



Application of TCN-biGRU neural network in $PM_{2.5}$ concentration prediction

Ting Shi¹ · Pengyu Li¹ · Wu Yang¹ · Ailin Qi¹ · Junfei Qiao¹

Received: 17 April 2023 / Accepted: 5 October 2023 / Published online: 6 November 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

Fine particulate matter ($PM_{2.5}$) poses a significant threat to human life and health, and therefore, accurately predicting $PM_{2.5}$ concentration is critical for controlling air pollution. Two improved types of recurrent neural networks (RNNs), the long short-term memory (LSTM) and gated recurrent unit (GRU), have been widely used in time series data prediction due to their ability to capture temporal features. However, both degrade into random guessing as the time length increases. In order to enhance the accuracy of $PM_{2.5}$ concentration prediction and address the issue of random guessing in RNNs neural networks, this study introduces a TCN-biGRU neural network model. This model is a hybrid prediction approach based on combining temporal convolutional networks (TCN) and bidirectional gated recurrent units (bi-GRU). TCN extracts higher-level feature information from longer time series data of $PM_{2.5}$ concentrations, while bi-GRU captures features from past and future data to achieve more accurate predictive outcomes. This case study utilizes data from monitoring stations in Beijing in 2021 for conducting $PM_{2.5}$ prediction experiments. The TCN-biGRU model achieves an average absolute error, root mean square error, and R^2 of 4.20, 7.71, and 0.961 in its predictive outcomes. When compared to the predictive outcomes of individual LSTM, GRU, and bi-GRU models, it is evident that the TCN-biGRU model exhibits smaller errors and superior predictive performance.

Keywords $PM_{2.5}$ · Neural networks · TCN · Bi-GRU

Introduction

With the acceleration of China's industrialization process, severe climate problems have arisen. Among them, air pollution has received widespread public attention. According to data from the World Health Organization, about 4.2

million people die prematurely each year due to exposure to environmental air pollution (Song et al. 2017). The particulate matter 2.5 ($PM_{2.5}$) concentration, which is the main pollutant, is widely used as an air quality monitoring and regulatory indicator (Qi et al. 2019). $PM_{2.5}$ can remain suspended in the air for extended periods and penetrate deep into the lungs (Liu et al. 2023). Besides, epidemiological studies have shown that $PM_{2.5}$ particulate matter severely affects people's health (Dong et al. 2019), and long-term exposure to high $PM_{2.5}$ concentrations increases the risk of respiratory diseases, lung cancer, and cardiovascular diseases (Zou et al. 2016; Li et al. 2015). Since one of the pathways for the respiratory system's exposure to $PM_{2.5}$ is through the nasal-brain route, the central nervous system is highly vulnerable to harm (Liang et al. 2023). Therefore, individuals in areas with high concentrations of $PM_{2.5}$ pollution are more prone to developing neurodegenerative diseases (Younan et al. 2020). Currently, Beijing has set up 35 monitoring stations to monitor hourly air quality data, including the concentration of particulate matter 2.5 ($PM_{2.5}$), particulate matter 10 (PM_{10}), air quality index (AQI), carbon monoxide (CO), nitrogen

Responsible editor: Marcus Schulz

✉ Ting Shi
tingshi@bjut.edu.cn

Pengyu Li
purylee@163.com

Wu Yang
S202073067@emails.bjut.edu.cn

Ailin Qi
s202173076@emails.bjut.edu.cn

Junfei Qiao
adqiao@bjut.edu.cn

¹ Faculty of Information Technology, Beijing University of Technology, Nanmofang, 100124 Beijing, China

dioxide(NO_2), ozone(O_3), and sulfur dioxide(SO_2), which is beneficial for people to understand the pollution concentration information in their areas. $PM_{2.5}$ concentration exhibits nonlinear characteristics regarding time and space, implying that monitoring stations have limited effectiveness in preventing and controlling $PM_{2.5}$ pollution. Therefore, accurate prediction of $PM_{2.5}$ concentration holds significant importance for air pollution prevention and control (Yule 1927). Air quality prediction methods involve statistical regression, machine learning, and deep learning techniques. In Park et al. (2023), the authors utilized outdoor $PM_{2.5}$ concentration, temperature, and humidity data near indoor target points as input to calculate indoor $PM_{2.5}$ concentration using a multiple linear regression model. Experimental results demonstrated the feasibility of this approach in predicting indoor $PM_{2.5}$ concentration. Furthermore, the article evaluated the model's performance based on seasons, revealing that seasonal characteristics significantly influence indoor $PM_{2.5}$ concentration and the predictive model's performance. In Song et al. (2015), the authors constructed a generalized additive model (GAM) to estimate the statistical relationship between latent variables and $PM_{2.5}$ concentration. Experimental results demonstrated an increase of 18.73% in the R^2 value of this model compared to stepwise linear regression, indicating its applicability for $PM_{2.5}$ prediction. Due to the often complex and nonlinear relationships between variables, there is a need for improvement in the predictive accuracy of these methods. For example, Li et al. (2022b) constructed a random forest regression model incorporating MAIAC AOD, meteorological, topographical, date, and location data to estimate daily $PM_{2.5}$ concentrations in the Huaihai Economic Zone from 2000 to 2020. The results demonstrated the effectiveness of this approach in accurately predicting $PM_{2.5}$ concentrations in a separate study. Besides, in Wang et al. (2021), the authors employed PSO-SVR, GWO-SVR, PSO-GSA-SVR, and GRNN (with spread=0.4 and spread=0.5) to fit three intrinsic mode feature functions (imf) based on CEEMD. By randomly combining the predictive results of the three imfs to generate 125 individual models and subsequently selecting these individual models through the DPC method for combined predictions, the authors achieved significantly accurate forecasts for $PM_{2.5}$ time series data in four Chinese cities. However, the methods mentioned above did not address the temporal correlations in $PM_{2.5}$ concentration data. Additionally, these models exhibit limited capability to represent complex functions, and there is room for improvement in their generalization capabilities for complex prediction problems. In recent years, neural networks have significantly developed with improved computer computing power. Compared with traditional methods, neural networks can deal with complex

nonlinear relationships (Zhang et al. 2021a). Therefore, an increasing number of scholars worldwide utilize deep learning for regression problem-solving. Among these, recurrent neural networks (RNNs) are designed to handle sequential information effectively (Zhang et al., 2019) and have found widespread application in fields like fault diagnosis, machine translation, and speech recognition, achieving promising results (Mansouri et al. 2022; Li et al. 2018; Kim and Lee 2020; Ackerson et al. 2021; Zhou et al. 2019). However, RNNs suffer from the vanishing gradient problem and the exploding gradient problem caused by long sequences.

