lu, Q.; Yu, L. The differential impact of 2015–2020 El Niño and El Niño Modoki on warm- and cold-season $PM_{2.5}$ concentration and distribution across China. *Atmos. Environ.* **2023**, *305*, 119816. [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.