

PM_{2.5} prediction based on attention mechanism and Bi-LSTM

Xin Huang^a, Xianping Hong^a, Zuhan Liu^{*a}, Qiming Zhang^{#b}, Qin Huang^c, Zurun Liu^d

^aSchool of Information Engineering, Nanchang Institute of Technology, Nanchang 330099, Jiangxi, China; ^bJiangxi Provincial Environmental Monitoring Center Station, Nanchang 330029, Jiangxi, China; ^cNanfeng Ecological and Environmental Bureau of Fuzhou City, Fuzhou 344599, Jiangxi, China; ^dJiangxi Ohi Technology Co., Ltd., Nanchang 330096, Jiangxi, China

ABSTRACT

This study enhances the bidirectional long short-term memory (Bi-LSTM) model by incorporating an attention mechanism, which could provide the model with stronger data generalization capabilities. Moreover, it can predict a broader range of data and exhibits enhanced handling and adaptability to anomalies. Through the utilization of the attention mechanism, this research partitions the weights of the feature values, precisely dividing the input LSTM's feature values based on their weights. This enables the Bi-LSTM to more accurately capture relationships between different feature values in time series and dependencies on various features. Given the diverse air quality conditions in different regions, the introduced attention mechanism in Bi-LSTM manages the weights of different feature values. The Bi-LSTM, enhanced with attention mechanisms, excels at handling relationships in time series data, allowing it to predict PM_{2.5} values in more complex air quality environments. It demonstrates improved capabilities in handling anomalies. Even in air quality scenarios with various complex conditions, the model maintains satisfactory predictive quality.

Keywords: Attentional mechanism, PM_{2.5} prediction, Bi-LSTM, LSTM, Air quality

1. INTRODUCTION

With the rapid development of industrialization and urbanization, the issue of air pollution has become increasingly severe. Air pollution poses significant threats to both human health and the environment^{1,2}. To begin with, air pollution has cast a multitude of adverse effects on our well-being. Pollutants such as fine particulate matter (PM_{2.5}), ozone (O₃), and sulfur dioxide (SO₂) infiltrate the respiratory system, potentially leading to respiratory diseases, cardiovascular issues, and even cancer³. This risk is particularly pronounced among the elderly, children, and individuals with pre-existing respiratory conditions⁴. Moreover, air pollution wreaks havoc on the environment. Major contributors, such as vehicle emissions, industrial effluents, and combustion byproducts, have emerged as primary sources of air pollution. These pollutants not only directly harm vegetation and soil but also pose a threat to the delicate balance and biodiversity of ecosystems. Furthermore, air pollution exerts irreversible impacts on the atmospheric ozone layer and contributes to global climate change⁵. Hence, there is a growing societal awareness regarding air quality concerns, leading to concerted efforts to address the issue. Initiatives formulated by governments and various sectors include bolstering monitoring and controlling pollution sources. This will promote the use of clean energy, foster sustainable transportation, and advocate for public environmental consciousness. These endeavors hold the promise of mitigating the detrimental impacts of air pollution on human health and the environment, paving the way for a fresher, healthier future⁶.

China Daily cited Peking University public health expert Xiaochuan Pan in describing smoke from barbecues as a 'very common' source of PM_{2.5}, shorthand for the tiny airborne particulate matter smaller than 2.5 micrometers in diameter that health experts say is particularly damaging to human health. Prolonged exposure to high concentrations of PM_{2.5} not only threatens human health but also has environmental impacts. Elevated concentrations of PM_{2.5} can reduce visibility, leading to haze and adversely affecting traffic safety and aviation operations. Additionally, PM_{2.5} is associated with the formation of acid rain and poses potential pollution risks to water and soil, causing disruptions to environmental ecosystems⁷. PM_{2.5} is also a crucial indicator for assessing air quality. The concentration of PM_{2.5} reflects the content and quality of suspended particles in the air, indirectly indicating the level of pollutants in the air. Therefore, monitoring, assessing, and predicting the concentration of PM_{2.5} contribute to determining air quality conditions⁸.