LSTM and GRU neural networks have been proposed to address this issue. Since $PM_{2.5}$ concentration data has dynamic characteristics over time and can be described using time series models, the LSTM and GRU neural networks have been applied in $PM_{2.5}$ concentration prediction research. For example, Li et al. (2022c) established a GRU-based $PM_{2.5}$ concentration prediction model that utilizes the mean relative error (MRE), root mean square error (RMSE), and Pearson correlation coefficient as evaluation criteria to evaluate the network's accuracy. Extensive experiments demonstrated the model's appealing predictive performance. Besides, Zhou et al. (2019) used hourly $PM_{2.5}$ concentration and weather information in Beijing as input and trained four models based on seasons using a GRU model. The authors demonstrated that the GRU-based model has a higher prediction accuracy and is suitable for time series prediction of atmospheric pollutants. Ge et al. (2019) used a deep bidirectional and unidirectional long short-term memory (DBU-LSTM) neural network to obtain feature information for $PM_{2.5}$ concentration data and relied on tensor decomposition to complete missing data. Experiments highlighted the model's feasibility. However, none of the existing works discussed the correlation between other data in the dataset and $PM_{2.5}$ concentration and the correlation between $PM_{2.5}$ concentrations.

Furthermore, Huang et al. (2021) proposed an EMD-GRU neural network based on empirical mode decomposition (EMD) for predicting $PM_{2.5}$ concentration. The $PM_{2.5}$ concentration sequence was decomposed using EMD, and the resulting sub-sequences and meteorological features were input into a constructed GRU neural network for training and prediction. Experimental results highlighted that this method accurately predicted $PM_{2.5}$ concentration. Zhang et al. (2022a) suggested a method for hourly prediction of Beijing's $PM_{2.5}$ concentration based on a Bi-LSTM neural network and discussed the effectiveness of incorporating meteorological features. Indeed, the corresponding experimental results revealed that exploiting meteorological features effectively reduces the prediction error of $PM_{2.5}$

concentration. Besides, Ding and Zhu (2022) constructed an LSTM model based on principal component analysis (PCA) and attention mechanism to eliminate the correlation effect between indicators and reduce model complexity, achieving good experimental prediction results. However, LSTM and GRU models degrade into random guessing as the length of the time series data increases.

This study proposes a hybrid neural network called TCN-biGRU to address this issue and preserve historical information to a greater extent. This model combines the advanced feature extraction capability of TCN neural networks with bi-GRU neural networks' time series prediction ability. Unlike other studies that focus on improving accuracy by optimizing model parameters or increasing model complexity, this model is designed based on the data feature analysis, taking into account the advantages of TCN and bi-GRU models that are consistent with the inherent features of the data. In the developed architecture, the TCN neural network achieves an exponentially large receptive field (Liu et al. 2020) due to the dilated convolutions and residual connections structures, where the neural network's input can be a long time series data segment. Compared to LSTM neural networks, GRU neural networks have a simpler architecture, less computational complexity, and faster training speed (Wang et al. 2021). Moreover, incorporating directional information can improve the model's accuracy by considering the strong correlation between $PM_{2.5}$ concentrations between the previous and subsequent periods. Additionally, the bi-GRU neural network involves two GRU models from the time series and the reverse time series directions, providing complete historical and future information for each time point in the input sequence of the output layer (Liang et al. 2020). Hence, the proposed TCN-biGRU neural network combines the advantages of both models and can be used in the prediction research of $PM_{2.5}$ concentrations. This research fills a gap by exploring the fusion of TCN and bi-GRU models. The main contributions of this paper are as follows:

- 1) Pollutants highly correlated with $PM_{2.5}$ concentrations are investigated as inputs to the neural network, which improved the prediction accuracy compared to solely relying on $PM_{2.5}$ concentrations.
- 2) The autocorrelation of $PM_{2.5}$ concentrations is explored, and a strong correlation between $PM_{2.5}$ concentrations and the concentrations in the previous and subsequent periods is verified, providing a basis for using bi-GRU neural networks.
- 3) A neural network model is proposed named TCN-biGRU, and comparative experiments are conducted using the Beijing air quality data from 2021/01/01 to 2021/12/31 to validate the effectiveness and performance advantages of the developed method.

Methods

To accurately predict the $PM_{2.5}$ concentration ($\mu\text{g}/\text{m}^3$), this paper proposes a neural network model based on TCN-biGRU implemented using measurement data from air quality monitoring stations in Beijing. Additionally, to improve the model's accuracy, the relationship between other factors (PM_{10} ($\mu\text{g}/\text{m}^3$), AQI , CO ($\mu\text{g}/\text{m}^3$), NO_2 ($\mu\text{g}/\text{m}^3$), O_3 (mg/m^3), and SO_2 ($\mu\text{g}/\text{m}^3$)) and $PM_{2.5}$ concentration is discussed, as well as the autocorrelation analysis of $PM_{2.5}$ concentration.

The model input is historical air quality data. However, the correlation between each pollutant and $PM_{2.5}$ concentration is calculated to increase the model's accuracy. Since the Pearson correlation coefficient measures the correlation between variables (Shi et al. 2021), it is utilized to analyze the correlation between each pollutant and $PM_{2.5}$ concentration. Thus, the auto-correlation coefficient of $PM_{2.5}$ concentration data is calculated. Moreover, a suitable number of timesteps is selected based on the calculation cost and accuracy by evaluating 4, 6, 12, 18, and 28 timesteps. Finally, the monitoring point data is input to the input layer of the TCN-biGRU neural network, and the predicted value of the monitoring point $PM_{2.5}$ concentration is obtained. Next, the process is described in detail.

Correlation analysis

The monitoring stations report data on multiple pollutants, with existing studies revealing a correlation between pollutants (Zhang et al. 2021b; Wu et al. 2022; Popescu et al. 2017). Therefore, in this study, the Pearson correlation coefficient represents the relationship between $PM_{2.5}$ concentration and the concentration of other pollutants. The Pearson correlation coefficient (Pearson 1900) formula is as follows:

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \quad (1)$$

$$\text{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \quad (2)$$

where X and Y represent the concentration of $PM_{2.5}$ and other pollutants, respectively, $\text{Cov}(X, Y)$ is the covariance of X, Y , and σ_X , and σ_Y denotes the standard deviation between $PM_{2.5}$ concentration and other pollutant concentrations.

Autocorrelation analysis

To validate that the $PM_{2.5}$ concentration at time T is influenced by the $PM_{2.5}$ concentration of the previous and subsequent time points and to demonstrate the importance

of bi-GRU in the TCN-biGRU model, the autocorrelation function (ACF) is utilized to prove that the $PM_{2.5}$ concentration time series has autocorrelation, i.e., ACF reveals the correlation at each lag (Flores et al. 2012). The concept and formula of autocorrelation function (ACF) were first introduced in Yule (1927). The concept and formula of ACF have gradually evolved and taken shape throughout the development of time series analysis. The ACF is defined as follows:

$$Corr = \frac{\sum_{i=1}^{n-k} (x_i - u)(x_{i+k} - u)}{\sum_{i=1}^n (x_i - u)^2} \tag{3}$$

where K is the order, u is the mean value of the sequence, and x_i and x_{i+k} correspond to item i of the two separated sequences.

TCN-biGRU neural network

We combine the TCN and bi-GRU neural networks into TCN-biGRU to benefit from the advantages of each network. This section will introduce the TCN and biGRU neural networks separately.

