



## Review

Deep-learning architecture for PM<sub>2.5</sub> concentration prediction: A reviewShiyun Zhou<sup>a, b</sup>, Wei Wang<sup>a, \*\*</sup>, Long Zhu<sup>c</sup>, Qi Qiao<sup>a</sup>, Yulin Kang<sup>a, \*</sup><sup>a</sup> Institute of Environmental Information, Chinese Research Academy of Environmental Sciences, Beijing 100012, China<sup>b</sup> School of Optics and Photonics, Beijing Institute of Technology, Beijing 100081, China<sup>c</sup> College of Water Sciences, Beijing Normal University, Beijing 100875, China

## ARTICLE INFO

## Article history:

Received 24 July 2023

Received in revised form

5 February 2024

Accepted 6 February 2024

## Keywords:

PM<sub>2.5</sub> concentration prediction

Deep-learning based model

Bibliometrics analysis

Evaluation framework

## ABSTRACT

Accurately predicting the concentration of fine particulate matter (PM<sub>2.5</sub>) is crucial for evaluating air pollution levels and public exposure. Recent advancements have seen a significant rise in using deep learning (DL) models for forecasting PM<sub>2.5</sub> concentrations. Nonetheless, there is a lack of unified and standardized frameworks for assessing the performance of DL-based PM<sub>2.5</sub> prediction models. Here we extensively reviewed those DL-based hybrid models for forecasting PM<sub>2.5</sub> levels according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. We examined the similarities and differences among various DL models in predicting PM<sub>2.5</sub> by comparing their complexity and effectiveness. We categorized PM<sub>2.5</sub> DL methodologies into seven types based on performance and application conditions, including four types of DL-based models and three types of hybrid learning models. Our research indicates that established deep learning architectures are commonly used and respected for their efficiency. However, many of these models often fall short in terms of innovation and interpretability. Conversely, models hybrid with traditional approaches, like deterministic and statistical models, exhibit high interpretability but compromise on accuracy and speed. Besides, hybrid DL models, representing the pinnacle of innovation among the studied models, encounter issues with interpretability. We introduce a novel three-dimensional evaluation framework, i.e., Dataset-Method-Experiment Standard (DMES) to unify and standardize the evaluation for PM<sub>2.5</sub> predictions using DL models. This review provides a framework for future evaluations of DL-based models, which could inspire researchers to standardize DL model usage in PM<sub>2.5</sub> prediction and improve the quality of related studies.

© 2024 The Authors. Published by Elsevier B.V. on behalf of Chinese Society for Environmental Sciences, Harbin Institute of Technology, Chinese Research Academy of Environmental Sciences. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Short-term exposure to ambient particulate matter with a diameter of 2.5 μm or less (PM<sub>2.5</sub>) is a leading contributor to the global burden of disease and mortality [1]. Precise prediction of the PM<sub>2.5</sub> concentration is essential to controlling air pollution, safeguarding public health, guiding urban planning decisions, and gaining insights into climate impacts. Previous studies have

investigated the prediction of PM<sub>2.5</sub> concentrations, with the majority relying on numeric or statistical learning models. Notably, methods based on DL are a cutting-edge and widely adopted facet of statistical learning. These methods are effective in addressing challenges that traditional models struggle with. The efficacy of DL in PM<sub>2.5</sub> prediction is attributed to the capacity of DL to handle extensive datasets, which is crucial to this type of prediction [2–5].

PM<sub>2.5</sub> time-series data encapsulate a dynamic functional relationship. DL is particularly adept at modeling such intricate connections and has exhibited remarkable performance across various time-series prediction tasks, positioning it as the preferred approach for tackling challenges in PM<sub>2.5</sub> concentration prediction. DL is adopted in PM<sub>2.5</sub> concentration predictions (i) as a core algorithm for prediction and (ii) to improve the performance of numerical simulation models. This review article provides a

\* Corresponding author. Chinese Research Academy of Environmental Sciences, China.

\*\* Corresponding author. Chinese Research Academy of Environmental Sciences, China.

E-mail addresses: [weiwang@craes.org.cn](mailto:weiwang@craes.org.cn) (W. Wang), [kangyulin@craes.org.cn](mailto:kangyulin@craes.org.cn) (Y. Kang).

comprehensive review of DL as the core algorithm for predicting the PM<sub>2.5</sub> concentration.

Several researchers have reviewed DL-based approaches for predicting PM<sub>2.5</sub> concentrations, providing insights from various perspectives. Ayturan et al. [6] surveyed DL techniques for air quality forecasting, covering convolutional neural networks (CNNs), long short-term memory (LSTM), and autoencoders. Despite reviewing only six articles, they laid the groundwork for exploring the potential of these techniques. Liao et al. [7] offered a concise overview of recent attempts at deep network architectures and their utility in capturing nonlinear spatiotemporal correlations of air pollution at various scales. Drewil et al. [8] conducted a comprehensive study of air pollution detection and prediction in smart cities, reviewing studies on DL techniques within the framework of smart cities. In addition, Istiana et al. [9] provided an in-depth review of LSTM networks for PM<sub>2.5</sub> concentration and prediction, delving into network architectures and potential applications. Zaini et al. [10] conducted a systematic review of DL techniques for time-series air quality forecasting, encompassing CNNs, LSTM, and hybrid models. Zhang et al. [11] reviewed DL architectures for air quality prediction. Their review offers a comprehensive overview and detailed discussions on the strengths, advantages, and limitations of various techniques. While highlighting challenges such as data scarcity and the need for explainable models, previous studies have lacked a detailed discussion on addressing these issues from more practical and holistic perspectives. Previous studies have only listed and summarized related methods without presenting a comprehensive evaluation or further analysis. No unified evaluation framework has been proposed to assess the quality of DL application in predicting PM<sub>2.5</sub> concentrations. In addition, there is a scarcity of reviews on the latest technologies for multi-model fusion forecasting.

In light of the aforementioned research landscape, this review article presents a comprehensive overview of cutting-edge deep-learning techniques for predicting PM<sub>2.5</sub> concentrations. Our review discusses the merits and limitations of a diverse array of DL models. We found three notable features. (i) A strict search strategy and large review depth: Our search strategy and inclusion/exclusion criteria are strict, transparent, and objective as they are based on PRISMA guidelines. In addition, a bibliometric analysis is conducted to provide a robust foundation for our comprehensive survey. (ii) A fine classification and summarization of existing DL frameworks: Unlike previous reviews, we take the initiative to classify and summarize DL frameworks used in air pollution prediction. Through meticulous categorization, we refine and define the use of hybrid models. Our survey outlines state-of-the-art methods for applying DL to various types of PM<sub>2.5</sub> concentration prediction, which will aid researchers and technicians in comprehending the current landscape of PM<sub>2.5</sub> concentration prediction. (iii) A proposal of a standard evaluation framework, Dataset-Method-Experiment Standard (DMES): The lack of standardization in applying DL to PM<sub>2.5</sub> concentration prediction hampers comparability. We address this gap by proposing a standard evaluation framework. Our work will encourage researchers to standardize their use of DL models in PM<sub>2.5</sub> concentration prediction and assist in measuring the quality of related research.

## 2. Bibliometrics analysis

### 2.1. Literature search and selected strategy

We conducted a literature review on DL-based PM<sub>2.5</sub> concentration prediction adopting the theory proposed by Kitcharoen

et al. [12], the methodology developed by Brereton et al. [13], and the PRISMA guidelines.

#### 2.1.1. Sourcing the articles

We searched scholarly databases, namely the Web of Science, Scopus, IEEE Xplore, and Springer, to identify peer-reviewed articles in well-known research journals and other academic publications. The literary investigation used the keywords “PM<sub>2.5</sub> prediction”, “air pollution estimation”, “air quality analysis”, “air pollutant concentration forecast”, “deep learning”, “convolutional neural networks”, and “artificial neural networks”. The keywords were merged adopting the search string [(“PM<sub>2.5</sub> prediction” OR “air pollution estimation” OR “air quality analysis” OR “air pollutant concentration forecast”) AND (“deep learning” OR “convolutional neural networks” OR “artificial neural networks”)]. The Boolean search operators (e.g., “OR,” “AND”) were utilized to integrate distinct keywords into a unified search string. Furthermore, internet search engines facilitated the collection of pertinent information regarding the advantages and disadvantages of DL-based methods in forecasting PM<sub>2.5</sub> concentrations. This approach resulted in the identification of over 1967 candidate research papers.

#### 2.1.2. Screening the articles

Recent articles on the most credible, authoritative, and reliable research having a worldwide scope were prioritized. The search process was repeated until relevant citations ended. In addition, the lists of references provided by articles were analyzed to identify other articles. Only articles published in the English language were selected. A total of 1967 articles were selected in the literature search. In the next step, we read the titles and abstracts of all articles and checked the quartile rankings of the journals. During this filtering process, 1509 papers were excluded from the selection list for the following reasons: (i) they did not pertain to the specified topic, (ii) they did not employ DL methodologies, and (iii) they were duplicative works, akin to other publications by the same authors. Furthermore, 77 papers were duplicates across different databases. Among the remaining 381 articles, 27 articles that were literature reviews instead of methodology papers were excluded. In the final phase of our selection process, we thoroughly reviewed the full texts of the articles. Subsequently, 236 full-text articles were excluded for not meeting our inclusion criteria. The reasons for exclusion were: (i) the articles were focused on a specific environment or location, such as a roadside or factory; (ii) the forecasts were derived from images; (iii) there was an insufficient dataset for effectively training a deep neural network; and (iv) DL techniques were applied for purposes other than air quality forecasting. Ultimately, following a meticulous evaluation, a total of 118 manuscripts were included for the qualitative and quantitative analysis of the review (Fig. 1). The articles were primarily published from 2016 to 2023, with 99 published from 2020 to 2023.

#### 2.1.3. Analyzing the selected articles

The selected articles were analyzed in depth to review the existing scenario of using DL-based techniques in PM<sub>2.5</sub> concentration prediction. The 118 papers selected for the literature review were analyzed concerning the year of publication, journal, and type of used DL-based model. We then exported the 118 articles in Research Information Systems (RIS) format to VOSviewer software (version 1.6.18). Bibliometric tools were used to extract information on the number and relationships of authors' publications and keywords. We performed cluster analysis for keywords from each year using VOSviewer software to generate social network maps. The social network maps indicate the importance of the size of

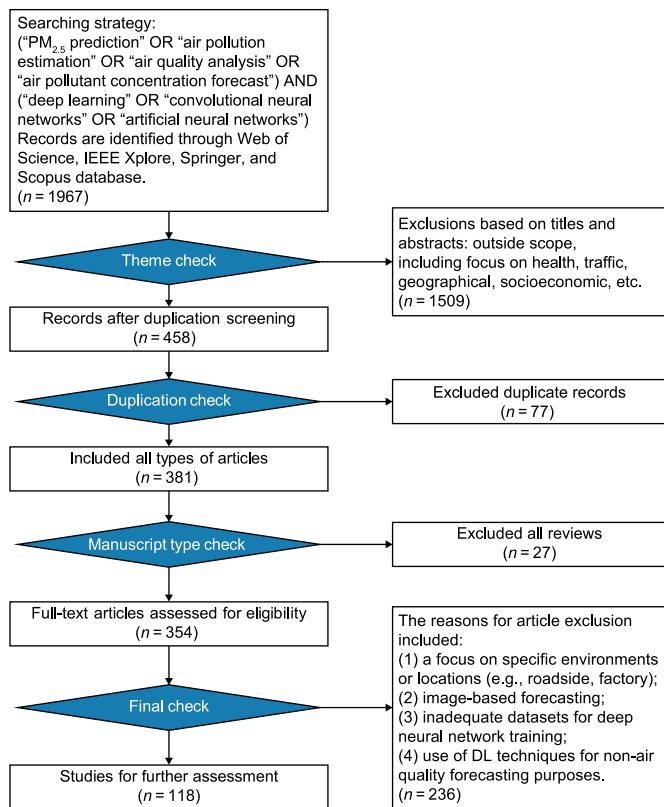


Fig. 1. Flowchart of the present review based on PRISMA guidelines.

nodes and the thickness of lines; i.e., the nodes represent the frequency of occurrence, and the lines represent associations between nodes [14]. Specifically, a thicker line indicates a stronger relationship [15]. Data aggregation and analysis were conducted in Microsoft Excel, and related figures were drawn with GraphPad Prism 9.4.0.

As shown in Fig. 2a, the authors of the articles are relatively independent of one another, which may explain the lack of unified standards in prediction and evaluation based on DL. Fig. 2b shows the keywords of the selected articles. There is no obvious clustering of keywords, indicating a lack of clarity in the current research lineage. Fig. 2c presents the JCR partitioning of the selected articles. This categorization indicates that the articles are of high quality, with more than half having been published in Science Citation Index (SCI) Q1. Fig. 2d classifies the articles according to the DL-related algorithm or model adopted in the articles. It is seen that most of the recent DL-based studies on PM<sub>2.5</sub> concentration prediction used algorithms relating to LSTM networks. In addition, many studies adopted hybrid model algorithms, especially in combination with conventional methods or other DL models.

