Prediction of the Short-Term PM2.5 Concentration Based on Informer

Jijing Cai[1,2], Chen Wang[3], Le Yu[2], Meilei Lv[4], and Kai Fang[2(✉)]

[1] Hangzhou Dianzi University, Hangzhou, China
[2] Zhejiang Agricultural and Forestry University, Hangzhou, China
kaifang@ieee.org
[3] Quzhou Meteorological Bureau, Quzhou, China
[4] Quzhou University, Quzhou, China

Abstract. PM2.5 poses a significant health risk to individuals. Hence, precise prediction of PM2.5 concentration is of utmost importance. This article proposed a novel hybrid deep learning model called WPI for short-term PM2.5 prediction. The model utilizes wiener filtering and principal component analysis to denoise and extract features from meteorological data consisting of 10 variables. By combining these techniques with the Informer model, the WPI model achieves accurate short-term prediction of PM2.5 concentration. To evaluate the performance of the WPI model, three types of metrics are used: root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2). In short-term prediction, the WPI model achieves the R^2 value at least 0.0418 higher than the comparison model, with RMSE at least 0.4647 lower and MAE at least 0.5507 lower.

Keywords: PM2.5 forecast · Informer · Wiener filter · Principal component analysis

1 Introduction

In recent years, with the accelerated urbanization process and the increase in industrialization levels, air pollution has become one of the global focal points of concern. Among them, PM2.5 (fine particulate matter) is an important indicator of air pollution. With a diameter of equal to or less than 2.5 μm, PM2.5 can easily be inhaled into the respiratory system, posing significant health risks. According to statistics, high concentrations of PM2.5 can lead to respiratory diseases, cardiovascular diseases, cancer, and other illnesses. The study by Zhang et al. [1] has discovered direct evidence of the impact of haze pollution on human health. They have also identified potential biomarkers, such as propanol and isoprene, for assessing exposure to haze air pollution. Xing et al. [2] found that PM2.5 can penetrate deep into the lungs, causing irritation, corrosion of alveolar walls, and significant damage to lung function, resulting in severe health implications. The research conducted by Xu et al. [3] revealed that short-term

V. C. M. Leung et al. (Eds.): Qshine 2023, LNICST 574, pp. 197–210, 2024.
https://doi.org/10.1007/978-3-031-65123-6_15

exposure to high concentrations of PM2.5 can lead to reduced lung function and induce inflammation and oxidative stress in healthy adults. Luo et al. [4] incorporated daily data on respiratory disease hospitalizations, daily concentrations of environmental air pollutants, and meteorological factors from October 1, 2014, to September 30, 2017, in Taiyuan City. They observed a clear association between PM2.5 concentrations and the number of hospitalizations due to respiratory diseases. The study by Hong et al. [5] indicated that climate change and more intense extreme weather events may increase the risk of severe pollution incidents in China. Therefore, accurate prediction of PM2.5 concentrations is of significant importance.

Currently, there are several methods for predicting PM2.5 concentrations, including numerical simulation methods, statistical modeling methods, and machine learning methods. Statistical modeling methods involve analyzing and processing historical observational data using statistical theories and methods to establish PM2.5 prediction models [6–8]. Numerical simulation methods, on the other hand, simulate atmospheric physical and chemical processes to predict the concentration changes of PM2.5 over a certain period of time [9]. This approach requires considering the interactions of various complex factors in the atmospheric environment, such as meteorological conditions, emission sources, and chemical reactions. Deep learning methods represent a novel machine learning approach that has gained prominence in recent years due to their ability to tackle high-dimensional, nonlinear, and intricate data. Deep learning methods can be used to build more accurate PM2.5 prediction models. Common deep learning models include convolutional neural network (CNN) [10–12] and recurrent neural network (RNN) [13,14]. S. Du et al. [15] proposed a new deep learning model for PM2.5 prediction, combining one-dimensional CNN(1D-CNN) with bidirectional long short-term memory networks (Bi-LSTM). This model has shown satisfactory accuracy. M.-C et al. [16] presented a composite neural network model for PM2.5 concentration prediction, which consists of a set of pre-trained and uninstantiated neural network models. Experimental results demonstrate that this composite neural network model performs well in PM2.5 prediction. Huang et al. [17] introduced an EMD-GRU (empirical mode decomposition-based gated recurrent unit) neural network model for PM2.5 concentration prediction, which outperforms a single GRU model. Qi et al. [18] proposed a hybrid model called GC-LSTM, which combines graph convolutional networks and long short-term memory networks, for PM2.5 concentration prediction using deep learning methods. The model effectively captures the spatiotemporal variations of PM2.5 by integrating both spatial and temporal information. The GC-LSTM model has shown promising results in PM2.5 prediction. Pak et al. [19] developed a deep learning-based hybrid model known as CNN-LSTM. This model combines convolutional neural networks (CNN) and long short-term memory networks (LSTM) to leverage the strengths of both architectures. The CNN component extracts inherent features from air quality and meteorological input data related to PM2.5, while the LSTM component captures the long-term dependencies in the input time series. Experimental results demonstrate that the CNN-LSTM