TCN neural network

TCN is a convolutional neural network proposed by Bai et al. (2018). Its design combines best practices, such as fully convolutional networks, dilated convolutions, residual connections, and causal convolutions (Hu et al. 2022). Experimental results have shown that TCN outperforms RNN and LSTM networks in predicting longer time series data (Yan et al. 2020) due to relying on dilated convolutions and residual layer modules that increase the network’s receptive field and obtain more historical information. Hence, the proposed model exploits this characteristic of the TCN network. Therefore, the following sections detail the dilated convolution and residual layer modules of TCN. Besides, the fully convolutional networks and causal convolutions ensure that the network’s input and output sequence lengths are the same and that information does not “leak” into the past data. It should be noted that these two parts will not be discussed in this

paper. Using dilated convolutions can achieve exponentially large receptive fields. The receptive field can be understood as the maximum number of steps back from the current data at time T . When the kernel size is $k=2$, and the dilation rate is $d = [1,2,4]$, dilation allows for the input to be sampled at spaced intervals during convolution. Figure 1 illustrates the structure diagram of a dilated convolution, which is formulated as follows (Bai et al. (2018)):

$$F(s) = (x *_d f)(s) = \sum_{i=0}^{k-1} f(i)\dot{x}_{s-di} \tag{4}$$

where $x \in R^n$ is the input of a one-dimensional sequence, f is the convolution kernel, and s is the element in the sequence. The receptive field’s formula is as follows:

$$R = (k - 1) * (\sum_i d_i + 1) \tag{5}$$

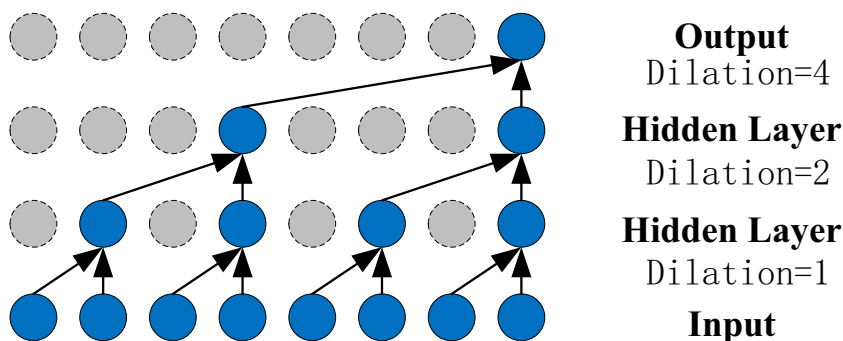
Therefore, to increase the receptive field, the convolution kernel size k is changed, or the dilation factor d is increased. However, as the network becomes deeper, this strategy increases the computation cost, gradient explosion, and gradient vanishing (Li et al. 2022a).

To solve these problems, residual layer modules are added to TCN, with Fig. 3 presenting the updated structure diagram. The residual layer module comprises two identical layers, including dilated causal convolution, weight normalization, ReLU, and dropout layers (not used). The 1×1 convolution layer ensures that the input and output of the residual have the same dimensions. The output o of input i is presented in Eq.6 (Bai et al. (2018)), and therefore, the receptive field R of the TCN neural network with N residual blocks is presented in Eq.7 (Bai et al. (2018)). Factor 2 is set because a cell has two layers.

$$o = Activation(i + F(i)) \tag{6}$$

$$R = 2 * (k - 1) * N * (\sum_{i=0} d_i + 1) = N * (k - 1) * 2^i \tag{7}$$

Fig. 1 The structure diagram of the dilated convolution



Bi-GRU neural network

Compared to the input, forget, and output gates in LSTM neural networks, the GRU neural network is optimized and has only a reset and an update gate. The former gate preserves useful past information, and the latter combines past and current information. Moreover, the reset gate captures short-term dependencies in time series, and the update gate captures long-term dependencies. The GRU neural network structure is simple and thus reduces processing and training time (Zhang et al. 2022b). The structure diagram of GRU is illustrated in Fig. 2, where H_{t-1} represents the hidden state of the previous timestep, H_t is the hidden state of timestep t , \tilde{H}_t is the candidate hidden state of timestep t , and R_t and Z_t represent the reset door and update door, respectively. W is the weight parameter, and b is the deviation parameter. The GRU core formulas are as follows (Cho et al.2014):

$$R_t = \sigma(X_t W_{xr} + H_{t-1} W_{hr}) \tag{8}$$

$$Z_t = \sigma(X_t W_{xz} + H_{t-1} W_{hz}) \tag{9}$$

$$\tilde{H}_t = \tanh(X_t W_{xh} + (R_t * H_{t-1}) W_{hh}) \tag{10}$$

$$H_t = Z_t * \tilde{H}_t + (1 - Z_t) * H_{t-1} \tag{11}$$

The bi-GRU neural network comprises a forward and a backward GRU (Ortega-Bueno et al. 2019), with one GRU processing the time series data, arranged forward and the other backward. The structure diagram is presented in Fig. 2, and the output formula is formulated in Eq. 12. Considering

that the $PM_{2.5}$ concentration is highly correlated with the concentration data of the previous and subsequent moments, using a bi-GRU neural network to train the network from both forward and backward directions improves the network’s predictive accuracy.

$$H_t = (\overleftarrow{H}_t + \overrightarrow{H}_t)/2 \tag{12}$$

TCN-biGRU neural network

Figure 3 illustrates the main network architecture of the proposed TCN-biGRU prediction model. The Dense 1 layer aims to change the shape of the TCN network output data from (batch_size, nb_filters) to (batch_size, timesteps * input_dim), where nb_filters is the number of filters used in the convolution layer. It should be noted that this work does not employ a dropout in the residual linking layer, as our trials revealed that it did not significantly improve the model’s performance. Besides, the TCN-biGRU neural network has a large receptive field and can extract features through time series and reverse time series, enhancing $PM_{2.5}$ concentration prediction.

Experiment

The deep learning models employed in this paper are built using the Keras framework and Python programming language. All experiments are conducted on a 64-bit Windows 10 operating system with an Intel Core i7-8750H CPU processor.