### 3. Method review

#### 3.1. Evaluation metrics

Four mainstream indicators are used to evaluate the forecasting performance of the PM<sub>2.5</sub> concentration prediction models in our review as follows:

Rooted mean squared error (RMSE) is a commonly used metric for measuring the difference between the actual and predicted

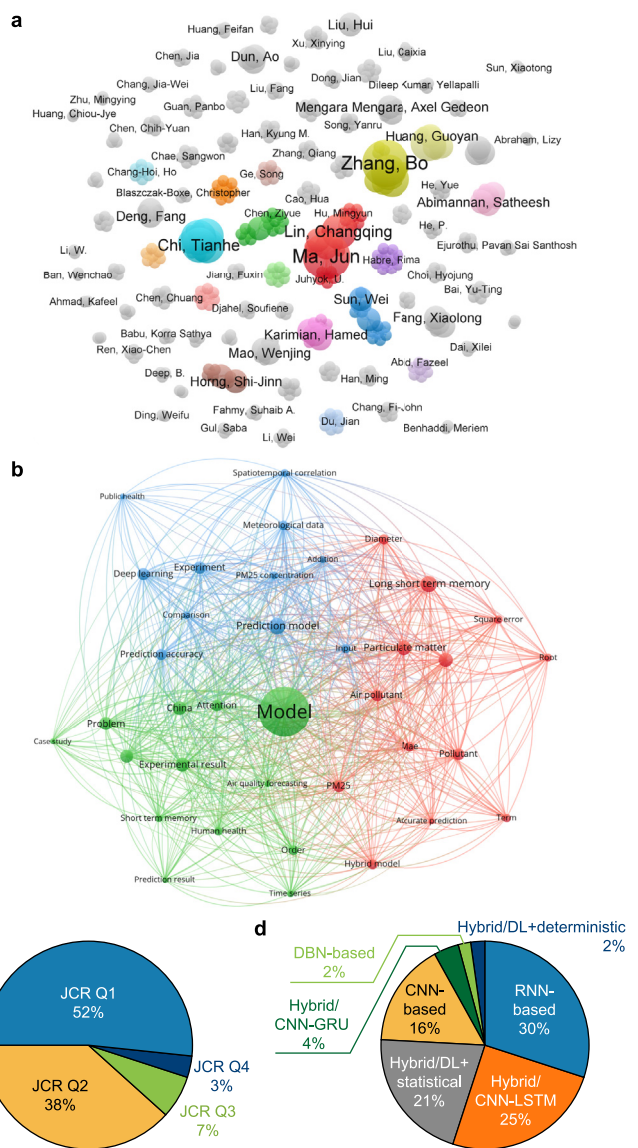


Fig. 2. The bibliometric analysis of the final selected articles. **a**, Relationship between the authors of the reviewed papers. The size of the displayed area represents the number of articles written by the author, the distance between authors represents the communication between them, and the color represents the degree of correlation obtained using a clustering algorithm. **b**, Keywords trend. The circle size represents the overall frequency of occurrence in the articles. A line segment is drawn between two words each time they appear simultaneously, and the number of line segments thus reflects the relationship between the words. High-frequency words that appear simultaneously are grouped and summarized through clustering, and the color indicates the word's frequency classification. **c**, Journal Citation Reports partitioning of selected papers, showing the percentage of articles in quartiles Q1, Q2, Q3, and Q4, respectively. **d**, Selected articles were classified according to their structures, showing the proportions of articles on various models.

values in regression analysis. It represents the square root of the average of the squared differences between the predicted and actual values and gives a sense of the magnitude of the errors in the predictions.

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

Mean absolute error (MAE) is another metric used in regression analysis to measure the average magnitude of the errors between the predicted and actual values. Unlike RMSE, MAE does not penalize large errors heavily and is a more robust measure of error for outliers.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2)$$

Mean absolute percentage error (MAPE) is a commonly used metric for measuring the accuracy of a prediction model in percentage terms. It measures the average percentage difference between the predicted and actual values and is often used in business forecasting and economic analysis.

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (3)$$

$R^2$  (coefficient of determination) is a measure of the proportion of variability in the dependent variable that is explained by the independent variables in a regression model. It ranges from 0 to 1, with higher values indicating a better fit between the predicted and actual values.  $R^2$  is often used to assess the goodness of fit of a regression model and can be interpreted as the percentage of the total variation in the dependent variable that is explained by the independent variables.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

In equations (1)–(4),  $y_i$  represents the actual value of the  $i$  th sample,  $\hat{y}_i$  represents the predicted value of the  $i$  th sample,  $\bar{y}$  represents the mean of  $y$ , and  $n$  is the total number of samples.

When different studies use various evaluation metrics that can be mutually converted, we standardize them, for example:

$$SMAPE = \frac{100\%}{n} \sum_{i=1}^n \frac{2|F_i - A_i|}{|F_i| + |A_i|} \quad (5)$$

where  $F_i$  represents the actual value of the  $i$  th sample,  $A_i$  represents the predicted value of the  $i$  th sample, and  $n$  is the total number of samples. It is evident that the SAMPE can be converted to MAPE using the following formula:

$$SMAPE = 2 \times MAPE \quad (6)$$

Therefore, in cases where the errors between the predicted values and actual values are small enough, for the purpose of facilitating result comparison, we unify the two metrics.

In order to ensure the unity and observability of the results, we conducted average operations on the results of some articles and did not adopt some indicators with low usage rates, like Recall, Precision, Information Gain Ratio (IA), etc.

### 3.2. Deep learning-based methods

DL is a subset of machine learning that uses neural networks with multiple layers to model complex patterns in data. From the

perspective of fulfilling PM<sub>2.5</sub> concentration prediction, some papers use deep belief networks (DBNs) or CNNs to extract spatial features from air quality data, while others use recurrent neural networks (RNNs) such as LSTMs or bidirectional long short-term memories (BiLSTMs) to capture temporal dependencies in the data. In terms of deep architecture, these DL-based forecasting strategies can be categorized into DBN-, CNN-, and RNN-based methods.

In this section, the most commonly used DL-based techniques (i.e., DBN, CNN, and RNN) in the field of PM<sub>2.5</sub> concentration prediction are briefly described. Moreover, particular attention is given to the core concepts and working of these techniques.

#### 3.2.1. Deep belief network-based methods

DBNs are a class of DL algorithms that are used for unsupervised learning tasks such as feature extraction, dimensionality reduction, and pattern recognition. They consist of multiple layers of latent variables, or hidden units, that are connected through probabilistic models. The foundation of DBNs is the idea of stacking multiple layers of restricted Boltzmann machines (RBMs), which are a type of generative stochastic artificial neural network. In a DBN, the first layer of hidden units is trained on the input data, and subsequent layers are trained on the output of the previous layer. This unsupervised learning approach allows the DBN to learn complex hierarchical representations of the input data, capturing both low- and high-level features. Once trained, a DBN can be fine-tuned for a specific supervised learning task, such as classification or regression, using backpropagation.

As shown in Table 1, H. Xing et al. [16] and Y. Xing et al. [17] both proposed a DBN approach for predicting PM<sub>2.5</sub> concentrations. The first research [16] highlighted the importance of considering environmental factors such as temperature in these models, while the second research [17] focused on using the grey wolf optimization (M-GWO) algorithm to search for the optimal solution.

The advantages of the above-mentioned DBN-based approaches lie in the DBN's capability to autonomously learn valuable features from raw data through unsupervised training. This is particularly relevant for PM<sub>2.5</sub> prediction, as future time points often lack labeled ground truth values, making unsupervised training a better fit for real-world applications. Additionally, the introduction of the M-GWO algorithm effectively optimizes the parameters of the DBN, leading to the efficient learning of dynamic relationships between PM<sub>2.5</sub> and meteorological variables, resulting in lower error rates compared to traditional statistical models.

However, DBN-based methods have their limitations. Unsupervised training requires significant computational resources and data, and the parameter iteration process can be time-consuming. Furthermore, the proposed approach only considers meteorological variables as inputs to the DBN without accounting for other potential factors that may influence PM<sub>2.5</sub> concentrations, such as traffic volume and industrial emissions. In summary, the DBN-based approach provides a novel method that combines DL and optimization algorithms for PM<sub>2.5</sub> concentration prediction and demonstrates promising results in initial studies. Nevertheless, further research is required to enhance the method's performance and incorporate other factors that could impact PM<sub>2.5</sub> concentrations. Exploring the interpretation of features learned by DBN is also a meaningful research direction.

#### 3.2.2. Convolutional neural network-based methods

One popular DL architecture used for air pollution prediction is CNN. A CNN is a type of neural network that uses convolutional layers to automatically learn and extract features from the input data. In air pollution prediction, the input data can be historical air quality data, meteorological data, or other relevant data. As shown



**Table 1**

The research used DBN-based methods.

Study	Year	Location	Model	Time step	RMSE ( $\mu\text{g m}^{-3}$ )	MAE ( $\mu\text{g m}^{-3}$ )	$R^2$
H.Xing et al. [16]	2021	Beijing, China	TDBN	D/S/T+1	11.197	12.298	0.862
Y.Xing et al. [17]	2019	Baoding, China	MGWO + DBN	-	20.260	17.604	0.884

**Table 2**

The research used CNN-based methods.

Study	Year	Location	Model	Time step	RMSE ( $\mu\text{g m}^{-3}$ )	MAE ( $\mu\text{g m}^{-3}$ )	MAPE (%)	$R^2$
Zhang et al. [33]	2023	Yangtze River Delta Region, China	STA-ResCNN	H/M/T+1	6.98	3.91	12.62	-
Yu et al. [34]	2023	Los Angeles, US	ST-Transformer	H/S/T+12	6.92	4.00	-	-
Choudhury et al. [26]	2022	Delhi, India	AGCTCN	H/M/T+(1-3)	11.76	8.75	-	0.64
Li et al. [18]	2016	Beijing, China	STD	S/S/-	14.96	9.00	21.75	-
Li et al. [19]	2020	California, US	ensemble-based DL	D/S/T+1	2.70	-	-	-
Samal et al. [20]	2021	Talcher, India	MTCAN	D/S/T+14	9.00	7.00	-	-
Luo et al. [19]	2020	Shanghai, China	CNN + GBM	H/S/T+1	10.02	-	-	0.85
Chae et al. [24]	2021	South Korea	ICNN	H/S/T+24	1.64	-	-	0.97
Zhang et al. [32]	2021	Beijing, China	ST-CausalConvNet	H/M/-	17.43	11.74	-	0.93
Shi et al. [25]	2021	Beijing, China	DSTP-FC(Encoder-Decoder)	H/M/T+(1-6)	32.51	19.50	-	-
Xiao et al. [27]	2022	-	DP-DDGCN	H/S/T+9	11.75	14.53	-	-
Zhao et al. [28]	2021	Jing-Jin-Ji Region, China	AQSTN-GCN	H/S/T+1	19.00	12.03	0.30	0.94
Wang et al. [36]	2022	China	STWC-DNN	H/S/T+1	12.70	-	-	0.92
Wang et al. [23]	2020	Shanghai, China	Sequence-to-Sequence	D/S/T+7	22.32	-	-	0.52
Ni et al. [22]	2022	Beijing/Tianjin, China	TL-DSTP-DANN	H/S/T+3	15.97	11.75	20.00	-
Dun et al. [35]	2022	Fushun, China	DGRA-STCN	H/S/T+2	12.50	8.21	88.40	-
Ouyang et al. [29]	2022	Beijing, China/London, UK	DC-STDGN	H/M/T+(1-3)	30.58/13.42/4.28	29.63/12.15/3.03	-	-
Ejurothu [31]	2022	New Delhi, India	HGNN	H/S/T+8	19.83	16.61	-	-
Dun et al. [30]	2022	Beijing/Fushun, China	DGC-MTCN	H/S/T+1	9.77/12.96	5.54/8.39	-	0.95/0.91

in Table 2, innovations for CNN-based methods in terms of feature selection, feature extraction, and feature fusion can generally be divided into three categories based on model structure: (i) improvements built upon the foundations of CNN/DNN, (ii) improvements based on graph convolutional neural networks (GCNNs), and (iii) improvements based on spatial-temporal (ST) correlations.

**CNN/DNN-based methods.** X. Li et al. [18] proposed a model that incorporates multiple features related to air quality, such as meteorological data and satellite imagery. The model performed well in predicting air quality for a large region, but the dataset used to train the model covered only a single year, which may limit the generalizability of the model to different years or seasons. L. Li et al. [19] presented an ensemble-based DL approach that combined the strengths of multiple models, including satellite imagery and meteorological data. The benefits of this approach are its ability to handle complex, multidimensional data and its performance in predicting  $\text{PM}_{2.5}$  concentrations during wildfire events. Samal et al. [20] utilized a multidirectional temporal convolutional neural network, allowing for both past and future temporal information to be incorporated into the predictions. The model performed well in predicting air quality, especially during high pollution events. However, the use of this model may be limited to the specific geographical region and environmental conditions considered in the study. Luo et al. [21] proposed an approach for  $\text{PM}_{2.5}$  concentration estimation using a CNN for feature extraction and a gradient boosting machine (GBM) for prediction. The benefits of this approach are its ability to handle nonlinear relationships between inputs and outputs and its performance in predicting  $\text{PM}_{2.5}$  concentrations with high accuracy. Ni et al. [22] proposed a model consisting of two parts: a feature extractor and a regression network. The feature extractor was pretrained on a large dataset and fine-tuned on the target dataset, while the regression network was trained from scratch on the target dataset. The use of transfer learning can help to reduce the amount of data needed for training and improve prediction accuracy. Wang et al. [23] utilized a CNN as

the encoder and an RNN as the decoder to capture the spatial-temporal patterns of air pollution. However, focusing only on roadside air quality forecasting may cause the model to not be applicable to other locations. In addition to these models, Chae et al. [24] proposed an interpolated CNN model for real-time prediction, and Shi et al. [25] combined attention mechanisms for feature selection in CNN architecture. Due to the complexity of CNN structures, these models are both limited by the lack of interpretability and high computational cost in training.

#### **Graph convolutional neural network (GCNN)-based methods.**

In previous studies [26–31], various studies have proposed GCNN models for  $\text{PM}_{2.5}$  prediction. Choudhury et al. [26] proposed an attention-enhanced hybrid model that combines a GCNN with an attention mechanism to capture spatial and temporal dependencies in PM concentration data. The proposed model can effectively capture both spatial and temporal dependencies in the PM concentration data, and the attention mechanism allows the model to focus on the most relevant features in the input data for accurate prediction. Xiao et al. [27] proposed a dual-path dynamic directed graph convolutional network (DP-DDGCN) that included a spatial path and a temporal path to capture both spatial and temporal correlations in air quality data. The dual-path architecture allows the model to effectively integrate spatial and temporal information for accurate prediction. Ouyang et al. [29] proposed a dual-channel spatial-temporal difference graph neural network (DSTGNN) that also accounts for both spatial and temporal dependencies among  $\text{PM}_{2.5}$  concentration data. Dun et al. [30] proposed a dynamic DGCNN that can update the graph structure based on spatial-temporal correlations among air pollutant concentrations. The proposed DGCNN can effectively capture both spatial and temporal dependencies and dynamically update the graph structure to capture changes in spatial-temporal correlations over time. However, neither of these articles provides a detailed analysis of the interpretability of the proposed model or the features it learns from the  $\text{PM}_{2.5}$  concentration data. Zhao et al. [28] proposed a near-surface  $\text{PM}_{2.5}$  prediction model that combined a complex network

characterization method with a GCNN. The model used complex network analysis to identify the most important air quality monitoring stations for prediction and then applied a GCNN to capture spatial and temporal correlations among  $PM_{2.5}$  concentrations. The proposed model can effectively identify the most important monitoring stations for accurate prediction. Ejurothu et al. [31] proposed a cluster-based hybrid graph neural network (CGNN) approach for  $PM_{2.5}$  concentration prediction in India. In this model, cluster analysis to group similar monitoring stations and a GCNN were combined to capture spatial and temporal dependencies among the  $PM_{2.5}$  concentration data. Cluster analysis can group similar monitoring stations to improve prediction accuracy. Both of these models can effectively capture both spatial and temporal dependencies in the  $PM_{2.5}$  concentration data. However, these two studies considered only  $PM_{2.5}$  concentration data from a single city, which may limit the generalizability of the results to other regions or cities.