*lzh512@nit.edu.cn; # 357397880@qq.com

With the development of machine learning technology, the application of LSTM network models in predicting $PM_{2.5}$ time series has gained increasing prominence. Its utility extends beyond mere time series prediction, demonstrating exceptional performance in tasks such as natural language processing and speech recognition. The LSTM's adeptness at capturing long-term dependencies within time series data, effectively handling and modeling intricate temporal patterns, has positioned it as a commonplace choice for dealing with time series data across diverse domains. For instance, Wu et al. (2023) in their innovative approach, amalgamated Bi-LSTM with trend feature extraction to forecast novel short-term household loads⁹. This hybrid predictive model seamlessly integrates wavelet threshold denoising (WTD), variational mode decomposition (VMD), and Bi-LSTM networks. Empirical evidence indicates that, particularly in short-term household load prediction, this model delivers more stable and precise forecasting results, significantly enhancing predictive accuracy. In a similar vein, Ying et al. (2023) addressed the chaotic characteristics of wind power time series by proposing a wind power short-term forecasting method based on phase space reconstruction and bidirectional long short-term memory neural networks (Re-Bi-LSTM)¹⁰. This method, by considering the unique characteristics of wind power data through data reconstruction and incorporating meteorological data, further elevates forecasting accuracy. Furthermore, Yang et al. (2023) employed a deep learning approach using TCN and SA-Bi-LSTM for reservoir logging identification, effectively improving the model's predictive performance¹¹. When it comes to the time series prediction of $PM_{2.5}$, bidirectional LSTM proves superior in deciphering temporal discrepancies and adeptly handling outliers.

In addition, $PM_{2.5}$ often coexists with pollutants such as NO and SO₂. Therefore, when predicting $PM_{2.5}$, considering the correlation with other environmental factors becomes essential. Introducing an attention mechanism becomes necessary in this context. The Attention Mechanism is a commonly used technique in machine learning and deep learning to address information extraction and weighting issues when dealing with sequential data. It simulates the human attention mechanism in information processing, allowing the model to adaptively focus on the importance of different positions or features in the input sequence, thus capturing crucial contextual information more effectively. By introducing the attention mechanism, the model can weigh different parts or features of the input based on their importance, enhancing the encoding or decoding of the input sequence. This enables the model to better handle long sequences and complex patterns, achieving significant success in tasks such as natural language processing, machine translation, image processing, and speech recognition. Kang et al. (2023) combined attention with LSTM to predict the advancing route of a tunnel boring machine¹². The improved framework demonstrated high feasibility and accuracy in predicting the posture and position of the tunnel boring machine. Chen et al. (2023) used an Attention-LSTM model to predict the flatness of tandem cold-rolled steel strips, showing high accuracy and reliability in cold-rolled steel strip flatness prediction¹³.

In $PM_{2.5}$ prediction, Bidirectional LSTM excels at capturing correlations before and after the time series. Meanwhile, the attention mechanism selectively assigns weights to various features in the captured values, enabling LSTM to achieve better predictive accuracy. Therefore, the proposed Attention Bi-LSTM in this paper can more accurately predict $PM_{2.5}$ values, and the model performs well in different air quality environments.

2. DATA

The dataset used in this study was obtained from the China Knowledge Center for Engineering Science and Technology (<https://www.ckcest.cn/entry/>), which covers air quality data from eleven cities in Jiangxi Province. The dataset recorded hourly air quality data from February 2, 2017, to December 22, 2018, with a total of 15,525 records.

Due to the large number of feature values used in this experiment, the dataset is preprocessed in order to avoid the impact of the differences in the magnitude of different feature values on the training results. The selected dataset is larger has fewer missing values, and does not affect the overall data distribution, thus the missing values are directly removed from the process. Subsequently, we use the normalization method to scale the data to between (-1, 1) to ensure that the trained data is within a magnitude to ensure the accuracy of the training.

For data input, we used a time-sliding window to input the data sequentially. Since the dataset of this study is hourly time series data, 24H is selected as a sliding window, and six items, $PM_{2.5}$, PM_{10} , SO₂, NO₂, CO, and O₃, are input as the eigenvalues, and $PM_{2.5}$ is used as the target value for testing the accuracy of the prediction results.

3. MODELS

3.1 Bi-LSTM

The Bi-LSTM is a type of Recurrent Neural Network (RNN) structure widely employed in the processing of time series

data¹⁴⁻¹⁶. Unlike traditional unidirectional LSTMs, the Bi-LSTM can leverage both past and future information, leading to enhanced modeling capabilities. In a traditional unidirectional LSTM, each time step of the hidden layer can only access information from the past, limiting its ability to capture long-term dependencies in the sequence. To address this limitation, the bidirectional LSTM introduces an additional inverse layer that considers both past and future information.