model outperforms traditional models such as multi-layer perceptron (MLP) and standalone LSTM in terms of stability and prediction performance. Jin et al. [20] proposed a hybrid deep learning model that combines empirical mode decomposition (EMD) and convolutional neural networks (CNN) for decomposing and classifying meteorological data. Experimental results demonstrate that this model achieves high accuracy in PM2.5 prediction. Li, X et al. [21] introduced a novel extension of the long short-term memory neural network model called LSTME, which has shown remarkable performance in PM2.5 prediction. Zhou, Y. et al. [22] proposed a deep multi-output neural network model known as DM-LSTM. By integrating DM-LSTM with three deep learning algorithms, this model effectively enhances the spatiotemporal stability and accuracy of regional multi-step ahead air quality forecasting. Jt A et al. [23] introduced a deep learning model called DBN-BP, which is based on a deep belief network with backpropagation. Compared to traditional backpropagation neural networks, DBN-BP demonstrates superior predictive performance.

The main contributions of this paper are as follows:

(1) We proposed a new hybrid deep learning framework (WPI), for efficient data feature extraction and dimensionality reduction, demonstrating exceptional performance in time series prediction.
(2) We combined wiener filtering and principal component analysis (PCA) for data denoising and feature extraction. Given the intricate nature of meteorological data, an analysis was conducted to explore the correlation between each component of meteorological factors and PM2.5. Pearson's coefficient was employed to quantify these relationships.
(3) The experimental results demonstrate that the WPI model exhibits excellent performance in PM2.5 concentration prediction. In the 6-hour prediction experiment, the WPI model achieves the R^2 value of 0.6738, surpassing other models by at least 0.0388. The MAE value is 5.5183, which is at least 0.5507 lower than other models. Additionally, the WPI model achieves an RMSE value of 8.0487, outperforming other models by at least 0.4647.

2 PM2.5 FORECAST

2.1 Data Acquisition and Processing

Data Acquisition. This research focuses on the prediction of PM2.5 concentration and requires a significant amount of experimental data to support network training and learning. The study utilizes real-time air quality data collected from the Quzhou Meteorological Station in China. The dataset includes 10 meteorological factors, such as PM2.5, PM10, SO2, CO, NO2, O3, atmospheric temperature, atmospheric humidity, rainfall, and wind speed. We put the pollutant data in samples x denoted as $X = \{x_{PM_{2.5}}, x_{PM_{10}}, x_{SO_2}, x_{CO}, x_{NO_2}, x_{O_3}\}$, and the meteorological data are represented by samples $X^* = \{x_{temp}, x_{hum}, x_{rain}, x_{wind}\}$. The experimental data include 30198 hourly data from 0:00 on January 1, 2016, to 19:00 on December

30, 2019. These data possess wide temporal coverage and high representativeness, making them valuable for research purposes.

Data Processing. Considering the disparate scales of various factors, it is essential to normalize the data. The objective of normalization is to mitigate the adverse impact of outlier data and expedite the convergence rate of the model. The normalization formula used is as follows:

$$y = \frac{x - \min\limits_{1 \le j \le n} \{x_j\}}{\max\limits_{1 \le j \le n} \{x_j\} - \min\limits_{1 \le j \le n} \{x_j\}} \tag{1}$$

where x is the original data, $\min\{x_j\}$ and $\max\{x_j\}$ are the minimum and maximum values of the data respectively.

In the experiment, the model leveraged various feature factors from the preceding 24 h to forecast the PM2.5 concentration values for the subsequent 1, 3, and 6 h. To ensure robustness, the dataset was partitioned into training, testing, and validation sets, following an 8:1:1 ratio. The model was then trained, fine-tuned, and validated using this carefully divided dataset.

2.2 PM2.5 Prediction Process

Wiener Filter. Wiener filtering [24–27] is a filtering method to extract the signal from the noise and the schematic diagram is shown in Fig. 1. Enter a random signal $x(n)$ denoted as:

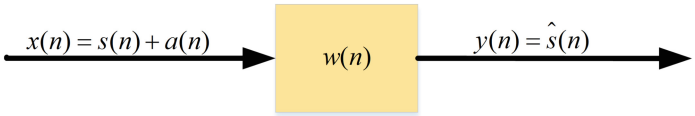

Fig. 1. Wiener filter structure diagram.

$$x(n) = s(n) + a(n) \tag{2}$$

where $s(n)$ denotes the signal, $a(n)$ denotes the noise, we want $y(n)$ to be close to after wiener filtering $w(n)$.