**Spatial-temporal (ST) correlations-based methods.** Zhang et al. [32], Zhang et al. [33], Yu et al. [34], Wang et al., and Dun et al. [35] all incorporated spatial and temporal information to accurately predict  $PM_{2.5}$  concentrations. Zhang et al. [32] proposed an ST-CausalConvNet that utilized spatial-temporal causal convolutional layers, Zhang et al. [33] applied spatial-temporal attention and residual learning in a multistep forecasting framework, and Yu et al. [34] proposed a spatiotemporal transformer model that allowed for capturing long-term dependencies and relationships. Dun et al. [35] and Wang et al. [36] captured the spatial and temporal features via CNN or DNN architecture. In summary, the strengths of these models are their ability to incorporate both spatial and temporal information to make accurate predictions of  $PM_{2.5}$  concentrations or air quality. However, most of these models are limited to only short-term predictions, and some are limited to specific regions or pollutants. Additionally, some models may be computationally expensive due to their complex architectures. Further research is needed to develop more efficient and accurate models that can be applied on a larger scale.

### 3.2.3. Recurrent neural network-based methods

Another type of DL architecture used for air pollution prediction is the RNN. RNNs are useful for modeling sequential data, such as time series data. Air pollution data often have a temporal nature, and RNNs can capture temporal dependencies in the data to make accurate predictions. The main characteristic of RNNs is that they have a feedback loop that allows them to maintain a “memory” of previous inputs as they process new inputs in a sequence. Through feedback loops, RNNs can remember historical information and pass it to the current time step, allowing the model to consider past context and better comprehend the current data point. Based on this, the model can better capture temporal relationships, making it particularly well-suited for tasks such as language modeling, speech recognition, and time series prediction. Dai et al. [37] developed a RNN for predicting indoor  $PM_{2.5}$  concentrations in residential buildings using historical data. The model was trained and tested using data collected from sensors installed in multiple apartments in a high-rise building in Beijing, China. The results showed that the proposed RNN model can accurately predict indoor  $PM_{2.5}$  concentrations up to 6 h in advance. The proposed RNN model accounts for both the temporal and spatial dependencies of the indoor  $PM_{2.5}$  data, which is important for accurate prediction in residential buildings. The use of real-world data collected from multiple sensors in a high-rise building in Beijing, China, makes the results highly relevant and applicable to similar indoor air quality prediction scenarios. Ayturan et al. [38] proposed an RNN-based model that utilized multiple input data and contained historical  $PM_{2.5}$  concentrations, meteorological data, and air quality index

(AQI) data. Both of these studies provided a detailed analysis of the interpretability of the proposed RNN model. However, Dai et al. [37] considered only indoor  $PM_{2.5}$  concentrations in a single high-rise building in Beijing, which may limit the generalizability of the results to other cities or types of buildings. Neither of these studies compares the performance of the proposed RNN model with other benchmark models for  $PM_{2.5}$  predictions. LSTM is a type of RNN. The gated recurrent unit (GRU) is a simplified version of LSTM that also uses gating mechanisms to control the information flow. LSTMs are a special type of RNN that are designed to address some limitations of traditional RNNs, such as vanishing gradients. LSTMs use a more complex architecture that includes “gates” to control the flow of information through the network, allowing it to selectively forget or remember information from previous time steps. This allows LSTMs to learn long-term dependencies in data and make accurate predictions even over long sequences.

Li et al. [39], Chang et al. [40], Xayasouk et al. [41], Karimian et al. [42], Mao et al. [43], Qadeer et al. [44], Kristiani et al. [45], Lin et al. [46], Park et al. [47], Peralta et al. [48], Waseem et al. [49], and Gul et al. [50] all directly used the LSTM model via partial fine-tuning of the structure or parameters. Using LSTM neural networks allowed for the modeling of temporal dependencies, and the large number of experiments at different real-world air quality monitoring datasets from multiple stations helped increase the generalizability of the results. However, most of these studies did not provide a detailed analysis of the interpretability of the LSTM model or the features it learned from air pollutant concentration data, which are both important for DL-based model prediction. Ma et al. [51], Tong et al. [52], Zhang et al. [53], and Deep et al. [54] used the BiLSTM model, which differed from regular LSTMs by having two separate hidden layers for processing the sequence in the forward and backward directions to handle both spatial and temporal correlations in the data and model complex nonlinear relationships between air quality parameters and meteorological factors. Mengara et al. [55,56] integrated CNN and AutoEncoder (AE) with BiLSTM in 2020 and 2022, respectively, allowing it to extract more meaningful and representative features. Xu et al. [57], Zhang et al. [58], and Zou et al. [59] developed AE-based LSTM neural networks. The models used both supervised and unsupervised learning to extract features from air quality data. However, these models were weakened by their complexity, which may lead to overfitting and longer training times. The prediction neural network always includes four portions: data preprocessing, feature extraction, model training, and evaluation. For data preprocessing, Shi et al. [60] proposed a balanced sampling approach to address imbalanced data to help  $PM_{2.5}$  concentration prediction. Most improvements based on the LSTM structure are within feature extraction. Ma et al. [61] incorporated the concept of lagged variables into the model. Ding et al. [62] used a combination of principal component analysis (PCA), an attention mechanism, and LSTM. Furthermore, Hu et al. [63] inset a one-dimensional layer to capture both local and global dependencies, while Wang et al. [64] used convolutional and recurrent layers. Both of these strategies help to extract more suitable features. For network structure innovations, Zhao et al. [65] proposed a fully connected LSTM neural network. Using a fully connected network allowed for the modeling of both temporal and spatial dependencies. Wen et al. [66] proposed a spatiotemporal convolutional LSTM neural network. The model used both spatial and temporal information in air quality data. Zhou et al. [67] combined multiple data sources, including satellite images, meteorological data, and air quality data, for LSTM-based  $PM_{2.5}$  prediction. The model used a feature fusion module to integrate different data sources and showed higher performance than other models. Sun et al. [68] proposed a deep residual learning framework for air quality prediction that used residual connections to improve the

flow of information. Wu et al. [69] proposed a novel DL model that combined an attention-based GRU and a convolutional encoder with adaptive gated activation (CE-AGA) for air quality prediction. Transfer learning was used in some papers to leverage pre-trained models for related tasks and improve the performance of air quality prediction models. Ma et al. [70] proposed a transfer learning approach for air quality prediction, where a pre-trained DL model on a location was fine-tuned for a different specific location.

The model performed better than models trained from scratch on the target location. Xiao et al. [72] introduce a weighted LSTM extended (WLSTME) model, specially designed to significantly enhance the accuracy and reliability of daily PM<sub>2.5</sub> concentration predictions. This innovative model leverages the power of artificial neural networks and cutting-edge techniques to improve forecasting capabilities. The article explores the key features and advantages of the WLSTME model, emphasizing its ability to capture intricate patterns and dependencies in time-series data. All results are listed in Table 3. However, as with most DL models, these methods may not be easily interpretable, making it difficult to understand the reasoning behind the predictions of the model. The performance of DL models may degrade when applied to data from new locations or with different characteristics than the training data.

### 3.2.4. Transformer-based methods

The Transformer [73] model is a DL architecture characterized

by its self-attention mechanism, originally developed for natural language processing tasks. It relies on self-attention mechanisms to effectively capture relationships between different positions in input sequences irrespective of the distance between them and without the need for recurrent or convolutional structures. The self-attention mechanism allows the model to consider all positions in the input sequence simultaneously and enables the model to capture dependencies between different positions in the input sequence without being constrained by distance, making it suitable for handling long-range dependencies. Transformers typically consist of an encoder and a decoder, which are used for encoding the input sequence and generating the output sequence, respectively. To differentiate between elements at different positions in the input sequence, transformers introduce positional embeddings, providing the model with information about the position of each element in the sequence. The Transformer model has seen significant success in the field of natural language processing, such as in machine translation, text generation, and sentiment analysis. Its capabilities have made it a foundational architecture in DL.

Based on the Transformer architecture, Informer [74] is specifically designed for time series forecasting. It combines ideas from the transformer architecture and optimizes them for time series data, emphasizes the importance of time steps in a sequence, and introduces time embeddings to capture time-related information. Informer proposed a prob-sparse self-attention mechanism, which can sort the most important query, significantly lowering the

**Table 3**  
The research used RNN-based methods.

Study	Year	Location	Model	Time step	RMSE ( $\mu\text{g m}^{-3}$ )	MAE ( $\mu\text{g m}^{-3}$ )	MAPE (%)	R <sup>2</sup>
Ayturan et al. [38]	2018	Ankara, Turkey	RNN	H/S/T+1	6.28	4.21	-	-
Dai et al. [37]	2021	Tianjin, China	RNN	H/-/T+1	11.87	-	-	-
Ma et al. [51]	2019	Guangdong, China	BiLSTM	H/S/T+1	8.24	4.80	9.01	-
Li et al. [39]	2017	Beijing, China	LSTM	H/S/T+1	12.60	5.46	11.93	-
Zhao et al. [65]	2019	Beijing, China	LSTM-FC	H/M/T+(1-6)	35.82	23.97	-	-
Wen et al. [66]	2019	Beijing, China	STCLSTM	H/M/T+1	12.08	5.82	17.09	-
Zhou et al. [67]	2019	Taiwan, China	DM-LSTM	H/S/T+1	4.49	-	-	-
				H/S/T+4	9.31	-	-	-
Ma et al. [70]	2019	Guangdong, China	TL-BLSTM	H/S/T+1	8.54	4.95	22.32	-
Chang et al. [40]	2020	Taiwan, China	LSTM	H/S/T+1	-	-	-	-
Xayasouk et al. [41]	2020	Seoul, South Korea	LSTM	H/S/T+1	11.11	-	-	-
Karimian et al. [42]	2019	Tehran, Iran	LSTM	H/S/T+12	10.32	7.41	-	0.74
Tong et al. [52]	2019	Florida, US	BiLSTM	H/S/T+1	3.65	1.62	18.48	-
Mao et al. [43]	2021	Jing-Jin-Ji Region, China	LSTM	H/M/T+(1-24)	20.68	14.56	-	0.74
Ma et al. [61]	2020	Wayne, US	Lag-LSTM	H/S/T+1	3.48	1.85	25.63	-
Zhang et al. [58]	2020	Beijing, China	AE + BiLSTM	H/S/T+24	2.19	-	-	-
Zou et al. [59]	2021	Yangtze River Delta Region, China	FDN (AE + LSTM)	H/S/T+1	4.32	3.31	-	-
Xu et al. [57]	2021	Beijing, China	AE + LSTM	H/S/T+1	14.52	8.22	45.40	-
				H/M/T+(1-3)	24.87	15.60	64.72	-
Qadeer et al. [44]	2020	Seoul, South Korea	LSTM	H/-/-	4.82	3.58	-	0.87
Zhang et al. [53]	2021	Beijing, China	BiLSTM	H/-/T+1	17.20	14.15	-	-
Wang et al. [64]	2021	Beijing, China	CR-LSTM	H/S/T+24	8.96	12.89	-	0.74
Shi et al. [60]	2022	Beijing, China	BS-LSTM	H/S/T+3	32.32/12.42	18.36/9.75	-	-
Kristiani et al. [45]	2022	-	LSTM	H/S/T+1	1.90	1.27	11.12	-
Deep et al. [54]	2022	Delhi, India	BiLSTM	H/S/T+1	15.59	-	-	-
Sun et al. [68]	2019	Liaoning, China	LSTM-DRSL	H/S/T+1	10.53	9.09	20.05	-
Lin et al. [46]	2020	Taiwan, China	LSTM	H/S/T+1	4.46	-	30.00	0.86
Park et al. [47]	2021	Seoul, South Korea	LSTM	H/S/T+3	-	-	-	-
Mengara et al. [56]	2022	Seoul, South Korea	AE + BiLSTM	H/S/T+1	7.48	5.02	30.48	-
Mengara et al. [55]	2020	Busan, South Korea	CNN + BiLSTM	H/S/T+1	6.93	5.07	30.90	-
Ding et al. [62]	2022	Ningxia, China	PCA-Attention-LSTM	D/S/T+1	7.57	4.93	-	0.91
Peralta et al. [48]	2022	Santiago, Chile	LSTM	H/S/T+1	9.85	4.40	-	0.74
Liu et al. [71]	2022	Jing-Jin-Ji Region, China	MGC-LSTM	H/S/T+1	2.91	2.16	12.96	-
Hu et al. [63]	2022	Beijing, China	Conv1D-LSTM	H/S/T+1	20.76	11.20	-	0.96
Wu et al. [69]	2022	Beijing, China	CE-AGA-LSTM	H/S/T+1	21.88	14.49	-	0.95
Waseem et al. [49]	2022	Lahore/Karachi/Islamabada, Pakistan	LSTM	H/S/T+1	-	-	11.70/7.40/9.50	-
				D/S/T+1	-	-	28.2/42.1/15.1	-
Gul et al. [50]	2022	Punjab, India	LSTM	H/S/T+1	0.19	-	-	-
				H/S/T+(1-24)	0.73	-	-	-
Xiao et al. [72]	2020	Jing-Jin-Ji Region, China	WLSTME	D/S/T+1	40.67	26.10	-	-

complexity. Informer can work in a sequence-to-sequence fashion, where it takes an input sequence of historical data and generates an output sequence of future predictions. This demonstrates strong performance in time series forecasting tasks, especially for weather forecasting.

M.A.A.A.-q et al. [75] proposed a ResInformer built upon the ideas from the above architecture. The novelty of this work lies in the incorporation of residual connections, a common feature in many deep neural networks. The residual connections allow the model to skip one or more layers, helping to mitigate the vanishing gradient problem and facilitating the training of very deep networks. In this article, a series of ablation experiments were conducted, and the proposed ResInformer achieved the best test results, demonstrating its superiority among the transformer-based methods. The comparison results are presented in Table 4.

In summary, the Transformer, as the latest DL architecture, holds the promise of significantly advancing PM<sub>2.5</sub> prediction methods based on DL. Its capabilities to capture intricate temporal and spatial patterns, effectively manage sequences, and adapt to diverse data types have the potential to enhance the accuracy and robustness of PM<sub>2.5</sub> forecasts. Moreover, its scalability and parallel processing capabilities position it as a fitting choice for handling the substantial datasets often encountered in air quality prediction.

### 3.3. Deep-learning based hybrid methods

As research progresses and massive data and complex learning objectives increase, simple models are becoming insufficient for supporting various conditions. In addition to simple DL-based predictors, other components have also been combined to form hybrid predictive models. These approaches can be divided into two structural categories: DL-based models combined with conventional methods and combinations of DL-based models. DL-based hybrid models can combine statistical and machine learning models with deep neural networks to exploit the strengths of each approach. For example, a hybrid model might use a statistical model to model the seasonal air pollution patterns, a machine learning model to model the relationships between meteorological variables and pollution levels, and a deep neural network to model the nonlinear interactions between different variables. One advantage of DL-based hybrid models is their ability to model highly nonlinear relationships and interactions between different variables, which can be difficult for other models to capture. DL-based models can also learn representations of data that are more compact and expressive than traditional feature engineering approaches, which can improve the accuracy of predictions.