3.2 Attention mechanism

The Attention Mechanism is a technique employed to enhance the expressive power of neural network models, especially in tasks involving sequence-based data and long-term dependencies¹⁷. This mechanism enables the model to focus on the most relevant parts of the input by dynamically assigning different weights to different positions in the input sequence. A common implementation of the attention mechanism includes dot product attention or weighted average attention. Dot product attention calculates attention weights based on the similarity between inputs and the query, while weighted average attention performs a weighted sum based on the importance of the inputs. The advantage of the attention mechanism lies in its ability to provide a richer representation capability and a more precise attention mechanism. This, in turn, helps the model better handle long-term dependencies within input sequences and capture key information more effectively. Consequently, the performance and generalization ability of the model are improved.

3.3 ATT-Bi-LSTM

ATT-Bi-LSTM (Bidirectional Attention LSTM) combines Bidirectional LSTM and Attention mechanism. This is a special type of recurrent neural network that starts with two independent LSTM layers, one processing the input sequence from front to back in time steps and the other processing the input sequence from back to front in the opposite order. In this way, the model can extract feature information from both past and future contexts, thus better capturing long-term dependencies in the sequences. Subsequently, with the introduction of the attention mechanism, it aggregates the information in the input sequence by computing a vector of weights indicating the importance of different input positions to the current hidden state and then using these weights to weight the information in the input sequence. In this way, the model can dynamically and selectively focus on certain parts of the input sequence according to different contexts, improving the modeling accuracy and generalization ability. After feature selection and weight aggregation, the selected data is again fed to Bi-LSTM, which further extracts and captures features from the filtered data, thus further capturing the correlation between each feature value. Finally, a Dropout layer is used to randomly filter the data to prevent overfitting, and then a fully connected layer is used to output the data. ATT-Bi-LSTM combines the temporal modeling capability of Bi-LSTM with the feature value selection capability of the Attention Mechanism, which allows better processing of sequential data. ATT-Bi-LSTM is shown in Figure 1.