$$y(n) = \widehat{s}(n) \tag{3}$$

Principal Component Analysis. Before conducting experiments, it is essential to conduct an analysis of the correlation between PM2.5 and various meteorological factors and to retain the meteorological data. Figure 2 presents the correlation analysis graph of PM2.5 with each pollutant factor. When combined with Table 1, a strong correlation between PM2.5 with PM10, CO, NO2, SO2,

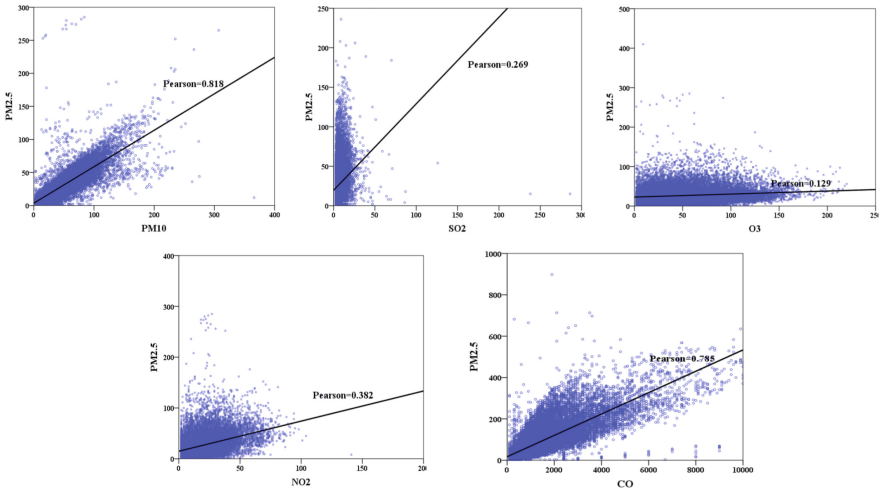

Fig. 2. PM2.5 and scatter plot of correlation with various pollutant factors.

and O3 becomes evident. As indicated in Table 2, PM2.5 exhibits a negative correlation with temperature, humidity, wind speed, and wind direction, and a weak positive correlation with precipitation. Consequently, it is imperative to downscale the meteorological data while preserving the pollutant data factors, all while using the meteorological data factors as auxiliary parameters. Principal component analysis (PCA) [28–31] is a linear dimensionality reduction technique that reduces the dimensionality of data by mapping high-dimensional data into a low-dimensional space and retaining the most important information while minimizing the loss of information. The core idea of PCA is to find the principal components in the data, which are the directions with the largest variance in the data. The direction contains the most important information in the data. PCA maps the data in these directions, which can achieve dimension reduction.

Table 1. Correlation coefficient of pollutant factors

Pollutant factors	PM10	SO2	CO	NO2	O3
Pearson coefficient	0.818	0.269	0.785	0.382	0.129

Table 2. Correlation coefficient of meteorological factors

Meteorological factors	Temperatures	Humidity	Wind Speed	Wind direction
Pearson coefficient	−0.300	−0.163	−0.140	0.04

Specifically, the PCA process can be divided into the following steps:

1) Data preprocessing: the data is centralized and the mean values are subtracted from $X_c = X - \overline{X}$, where X is the original m rows and n columns data matrix, $\overline{X}$ are the 1 row and n columns matrix composed of the mean values of each column of X, X_c is the centered m rows and n columns data matrix.
2) Calculate the covariance matrix: the covariance matrix is calculated for the centralized data $S = \frac{1}{m-1} {X_\mathbf{c}}^T X_\mathbf{c}$, where S is the covariance matrix of n rows and n columns.
3) Calculate the eigenvalues and eigenvectors: the covariance matrix is decomposed into eigenvalues to obtain the eigenvalues and the corresponding eigenvectors $Sv = \lambda v$, where λ is the eigenvalue and v is the eigenvector.
4) Select the principal components: sort the eigenvalues from largest to smallest, and select the eigenvectors corresponding to the top k eigenvalues as the principal components $P = [v_1, v_2, ...v_k]$, where $v_1, v_2, ...v_k$ is the eigenvector corresponds to the top eigenvalues.
5) Mapping data: The original data is mapped to the selected principal components to obtain the reduced dimensional data $Y = X_c P$, which Y is the reduced dimensional data matrix of m rows and k columns.