#### 3.3.1. Deep learning combines with conventional methods

Models combining DL with conventional methods can be

divided into two categories: DL plus deterministic methods and DL plus statistical methods.

**Deep learning plus deterministic methods.** Deterministic methods adopt meteorological principles and mathematical equations to simulate the process of pollutant emission, transformation, diffusion, and removal based on atmospheric physical and chemical reactions [39]. For example, weather research and forecasting (WRF) models [78] are used for atmospheric research and prediction application, and other deterministic methods, such as community multiscale air quality (CMAQ), are also applied to air pollution prediction. As shown in Table 5, Chang-Hoi et al. [76] combined the deterministic CMAQ model with RNN for PM<sub>2.5</sub> concentration prediction. The model accounted for various meteorological and environmental factors and incorporated them into the prediction process. Sun et al. [77] proposed a hybrid approach that combined the numerical simulation model WRF and CMAQ with DL techniques LSTM for PM<sub>2.5</sub> and O<sub>3</sub> forecasting. The simulation model provided a spatial distribution of PM<sub>2.5</sub> and O<sub>3</sub> concentrations, while DL techniques were used to predict the temporal variation in these concentrations. In this study, the hybrid approach outperformed both the simulation and DL models alone, indicating the potential benefits of combining different techniques for air quality forecasting.

**Deep learning plus statistical methods.** As shown in Table 6, statistical methods are well known to researchers because these researchers avoid sophisticated theoretical models and simply apply statistics-based models, which have gradually emerged in air pollution prediction [79]. These methods can also be classified into two categories: classic statistical methods and traditional machine learning methods. Classic statistical methods are those based on auto regression integrated moving average (ARIMA), wavelet transform (WT), or empirical mode decomposition (EMD). In contrast, traditional machine learning methods usually use random forest (RF), support vector machine-based regression (SVR), gradient boosting decision tree (GBDT), etc. These methods can capture nonlinear features from raw data to a certain extent, but they cannot fully extract complex spatiotemporal correlations in historical data (Table 7).

**Combined with classic methods.** The WT is a classic mathematical tool used to analyze signals and data. It breaks a signal or data down into smaller components called wavelets, which are basically small wave-like functions of varying frequencies and durations. These wavelets are then used to represent the original signals or data in a form that can be more easily analyzed. The WT is particularly useful for analyzing signals with nonstationary properties, such as signals that change over time or have sudden bursts of activity. The wavelet transform can capture these changes more accurately than traditional signal analysis techniques. Therefore, some studies have been devoted to incorporating the wavelet

**Table 4**  
The research used Transformer-based methods.

Study	Year	Location	Model	Time step	RMSE ( $\mu\text{g m}^{-3}$ )	MAE ( $\mu\text{g m}^{-3}$ )	MAPE (%)	R <sup>2</sup>
Zhou, H et al. [73]	2021	Beijing, China	Informer	D/S/T+1	0.2852	0.2159	0.80	0.8285
		Shijiazhuang, China		D/S/T+1	0.5112	0.2890	1.79	0.6433
		Wuhan, China		D/S/T+1	0.5225	0.4180	1.44	0.5329
	2021	Beijing, China	InformerStack	D/S/T+1	0.2692	0.2012	0.89	0.8472
		Shijiazhuang, China		D/S/T+1	0.5408	0.3081	2.25	0.6020
		Wuhan, China		D/S/T+1	0.3716	0.2911	1.49	0.7621
M.A.A.A.-q et al. [74]	2023	Beijing, China	ResInformer	D/S/T+1	0.2822	0.2130	0.85	0.8320
		Shijiazhuang, China		D/S/T+1	0.4646	0.3138	2.00	0.5857
		Wuhan, China		D/S/T+1	0.4706	0.3782	1.54	0.6142
	2023	Beijing, China	ResInformerStack	D/S/T+1	0.2623	0.1964	0.75	0.8549
		Shijiazhuang, China		D/S/T+1	0.5343	0.3055	1.91	0.4937
		Wuhan, China		D/S/T+1	0.3712	0.2982	1.39	0.7656



**Table 5**  
The research used DL plus deterministic methods.

Study	Year	Location	Model	Time step	RMSE ( $\mu\text{g m}^{-3}$ )	MAE ( $\mu\text{g m}^{-3}$ )	MAPE (%)	R <sup>2</sup>
Chang-Hoi et al. [76]	2021	South Korea	RNN-CMAQ	H/S/T+6	-	5.10	-	0.862
Sun et al. [77]	2021	China	LSTM-WRF-CMAQ	H/S/T+48	11.03	8.08	68.83	0.884

**Table 6**  
The description of statistical methods.

Category	Model	Details
Classic method	ARIMA	Autoregressive integrated moving average model
	WT	Wavelet transform
	EWT	Empirical wavelet transforms
	EMD	Empirical mode decomposition
	EEMD	Ensemble empirical mode decomposition
	CEEMD	Complementary ensemble empirical mode decomposition
Machine learning	VMD	Variational mode decomposition
	RF	Random forest
	SVR	Support vector machine-based regression
	GBDT	Gradient boosting decision tree
	ANN	Artificial neural networks
	PSO	Particle swarm optimization
	DBSCAN	Density-based spatial clustering of applications with noise
	FE	Fuzzy entropy
	Kalman-filter	Kalman filter
	GWO	Grey wolf optimizer
	mRMR	Max-relevance and min-redundancy
Q	Q-learning	

transform into DL algorithms. Qiao et al. [80] combined the wavelet transform with an SAE and LSTM hybrid model. The wavelet transform decomposes the PM<sub>2.5</sub> time series into multiple subseries, and the DL algorithm is used to model and predict each subseries. Benhaddi et al. [81] combined the wavelet transform with a CNN-based model for multivariate time series forecasting in urban air quality prediction. The model used dilated residual convolutional neural networks (DRCNs) to capture temporal dependencies and interactions among multiple air quality indicators. In addition, Kim et al. [82] integrated clustering, feature selection, and empirical wavelet transform (EWT) into an LSTM-based framework to capture the spatiotemporal dynamics of air pollutant concentrations. The clustering method was used to identify the spatial clusters of air quality monitoring stations, and the EWT was designed to decompose the time series data into different frequency components. Then, a DL model was developed to predict the air pollutant concentration using the selected features. Both the proposed hybrid models outperformed other models regarding forecasting accuracy and efficiency.

EMD (empirical mode decomposition), EEMD (ensemble empirical mode decomposition), and CEEMD (complete ensemble empirical mode decomposition with adaptive noise) are three signal decomposition algorithms used within the EMD method or

**Table 7**  
The research used DL plus statistical methods.

Study	Year	Location	Model	Time step	RMSE ( $\mu\text{g m}^{-3}$ )	MAE ( $\mu\text{g m}^{-3}$ )	MAPE (%)	R <sup>2</sup>
Qiao et al. [80]	2019	China	WT-SAE-LSTM	D/S/T+1	-	3.88	-	-
Huang et al. [83]	2021	Beijing, China	EMD-GRU	H/M/T+(1-4)	11.37	6.53	25.79	0.98
Jin et al. [84]	2020	Beijing, China	EMD-CNN-GRU	H/M/T+(1-24)	42.26	34.95	65.30	0.67
Zaini et al. [85]	2022	Cheras/Batu Muda, Malaysia	EEMD-LSTM	H/S/T+1	4.21/4.89	2.81/2.77	14.15/14.64	0.97/0.96
Zhang et al. [86]	2021	Beijing, China	VMD-BiLSTM	H/S/T+1	9.39	5.35	16.40	0.99
Chang et al. [87]	2020	Taiwan, China	GBDT-SVR-LSTM	H/S/T+1	7.67	5.00	-	-
Liu et al. [88]	2021	Changsha, China	GCN-LSTM-GRU-Q	M/-/-	17.63	14.24	2.91	-
Liu et al. [89]	2020	Shanghai, China	CEEMD-LSTM	H/S/T+3	3.28	2.23	5.74	0.99
Jiang et al. [90]	2021	Beijing, China	CEEMD + DeepTCN	H/S/T+1	1.11	0.65	2.65	-
Kim et al. [82]	2021	Beijing, China	FC-DTWD-EWT-CBLSTM	H/S/T+1	2.29	1.51	4.03	0.94
				H/S/T+10	5.17	3.37	8.98	0.85
Lu et al. [91]	2021	Yangtze River Delta Region, China	DBSCAN-DNN	H/S/T+1	13.29	-	0.90	-
Teng et al. [92]	2022	Shanghai, China	EMD-SE-BiLSTM	H/S/T+1	2.77	1.88	-	0.98
				H/S/T+3	5.04	3.56	-	0.95
Fu et al. [93]	2021	Hangzhou, China	CEEMD-LSTM	H/S/T+1	6.48	4.76	15.76	-
Zhang et al. [94]	2020	Gansu, China	ESN-PSO	H/S/T+1	8.73	5.47	8.20	0.93
Wang et al. [95]	2022	China	LSTM-RF-PSO	H/S/T+1	4.93	2.91	24.36	-
Wang et al. [96]	2022	-	CEEMD-FE-mRMR-GWO-LSTM	H/-/-	8.26	6.60	19.77	0.95
Wei Sun et al. [97]	2022	Jing-Jin-Ji Region, China	LSTM-CEEMADN	D/S/T+1	3.52	2.73	-	0.97
Xu et al. [98]	2022	China	CEEMD-CNN-LSTM	H/S/T+2	12.67	9.60	-	0.87
Zhou et al. [99]	2022	Chongqing, China	Kalman-Filter-LSTM	H/S/T+1	8.45	7.30	-	0.96
Zhao et al. [100]	2022	Beijing/Guangzhou, China	RF-BiLSTM	H/S/T+1	7.26/1.77	3.73/1.33	-	1.00/0.99
		Xi'an/Shenyang, China			2.75/4.51	1.37/2.20	-	1.00
Zhang et al. [101]	2023	Pingqiao/South bay/Brewing, China	CEEMD-FCN-LSTM	-	3.81/5.39/4.02	2.47/2.81/4.55	4.48/6.48/2.37	0.98/0.97/0.98
Masood et al. [102]	2023	Delhi, India	ANN	-	24.12	-	-	0.94
Liu et al. [103]	2022	Shenyang/Changsha/Shenzhen, China	VMD-LSTM-ESN-TCN-GBDT	-	1.98/2.20/1.68	1.58/1.71/1.32	3.95/4.11/4.53	-
Benhaddi et al. [81]	2021	Marrakesh, Morocco	WT-CNN	H/-/-	0.01	-	99.10	-
Ban et al. [104]	2022	Hangzhou, China	CEEMD-LSTM-BP-ARIMA	D/S/T+1	4.55	3.66	-	0.79
M.A.A.A.-q et al. [105]	2021	Wuhan, China	PSO-SMA-ANFIS	H/S/T+1	22.39	17.50	16.83	0.51

its variants. EMD is a fully adaptive signal decomposition algorithm, and EEMD is an improved version of EMD that introduces randomness into the decomposition process to improve robustness and reduce mode mixing. CEEMD is a further improvement of EEMD that incorporates adaptive noise reduction technology into the decomposition process. In addition, variational mode decomposition (VMD) is a signal decomposition method commonly used in time series analysis. The benefits of these signal decomposition algorithms are that they are data-driven and do not require any a priori knowledge of the frequency or time scale of the signal. Using EMD and VMD allows the model to capture the different frequency components of the input data, which is useful for forecasting PM<sub>2.5</sub> concentrations that exhibit complex and nonlinear relationships with their influencing factors. Therefore, some researchers take it as a priority of DL methods, which can help the training process. Huang et al. [83], Jin et al. [84], and Teng et al. [92] proposed a hybrid model that combined EMD with GRU, CNN, and BiLSTM, respectively. Zaini et al. [85] applied EEMD to decompose the input data for LSTM, while Liu et al. [89], Fu et al. [93], and Sun et al. [97] utilized CEEMD for LSTM input decomposition. Jiang et al. [90] combined the CEEMD with deep-TCN. Zhang et al. [86] used VMD and BiLSTM for PM<sub>2.5</sub> concentration prediction. These models can capture nonlinear relationships between variables using EMD or VMD. However, the model's performance heavily depends on the decomposition quality, which can be affected by noise and other factors.

Furthermore, hybrid models based on the combination of signal processing and DL algorithms improve the model's performance in more ways. Xu et al. [98] and Zhang et al. [101] proposed improved CEEMD-LSTM models combined with CNN and FCN, respectively. The convolutional layers helped in feature selection to enhance prediction accuracy. Wang et al. [96] proposed a hybrid model based on CEEMD-LSTM and optimized their weights using multiple machine-learning algorithms (FE, mRMR, and GWO). Ban et al. [104] incorporated CEED, BP, LSTM, and ARIMA to build a comprehensive hybrid framework that considered multiple factors and scaled for air pollutant prediction and early warning. Liu et al. [103] proposed an enhanced hybrid model that combined multiple DL models (LSTM, ESN, and TCN) with statistical algorithms (VMD and GBDT). The ESN and TCN were extracted features, while LSTM was used to predict the PM<sub>2.5</sub> concentration. The VMD and GBDT can enhance the model's performance by optimizing their weights. The strengths of this model include the use of multiple DL and statistical techniques, which can capture complex temporal and spatiotemporal patterns in the data. However, one potential weakness of the model is its reliance on meteorological data, which may limit its accuracy in areas with sparse meteorological monitoring stations.