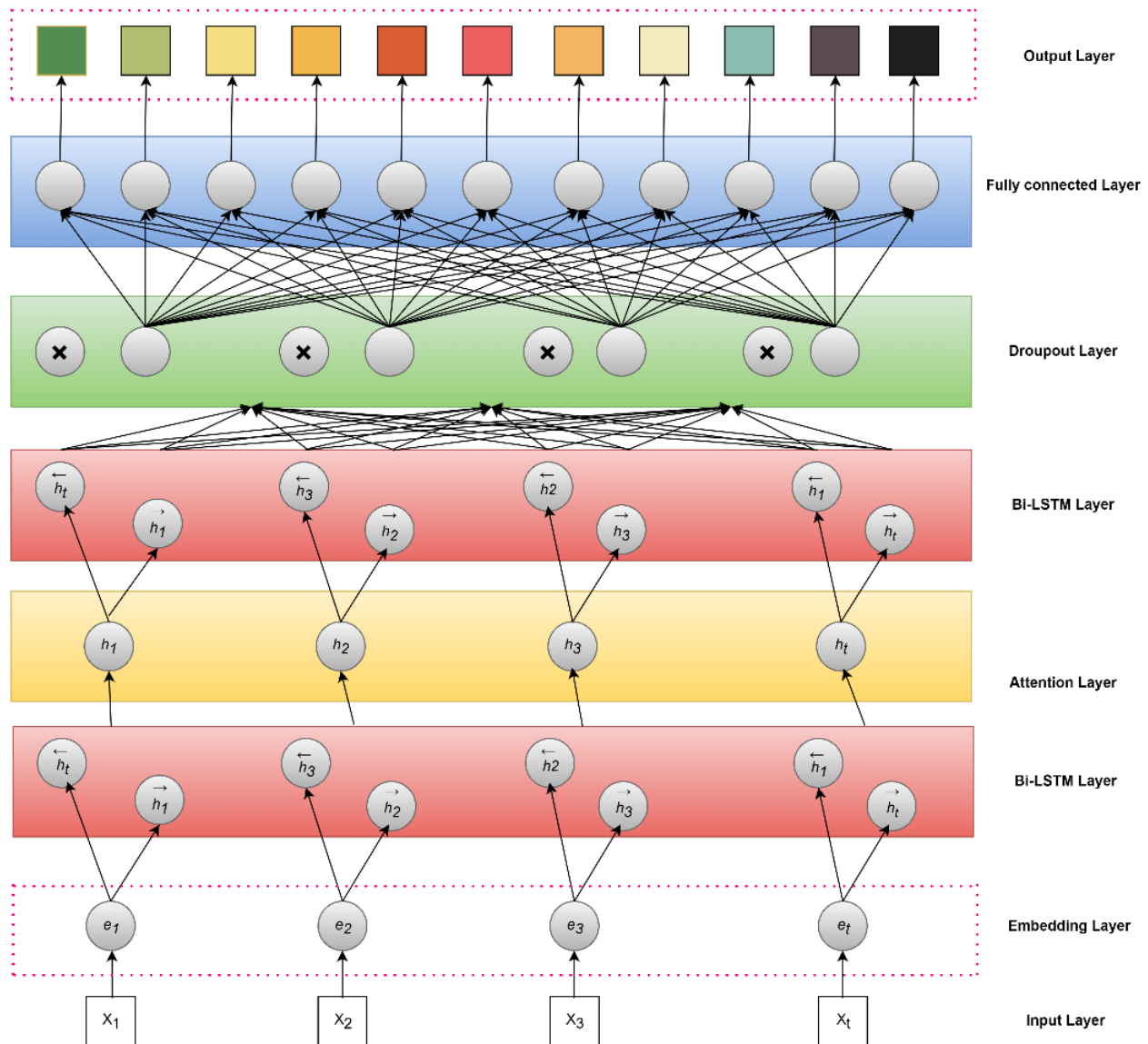


Figure 1. ATT-Bi-LSTM

4. SIMULATION EXPERIMENT

4.1 Experimental procedures

The dataset is initially processed, followed by the creation of a time-sliding window for data input, as described above. Moving to the modeling phase, we first construct a Bi-LSTM neural network model incorporating the attention mechanism, setting the appropriate parameters such as the number of neurons, batch size, and training generations. Different weights for input eigenvalues are selected through the attention mechanism's weight selection, feeding them into the bidirectional LSTM network. Subsequently, the final predicted values are obtained by selecting the output of the bidirectional LSTM neural network model. Finally, the predicted values are evaluated against the real values. The specific process and model shape parameters are outlined below, as shown in Figure 2 and Table 1.