Informer. Informer is a supervised learning model based on the attention mechanism [32], which mainly consists of two parts, encoder and decoder. The informer model is shown in Fig. 3. In the Informer framework, the encoder and decoder adopt a structure similar to Seq2Seq to predict values for future time series. The encoder of the Informer framework uses a multi-channel convolutional neural network (CNN) to extract features from time series data, CNN is

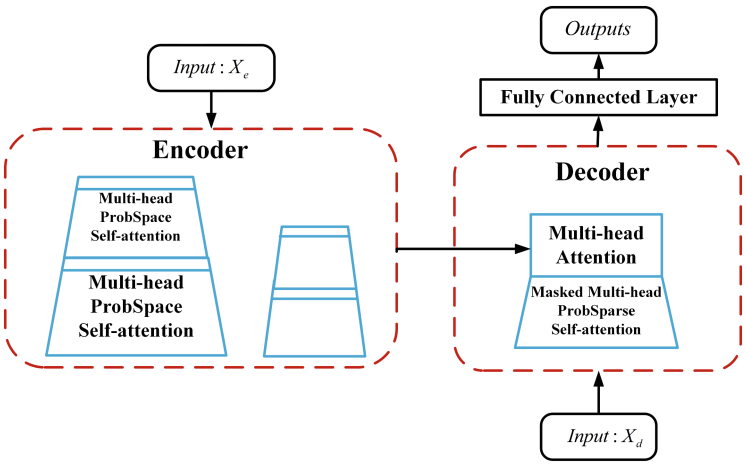

Fig. 3. Architecture of Informer model.

a common deep learning model with the advantages of efficiency, accuracy, and interpretability. The encoder uses a self-attentive mechanism similar to Transformer [33] with the input sequence $X = [x_1, x_2, ...x_n]$, and the self-attentive mechanism is calculated as:

$$\text{Self} - \text{Attention}(X) = \text{softmax}(\frac{XW_q(XW_k)^T}{\sqrt{d_k}})XW_v \tag{4}$$

where W_q, W_k and W_v are learnable weight matrices, d_k is the vector dimension of one equal fraction of the attention head count. The decoder of the Informer framework uses an auto-regressive model to predict the values of future time series. It is a multi-layer recurrent neural network structure. The output of the encoder Encoder is given by:

$$\text{En}(X) = \text{LaN}(X + \text{SelfAttention}(X) + \text{FeedForward}(X)) \tag{5}$$

where Feed-Forward is a feed-forward neural network and LayerNorm is a normalization method. Specifically, the encoder converts the input sequence into a series of hidden states, which are fed into the decoder, and the output of the decoder is given by:

$$\text{De}(X, H) = \text{LaN}(X + \text{SelfAttention}(X) + \text{CrossAttention}(H, X)) \tag{6}$$

where H is the output of the encoder and CrossAttention is a cross-attention mechanism to capture the information from the encoder in the decoder.

The Informer framework employs a multi-head self-attention mechanism to construct long-term dependencies in time-series data. For each multi-head attention, the input sequence $X \in R^{L\times d}$ is divided into h sub-sequences, each sub-sequence has the same dimension d/h, denoted as $X^k \in R^{L\times d/h}$, where $k \in [1, h]$.

Then, multiply each X^k by the projection matrix of queries, keys, and values to obtain their three representations:

$$Q^k = X^k W_Q \in R^{L\times d_{model}}, K^k = X^k W_k \in R^{L\times d_{model}}, V^k = X^k W_V \in R^{L\times d_{model}}$$

where $W_Q, W_K, W_V \in R^{d/h\times d_{\text{model}}}$ is the projection matrix. Calculate the attention weight of the dot product A^k between Q^k and K^k :

$$A^k = soft\max\left(\frac{Q^k\left(K^k\right)^T}{\sqrt{d_{\text{mod el}}}}\right) \in {}^{L\times L} \tag{7}$$

where $\sqrt{d_{\text{model}}}$ is a scaling factor that helps to make the value of the dot product smoother and avoid gradient disappearance or explosion.

Finally, multiply A^k by the corresponding value matrix V^k and concatenate the results of all multi-heads to obtain the output of the multi-head attention:

$$MultiHead\,(X) = concat\,(head_1, ..., head_h)\,W_O \in {}^{L\times d_{\text{mod el}}} \tag{8}$$

where $head_K = A^k V^k$ represents the output of the $k-th$ header, $W_O \in {}_{hd\text{model}\times d_{\text{model}}}$ is the output projection matrix, and *concat* represents multiple tensors connected by columns.

The Informer framework can efficiently handle missing values in time-series data. It uses a mask-based mechanism to handle missing values. In the self-attentive computation, the Informer framework applies the mask to the softmax function to ignore the missing values before computing the features. This gives the Informer framework high accuracy and robustness in handling missing values.

The Informer framework uses a new prediction and training strategy called random positional encoding (RPE). Random positional encoding enhances the robustness of the model, enabling it to better handle noise and outliers in time-series data.

2.3 WPI Prediction Model

The overall architecture of the PM2.5 concentration prediction model proposed in this paper is shown in Fig. 4. The hybrid model consists of 3 parts: Wiener filtering is used for noise reduction of the data; Principal component analysis is used for feature extraction and dimensionality reduction of the data; Informer model is used for training and prediction.