**Methods combined with machine learning.** Machine learning approaches often require more meticulously designed features, which are typically manually annotated to help increase model interpretability, while DL approaches can automatically learn features without the need for manual design, giving them an advantage in handling large-scale, high-dimensional data but lacking model interpretability. By combining these approaches, we can exploit both the feature engineering interpretability of machine learning and the powerful feature-learning abilities of DL. Lu et al. [91] combined the DNN and DBSCAN clustering algorithms to improve the PM<sub>2.5</sub> concentration forecasting accuracy. Incorporating DBSCAN showcases strengths in handling spatial clustering and outlier detection, contributing to the enhanced accuracy of PM<sub>2.5</sub> concentration forecasting. Chang et al. [87] proposed an ensemble learning-based hybrid model that integrated multiple machine learning algorithms (GBDT and SVR) to improve the performance of the LSTM model. The GBDT and SVR provide insights

into feature importance, aiding in the identification of key variables that significantly contribute to the predictive power of the LSTM model. Masood et al. [102] provided a data-driven predictive modeling approach based on ANN, and Zhao et al. [100] proposed a forecasting model for fine particulate matter concentrations using RF and BiLSTM. The fusion of LSTM and RF could enhance the model's ability to model complex data, capitalizing on LSTM's long short-term memory and the decision tree advantages of RF. Zhou et al. [99] integrated Kalman filtering, an attention mechanism, and an LSTM neural network. The inclusion of Kalman filtering serves to reduce the noise of input data and handle the missing data. This integration enhances the robustness of the algorithm. Zhang et al. [94] proposed a novel combined model based on an echo state network (ESN) and PSO, while Wang et al. [95] proposed a model based on LSTM, RF, and PSO. Using PSO helps the model quickly converge to an optimal solution, even in high-dimensional search spaces with complex, nonlinear relationships between variables. The combination of ESN and PSO might excel in time series prediction, leveraging ESN's strengths in processing sequential data and PSO's capabilities in global search. Additionally, M.A.A.A.-q et al. introduced a novel DL hybrid method that combines the power of the slime mould algorithm (SMA) and PSO within the adaptive neuro-fuzzy inference system (ANFIS) framework for PM<sub>2.5</sub> prediction [105]. The proposed model, known as PSOSMA-ANFIS, effectively harnesses the strengths of both SMA and PSO to optimize the parameters of the ANFIS model, a widely used tool for air quality prediction. SMA, a metaheuristic algorithm, is modified within the PSOSMA method. PSO plays a pivotal role in generating the initial population of solutions, greatly influencing their convergence toward the optimal solution. The synergy of SMA and PSO in the PSOSMA method results in improved algorithmic exploitation capabilities, ultimately enhancing the ANFIS model's performance. Similarly, Liu et al. [88] utilized Q-learning to guarantee that the proposed GCN-LSTM-GRU DL model converged to an optimal policy under certain conditions. Importantly, Q-learning is a reinforcement learning method that is effective in dealing with environments characterized by large or continuous state spaces. While combining machine learning and DL can lead to powerful models for many applications, these models still have some limitations, such as a lack of interpretability and expensive computation, making it harder to trace back on and employ resource-constrained devices for further use.

### 3.3.2. Deep neural network ensembles methods

In addition to DCNs, CNNs, and RNNs, some researchers have used hybrid DL architectures. These include the combinations of CNNs with LSTM networks, now known as "CNN-LSTM" models, as well as the combination of CNN and GRU.

**CNN + LSTM methods.** The most direct way to combine CNN with LSTM is to use CNN to obtain the features of the input data and then use LSTM to model the temporal dependencies with the sequence of features. The hybrid models are trained as a whole. The CNN + LSTM hybrid model determines whether the CNN interacts with the LSTM to form an information fusion. If it is only used as a module via several layers in the data processing, the model is classified as CNN-based or LSTM-based.

As shown in Table 8 [92,106–115], all directly used the "CNN-LSTM" model to capture both spatial and temporal dependencies in PM<sub>2.5</sub> data. The CNN layers were applied to the spatial dimensions of the input tensor, resulting in a feature map for each time step. The LSTM layers were then applied to the resulting feature maps along the time axis to capture the temporal dependencies. Many experiments involving different data at different sites illustrated the high feasibility of the CNN + LSTM model.

Improvements to the basic CNN-LSTM model based on CNNs or

**Table 8**  
The research used CNN-LSTM methods.

Study	Year	Location	Model	Time step	RMSE ( $\mu\text{g m}^{-3}$ )	MAE ( $\mu\text{g m}^{-3}$ )	MAPE (%)	R <sup>2</sup>
Huang et al. [106]	2018	Beijing/Shanghai, China	CNN-LSTM	H/S/T+1	24.22	14.63	-	-
Qin et al. [107]	2019	Shanghai, China	CNN-LSTM	H/S/T+24	14.30	-	-	-
Li et al. [108]	2020	Beijing, China	CNN-LSTM	D/S/T+1	18.99	16.81	-	-
Zhang et al. [109]	2020	Shijiazhuang, China	CNN-LSTM	H/S/T+1	14.94	-	-	-
Yang et al. [110]	2021	Beijing, China	CNN-LSTM	H/S/T+1	19.09	-	-	0.92
Wei et al. [111]	2021	Beijing, China	CNN-LSTM	H/S/T+6	-	19.54	-	0.62
Bekkar et al. [112]	2021	Beijing, China	CNN-LSTM	D/S/T+1	12.92	6.74	-	0.98
Wardana et al. [113]	2021	Beijing, China	CNN-LSTM	H/-/T+1	15.26	8.77	-	-
Tsokov et al. [114]	2022	Beijing, China	CNN-LSTM	H/S/T+1	14.95	8.48	-	-
Teng et al. [92]	2022	Beijing, China	CNN-LSTM	H/S/T+1	8.93	6.52	-	0.92
Kim et al. [115]	2022	South Korea	CNN-LSTM	H/S/T+1	10.52	-	-	0.37
Shao et al. [116]	2022	Seoul, South Korea	SCNN-LSTM	H/M/T+(1-10)	8.05	5.04	23.96	0.70
Choi et al. [117]	2022	Beijing, China	ResNet-LSTM	H/S/T+1	0.02	0.01	9.02	-
Zhang et al. [118]	2022	Yangtze River Delta Region, China	ResNet-LSTM	H/S/T+1	5.47	3.89	-	-
Cheng et al. [119]	2022	Beijing, China	SResCNN-LSTM	D/S/T+5	40.67	23.74	-	0.80
Zhao et al. [120]	2019	Beijing/Tianjin, China	STCNN-LSTM	H/S/T+6	19.36	15.53	26.00	0.70
Qi et al. [121]	2019	Jing-Jin-Ji Region, China	GCNN-LSTM	H/S/T+1	22.41	13.72	-	-
Soh et al. [123]	2018	Taiwan/Beijing, China	ANN-CNN-LSTM	H/S/T+6	-	-	-	-
Yang et al. [124]	2019	Beijing, China	DWFD-CNN-LSTM	H/M/T+(1-6)	43.90	29.17	-	-
Li et al. [122]	2020	Taiyuan, China	Attention-CNN-LSTM	H/M/T+(1-24)	14.83	8.98	-	0.99
Li et al. [125]	2022	Beijing, China	CBAM-CNN-BiLSTM	H/M/T+(13-18)	31.47	21.86	-	0.81
				H/M/T+(25-48)	32.34	22.30	-	0.79
Moursi et al. [126]	2022	Beijing, China	NARX-CNN-LSTM	H/S/T+1	23.64	-	-	0.92
Zhu et al. [127]	2023	Shanghai, China	1D-CNN + BiLSTM	H/S/T+1	3.88	2.52	-	0.94
Pak et al. [128]	2020	Beijing, China	PM predictor	D/S/T+1	2.99	2.21	3.90	-
Du et al. [129]	2021	Beijing, China	DAQFF	H/S/T+1	8.20	6.19	-	-
Zhu et al. [130]	2021	Jing-Jin-Ji Region, China	APNet	H/-/T+1	17.93	9.93	-	0.95
				H/-/T+72	29.11	20.07	-	0.87
Zhang et al. [131]	2022	Hong Kong/Beijing, China	Deep-AIR	H/S/T+1	-	-	21.10/23.90	-
Mohan et al. [132]	2022	Kerala, India	EDPF	H/M/T+24	12.96	9.28	56.73	0.44
Li et al. [133]	2022	Beijing, China	FPHFA	H/M/T+(1-12)	28.15	19.19	56.10	0.87
				H/M/T+(13-24)	22.12	15.27	43.80	0.93
Gunasekar et al. [134]	2022	Chennai, Tamandu	ARTOCL	NA	0.50	0.32	-	0.69

LSTMs are constantly being integrated. Shao et al. [116] proposed a space-shared CNN-LSTM model for multisite daily ahead PM<sub>2.5</sub> concentration forecasting. The model was designed to consider the correlation between different sites and the spatial information of each site. The results showed that the proposed model had higher prediction accuracy than several baseline models. Choi et al. [117] and Zhang et al. [118] both proposed ResNet-based CNN-LSTM models. Choi et al. [117] incorporated gradient-based feature attribution methods for a more explainable prediction. Cheng et al. [119] proposed a fixed ResNet-LSTM that was designed to consider the spatial and temporal correlation of air quality data. Zhao et al. [120] proposed a regional spatiotemporal collaborative CNN-LSTM prediction model. Qi et al. [121] proposed a hybrid model based on a graph convolutional neural network (GCN) and LSTM for spatiotemporal forecasting of PM<sub>2.5</sub>. This model incorporated an attention mechanism to emphasize the most important features in the input data and used CNN for feature extraction and LSTM for temporal modeling. Li et al. [122] also proposed an attention-based CNN-LSTM model for urban PM<sub>2.5</sub> concentration prediction. The model used an attention mechanism to weigh the importance of input features and combined CNN and LSTM for feature extraction and temporal modeling. Soh et al. [123] proposed a model ST-DNN that comprised ANN, LSTM, and CNN. The proposed model was designed to capture both spatial and temporal correlations in PM<sub>2.5</sub> data. Yang et al. [124] proposed a novel multistep-ahead forecasting CNN-LSTM model based on dynamic wind field distance for PM<sub>2.5</sub> prediction. Li et al. [125] proposed a DL model based on CNN and bidirectional LSTM with a convolutional block attention module (CBAM). Moursi et al. [126] proposed a combined CNN and LSTM hybrid model based on a nonlinear autoregressive network with exogenous inputs (NARX) for enhancing PM<sub>2.5</sub> prediction. Zhu et al. [127] used multiple input streams and parallel processing to

improve PM<sub>2.5</sub> prediction accuracy.

In addition to the hybrid of CNN and LSTM, the range of mixing methods has expanded. Mohan et al. [132] proposed an EDPF model combining a long short-term memory network, convolutional neural network, and random forest algorithm. Gunasekar et al. [134] proposed a sustainable, optimized hybrid intelligent system named ARTOCL; the combination of CNN and LSTM could improve air quality prediction accuracy and reduce the number of false alarms. Pak et al. [128], Du et al. [129], and Zhu et al. [130] all considered spatiotemporal correlations by using a combination of CNN and LSTM to capture spatial and temporal patterns, respectively. The proposed PM predictor [128] and DAQFF [129] directly utilized CNN to extract spatial features, while APNet [130] used parallel CNN and attention mechanisms to weigh the importance of different spatial features. In addition, both Zhang et al. [131] and Li et al. [133] proposed hybrid frameworks that combined a CNN, an LSTM, and an attention mechanism, and Zhang et al. [131] considered fine-grained air pollution estimation. Using an attention mechanism can help the model focus on important features. The strengths of DL hybrid models are their ability to capture complex relationships and patterns from the PM<sub>2.5</sub> concentration data, incorporate spatiotemporal information, and adapt to different environmental conditions. However, DL hybrid approaches may require significant computational resources, much data, and careful model tuning to achieve optimal results.

**CNN + GRU methods.** CNNs are commonly used for image recognition tasks, where they are able to learn hierarchical features from raw pixel data. CNNs can be used in time-series data to learn patterns in the temporal sequences of values, such as daily or hourly pollutant concentrations. GRUs are recurrent neural networks (RNNs) designed to capture long-term dependencies in sequential data. Unlike traditional RNNs, GRUs are able to

**Table 9**  
The research used CNN + GRU methods.

Study	Year	Location	Model	Time step	RMSE ( $\mu\text{g m}^{-3}$ )	MAE ( $\mu\text{g m}^{-3}$ )	MAPE (%)	R <sup>2</sup>
Tao et al. [135]	2019	Beijing, China	CBGRU	H/S/T+2	14.53	10.47	34.09	-
Zhang et al. [136]	2020	Lanzhou, China	MTD-CNN-GRU	H/S/T+1	7.96	4.54	-	-
Faraji et al. [137]	2022	Tehran, Iran	3D CNN-GRU	H/S/T+1	-	-	-	0.84
				D/S/T+1	-	-	-	0.78
Chiang et al. [138]	2021	Taiwan, China	AE + CNN + GRU	D/S/T+1	5.03	3.10	-	-
Mao et al. [139]	2022	Taiwan, China	CNN + GRU	H/S/T+1	4.78	3.56	-	0.89
		Kennedy/Simon Bolivar, US		D/S/T+1	6.83/6.15	5.29/4.58	-	0.44/0.56

selectively update and reset their internal state, making them more effective at learning long-term dependencies.

As shown in Table 9, several papers are based on the CNN and GRU hybrid model. Tao et al. [135] combined two powerful DL techniques, 1D CNN and BiGRU, to capture both local and temporal patterns for predicting air pollution levels. Zhang et al. [136] proposed a multitask DL model based on a CNN and GRU hybrid model. This approach modeled the complex relationships between PM<sub>2.5</sub> concentrations and meteorological variables well, and it could predict PM<sub>2.5</sub> concentrations at multiple monitoring stations simultaneously. However, in this study, the authors only considered using meteorological variables as input to the model, which could potentially lead to insufficient sample information. Furthermore, this may also limit the performance of the CNN + GRU model as it relies on a large dataset. Faraji et al. [137] combined a 3D CNN with GRUs for predicting short-term PM<sub>2.5</sub> concentrations in urban environments. Mao et al. [139] proposed a hybrid DL model that combines CNN, BiGRU, and a fully connected layer. The advantage of these two approaches is their ability to model both spatial and temporal patterns. Chiang et al. [138] proposed a hybrid DL model based on a stacked autoencoder (AE), CNN, and GRU. The model was trained on a large dataset of air pollutant data from Beijing and was able to predict hourly concentrations of multiple air pollutants up to 24 h in advance. The strengths of this model include its ability to handle missing data and its high prediction accuracy. Overall, by combining the strengths of CNN and GRU, the hybrid model is able to learn both local and global patterns in time-series data, making it well-suited for air pollutant concentration prediction. The CNN component of the model learns local patterns in the temporal sequence of values, while the GRU component captures longer-term dependencies and trends.

**Table 10**  
Description of the proposed indicators.

Standard	Indicator	Description	
Dataset	Open source	Given the available dataset link or declare the process of data collection.	
	Data feature	Predict step	The step size of the prediction task, i.e., single step or multi-step.
		Time resolution	The time resolution of dataset, i.e., hourly, daily, or monthly.
		Data size	The size of the using dataset.
	Data dimensions	Multiple data inputs include meteorological data.	
	Dataset split	The division of training, valid, and testing set. The test set needs to contain all kinds of samples.	
Pre-processing	Normalize	Min-max scaling, z-score normalization, and decimal scaling, etc.	
	Missing value	Describes how to handle missing or outlier values to ensure data continuity.	
Method	Open source	Provide a link where the code will be available.	
	Architecture	Whether to describe the network structure and give the parameters.	
	Training process	The design or trend of loss or the learning objectives.	
	Visual analysis	Visual visualization of predicted and ground truth.	
Novelty		Whether to innovate or apply the model to a domain for the first time.	
	Experimental setting	The setting of design parameters, such as the convolution kernel size.	
Experiments	Model config	The setting of design parameters, such as the convolution kernel size.	
	Computation setup	Basic information about the used CPU and GPU in the experiment.	
Results metrics		The evaluation metrics, usually RMSE, MAE, and MAPE, also have SSIM, ACC, R <sup>2</sup> , etc.	
Modeling metrics	FLOPs	The Floating Point Operations (FLOPs).	
	Params	The number of trainable parameters.	
Comparison with SOTAs		The results are compared with the advanced algorithm under the same experimental setting.	
Ablation study		Removing or disabling different components or features to see how it affect the model's performance.	