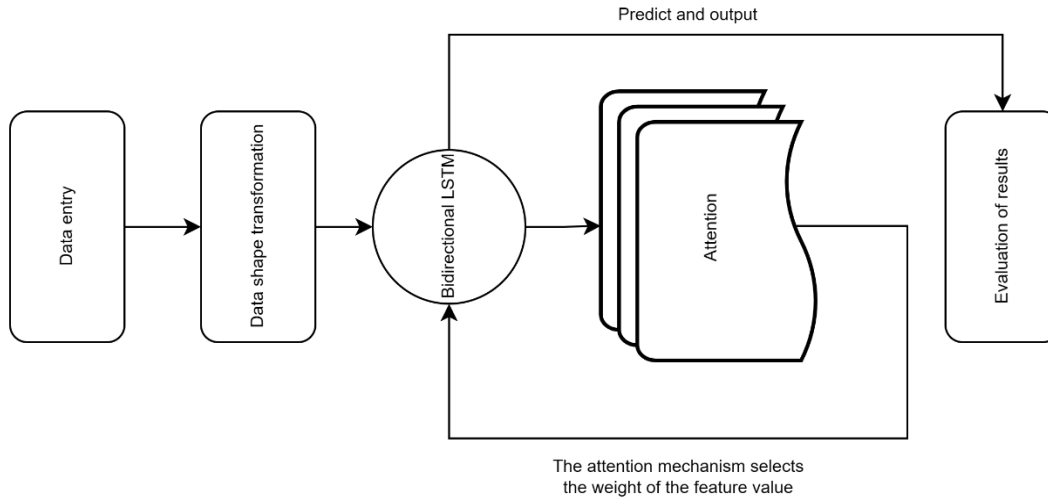


Figure 2. Flowchart of the model run

Table 1. Model shape

Layer (type)	Output shape	Param #	Connected to
Input_1 (inputlayer)	[(None, 24, 5)]	0	
Bidirectional (Bidirectional)	(None, 24, 256)	137216	Input_1[0][0]
Leaky_re_lu (leakyrelu)	(None, 24, 256)	0	Bidirectional[0][0]
Dropout (Dropout)	(None, 24, 256)	0	Leaky_re_lu[0][0]
Time_distributed (timedistributed)	(None, 24, 1)	257	Dropout[0][0]
Activation (Activation)	(None, 24, 1)	0	Time_distributed[0][0]
Dot (Dot)	(None, 1, 256)	0	Activation[0][0] Dropout[0][0]
Bidirectional_2(Bidirectional)	(None, 512)	1050624	Dot[0][0]
Leaky_re_lu_1(leakyrelu)	(None, 512)	0	Bidirectional_2[0][0]
Dropout_1 (Dropout)	(None, 512)	0	Leaky_re_lu_1[0][0]
Dense_1 (Dense)	(None, 1)	513	Dropout_1[0][0]

4.2 Experimental results

In this experiment, the proposed ATT-Bi-LSTM model is employed to predict the hourly PM_{2.5} concentration using the same time series data from eleven districts: Fuzhou, Ganzhou, Ji'an, Jingdezhen, Jiujiang, Nanchang, Pingxiang, Shangrao, Xinyu, Yichun, and Yingtian in Jiangxi Province. We compare the predictive performance with ATT-LSTM, BI-LSTM, and BI-ATT-LSTM models. Model performance is evaluated using metrics such as MSE, RMSE, MAE, R², and loss to assess its effectiveness. Ultimately, ATT-Bi-LSTM demonstrates superior performance. The specific results are depicted in Figure 3.