3 Experimental Analysis

3.1 Evaluation Indicators

To quantitatively evaluate the accuracy of the experimental prediction results, we adopt root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2) as evaluation metrics. Smaller values of RMSE and MAE indicate higher prediction accuracy. A value of R^2 closer to 1 indicates higher prediction accuracy, while a value closer to 0 indicates lower prediction accuracy. The formulas for these metrics are as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(y_t - y_p)^2} \tag{9}$$

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|y_t - y_p| \tag{10}$$

$$R^2 = 1 - \frac{\sum_{t=1}^{N}(y_t - y_p)^2}{\sum_{t=1}^{N}(y_t - \bar{y}_t)^2} \tag{11}$$

where y_t is the actual value, y_p is the predicted value, $\overline{y_t}$ is the average value, and $t = 1, 2, 3, ..., N$.

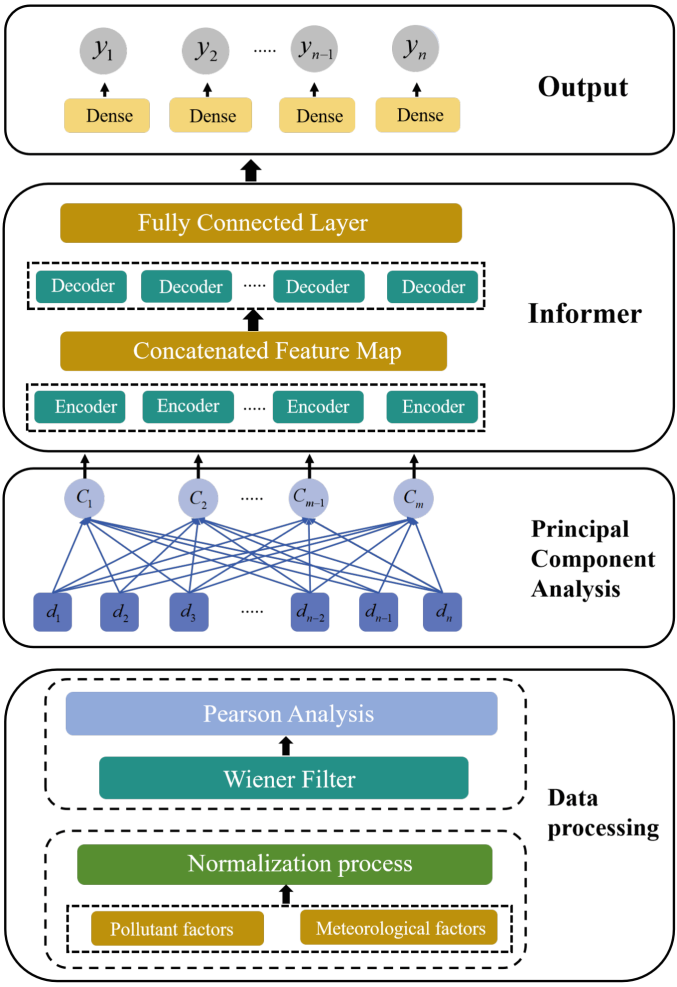

Fig. 4. Architecture of the proposed WPI model.

3.2 Experimental Environment and Hyperparameter Design

To explore the superiority of the WPI model. The experiment aims to compare the prediction performance and accuracy of short-term PM2.5 concentration sequences. To ensure fairness in the comparison of experimental results, the experiment was conducted on the same computer. The dataset was divided into training, testing, and validation sets in an 8:1:1 ratio, and the hyperparameter settings were consistent for the same model. The detailed specifications of the computer are provided in Table 3, and the hyperparameter settings for the model are described in Table 4.

Table 3. Computer configuration

Name	Settings
Processor	Intel(R) Core(TM) i5-10400U CPU @2.90GHz
Hard drive	SSD 1000 GB
Graphics card	NVIDIA GeForce RTX 2060 SUPER
Language	Python 3.9

Table 4. Model hyperparameter setting

Model	Settings
WPI	enc_in=6,dec_in=6,c_out=1,d_model=512,n_heads=8, batch_size=16,e_layers=2,d_layers=1,d_ff=2048, factor=6,padding=0,distil=true,attn=prob,embed=timeF, mix=true,Dropout=0.05,epochs=20,batch_size=128, optimizer=Adam,lr=0.0005,activation=relu,loss=mse;
PCA-BiLSTM PCA-LSTM	n_components=0.9, timesteps=24, Output steps=1, Hidden layers=2, Hidden layer cells=128, Hidden layers=16, Dropout=0.2, optimizer=Adam, epochs= 20, lr=0.0005, activation=relu, loss =mse;
BiLSTM LSTM RNN	timesteps=24, Output steps=1, Hidden layers=2, Hidden layer cells=128, batch_size=16, Dropout=0.2, optimizer=Adam, epochs=20, lr=0.0005, activation=relu, loss =mse;