## 4. Discussion

Building upon the above review, we further established a novel evaluation framework designed to assess the quality of DL articles in PM<sub>2.5</sub> prediction, dubbed the Dataset-Method-Experiment Standard (DMES). This proposed standard offers a comprehensive evaluation from three critical perspectives: the datasets established, the methodologies processed, and the experimental results, as detailed in Table 10. This approach not only underscores the importance of each component but also facilitates a more holistic understanding of the strengths and limitations of current DL applications in this field.

### 4.1. Dataset

Initially, we identified five key indicators to assess the quality of dataset descriptions in the literature. Firstly, the availability of open-source data is paramount; open data facilitates code reproducibility, model validation, and propels industry advancement. Thus, the openness of data serves as the fundamental criterion in our evaluation of the dataset sections within each paper. Secondly, a detailed description of the data features is essential. This includes information on prediction step size, training data resolution, and overall data volume. Articles that provide a more thorough representation of the data, especially those dealing with complex prediction data types, will accordingly receive higher scores in dataset description. Thirdly, in recognition of the growing trend towards multi-dimensional data prediction, we evaluate the dimensionality of the processing data. We particularly advocate for the integration of multi-dimensional data to enhance prediction accuracy. Fourthly, dataset partitioning is critical for DL model configuration;



appropriate division of datasets underpins the reliability of the training outcomes of the proposed models. Articles that reasonably partition datasets will be awarded points for this aspect. Finally, the methods of data preprocessing must be meticulously detailed, including standardization techniques and the treatment of missing or anomalous values. Articles addressing relevant data preprocessing parts will receive points for this criterion. Given the sensitivity of DL models to data quality, providing a transparent and detailed account of data preprocessing is vital for ensuring efficient training and reliable results. Furthermore, such clarity is essential to guarantee reproducibility and to support algorithmic innovation by other researchers.

#### 4.2. Method

Furthermore, we have established five criteria for evaluating the methods section of research papers. Analogous to the significance of open-source datasets, the provision of open-source code link substantially aids in the replication of algorithms, thereby affirming the model's reproducibility and enhancing the reliability of modeling outcomes, giving available code link or source will receive points for this criterion. Then, the detailed structural description of the model, encompassing the architecture of each network component and the layer structures such as convolution kernel sizes and pooling layers, is imperative. Papers that contain such descriptions will be awarded points for this aspect. Such comprehensive structural elucidation facilitates the comparative assessment of the model's performance against others. Next, the training process, including configuration settings, articulation of learning objectives, and the formulation of loss functions, constitute critical facets of a DL model's architecture that ought to be thoroughly described within the manuscript. Besides, The visualization analysis offers an intuitive depiction of the prediction outcomes, enabling readers to directly evaluate the model's predictive accuracy. High-quality result visualization will lead to higher scores. Lastly, the novelty criterion appraises the originality and significance of the research, examining whether the study introduces innovative approaches, enhances existing methodologies, or addresses pivotal challenges within the DL domain. Articles that demonstrate such innovation will receive higher scores. This criterion also facilitates the contextual comparison of the presented work with prior studies, among other considerations.

#### 4.3. Experiment

To rigorously assess the quality of experimental descriptions in research papers, we established five key indicators focused on experimental settings and results. These indicators are designed to ensure a comprehensive and fair evaluation of studies in PM<sub>2.5</sub> concentration prediction using DL models.

- (i) **Experimental setting:** This encompasses the specific configuration of the experimental models, such as the learning rate, training epochs, utilization of pretrained models, and uniformity in the test sets applied. This indicator aims to ensure that experimental comparisons are conducted under equitable conditions. Indeed, maintaining consistency in the computational setup across experiments is vital to ensure the fairness and reliability of comparisons. This consistency encompasses both hardware specifications and software configurations, including the type and model of GPUs (graphics processing units) and CPUs (central processing units) used in the experiments. By providing these information, other researchers can accurately assess the performance and efficiency of the article's proposed DL models

under equivalent computational conditions, thereby eliminating variables that could potentially skew results. Such transparency in reporting computational resources also facilitates reproducibility and enables other researchers to replicate findings with similar setups, further contributing to the integrity and credibility of scientific research in the field of DL and PM<sub>2.5</sub> concentration prediction. Providing the information of experimental settings will result in corresponding scores.

- (ii) **Results metrics:** Through our literature review, we identified RMSE, MAE, and MAPE as the prevalent evaluation metrics for PM<sub>2.5</sub> concentration prediction. The inclusion of these three metrics offers a holistic view of the model's performance. Besides, recognizing the diversity in research objectives and methodologies, we also introduced an "Others" category to accommodate the application of specialized evaluation indicators that may be pertinent to specific studies. Conducting performance evaluations of models is crucial in research articles, as using widely accepted metrics enables effective horizontal comparisons between different models. Consequently, articles that employ more popular evaluation metrics for assessing model performance are likely to receive higher evaluation scores, reflecting the ease with which their results can be compared and understood within the broader research community.
- (iii) **Modeling metrics:** These metrics evaluate the computational and energy efficiency of the proposed methods, measuring the model's size and computational speed. A comprehensive model evaluation should consider not only prediction accuracy but also the resource efficiency of the model, balancing performance with computational cost. Articles that conduct model performance evaluations surpass those that solely assess the ability of predictive performance under our evaluation framework.
- (iv) **Comparison with state-of-the-art SOTA:** Most articles typically include comparisons with SOTA performance metrics to demonstrate the effectiveness of a given algorithm. Only a minority of studies test their models and data in isolation, without engaging in horizontal comparisons. Conducting a performance comparison with state-of-the-art algorithms is crucial for illustrating the efficacy of an algorithm. Such comparisons not only underscore the advancements achieved by the new algorithm but also encourage deeper interactions and discussions within the research community.
- (v) **Ablation study:** The inclusion of a comprehensive ablation study in a research article significantly enhances the credibility of the proposed model. Many articles overlook this aspect; however, in the field of DL, ablation studies are often essential. It provides concrete evidence that the improvements claimed by the researchers are indeed effective and not the result of external factors or coincidences. For articles that include ablation studies, we consider their DL models to be more comprehensive. Such articles are likely to receive higher evaluations because ablation studies demonstrate a thorough understanding of the model's components and their contributions to overall performance. This approach not only validates the effectiveness of the model but also highlights the authors' commitment to transparency and scientific rigor, enhancing the credibility and reproducibility of their research.

By applying these indicators, we have developed a three dimensional evaluation framework, i.e., Dataset-Method-Experiment Standard (DMES) for articles related to DL architecture for PM<sub>2.5</sub> concentration prediction. We aim to foster a more

standardized and equitable evaluation of DL-based research in the field. Clearly, this comprehensive evaluation system facilitates a nuanced assessment of the contributions and strengths of each article in the context of DL-based PM<sub>2.5</sub> prediction, providing a structured approach to understanding the current state of research in this crucial area.

## 5. Conclusion

In this review article, we conducted an objective, rigorous and comprehensive review of DL-based architecture for PM<sub>2.5</sub> concentration prediction, specially focus on the types and structures of DL-based models applied. 118 papers were meticulously selected in adherence to the PRISMA guidelines. From the perspective of utilized model architectures, we categorized and summarized seven types of DL-based model structures. Our classification offers a high-level overview of the current research landscape in this domain, enabling readers to quickly grasp the state of the art. By categorizing these methods, we present the results of various models in a tabulated form, facilitating clear and effective comparisons. Through a detailed classification of DL-based models, we have critically examined and synthesized performance indicators and application conditions for various PM<sub>2.5</sub> prediction methodologies. Our analysis provides an in-depth exploration of their strengths and weaknesses, enriching the discourse on the efficacy and adaptability of these models in addressing the complex challenge of air quality prediction. Moreover, we have introduced a novel evaluation framework, the DMES, specifically designed to assess and standardize the evaluation of articles on similar topics. This framework represents a significant stride towards enhancing the consistency and comparability of DL-based research papers, ultimately facilitating more reliable and equitable evaluations. Furthermore, we applied this three-dimensional evaluation framework, DMES, to the 118 reviewed articles. The introduction of this standard aims to improve the comparability of research outcomes and promote a more unified methodology in the assessment of DL-based models for PM<sub>2.5</sub> concentration prediction. As the forecasting of environmental PM<sub>2.5</sub> levels continues to be a critical concern, our work seeks to lay the groundwork for future research, moving towards a more integrated and standardized framework in this vital area of study.

Conclusively, our work not only contributes to the existing body of knowledge by providing a comprehensive review and a systematically categorized critical analysis of DL-based PM<sub>2.5</sub> prediction methodologies but also pioneers a structured approach for future evaluations. The establishment of the DMES framework marks a pivotal advancement in the standardization of research evaluations, paving the way for more rigorous, transparent, and reproducible scientific inquiry in the realm of DL and environmental prediction.

## CRedit authorship contribution statement

**Shiyun Zhou:** Investigation, Writing - Original Draft, Writing - Review & Editing. **Wei Wang:** Methodology, Writing - Review & Editing, Supervision. **Long Zhu:** Data Curation, Software. **Qi Qiao:** Supervision. **Yulin Kang:** Writing - Original Draft, Writing - Review & Editing, Supervision, Conceptualization, Funding Acquisition.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This work was supported by the Fundamental Research Funds for the Central Public-interest Scientific Institution (2022YSKY-73).

## References

- [1] Global Air Quality Guidelines, World health organization, 2021. <https://www.who.int/>.
- [2] C. Li, A. van Donkelaar, M.S. Hammer, E.E. McDuffie, R.T. Burnett, J.V. Spadaro, D. Chatterjee, A.J. Cohen, J.S. Apte, V.A. Southerland, Reversal of trends in global fine particulate matter air pollution, *Nat. Commun.* 14 (2023) 5349, <https://doi.org/10.1038/s41467-023-41086-z>.
- [3] J. Rentschler, N. Leonova, Global air pollution exposure and poverty, *Nat. Commun.* 14 (2023) 4432.
- [4] X. Xu, K. Shi, Z. Huang, J. Shen, What factors dominate the change of PM<sub>2.5</sub> in the world from 2000 to 2019? A study from multi-source data, *Int. J. Environ. Res. Publ. Health* 20 (2023) 2282, <https://doi.org/10.3390/ijerph20032282>.
- [5] W. Yu, T. Ye, Y. Zhang, R. Xu, Y. Lei, Z. Chen, Z. Yang, Y. Zhang, J. Song, X. Yue, Global estimates of daily ambient fine particulate matter concentrations and unequal spatiotemporal distribution of population exposure: a machine learning modelling study, *Lancet Planet. Health* 7 (2023) e209–e218, [https://doi.org/10.1016/S2542-5196\(23\)00008-6](https://doi.org/10.1016/S2542-5196(23)00008-6).
- [6] Y.A. Ayturan, Z.C. Ayturan, H.O. Altun, Air pollution modelling with deep learning: a review, *Int. J. Environ. Pollution and Environ. Modelling* 1 (2018) 58–62, <https://doi.org/10.1016/j.atmosenv.2022.119347>.
- [7] Q. Liao, M. Zhu, L. Wu, X. Pan, X. Tang, Z. Wang, Deep learning for air quality forecasts: a review, *Current Pollution Reports* 6 (2020) 399–409, <https://doi.org/10.1007/s40726-020-00159-z>.
- [8] G. Drewll, R. Al-Bahadili, Forecast air pollution in smart city using deep learning techniques: a review, *Multicult. Educ.* 7 (2021), <https://doi.org/10.5281/zenodo.4737746>.
- [9] T. Istiana, B. Kurniawan, S. Soekirno, B. Prakoso, Deep learning implementation using long short term memory architecture for PM<sub>2.5</sub> concentration prediction: a review, *IOP Conf. Ser. Earth Environ. Sci.* 1105 (2022) 012026, <https://doi.org/10.1088/1755-1315/1105/1/012026>.
- [10] N.a. Zaini, L.W. Ean, A.N. Ahmed, M.A. Malek, A systematic literature review of deep learning neural network for time series air quality forecasting, *Environ. Sci. Pollut. Control Ser.* 29 (2022) 4958–4990, <https://doi.org/10.1007/s11356-021-17442-1>.
- [11] W. Zhang, Y. Wu, J.K. Calautit, A review on occupancy prediction through machine learning for enhancing energy efficiency, air quality and thermal comfort in the built environment, *Renew. Sustain. Energy Rev.* 167 (2022) 112704, <https://doi.org/10.1016/j.rser.2022.112704>.
- [12] K. Kitcharoen, The importance-performance analysis of service quality in administrative departments of private universities in Thailand, *ABAC Journal* 24 (2004).
- [13] P. Brereton, B.A. Kitchenham, D. Budgen, M. Turner, M. Khalil, Lessons from applying the systematic literature review process within the software engineering domain, *J. Syst. Software* 80 (2007) 571–583, <https://doi.org/10.1016/j.jss.2006.07.009>.
- [14] F.M. Padilla, M. Gallardo, F. Manzano-Agugliaro, Global trends in nitrate leaching research in the 1960–2017 period, *Sci. Total Environ.* 643 (2018) 400–413, <https://doi.org/10.1016/j.scitotenv.2018.06.215>.
- [15] Y. Gao, L. Ge, S. Shi, Y. Sun, M. Liu, B. Wang, Y. Shang, J. Wu, J. Tian, Global trends and future prospects of e-waste research: a bibliometric analysis, *Environ. Sci. Pollut. Control Ser.* 26 (2019) 17809–17820, <https://doi.org/10.1007/s11356-019-05071-8>.
- [16] H. Xing, G. Wang, C. Liu, M. Suo, PM<sub>2.5</sub> concentration modeling and prediction by using temperature-based deep belief network, *Neural Network.* 133 (2021) 157–165, <https://doi.org/10.1016/j.neunet.2020.10.013>.
- [17] Y. Xing, J. Yue, C. Chen, Y. Xiang, Y. Chen, M. Shi, A deep belief network combined with modified grey wolf optimization algorithm for PM<sub>2.5</sub> concentration prediction, *Appl. Sci.* 9 (2019) 3765, <https://doi.org/10.3390/app9183765>.
- [18] X. Li, L. Peng, Y. Hu, J. Shao, T. Chi, Deep learning architecture for air quality predictions, *Environ. Sci. Pollut. Control Ser.* 23 (2016) 22408–22417, <https://doi.org/10.1007/s11356-016-7812-9>.
- [19] L. Li, M. Girusis, F. Lurmann, N. Pavlovic, C. McClure, M. Franklin, J. Wu, L.D. Oman, C. Breton, F. Gilliland, R. Habre, Ensemble-based deep learning for estimating PM<sub>2.5</sub> over California with multisource big data including wildfire smoke, *Environ. Int.* 145 (2020) 106143, <https://doi.org/10.1016/j.envint.2020.106143>.
- [20] K.K.R. Samal, K.S. Babu, S.K. Das, Multi-directional temporal convolutional artificial neural network for PM<sub>2.5</sub> forecasting with missing values: a deep learning approach, *Urban Clim.* 36 (2021) 100800, <https://doi.org/10.1016/j.uclim.2021.100800>.
- [21] Z. Luo, F. Huang, H. Liu, PM<sub>2.5</sub> concentration estimation using convolutional neural network and gradient boosting machine, *J. Environ. Sci.* 98 (2020) 85–93, <https://doi.org/10.1016/j.jes.2020.04.042>.
- [22] J. Ni, Y. Chen, Y. Gu, X. Fang, P. Shi, An improved hybrid transfer learning-based deep learning model for PM<sub>2.5</sub> concentration prediction, *Appl. Sci.*