Fig. 2 The structure diagram of GRU model and bi-GRU model

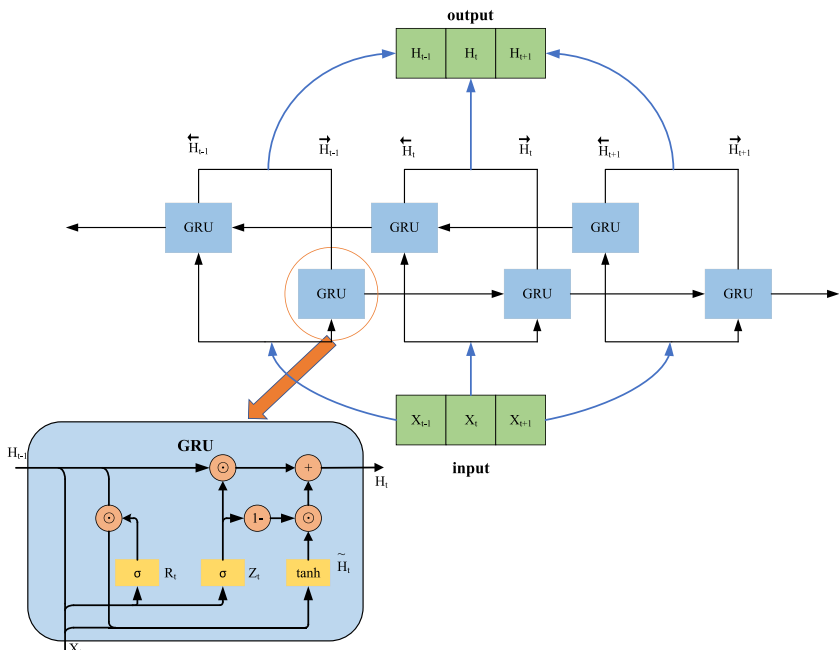
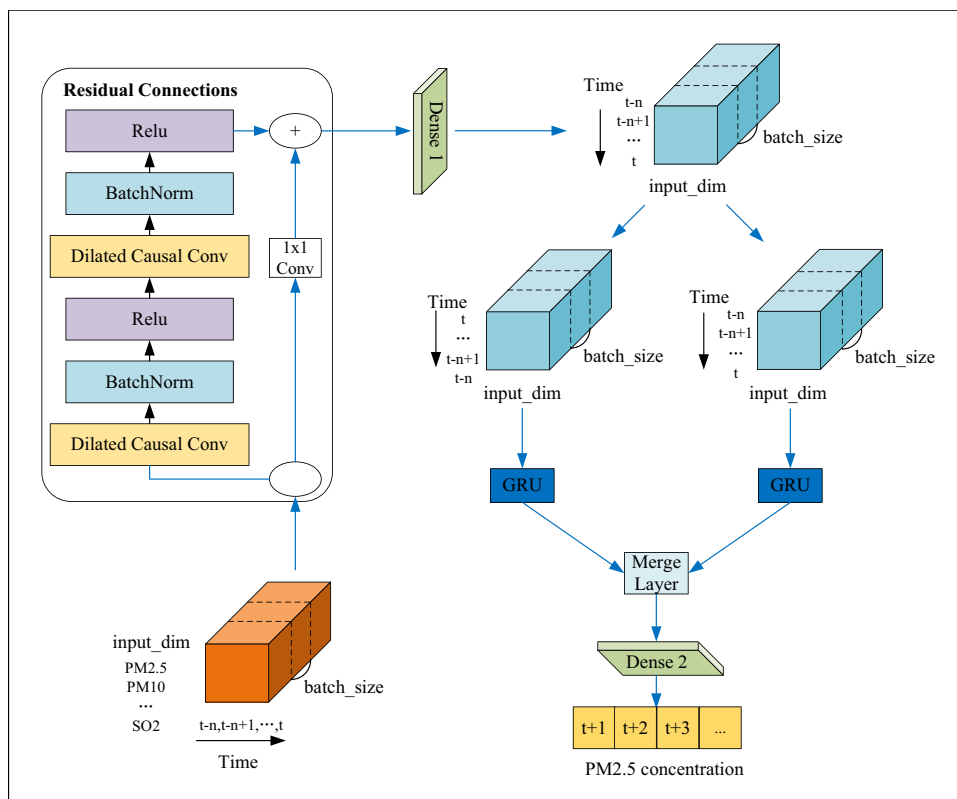


Fig. 3 The structure diagram of the dilated convolution



Data preprocessing

Data source

To verify the prediction accuracy of the proposed TCN-biGRU model, the air quality data of Beijing were collected and released by the National Environmental Monitoring Station for time series prediction. The dataset covers hourly data from January 1st, 2021, to December 31st, 2021, including 7 features: $PM_{2.5}$ ($\mu g/m^3$), PM_{10} ($\mu g/m^3$), AQI , CO ($\mu g/m^3$), NO_2 ($\mu g/m^3$), O_3 (mg/m^3), and SO_2 ($\mu g/m^3$). For data for each contaminant, if three or more samples (rows) are missing in a row, samples (rows) with missing information attribute values are deleted. If there are fewer than three consecutive missing data, the mean is used for completion. We use 80% of the data as the training set and 20% as the test set of the neural network.

Correlation analysis

The correlation between $PM_{2.5}$ concentration and the concentrations of other pollutants is verified based on the Pearson correlation coefficients between the $PM_{2.5}$ concentration and the remaining six pollutant concentrations. Figure 4 depicts the corresponding heatmap highlighting that $PM_{2.5}$ concentration positively correlates with PM_{10} , AQI , NO_2 , and SO_2 concentrations and negatively with O_3 concentra-

tion. The Pearson correlation coefficient ranges from -1 to 1 , with each correlation coefficient corresponding to a specific color (shown on the right side of the figure). As the correlation coefficient approaches 1 , the color becomes lighter, and as it approaches -1 , it becomes darker. A larger absolute value of the Pearson correlation coefficient indicates a stronger correlation. Additionally, the correlation with CO and O_3 is relatively small. Therefore, our neural network inputs are $PM_{2.5}$, PM_{10} , AQI , NO_2 , and SO_2 . Note that the same neural network structure and parameters are used to conduct a comparative experiment against solely utilizing the $PM_{2.5}$ concentration as input. The corresponding results are presented in Sect. 3.3.

Autocorrelation analysis

The effectiveness of the developed bi-GRU neural network is demonstrated by employing the $PM_{2.5}$ concentration data. The results are depicted in Fig. 5, where the vertical axis is the confidence coefficient, the horizontal axis is the lag k , and the blue area is the confidence interval. Figure 5 infers that the $PM_{2.5}$ concentration has a high correlation with itself, and the correlation coefficient gradually decreases as the lag k increases. Specifically, the autocorrelation coefficient is still around 0.9 when $k=4$, 0.8 when $k=6$, 0.7 when $k=12$, 0.6 when $k=18$, and 0.5 when $k=28$. The lag k provides a basis for the timestep values in Section 3.5.1. In addition, the bi-GRU

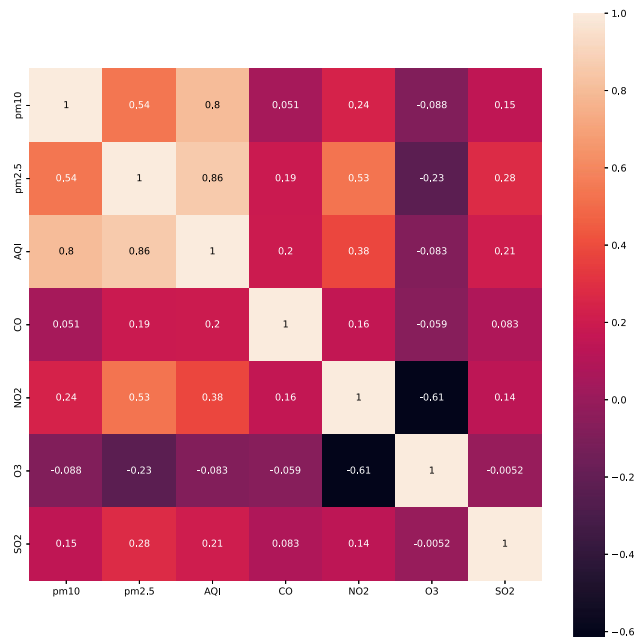


Fig. 4 A heatmap of the correlation between $PM_{2.5}$ concentration and other pollutant concentrations

neural network obtains more feature information from the forward and reverse sequences. Therefore, our model uses a bi-GRU neural network and fuses it with the TCN neural network.