3.3 Experiment: Comparison Experiment of Prediction Accuracy for Short-Term of PM2.5 Concentration

In this paper, to demonstrate the capability and efficiency of our proposed WPI model, we collected seven mainstream relevant prediction models, including LSTM, BiLSTM, RNN, PCA-LSTM, PCA-BiLSTM, CNN-BILSTM, and INFORMER. The parameters of each model were set according to Table 4 in the previous section, and they were applied to the same task scenarios to predict PM2.5 concentration in the next 1, 3, and 6 h for the experiments. Figure 5 shows the visualization of WPI, LSTM, BiLSTM, RNN, PCA-LSTM, PCA-BiLSTM, INFORMER, and CNN-BILSTM models in predicting the results for the next 1 h, and Fig. 6 shows the visualization of each model in predicting the results for the next 6 h.

The prediction performance of each model was evaluated using the evaluation metrics given above, as shown in Table 5. Analysis of Fig. 5, 6, and Table 5 revealed the following:

(1) Compared to other prediction models in the graph, the WPI model obtains prediction values that are closest to the original PM2.5 concentration samples. Moreover, according to the comparative results in Table 5, in the experiment of predicting the future 1 h, the model achieves R^2 of 0.9392, MAE of 2.1297, and RMSE of 3.4747, indicating the highest prediction accuracy of the model.

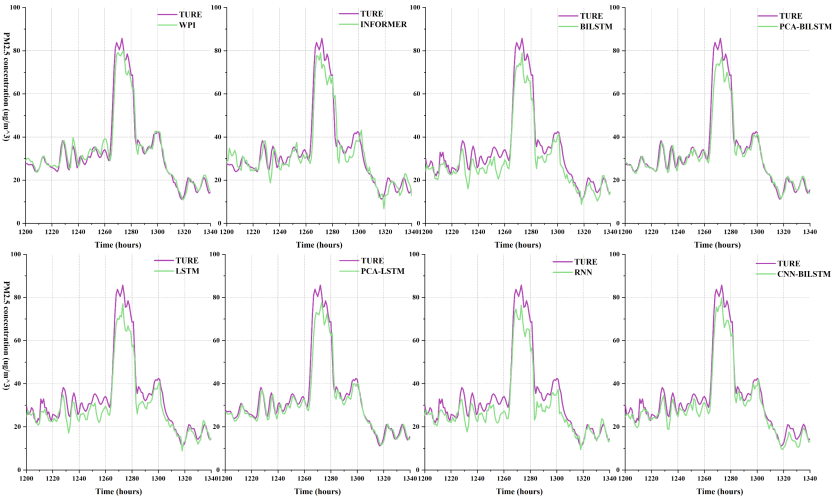

Fig. 5. Prediction of experimental results for the next 1 h.

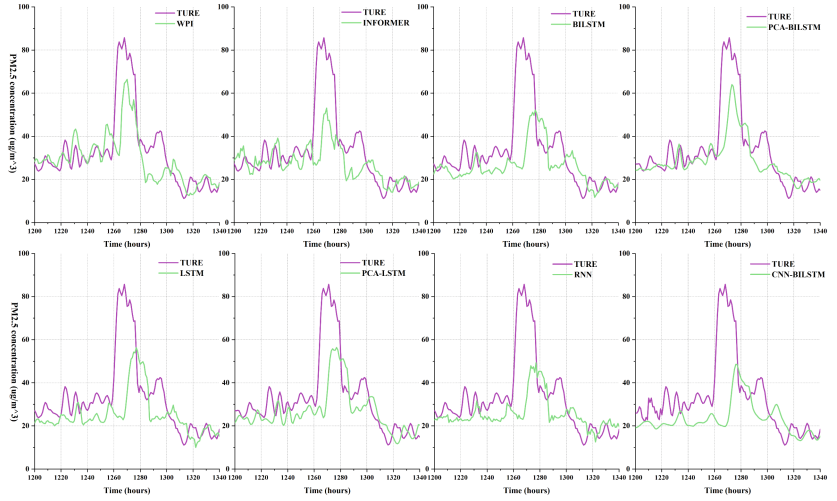

Fig. 6. Prediction of experimental results for the next 6 h.

(2) In the experiment of predicting the future 3 h, the WPI model achieves R^2 of 0.8037, MAE of 4.099, and RMSE of 6.2453, demonstrating the highest prediction accuracy of the model.