- 12 (2022) 3597, <https://doi.org/10.3390/app12073597>.
- [23] D. Wang, H.-W. Wang, C. Li, K.-F. Lu, Z.-R. Peng, J. Zhao, Q. Fu, J. Pan, Roadside air quality forecasting in Shanghai with a novel sequence-to-sequence model, *Int. J. Environ. Res. Publ. Health* 17 (2020) 9471, <https://doi.org/10.3390/ijerph17249471>.
- [24] S. Chae, J. Shin, S. Kwon, S. Lee, S. Kang, D. Lee, PM10 and PM2.5 real-time prediction models using an interpolated convolutional neural network, *Sci. Rep.* 11 (2021) 11952, <https://doi.org/10.1038/s41598-021-91253-9>.
- [25] P. Shi, X. Fang, J. Ni, J. Zhu, An improved attention-based integrated deep neural network for PM2.5 concentration prediction, *Appl. Sci.* 11 (2021) 4001, <https://doi.org/10.3390/app11094001>.
- [26] A. Choudhury, A.I. Mridha, S. Roy, Attention enhanced hybrid model for spatiotemporal short-term forecasting of particulate matter concentrations, *Sustain. Cities Soc.* 86 (2022) 104112, <https://doi.org/10.1016/j.scs.2022.104112>.
- [27] X. Xiao, Z. Jin, S. Wang, J. Xu, Z. Peng, R. Wang, W. Shao, Y. Hui, A dual-path dynamic directed graph convolutional network for air quality prediction, *Sci. Total Environ.* 827 (2022) 154298, <https://doi.org/10.1016/j.scitotenv.2022.154298>.
- [28] G. Zhao, H. He, Y. Huang, J. Ren, Near-surface PM2.5 prediction combining the complex network characterization and graph convolution neural network, *Neural Comput. Appl.* 33 (2021) 17081–17101, <https://doi.org/10.1007/s00521-021-06300-3>.
- [29] X. Ouyang, Y. Yang, Y. Zhang, W. Zhou, D. Guo, Dual-channel spatial-temporal difference graph neural network for PM2.5 forecasting, *Neural Comput. Appl.* 35 (2023) 7475–7494, <https://doi.org/10.1007/s00521-022-08036-0>.
- [30] A. Dun, Y. Yang, F. Lei, Dynamic graph convolution neural network based on spatial-temporal correlation for air quality prediction, *Ecol. Inf.* 70 (2022) 101736, <https://doi.org/10.1016/j.ecoinf.2022.101736>.
- [31] P.S.S. Ejurothu, S. Mandal, M. Thakur, Forecasting PM2.5 concentration in India using a cluster based hybrid graph neural network approach, *Asia-Pacific Journal of Atmospheric Sciences* 59 (2023) 545–561, <https://doi.org/10.1007/s13143-022-00291-4>.
- [32] L. Zhang, J. Na, J. Zhu, Z. Shi, C. Zou, L. Yang, Spatiotemporal causal convolutional network for forecasting hourly PM2.5 concentrations in Beijing, China, *Comput. Geosci.* 155 (2021) 104869, <https://doi.org/10.1016/j.cageo.2021.104869>.
- [33] K. Zhang, X. Yang, H. Cao, J. Thé, Z. Tan, H. Yu, Multi-step forecast of PM2.5 and PM10 concentrations using convolutional neural network integrated with spatial-temporal attention and residual learning, *Environ. Int.* 171 (2023) 107691, <https://doi.org/10.1016/j.envint.2022.107691>.
- [34] M. Yu, A. Masrur, C. Blaszcak-Boxe, Predicting hourly PM2.5 concentrations in wildfire-prone areas using a SpatioTemporal Transformer model, *Sci. Total Environ.* 860 (2023) 160446, <https://doi.org/10.1016/j.scitotenv.2022.160446>.
- [35] A. Dun, Y. Yang, F. Lei, A novel hybrid model based on spatiotemporal correlation for air quality prediction, *Mobile Inf. Syst.* (2022) 2022, <https://doi.org/10.1155/2022/9759988>.
- [36] Z. Wang, R. Li, Z. Chen, Q. Yao, B. Gao, M. Xu, L. Yang, M. Li, C. Zhou, The estimation of hourly PM2.5 concentrations across China based on a spatial and temporal weighted continuous deep neural network (STWC-DNN), *ISPRS J. Photogrammetry Remote Sens.* 190 (2022) 38–55, <https://doi.org/10.1016/j.isprsjprs.2022.05.011>.
- [37] X. Dai, J. Liu, Y. Li, A recurrent neural network using historical data to predict time series indoor PM2.5 concentrations for residential buildings, *Indoor Air* 31 (2021) 1228–1237, <https://doi.org/10.1111/ina.12794>.
- [38] A. Ayturan, Z. Ayturan, H. Altun, C. Kongoli, F. Tunçez, S. Dursun, A. Ozturk, Short-term prediction of PM2.5 pollution with deep learning methods, *Global Nest J.* 22 (2020) 126–131, <https://doi.org/10.30955/gnj.003208>.
- [39] X. Li, L. Peng, X. Yao, S. Cui, Y. Hu, C. You, T. Chi, Long short-term memory neural network for air pollutant concentration predictions: method development and evaluation, *Environ. Pollut.* 231 (2017) 997–1004, <https://doi.org/10.1016/j.envpol.2017.08.114>.
- [40] Y.-S. Chang, H.-T. Chiao, S. Abimannan, Y.-P. Huang, Y.-T. Tsai, K.-M. Lin, An LSTM-based aggregated model for air pollution forecasting, *Atmos. Pollut. Res.* 11 (2020) 1451–1463, <https://doi.org/10.1016/j.apr.2020.05.015>.
- [41] T. Yayasouk, H. Lee, C. Lee, Air pollution prediction using long short-term memory (LSTM) and deep autoencoder (DAE) models, *Sustainability* 12 (2020) 2570, <https://doi.org/10.3390/su12062570>.
- [42] H. Karimian, Q. Li, C. Wu, Y. Qi, Y. Mo, G. Chen, X. Zhang, S. Sachdeva, Evaluation of different machine learning approaches to forecasting PM2.5 mass concentrations, *Aerosol Air Qual. Res.* 19 (2019) 1400–1410, <https://doi.org/10.4209/aaqr.2018.12.0450>.
- [43] W. Mao, W. Wang, L. Jiao, S. Zhao, A. Liu, Modeling air quality prediction using a deep learning approach: method optimization and evaluation, *Sustain. Cities Soc.* 65 (2021) 102567, <https://doi.org/10.1016/j.scs.2020.102567>.
- [44] K. Qadeer, W.U. Rehman, A.M. Sheri, I. Park, H.K. Kim, M. Jeon, A long short-term memory (LSTM) network for hourly estimation of PM2.5 concentration in two cities of South Korea, *Appl. Sci.* 10 (2020) 3984, <https://doi.org/10.3390/app10113984>.
- [45] E. Kristiani, H. Lin, J.-R. Lin, Y.-H. Chuang, C.-Y. Huang, C.-T. Yang, Short-term prediction of PM2.5 using LSTM deep learning methods, *Sustainability* 14 (2022) 2068, <https://doi.org/10.3390/su14042068>.
- [46] L. Lin, C.Y. Chen, H.Y. Yang, Z. Xu, S.H. Fang, Dynamic system approach for improved PM2.5 prediction in taiwan, *IEEE Access* 8 (2020) 210910–210921, <https://doi.org/10.1109/ACCESS.2020.3038853>.
- [47] J. Park, S. Chang, A particulate matter concentration prediction model based on long short-term memory and an artificial neural network, *Int. J. Environ. Res. Publ. Health* 18 (2021) 6801, <https://doi.org/10.3390/ijerph18136801>.
- [48] B. Peralta, T. Sepúlveda, O. Nocolis, L. Caro, Space-Time Prediction of PM2.5 Concentrations in Santiago de Chile Using LSTM Networks, *Appl. Sci.* 12 (2022) 11317, <https://doi.org/10.3390/app122211317>.
- [49] K.H. Waseem, H. Mushtaq, F. Abid, A.M. Abu-Mahfouz, A. Shaikh, M. Turan, J. Rasheed, Forecasting of air quality using an optimized recurrent neural network, *Processes* 10 (2022) 2117, <https://doi.org/10.3390/pr10102117>.
- [50] S. Gul, G.M. Khan, S. Yousaf, Multi-step short-term PM2.5 forecasting for enactment of proactive environmental regulation strategies, *Environ. Monit. Assess.* 194 (2022) 386, <https://doi.org/10.1007/s10661-022-10029-4>.
- [51] J. Ma, Y. Ding, V.J.L. Gan, C. Lin, Z. Wan, Spatiotemporal prediction of PM2.5 concentrations at different time granularities using IDW-BLSTM, *IEEE Access* 7 (2019) 107897–107907, <https://doi.org/10.1109/ACCESS.2019.2932445>.
- [52] W. Tong, L. Li, X. Zhou, A. Hamilton, K. Zhang, Deep learning PM2.5 concentrations with bidirectional LSTM RNN, *Air Quality, Atmosphere & Health* 12 (2019) 411–423, <https://doi.org/10.1007/s11869-018-0647-4>.
- [53] M. Zhang, D. Wu, R. Xue, Hourly prediction of PM2.5 concentration in Beijing based on Bi-LSTM neural network, *Multimed. Tool. Appl.* 80 (2021) 24455–24468, <https://doi.org/10.1007/s11042-021-10852-w>.
- [54] B. Deep, I. Mathur, N. Joshi, An approach to forecast pollutants concentration with varied dispersion, *Int. J. Environ. Sci. Technol.* 19 (2022) 5131–5138, <https://doi.org/10.1007/s13762-021-03378-z>.
- [55] A.G. Mengara Mengara, Y. Kim, Y. Yoo, J. Ahn, Distributed deep features extraction model for air quality forecasting, *Sustainability* 12 (2020) 8014, <https://doi.org/10.3390/su12198014>.
- [56] A.G. Mengara Mengara, E. Park, J. Jang, Y. Yoo, Attention-based distributed deep learning model for air quality forecasting, *Sustainability* 14 (2022) 3269, <https://doi.org/10.3390/su14063269>.
- [57] X. Xu, M. Yoneda, Multitask air-quality prediction based on LSTM-autoencoder model, *IEEE Trans. Cybern.* 51 (2021) 2577–2586, <https://doi.org/10.1109/TCYB.2019.2945999>.
- [58] B. Zhang, H. Zhang, G. Zhao, J. Lian, Constructing a PM2.5 concentration prediction model by combining auto-encoder with Bi-LSTM neural networks, *Environ. Model. Software* 124 (2020) 104600, <https://doi.org/10.1016/j.envsoft.2019.104600>.
- [59] G. Zou, B. Zhang, R. Yong, D. Qin, Q. Zhao, FDN-Learning: urban pm2.5-concentration spatial correlation prediction model based on fusion deep neural network, *Big Data Research* 26 (2021) 100269, <https://doi.org/10.1016/j.bdr.2021.100269>.
- [60] L. Shi, H. Zhang, X. Xu, M. Han, P. Zuo, A balanced social LSTM for PM2.5 concentration prediction based on local spatiotemporal correlation, *Chemosphere* 291 (2022) 133124, <https://doi.org/10.1016/j.chemosphere.2021.133124>.
- [61] J. Ma, Y. Ding, J.C.P. Cheng, F. Jiang, V.J.L. Gan, Z. Xu, A Lag-FLSTM deep learning network based on Bayesian Optimization for multi-sequential-variant PM2.5 prediction, *Sustain. Cities Soc.* 60 (2020) 102237, <https://doi.org/10.1016/j.scs.2020.102237>.
- [62] W. Ding, Y. Zhu, Prediction of PM2.5 concentration in ningxia hui autonomous region based on PCA-attention-LSTM, *Atmosphere* 13 (2022) 1444, <https://doi.org/10.3390/atmos13091444>.
- [63] K. Hu, X. Guo, X. Gong, X. Wang, J. Liang, D. Li, Air quality prediction using spatio-temporal deep learning, *Atmos. Pollut. Res.* 13 (2022) 101543, <https://doi.org/10.1016/j.apr.2022.101543>.
- [64] W. Wang, W. Mao, X. Tong, G. Xu, A novel recursive model based on a convolutional long short-term memory neural network for air pollution prediction, *Rem. Sens.* 13 (2021) 1284, <https://doi.org/10.3390/rs13071284>.
- [65] J. Zhao, F. Deng, Y. Cai, J. Chen, Long short-term memory - fully connected (LSTM-FC) neural network for PM2.5 concentration prediction, *Chemosphere* 220 (2019) 486–492, <https://doi.org/10.1016/j.chemosphere.2018.12.128>.
- [66] C. Wen, S. Liu, X. Yao, L. Peng, X. Li, Y. Hu, T. Chi, A novel spatiotemporal convolutional long short-term neural network for air pollution prediction, *Sci. Total Environ.* 654 (2019) 1091–1099, <https://doi.org/10.1016/j.scitotenv.2018.11.086>.
- [67] Y. Zhou, F.-J. Chang, L.-C. Chang, I.F. Kao, Y.-S. Wang, Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts, *J. Clean. Prod.* 209 (2019) 134–145, <https://doi.org/10.1016/j.jclepro.2018.10.243>.
- [68] X. Sun, W. Xu, Deep random subspace learning: a spatial-temporal modeling approach for air quality prediction, *Atmosphere* 10 (2019) 560, <https://doi.org/10.3390/atmos10090560>.
- [69] X. Wu, C. Zhang, J. Zhu, X. Zhang, Research on PM2.5 concentration prediction based on the CE-AGA-LSTM model, *Appl. Sci.* 12 (2022) 7009, <https://doi.org/10.3390/app12147009>.
- [70] J. Ma, J.C.P. Cheng, C. Lin, Y. Tan, J. Zhang, Improving air quality prediction accuracy at larger temporal resolutions using deep learning and transfer learning techniques, *Atmos. Environ.* 214 (2019) 116885, <https://doi.org/10.1016/j.atmosenv.2019.116885>.
- [71] X. Liu, W. Li, MGC-LSTM: a deep learning model based on graph convolution of multiple graphs for PM2.5 prediction, *Int. J. Environ. Sci. Technol.* 20 (2023) 10297–10312, <https://doi.org/10.1007/s13762-022-04553-6>.
- [72] F. Xiao, M. Yang, H. Fan, G. Fan, M.A.A. Al-qaness, An improved deep learning