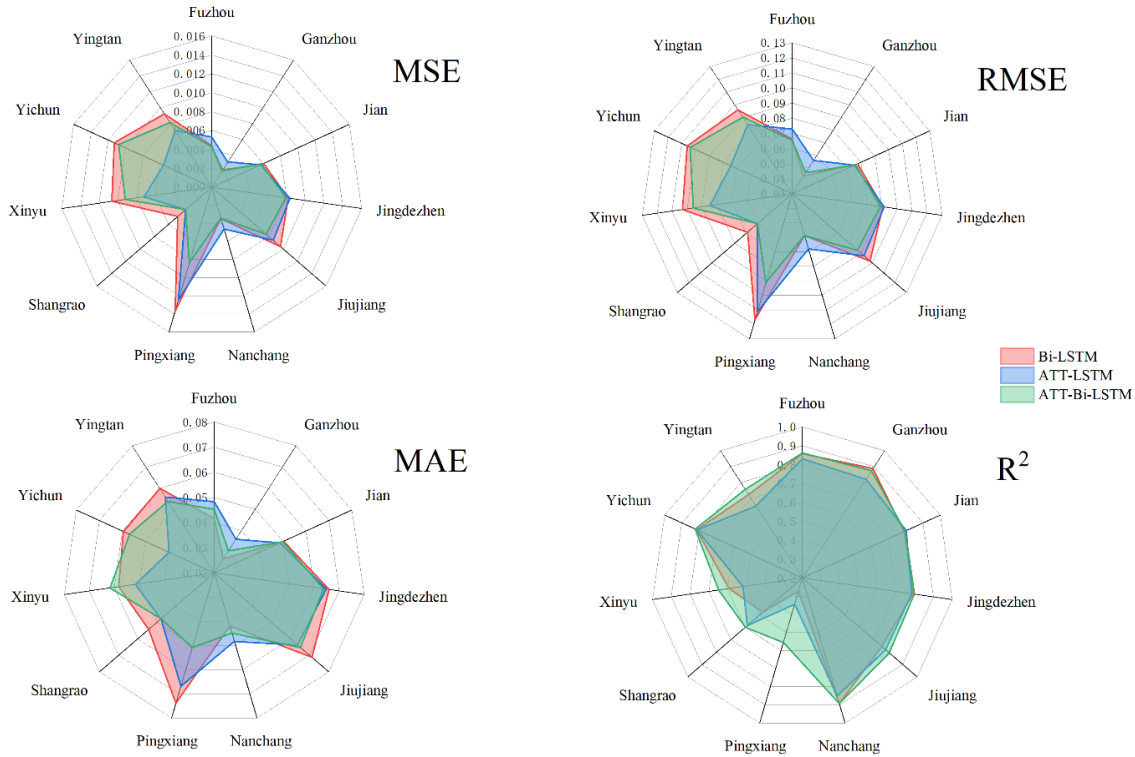


Figure 3. Performance evaluation radar

Initially, we utilize the Mean Squared Error (MSE) to assess the data for each city. The radar chart illustrates that in the case of Pingxiang, where air quality data exhibits instability, the prediction results of ATT-LSTM and Bi-LSTM models display significant fluctuations. However, the ATT-Bi-LSTM model consistently maintains a small error. To gain a more nuanced understanding of the error magnitude, we employ the Root Mean Square Error (RMSE) for further testing. In the same context, the radar chart highlights that the ATT-Bi-LSTM model continues to exhibit a small error. While MSE and RMSE provide valuable insights, Mean Absolute Error (MAE) is often used as a key metric to assess the performance of regression models. In the case of Pingxiang, we observe that the ATT-Bi-LSTM model yields a smaller error compared to both the Bi-LSTM and ATT-LSTM models. However, acknowledging that MAE represents the average of errors and may not fully reflect prediction accuracy, we also incorporate the coefficient of determination R² (R-squared) to gauge the model fit. The radar chart depicting R² performance evaluation demonstrates that the ATT-Bi-LSTM model inherits the strengths of both BI-LSTM and ATT-LSTM, exhibiting a higher degree of fit.

The prediction performance of the ATT-Bi-LSTM model is shown in Figure 4.

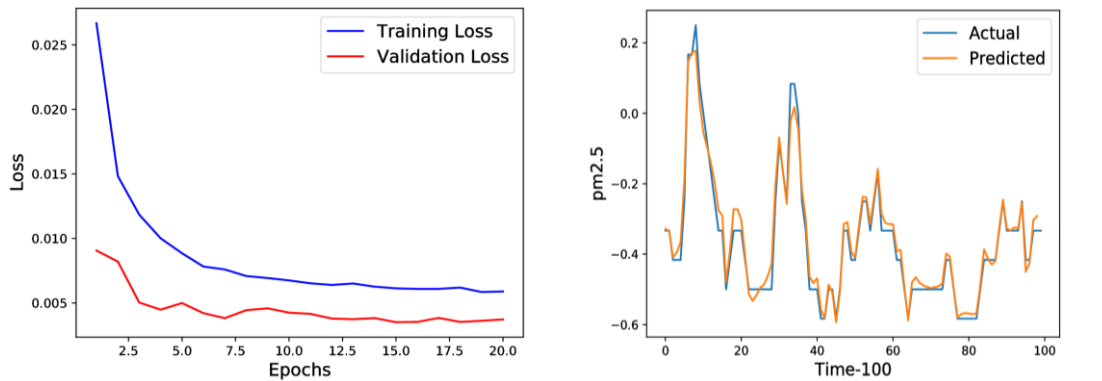


Figure 4. LOSS function of ATT-BI-LSTM model and its prediction

5. CONCLUSION

This study leveraged air quality data from eleven regions in Jiangxi Province, focusing on $PM_{2.5}$ as the predictive indicator, and subjected the model's data forecasting capabilities to thorough testing and validation. Performance analysis of the model utilized four indicators: MSE, RMSE, MAE, and R^2 . The results highlight that the ATT-Bi-LSTM model consistently demonstrates stability and low errors across various complex air quality scenarios while maintaining a high level of fit. This outstanding performance positions the model as a crucial tool for both research and practical applications, providing a deeper understanding and effective means to address issues related to abnormal air quality.

Firstly, the ATT-Bi-LSTM model excels in handling abnormal air quality situations, which can be influenced by diverse factors, including meteorological conditions and anthropogenic pollution. The model's unique design, incorporating Bi-LSTM networks and attention mechanism, allows it to effectively capture and leverage contextual information from input sequences, leading to accurate predictions and anomaly detection. This processing capability establishes the ATT-Bi-LSTM model as an effective tool for addressing abnormal air quality issues.

Secondly, the model demonstrates excellent fitting performance, a critical indicator of its ability to explain observed data. Utilizing the coefficient of determination (R-squared) to measure the degree of fit, the ATT-Bi-LSTM model yields a high R-squared value, indicating its proficiency in explaining the variability in the data. This suggests that the model accurately fits the observed data, enhancing our understanding of patterns and trends in the dataset.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (42261077).