(3) In the experiment of predicting the future 6 h, the WPI model attains R^2 of 0.6738, MAE of 5.5183, and RMSE of 8.0487, indicating the highest prediction accuracy of the model. In summary, in the experiments of predicting the future 1, 3, and 6 h, the WPI model achieves the highest prediction accuracy. The reason for this is that the Informer model utilizes convolutional

Table 5. Evaluation indicators of each model

Model	Forecast 1 h			Forecast 3 h			Forecast 6 h		
	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2
WPI	3.4747	2.1297	0.9392	6.2453	4.0990	0.8037	8.0487	5.5183	0.6738
Pca-Bilstm	3.5836	2.1541	0.9353	6.7088	4.5540	0.7734	8.5134	6.0690	0.6350
Pca-Lstm	4.2924	2.5476	0.9072	7.7511	5.0195	0.6974	10.3788	7.0423	0.4576
Informer	4.7914	3.2962	0.8903	7.5787	5.0969	0.7255	9.6111	6.5951	0.5585
Cnn-Bilstm	6.0297	4.1231	0.8262	9.2660	6.9776	0.5896	11.5576	7.9812	0.3615
Bilstm	6.2393	4.2946	0.8139	8.2650	5.6038	0.6734	9.9726	6.8189	0.5246
Rnn	6.3601	4.4180	0.8066	9.0354	6.2373	0.6097	10.6754	7.4987	0.4552
Lstm	6.5781	4.5364	0.7931	8.0767	5.4978	0.6882	10.4582	7.0283	0.4771

neural networks to model local features in the sequence, reducing sensitivity to noise and outliers in the sequence. The Informer model can handle multi-task prediction problems by employing a multi-head self-attention mechanism to simultaneously learn features from multiple time series data. The Informer model can be expanded by increasing the number of layers and feature dimensions to adapt to different datasets and tasks. Additionally, in the WPI model, wiener filtering is applied to denoise the data, and principal component analysis is used for feature extraction and dimensionality reduction, leading to excellent performance in PM2.5 concentration prediction.

4 Conclusion

We have explored a novel PM2.5 concentration prediction model known as WPI. This model utilizes state-of-the-art deep learning techniques, specifically the Informer deep learning technology, in the field of time series data prediction. Additionally, it incorporates wiener filtering for data denoising and employs principal component analysis for feature extraction and dimensionality reduction. These enhancements significantly improve the model's predictive performance. Experimental results have demonstrated that the WPI model exhibits outstanding performance in short-term PM2.5 concentration prediction, outperforming other models in terms of data metrics.

References

1. Zhang, L., Li, X., Chen, H.: Haze air pollution health impacts of breath-borne VOCs. Environ. Sci. Technol. ES&T (12), 56 (2022)
2. Xing, Y.F., Xu, Y.H., Shi, M.H., Lian, Y.X.: The impact of PM2.5 on the human respiratory system. J. Thorac. Dis. **8**(1), E69–E74 (2016)
3. Xu, R., et al.: Acute effects of exposure to fine particulate matter and ozone on lung function, inflammation and oxidative stress in healthy adults. Ecotoxicol. Environ. Saf. **243**, 114013 (2022). https://doi.org/10.1016/j.ecoenv.2022.114013

4. Luo, L., et al.: Short-term effects of ambient air pollution on hospitalization for respiratory disease in Taiyuan, China: a time-series analysis. Int. J. Environ. Res. Public Health **15**(10), 2160 (2018)
5. Hong, C., Zhang, Q., Zhang, Y., Davis, S.J., Schellnhuber, H.J.: Impacts of climate change on future air quality and human health in china. Proc. Natl. Acad. Sci. **116**(35), 201812881 (2019)
6. Zca, B., et al.: Influence of meteorological conditions on PM 2.5 concentrations across china: a review of methodology and mechanism. Environ. Int. **139**, 105558 (2020)
7. Wei, J., Huang, W., Li, Z., Xue, W., Cribb, M.: Estimating 1-KM-resolution PM2.5 concentrations across china using the space-time random forest approach. Remote Sens. Environ. **231**, 111221 (2019)
8. Fang, K., Wang, T., Zhou, X., Ren, Y., Guo, H., Li, J.: A topsis-based relocalization algorithm in wireless sensor networks. IEEE Trans. Industr. Inf. **18**(2), 1322–1332 (2022). https://doi.org/10.1109/TII.2021.3076770
9. Van Donkelaar, A., et al.: Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors. Environ. Sci. Technol. **50**(7), 3762–3772 (2016)
10. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. IEEE Trans. Pattern Anal. Mach. Intell. **42**(2), 318–327 (2020). https://doi.org/10.1109/TPAMI.2018.2858826
11. Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Yu, P.S.: A comprehensive survey on graph neural networks. IEEE Trans. Neural Netw. Learn. Syst. **32**(1), 4–24 (2021). https://doi.org/10.1109/TNNLS.2020.2978386
12. Zhao, Z.Q., Zheng, P., Xu, S.T., Wu, X.: Object detection with deep learning: a review. IEEE Trans. Neural Netw. Learn. Syst. **30**, 3212–3232 (2019)
13. Fei, J., Liu, L.: Real-time nonlinear model predictive control of active power filter using self-feedback recurrent fuzzy neural network estimator. IEEE Trans. Industr. Electron. **69**(8), 8366–8376 (2022). https://doi.org/10.1109/TIE.2021.3106007
14. Fang, K., Wang, T., Yuan, X., Miao, C., Pan, Y., Li, J.: Detection of weak electromagnetic interference attacks based on fingerprint in IIOT systems. Futur. Gener. Comput. Syst. **126**, 295–304 (2022)
15. Du, S., Li, T., Yang, Y., Horng, S.J.: Deep air quality forecasting using hybrid deep learning framework. IEEE Trans. Knowl. Data Eng. **33**(6), 2412–2424 (2021). https://doi.org/10.1109/TKDE.2019.2954510
16. Yang, M.C., Chen, M.C.: Composite neural network: theory and application to PM2.5 prediction. IEEE Trans. Knowl. Data Eng. **35**(2), 1311–1323 (2023). https://doi.org/10.1109/TKDE.2021.3099135
17. 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. **768**(3), 144516 (2021)
18. Qi, Y., Li, Q., Karimian, H., Liu, D.: A hybrid model for spatiotemporal forecasting of PM2.5 based on graph convolutional neural network and long short-term memory. Sci. Total Environ. **664**, 1–10 (2019)
19. Pak, U., et al.: Deep learning-based PM2.5 prediction considering the spatiotemporal correlations: a case study of Beijing, China. Sci. Total Environ. **699**, 133561 (2020)
20. Jin, X.B., et al.: Deep hybrid model based on EMD with classification by frequency characteristics for long-term air quality prediction. Mathematics **8**(2), 214 (2020)