Determination of model parameters

Due to the critical roles that activation function types and neuron counts play in the accuracy of artificial neural network models, it is ensured that other parameters were kept constant and separately investigated the effects of activation

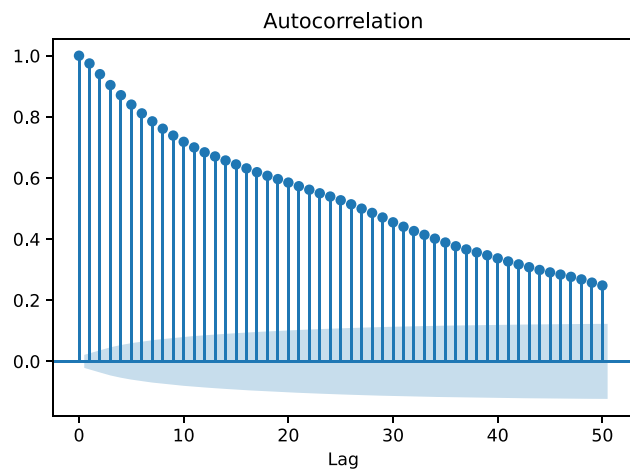


Fig. 5 A graph showing the results of autocorrelation analysis on $PM_{2.5}$ concentration

function types and neuron counts on the model. Four activation functions—linear, tanh, sigmoid, and ReLU—were selected for experimentation. In the case of neuron count analysis, the TCN network is chosen as a baseline. While maintaining the same structure for the bi-GRU neural network, the neuron count of the TCN neural network is altered within the bounds of its predictive capacity. Following attaining the TCN’s predictive limit, the TCN-biGRU model is established based on this baseline structure.

Data standardization

The Z-score normalization method is utilized to unify the different scales of the data and improve comparability. The dataset is normalized through Z-score normalization by setting the mean to 0 and the standard deviation to 1. Using the entire dataset’s overall mean and standard deviation mitigates the impact of outliers. This technique enhances model predictive performance (Tanaka et al. 2022). This common normalization method standardizes the data by subtracting and dividing the mean by the standard deviation. Z-Score is formulated as follows:

$$Z = \frac{x - \mu}{\sigma} \tag{13}$$

where μ and σ are the average and the standard deviation of all data, respectively.

Evaluating criterion

Our model’s prediction accuracy is evaluated based on the mean squared error (MSE), used as the loss function, and the MAE, RMSE, and R^2 evaluation metrics. These metrics are commonly used to evaluate the variability and accuracy of data, with the corresponding formulas presented in Eqs. 14–17, where N is the number of samples, y_i is the actual value, \check{y}_i is the predicted value, and \bar{y}_i is the average value. Note that the smaller the MSE, MAE, and RMSE, the higher

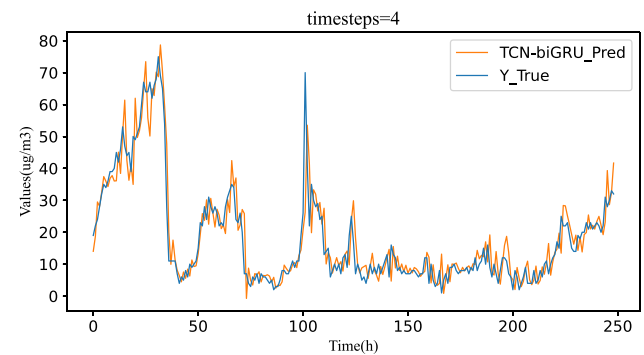


Fig. 6 Prediction results of the model when timesteps=4

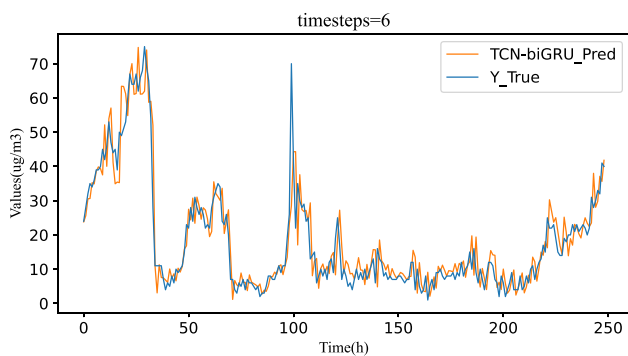


Fig. 7 Prediction results of the model when timesteps=6

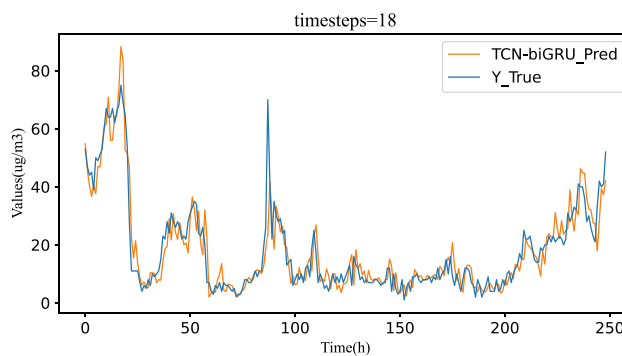


Fig. 9 Prediction results of the model when timesteps=18

the model’s accuracy. Moreover, the closer R^2 is to 1, the higher the model’s accuracy.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \tag{14}$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \tag{15}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \tag{16}$$

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (\bar{y}_i - y_i)^2} \tag{17}$$

Model prediction experiments

Determination of timesteps and input data

The “timesteps” parameter indicates how many timesteps of input data should be included in each unit model’s input, representing how many previous consecutive input data points are relevant to the current data. Therefore, the