REFERENCES

- [1] Jiang, W., Gao, W. D., Gao, X. M., Ma, M. C., Zhou, M. M., Du, K. and Ma, X., "Spatio-temporal heterogeneity of air pollution and its key influencing factors in the yellow river economic belt of China from 2014 to 2019," *J. Environ. Manage.*, 296, 113172 (2021).
- [2] Kangas, T., Gadeyne, S., Lefebvre, W., Vanpoucke, C. and Rodríguez-Loureiro, L., "Are air quality perception and $PM_{2.5}$ exposure differently associated with cardiovascular and respiratory disease mortality in brussels? Findings from a census-based study," *Environ. Res.*, 219, 115180 (2022).
- [3] Han, X. D., Li, H. J., Liu, Q., Liu, F. Z. and Arif, A., "Analysis of influential factors on air quality from global and local perspectives in China," *Environ Pollut.*, 248, 965-979 (2019).
- [4] Liu, Z. H., Wang, L. L. and Zhu, H. S., "A time-scaling property of air pollution indices: A case study of Shanghai, China," *Atmos. Pollut. Res.*, 6, 457-486 (2015).
- [5] Schwartz, J., Wei, Y., Dominici, F. and Yazdi, M., "Effects of low-level air pollution exposures on hospital admission for myocardial infarction using multiple causal models," *Environ. Res.*, 232, 116203 (2023).
- [6] Sun, T. T., Zhang, T. S., Xiang, Y., Fan, G. Q., Fu, Y. B. and Lv, L. H., "Investigation on the vertical distribution and transportation of $PM_{2.5}$ in the Beijing-Tianjin-Hebei region based on stereoscopic observation network," *Atmos. Environ.*, 294, 119511 (2023).
- [7] Zhong, L. J., Liu, X. L., Hu, X., Chen, Y. L. and Lian, H. Z., "*In vitro* inhalation bioaccessibility procedures for lead in $PM_{2.5}$ size fraction of soil assessed and optimized by *in vivo-in vitro* correlation," *J. Hazard Mater.*, 381, 121202 (2019).
- [8] Hou, X.Y., Guo, Q., Hong, Y. F., Yang, Q. W., Wang, X. K., Zhou, S. Y. and Liu, H. Q., "Assessment of $PM_{2.5}$ -related health effects: A comparative study using multiple methods and multi-source data in China," *Environ. Pollut.*, 306, 119381 (2022).
- [9] Wu, K. T., Peng, X. G., Chen, Z. W., Su, H. K., Quan, H. and Liu, H. Y., "A novel short-term household load forecasting method combined Bi-LSTM with trend feature extraction," *Energy Rep.*, 9, 1013-1022 (2023).
- [10] Ying, H. M., Deng, C. H., Xu, Z. H., Huang, H. X., Deng, W. S. and Yang, Q. L., "Short-term prediction of wind power based on phase space reconstruction and Bi-LSTM," *Energy Rep.*, 9, 474-482 (2023).
- [11] Yang, W. B., Xia, K. W. and Fan, S. R., "Oil logging reservoir recognition based on TCN and SA-Bi-LSTM deep learning method. Eng," *Appl. Artif. Intell.*, 121,105950 (2023).

- [12]Kang, Q., Chen, E. J., Li, Z. C., Luo, H. B. and Liu, Y., “Attention-based LSTM predictive model for the attitude and position of shield machine in tunneling,” *Undergr Space*, 13, 335-350 (2023).
- [13]Chen, Y. F., Peng, L. G., Wang, Y., Zhou, Y. L. and Li, C. S., “Prediction of tandem cold-rolled strip flatness based on Attention-LSTM model,” *J. Manuf. Process.*, 91, 110-121 (2023).
- [14]Wang, X. Y., Liu, H., Yang, Z. H., Du, J. Z. and Dong, X.Y., “CNformer: a convolutional transformer with decomposition for long-term multivariate time series forecasting,” *Appl. Intell.*, 53, 20191-20205 (2023).
- [15]Prihatno, A. T., Nurcahyanto, H., Ahmed, M. F., Rahman, M. H., Alam, M. M. and Jiang, Y. M., “Forecasting PM_{2.5} concentration using a single-dense layer Bi-LSTM method,” *Electronics*, 10,1808 (2021).
- [16]Shu, W. N., Cai, K. and Xiong, N. N., “A short-term traffic flow prediction model based on an improved gate recurrent unit neural network,” *IEEE T. Intell. Transp.*, 23,16654-16665 (2023).
- [17]Abbasimehr, H. and Paki, R., “Improving time series forecasting using LSTM and attention models,” *J. Ambient Intell. Human Comput.*, 13, 673-691 (2022).