21. Li, X., et al.: Long short-term memory neural network for air pollutant concentration predictions: method development and evaluation. Environ. Pollut. **231**, 997–1004 (2017)
22. 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. **209**, 134–145 (2018)
23. Tian, J., Liu, Y., Zheng, W., Yin, L.: Smog prediction based on the deep belief-BP neural network model (DBN-BP). Urban Clim. **41**, 101078 (2022)
24. Yağan, A.C., Özgen, M.T.: Spectral graph based vertex-frequency wiener filtering for image and graph signal denoising. IEEE Trans. Signal Inf. Process. Netw. **6**, 226–240 (2020). https://doi.org/10.1109/TSIPN.2020.2976704
25. Chang, S.Y., Wu, H.C.: Tensor wiener filter. IEEE Trans. Signal Process. **70**, 410–422 (2022). https://doi.org/10.1109/TSP.2022.3140722
26. Li, N., Lei, Y., Yan, T., Li, N., Han, T.: A wiener-process-model-based method for remaining useful life prediction considering unit-to-unit variability. IEEE Trans. Industr. Electron. **66**(3), 2092–2101 (2019). https://doi.org/10.1109/TIE.2018.2838078
27. Wang, W., Chen, J., Wang, J., Chen, J., Liu, J., Gong, Z.: Trust-enhanced collaborative filtering for personalized point of interests recommendation. IEEE Trans. Industr. Inf. **16**(9), 6124–6132 (2020). https://doi.org/10.1109/TII.2019.2958696
28. Singh, V., Verma, N.K., Cui, Y.: Type-2 fuzzy PCA approach in extracting salient features for molecular cancer diagnostics and prognostics. IEEE Trans. Nanobiosci. **18**(3), 482–489 (2019). https://doi.org/10.1109/TNB.2019.2917814
29. Xia, Z., Chen, Y., Xu, C.: Multiview PCA: a methodology of feature extraction and dimension reduction for high-order data. IEEE Trans. Cybern. **52**(10), 11068–11080 (2022). https://doi.org/10.1109/TCYB.2021.3106485
30. Zhang, X., Jiang, X., Jiang, J., Zhang, Y., Liu, X., Cai, Z.: Spectral-spatial and superpixelwise PCA for unsupervised feature extraction of hyperspectral imagery. IEEE Trans. Geosci. Remote Sens. **60**, 1–10 (2022). https://doi.org/10.1109/TGRS.2021.3057701
31. Fu, H., Sun, G., Ren, J., Zhang, A., Jia, X.: Fusion of PCA and segmented-PCA domain multiscale 2-D-SSA for effective spectral-spatial feature extraction and data classification in hyperspectral imagery. IEEE Trans. Geosci. Remote Sens. **60**, 1–14 (2022). https://doi.org/10.1109/TGRS.2020.3034656
32. Zhou, H., Zhang, S., Peng, J., Zhang, S., Zhang, W.: Informer: beyond efficient transformer for long sequence time-series forecasting (2020)
33. Vaswani, A., et al.: Attention is all you need. arXiv (2017)