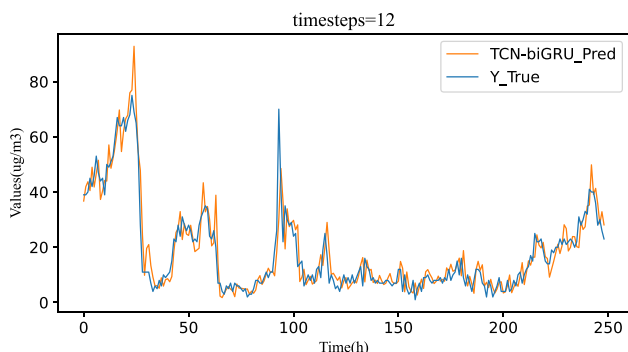


Fig. 8 Prediction results of the model when timesteps=12

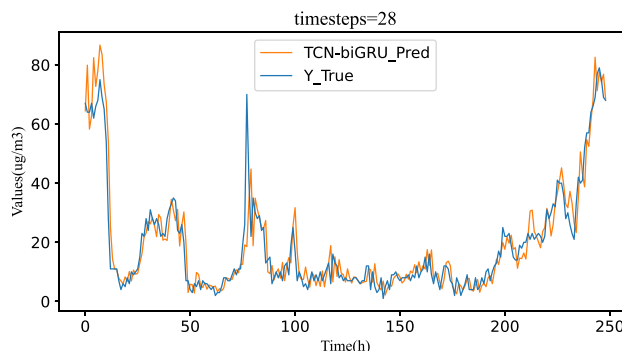


Fig. 10 Prediction results of the model when timesteps=28

Table 1 Performance comparison of models with different timestep values

Timesteps	MAE	RMSE	R^2	Time (s)
Timesteps=4	4.19	8.13	0.953	158
Timesteps=6	4.76	10.16	0.930	146
Timesteps=12	4.73	9.32	0.940	182
Timesteps=18	4.69	8.35	0.952	214
Timesteps=28	4.93	8.97	0.949	270

Table 2 Performance comparison of models with different input dimension values

Input dimension	MAE	RMSE	R^2
input_dim=1	4.12	8.89	0.949
input_dim=5	4.19	8.13	0.955

Table 3 Predictive results of different activation function models

Activation	MAE	RMSE	R^2
ReLu	4.19	8.13	0.953
tanh	5.09	9.97	0.935
sigmoid	5.47	11.11	0.920
linear	4.45	8.35	0.954

Table 4 Predictive results of models with different neuron counts

Neuron	MAE	RMSE	R^2
32	4.62	8.72	0.942
50	4.19	8.13	0.953
64	4.52	8.57	0.944
100	4.56	8.84	0.941
128	4.65	9.35	0.941

“timesteps” value is determined by referring to the lag order k in autocorrelation analysis (Section 3.1.3). Timesteps are a very important hyperparameter in neural networks, and an appropriate value can improve the accuracy of time series prediction models. To select a suitable value for timesteps, comparative experiments are conducted with the timestep value set to 4, 6, 12, 18, and 28 while keeping other neural network parameters the same. The prediction results for different “timesteps” values are depicted in the figure below (Fig. 6, 7, 8, 9, 10, partial results are displayed), revealing that when the timesteps are 4, the model can track the $PM_{2.5}$ concentration during dramatic changes better than in the other four timestep values. The experimental results are reported in Table 1, highlighting that the proposed model performs best when the timestep is 4, while simultaneously, the computation and training time of the system are relatively short. Therefore, 4 timesteps are selected. Additionally, various input data setups are compared, including $PM_{2.5}$, PM_{10} , AQI , NO_2 , and SO_2 , and using only $PM_{2.5}$, to verify that the neural network’s prediction accuracy is higher when using the five input data. Similarly, the input dimensions of the data are modified while preserving the other parameters, with the corresponding results reported in Table 2. The results infer that the MAE is better when $input_dim=1$, but the RMSE and R^2 are inferior when $input_dim=5$. Hence, to enhance the model’s accuracy, $input_dim = 5$. It is observed that insufficiently comprehensive input parameters can impact the accuracy of the model’s predictions.

Determination of activation functions and neuron count

In order to achieve optimal performance of the model and achieve more accurate predictions, we separately investigated the impact of different activation functions and neuron counts on the model’s accuracy. While keeping other net-

work parameters consistent, four commonly used activation functions are explored: linear, tanh, sigmoid, and ReLU. The experimental results are presented in Table 3, revealing that the model performs best when the ReLU activation function is chosen for the neural network.

For the selection of neuron counts, five different cases of neuron counts are evaluated: 32, 50, 64, 100, and 128, as shown in Table 4. The results indicate that as the number of neurons in the TCN hidden layers increases, the prediction performance improves significantly and gradually decreases. The model’s performance reaches optimal when the neuron count is set to 50. Therefore, the neuron count is set to 50 in the experiments.

Prediction results

It is compared against the LSTM and GRU neural networks to validate the proposed model’s accuracy. Since the LSTM, GRU, and TCN-biGRU neural networks converged after 150 iterations (see Fig. 11), we set 150 epochs for each model. The predicted results of all models are reported in Table 5. Some of the predicted results are visualized in Fig. 12.

Furthermore, it is compared against the bi-GRU and TCN-GRU neural networks to validate the proposed model better. In the test data, the TCN-biGRU neural network had a lower mean absolute error, root mean square error, and a higher R^2 than the competitor neural networks. Figure 12 reveals that the TCN-biGRU-based predicted values follow the trend of the actual values to a certain extent, and it performs better than the competitor networks when the data have larger variations. Additionally, the results demonstrate that the TCN-biGRU neural network converged significantly faster than the LSTM and GRU neural networks. Therefore, the suggested network can be effectively applied for $PM_{2.5}$ concentration prediction.

Furthermore, to validate the effectiveness of our proposed model, its predictive results are compared with two different hybrid models. The predictive outcomes of the CNN-LSTM model were sourced from the study by Xie et al. (2023), while the predictive outcomes of the LSTM-Attention model were sourced from the study by Gao and Li (2022), ensuring consistency in the dataset and model settings. The predictive results are presented in Table 6, revealing that the TCN-biGRU neural network outperforms the other two models. Hence, establishing the TCN-biGRU neural network model

Fig. 11 With the increase of the number of iterations, the loss values of the three models reach a stable level in about 150 iterations

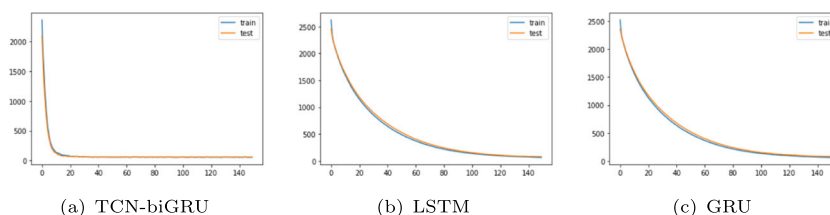


Table 5 Performance comparison of five different neural networks under the same conditions

Model	MAE	RMSE	R^2
LSTM	4.80	10.14	0.938
GRU	4.85	10.60	0.932
Bi-GRU	4.72	10.36	0.935
TCN-GRU	4.61	9.31	0.947
TCN-biGRU	4.19	8.13	0.955

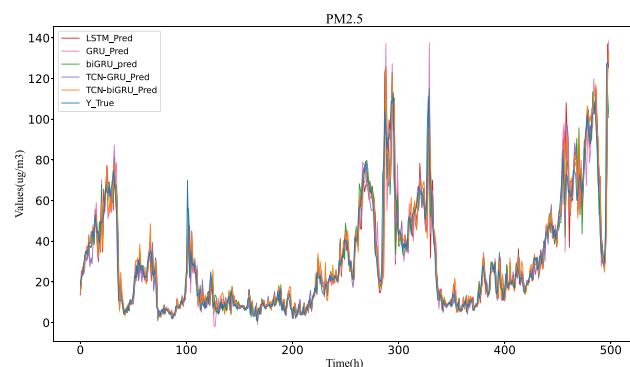
is highly necessary for enhancing the accuracy of $PM_{2.5}$ concentration prediction.

Discussion and conclusion

Discussion

This study introduces the TCN-biGRU neural network for predicting $PM_{2.5}$ concentration, with various factors and parameters influencing the model's accuracy. The paper investigates the impact of single-input versus multiple-input variables on predictive accuracy and the influence of neural network hyperparameters such as timestep, activation function, and number of neurons on predictive accuracy.

Table 1 highlights that using five input variables yields a prediction result with a 0.76 lower RMSE and a 0.009 higher R^2 compared to using only $PM_{2.5}$ concentration as input. Table 2 reveals that the predictive accuracy is highest when timesteps=4, with a significant reduction in MAE and RMSE by 0.74 and 2.03, respectively, and an improvement in R^2 by 0.023, compared to other scenarios. Table 3 demonstrates that using the ReLu activation function results in the largest reduction in MAE and RMSE by 1.28 and 2.98, respectively, compared to the reference experiment. The impact of TCN neural neuron count on predictive accuracy is presented in Table 4, where using 50 neurons yields the maximum reduc-

**Fig. 12** The prediction results of $PM_{2.5}$ concentration using five types of neural networks**Table 6** Performance comparison of two different neural networks under the same conditions

Model	MAE	RMSE	R^2
TCN-biGRU	4.19	8.13	0.955
CNN-LSTM	4.53	9.24	0.948
LSTM-Attention	4.64	9.52	0.945

tion in MAE and RMSE by 0.46 and 1.22, respectively, compared to the reference experiment. Both input variables and hyperparameters influence the model's accuracy. Among these, the choice of timesteps and activation function has a more pronounced impact on the model's output than the effects of input variables and the number of neurons on the model's accuracy.

Figure 12 reveals that all five models can effectively track the variations in $PM_{2.5}$ concentration data. However, when confronted with larger fluctuations, the TCN-biGRU neural network exhibits superior tracking performance compared to the other four models. Additionally, by comparing Tables 5 and 6, it is evident that fused neural networks like CNN-LSTM and LSTM-Attention show better predictive accuracy than standalone LSTM and GRU neural networks. Nonetheless, compared to the TCN-biGRU neural network proposed in this study, there remains some disparity. From the above discussions, it is evident that the TCN-biGRU neural network can be utilized for predicting $PM_{2.5}$ concentration.

Conclusion

This paper introduces the TCN-biGRU model for predicting $PM_{2.5}$ concentration in Beijing based on a combination of temporal convolutional neural network (TCN) and bidirectional gated recurrent unit (bi-GRU) neural network. The model considers the relationship between meteorological features and $PM_{2.5}$ concentration and the impact of $PM_{2.5}$ concentration's self-correlation on predictions. It also investigates the influence of input parameters, neuron counts, and activation functions on prediction accuracy. Based on the experimental results, it is evident that the input variables of the neural network and the neural network's hyperparameters influence the model's accuracy. Furthermore, the TCN-biGRU model is also compared with other prediction models. The experimental results indicate that this approach outperforms the other comparative models with MAE, RMSE, and R^2 values of 4.19, 8.13, and 0.955, respectively. This research offers valuable insights for $PM_{2.5}$ concentration prediction and environmental control.

Author contribution All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by WY and AQ. The first draft of the manuscript was written by PL. TS helped perform the analysis with constructive discussions,

and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was supported in part by the Organization Department of Beijing Municipal Committee under Grant Z2020549, in part by the Ministry of Education of China under Grant 202102341001, 202102165002, in part by the National Science Foundation of China under Grant 62273011 and Grant 62076013, and in part by the Beijing Natural Science Foundation under Grant JQ21014.

Data availability Not applicable

Declarations

Ethical approval Not applicable

Consent to participate Not applicable

Consent for publication Not applicable

Competing interests The authors declare no competing interests.

References

- Ackerson JM, Dave R, Seliya N (2021) Applications of recurrent neural network for biometric authentication & anomaly detection. [arXiv:2109.05701](https://arxiv.org/abs/2109.05701)
- Bai S, Kolter JZ, Koltun V (2018) An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. [arXiv:1803.01271](https://arxiv.org/abs/1803.01271)
- Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y (2014) Learning phrase representations using RNN encoder-decoder for statistical machine translation. [arXiv:1406.1078](https://arxiv.org/abs/1406.1078)
- Ding W, Zhu Y (2022) Prediction of PM_{2.5} concentration in Ningxia Hui autonomous region based on PCA-Attention-LSTM. *Atmosphere*, 13:1444
- Dong F, Yu B, Pan Y (2019) Examining the synergistic effect of CO₂ emissions on PM_{2.5} emissions reduction: evidence from China. *J Clean Prod* 223:759–771
- Flores JHF, Engel PM, Pinto RC (2012) Autocorrelation and partial autocorrelation functions to improve neural networks models on univariate time series forecasting. In: The 2012 international joint conference on neural networks (IJCNN), IEEE, pp 1–8
- Gao Y, Li Y (2022) Prediction of rainfall-type debris flow in Jiangjiagou based on LSTM-Attention. 2022 3rd International conference on computer vision, image and deep learning & international conference on computer engineering and applications (CVIDL & ICCEA), IEEE, pp 1–6
- Ge L, Zhou A, Li H, Liu J (2019) Spatially fine-grained air quality prediction based on DBU-LSTM. In: Proceedings of the 16th ACM international conference on computing frontiers, pp 202–205
- Hu C, Cheng F, Ma L, Li B (2022) State of charge estimation for lithium-ion batteries based on TCN-LSTM neural networks. *J Electrochem Soc* 169(3):030544
- Huang G, Li X, Zhang B, Ren J (2021) PM_{2.5} concentration forecasting at surface monitoring sites using GRU neural network based on empirical mode decomposition. *Sci Total Environ* 768:144516
- Kim J, Lee Y (2020) Accelerating RNN transducer inference via one-step constrained beam search. [arXiv:2002.03577](https://arxiv.org/abs/2002.03577)
- Li M, Xiang Q, Chen Z, Wang M (2018) A unified neural network for quality estimation of machine translation. *IEICE Trans Inf Syst* 101(9):2417–2421
- Li R, Kou X, Geng H, Xie J, Yang Z, Zhang Y, Cai Z, Dong C (2015) Effect of ambient PM_{2.5} on lung mitochondrial damage and fusion/fission gene expression in rats. *Chem Res Toxicol* 28(3):408–418
- Li W, Wei Y, An D, Jiao Y, Wei Q (2022) LSTM-TCN: dissolved oxygen prediction in aquaculture, based on combined model of long short-term memory network and temporal convolutional network. *Environ Sci Pollut Res Int* 29(26):39545–39556
- Li X, Li L, Chen L, Zhang T, Xiao J, Chen L (2022b) Random forest estimation and trend analysis of PM_{2.5} concentration over the Huaihai economic zone, China (2000–2020). *Sustainability* 14(14):8520
- Li Y, Zhu Z, Xin C, Chen Z, Wang S, Liang Z, Zou X (2022c) Research on the spatiotemporal characteristics and concentration prediction model of PM_{2.5} during Winter in Jiangbei New District, Nanjing, China. *Atmosphere*, 13:1542
- Liang R, Chang X, Jia P, Xu C (2020) Mine gas concentration forecasting model based on an optimized BiGRU network. *ACS omega* 5(44):28579–28586
- Liang C, Jiang Y, Zhang T, Ji Y, Zhang Y, Sun Y, Li S, Qi Y, Wang Y, Cai Y et al (2023) Atmospheric particulate matter impairs cognition by modulating synaptic function via the nose-to-brain route. *Sci Total Environ* 857:159600
- Liu Y, Ma J, Tao Y, Shi L, Wei L, Li L (2020) Hybrid neural network text classification combining TCN and GRU. In: 2020 IEEE 23rd international conference on computational science and engineering (CSE), pp 30–35
- Liu F, Liu C, Liu Y, Wang J, Wang Y, Yan B (2023) Neurotoxicity of the air-borne particles: from molecular events to human diseases. *J Hazard Mater* 131827
- Mansouri M, Dhibi K, Hajji M, Bouzara K, Nounou H, Nounou M (2022) Interval-valued reduced RNN for fault detection and diagnosis for wind energy conversion systems. *IEEE Sensors J* 22(13):13581–13588
- Ortega-Bueno R, Rosso P, Pagola JEM (2019) UO UPV2 at HAH 2019: BiGRU neural network informed with linguistic features for humor recognition. In: Proceedings of the Iberian languages evaluation forum (IberLEF 2019). CEUR Workshop Proceedings, CEUR-WS, Bilbao, Spain (9 2019)
- Park S-Y, Yoon D-K, Park S-H, Jeon J-I, Lee J-M, Yang W-H, Cho Y-S, Kwon J, Lee C-M (2023) Proposal of a methodology for prediction of indoor PM_{2.5} concentration using sensor-based residential environments monitoring data and time-divided multiple linear regression model. *Toxics* 11(6):526
- Pearson, K (1900) Mathematical contributions to the theory of evolution. VIII. On the correlation of characters not quantitatively measurable. *Proceedings of the Royal Society of London* 66(424-433):241–244
- Popescu M, Mihalache SF, Oprea M (2017) Air pollutants and meteorological parameters influence on PM_{2.5} forecasting and performance assessment of the developed artificial intelligence-based forecasting model. *Revista De Chimie* 68:864–868
- Qi Z, Song Y, Ding Q, Liao X, Li R, Liu G, Tsang S, Cai Z (2019) Water soluble and insoluble components of PM_{2.5} and their functional cardiotoxicities on neonatal rat cardiomyocytes in vitro. *Ecotoxicol Environ Saf* 168:378–387
- Shi P, Fang X, Ni J, Zhu J (2021) An improved attention-based integrated deep neural network for PM_{2.5} concentration prediction. *Appl Sci* 11:4001
- Song Y-Z, Yang H-L, Peng J-H, Song Y-R, Sun Q, Li Y (2015) Estimating PM_{2.5} concentrations in Xi'an city using a generalized additive model with multi-source monitoring data. *PloS one* 10(11):e0142149
- Song C, He J, Wu L, Jin T, Chen X, Li R, Ren P, Zhang L, Mao H (2017) Health burden attributable to ambient PM_{2.5} in China. *Environ Pollut* 223:575–586

- Tanaka T, Nambu I, Maruyama Y, Wada Y (2022) Sliding-window normalization to improve the performance of machine-learning models for real-time motion prediction using electromyography. *Sensors* 22(13):5005
- Wang B, Kong W, Zhao P (2021) An air quality forecasting model based on improved convnet and RNN. *Soft Comput* 25(14):9209–9218
- Wu X, Zhang C, Zhu J, Zhang X (2022) Research on PM_{2.5} concentration prediction based on the CE-AGA-LSTM model. *Appl Sci* 12(14):7009
- Xie Y, Chen X, Zhang L (2023) Prediction of PM_{2.5} concentration based on CNNLSTM deep learning model. 2023 Asia-Europe conference on electronics, data processing and informatics (ACEDPI), IEEE, pp 229–233
- Yan J, Mu L, Wang L, Ranjan R, Zomaya AY (2020) Temporal convolutional networks for the advance prediction of ENSO. *Sci Rep* 10(1):1–15
- Younan D, Petkus AJ, Widaman KF, Xinhui W, Ramon C, Espeland MA, Margaret G, Henderson VW, Manson JE, Rapp SR, Sachs BC, Serre ML, Gaussoin SA, Barnard R, Saldana S, Vizuete W, Beavers DP, Salinas JA, Chui HC, Resnick SM, Shumaker SA, Chen JC (2020) Particulate matter and episodic memory decline mediated by early neuroanatomic biomarkers of Alzheimer's disease. *Brain* 143(1):289–302
- Yule GU (1927) On a method of investigating periodicities in distributed series, with special reference to Wolfer's sunspot numbers. *Phil Trans R Soc London A* 226:267–298
- Zhang B, Jia M, Xu J, Zhao W, Deng L (2022a) Network security situation prediction model based on EMD and ELPSO optimized BiGRU neural network. *Computational Intelligence and Neuroscience*, 2022
- Zhang M, Wu D, Xue R (2021) Hourly prediction of PM_{2.5} concentration in Beijing based on Bi-LSTM neural network. *Multimed Tools Appl* 80:24455–24468
- Zhang P, Xue J, Lan C, Zeng W, Gao Z, Zheng N (2019) EleAtt-RNN: adding attentiveness to neurons in recurrent neural networks. *IEEE Trans Image Process* 29:1061–1073
- Zhang P, Xue J, Lan C, Zeng W, Gao Z, Zheng N (2019) EleAtt-RNN: adding attentiveness to neurons in recurrent neural networks. *IEEE Trans Image Process* 99:1–1
- Zhang X, Fan Y, Wei S, Wang H, Zhang J (2021) Spatiotemporal distribution of PM_{2.5} and its correlation with other air pollutants in winter during, (2016) 2018 in Xi'an, China. *Pol J Environ Stud* 30(2):1457–1464
- Zhang Z, Zhang S, Zhao X, Chen L, Yao J (2022b) Temporal difference-based graph transformer networks for air quality PM_{2.5} prediction: a case study in China. *Front Environ Sci* 10:924986. <https://doi.org/10.3389/fenvs.2022.924986>
- Zhou X, Xu J, Zeng P, Meng X (2019) Air pollutant concentration prediction based on GRU method. *J Phys Conf Ser* 1168(3):032058
- Zhu J, Jiang Q, Shen Y, Qian C, Xu F, Zhu Q (2022) Application of recurrent neural network to mechanical fault diagnosis: a review. *J Mech Sci Technol* 36:527–542
- Zou Y, Jin C, Su Y, Li J, Zhu B (2016) Water soluble and insoluble components of urban PM_{2.5} and their cytotoxic effects on epithelial cells (A549) in vitro. *Environ Pollut* 212:627–635

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.