- model for predicting daily PM<sub>2.5</sub> concentration, *Sci. Rep.* 10 (2020) 20988, <https://doi.org/10.1038/s41598-020-77757-w>.
- [73] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, *Adv. Neural Inf. Process. Syst.* 30 (2017), <https://doi.org/10.48550/arXiv.1706.03762>.
- [74] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, W. Zhang, Informer: beyond efficient transformer for long sequence time-series forecasting, in: *Proceedings of the AAAI Conference on Artificial Intelligence, 2021*, pp. 11106–11115.
- [75] M.A.A. Al-qaness, A. Dahou, A.A. Ewees, L. Abualigah, J. Huai, M. Abd Elaziz, A.M. Helmi, ResInformer: residual transformer-based artificial time-series forecasting model for PM<sub>2.5</sub> concentration in three major Chinese cities, *Mathematics* 11 (2023) 476, <https://doi.org/10.3390/math11020476>.
- [76] H. Chang-Hoi, I. Park, H.-R. Oh, H.-J. Gim, S.-K. Hur, J. Kim, D.-R. Choi, Development of a PM<sub>2.5</sub> prediction model using a recurrent neural network algorithm for the Seoul metropolitan area, Republic of Korea, *Atmos. Environ.* 245 (2021) 118021, <https://doi.org/10.1016/j.atmosenv.2020.118021>.
- [77] H. Sun, J.C.H. Fung, Y. Chen, W. Chen, Z. Li, Y. Huang, C. Lin, M. Hu, X. Lu, Improvement of PM<sub>2.5</sub> and O<sub>3</sub> forecasting by integration of 3D numerical simulation with deep learning techniques, *Sustain. Cities Soc.* 75 (2021) 103372, <https://doi.org/10.1016/j.scs.2021.103372>.
- [78] W. Skamarock, J. Klemp, J. Dudhia, D. Gill, D. Barker, W. Wang, J. Powers, *A Description of the Advanced Research WRF Version 2*, 2005.
- [79] B. Zhang, Y. Rong, R. Yong, D. Qin, M. Li, G. Zou, J. Pan, Deep learning for air pollutant concentration prediction: a review, *Atmos. Environ.* 290 (2022) 119347, <https://doi.org/10.1016/j.atmosenv.2022.119347>.
- [80] W. Qiao, W. Tian, Y. Tian, Q. Yang, Y. Wang, J. Zhang, The forecasting of PM<sub>2.5</sub> using a hybrid model based on wavelet transform and an improved deep learning algorithm, *IEEE Access* 7 (2019) 142814–142825, <https://doi.org/10.1109/ACCESS.2019.2944755>.
- [81] M. Benhaddi, J. Ouarzazi, Multivariate time series forecasting with dilated residual convolutional neural networks for urban air quality prediction, *Arabian J. Sci. Eng.* 46 (2021) 3423–3442, <https://doi.org/10.1007/s13369-020-05109-x>.
- [82] J. Kim, X. Wang, C. Kang, J. Yu, P. Li, Forecasting air pollutant concentration using a novel spatiotemporal deep learning model based on clustering, feature selection and empirical wavelet transform, *Sci. Total Environ.* 801 (2021) 149654, <https://doi.org/10.1016/j.scitotenv.2021.149654>.
- [83] G. Huang, X. Li, B. Zhang, J. Ren, PM<sub>2.5</sub> concentration forecasting at surface monitoring sites using GRU neural network based on empirical mode decomposition, *Sci. Total Environ.* 768 (2021) 144516, <https://doi.org/10.1016/j.scitotenv.2020.144516>.
- [84] X.-B. Jin, N.-X. Yang, X.-Y. Wang, Y.-T. Bai, T.-L. Su, J.-L. Kong, Deep hybrid model based on EMD with classification by frequency characteristics for long-term air quality prediction, *Mathematics* 8 (2020) 214, <https://doi.org/10.3390/math8020214>.
- [85] N.a. Zaini, L.W. Ean, A.N. Ahmed, M. Abdul Malek, M.F. Chow, PM<sub>2.5</sub> forecasting for an urban area based on deep learning and decomposition method, *Sci. Rep.* 12 (2022) 17565, <https://doi.org/10.1038/s41598-022-21769-1>.
- [86] Z. Zhang, Y. Zeng, K. Yan, A hybrid deep learning technology for PM<sub>2.5</sub> air quality forecasting, *Environ. Sci. Pollut. Control Ser.* 28 (2021) 39409–39422, <https://doi.org/10.1007/s11356-021-12657-8>.
- [87] Y.-S. Chang, S. Abimannan, H.-T. Chiao, C.-Y. Lin, Y.-P. Huang, An ensemble learning based hybrid model and framework for air pollution forecasting, *Environ. Sci. Pollut. Control Ser.* 27 (2020) 38155–38168, <https://doi.org/10.1007/s11356-020-09855-1>.
- [88] X. Liu, M. Qin, Y. He, X. Mi, C. Yu, A new multi-data-driven spatiotemporal PM<sub>2.5</sub> forecasting model based on an ensemble graph reinforcement learning convolutional network, *Atmos. Pollut. Res.* 12 (2021) 101197, <https://doi.org/10.1016/j.apr.2021.101197>.
- [89] H. Liu, S. Dong, A novel hybrid ensemble model for hourly PM<sub>2.5</sub> forecasting using multiple neural networks: a case study in China, *Air Quality, Atmosphere & Health* 13 (2020) 1411–1420, <https://doi.org/10.1007/s11869-020-00895-7>.
- [90] F. Jiang, C. Zhang, S. Sun, J. Sun, Forecasting hourly PM<sub>2.5</sub> based on deep temporal convolutional neural network and decomposition method, *Appl. Soft Comput.* 113 (2021) 107988, <https://doi.org/10.1016/j.asoc.2021.107988>.
- [91] X. Lu, J. Wang, Y. Yan, L. Zhou, W. Ma, Estimating hourly PM<sub>2.5</sub> concentrations using Himawari-8 AOD and a DBSCAN-modified deep learning model over the YRDUA, China, *Atmos. Pollut. Res.* 12 (2021) 183–192, <https://doi.org/10.1016/j.apr.2020.10.020>.
- [92] M. Teng, S. Li, J. Xing, G. Song, J. Yang, J. Dong, X. Zeng, Y. Qin, 24-Hour prediction of PM<sub>2.5</sub> concentrations by combining empirical mode decomposition and bidirectional long short-term memory neural network, *Sci. Total Environ.* 821 (2022) 153276, <https://doi.org/10.1016/j.scitotenv.2022.153276>.
- [93] M. Fu, C. Le, T. Fan, R. Prakashovich, D. Manko, O. Dmytrenko, D. Lande, S. Shahid, Z.M. Yaseen, Integration of complete ensemble empirical mode decomposition with deep long short-term memory model for particulate matter concentration prediction, *Environ. Sci. Pollut. Control Ser.* 28 (2021) 64818–64829, <https://doi.org/10.1007/s11356-021-15574-y>.
- [94] H. Zhang, Z. Shang, Y. Song, Z. He, L. Li, A novel combined model based on echo state network – a case study of PM<sub>10</sub> and PM<sub>2.5</sub> prediction in China, *Environ. Technol.* 41 (2020) 1937–1949, <https://doi.org/10.1080/09593330.2018.1551941>.
- [95] C. Wang, J. Zheng, J. Du, G. Wang, J.J. Klemeš, B. Wang, Q. Liao, Y. Liang, Weather condition-based hybrid models for multiple air pollutants forecasting and minimisation, *J. Clean. Prod.* 352 (2022) 131610, <https://doi.org/10.1016/j.jclepro.2022.131610>.
- [96] J. Wang, W. Xu, J. Dong, Y. Zhang, Two-stage deep learning hybrid framework based on multi-factor multi-scale and intelligent optimization for air pollutant prediction and early warning, *Stoch. Environ. Res. Risk Assess.* 36 (2022) 3417–3437, <https://doi.org/10.1007/s00477-022-02202-5>.
- [97] W. Sun, Z. Xu, A hybrid Daily PM<sub>2.5</sub> concentration prediction model based on secondary decomposition algorithm, mode recombination technique and deep learning, *Stoch. Environ. Res. Risk Assess.* 36 (2022) 1143–1162, <https://doi.org/10.1007/s00477-021-02100-2>.
- [98] S. Xu, W. Li, Y. Zhu, A. Xu, A novel hybrid model for six main pollutant concentrations forecasting based on improved LSTM neural networks, *Sci. Rep.* 12 (2022) 14434, <https://doi.org/10.1038/s41598-022-17754-3>.
- [99] H. Zhou, T. Wang, H. Zhao, Z. Wang, Updated prediction of air quality based on kalman-attention-LSTM network, *Sustainability* 15 (2023) 356, <https://doi.org/10.3390/su15010356>.
- [100] J. Zhao, L. Yuan, K. Sun, H. Huang, P. Guan, C. Jia, Forecasting fine particulate matter concentrations by in-depth learning model according to random forest and bilateral long- and short-term memory neural networks, *Sustainability* 14 (2022) 9430, <https://doi.org/10.3390/su14159430>.
- [101] L. Zhang, L. Xu, M. Jiang, P. He, A novel hybrid ensemble model for hourly PM<sub>2.5</sub> concentration forecasting, *Int. J. Environ. Sci. Technol.* 20 (2023) 219–230, <https://doi.org/10.1007/s13762-022-03940-3>.
- [102] A. Masood, K. Ahmad, Data-driven predictive modeling of PM<sub>2.5</sub> concentrations using machine learning and deep learning techniques: a case study of Delhi, India, *Environ. Monit. Assess.* 195 (2022) 60, <https://doi.org/10.1007/s10661-022-10603-w>.
- [103] H. Liu, D.-h. Deng, An enhanced hybrid ensemble deep learning approach for forecasting daily PM<sub>2.5</sub>, *J. Cent. S. Univ.* 29 (2022) 2074–2083, <https://doi.org/10.1007/s11771-022-5051-4>.
- [104] W. Ban, L. Shen, PM<sub>2.5</sub> prediction based on the CEEMDAN algorithm and a machine learning hybrid model, *Sustainability* 14 (2022) 16128, <https://doi.org/10.3390/su142316128>.
- [105] M.A.A. Al-qaness, H. Fan, A.A. Ewees, D. Yousri, M. Abd Elaziz, Improved ANFIS model for forecasting Wuhan City Air Quality and Analysis COVID-19 lockdown impacts on air quality, *Environ. Res.* 194 (2021) 110607, <https://doi.org/10.1016/j.envres.2020.110607>.
- [106] C.-J. Huang, P.-H. Kuo, A deep CNN-LSTM model for particulate matter (PM<sub>2.5</sub>) forecasting in smart cities, *Sensors* 18 (2018) 2220, <https://doi.org/10.3390/s18072220>.
- [107] D. Qin, J. Yu, G. Zou, R. Yong, Q. Zhao, B. Zhang, A novel combined prediction scheme based on CNN and LSTM for urban PM<sub>2.5</sub> concentration, *IEEE Access* 7 (2019) 20050–20059, <https://doi.org/10.1109/ACCESS.2019.2897028>.
- [108] T. Li, M. Hua, X. Wu, A hybrid CNN-LSTM model for forecasting particulate matter (PM<sub>2.5</sub>), *IEEE Access* 8 (2020) 26933–26940, <https://doi.org/10.1109/ACCESS.2020.2971348>.
- [109] G. Zhang, H. Lu, J. Dong, S. Poslad, R. Li, X. Zhang, X. Rui, A framework to predict high-resolution spatiotemporal PM<sub>2.5</sub> distributions using a deep-learning model: a case study of shijiazhuang, China, *Rem. Sens.* 12 (2020) 2825, <https://doi.org/10.3390/rs12172825>.
- [110] J. Yang, R. Yan, M. Nong, J. Liao, F. Li, W. Sun, PM<sub>2.5</sub> concentrations forecasting in Beijing through deep learning with different inputs, model structures and forecast time, *Atmos. Pollut. Res.* 12 (2021) 101168, <https://doi.org/10.1016/j.apr.2021.101168>.
- [111] J. Wei, F. Yang, X.-C. Ren, S. Zou, A short-term prediction model of PM<sub>2.5</sub> concentration based on deep learning and mode decomposition methods, *Appl. Sci.* 11 (2021) 6915, <https://doi.org/10.3390/app11156915>.
- [112] A. Bekkar, B. Hssina, S. Douzi, K. Douzi, Air-pollution prediction in smart city, deep learning approach, *Journal of Big Data* 8 (2021) 161, <https://doi.org/10.1186/s40537-021-00548-1>.
- [113] I.N.K. Wardana, J.W. Gardner, S.A. Fahmy, Optimising deep learning at the edge for accurate hourly air quality prediction, *Sensors* 21 (2021) 1064, <https://doi.org/10.3390/s21041064>.
- [114] S. Tsokov, M. Lazarova, A. Aleksieva-Petrova, A hybrid spatiotemporal deep model based on CNN and LSTM for air pollution prediction, *Sustainability* 14 (2022) 5104, <https://doi.org/10.3390/su14095104>.
- [115] H.S. Kim, K.M. Han, J. Yu, J. Kim, K. Kim, H. Kim, Development of a CNN+LSTM hybrid neural network for daily PM<sub>2.5</sub> prediction, *Atmosphere* 13 (2022) 2124, <https://doi.org/10.3390/atmos13122124>.
- [116] X.-S. Kim, Chang-Soo, Accurate multi-site daily-ahead multi-step PM<sub>2.5</sub> concentrations forecasting using space-shared CNN-LSTM, *Computers, Materials & Continua* 70 (2022) 5143–5160, <https://doi.org/10.32604/cmc.2022.020689>.
- [117] H. Choi, C. Jung, T. Kang, H.J. Kim, I.Y. Kwak, Explainable time-series prediction using a residual network and gradient-based methods, *IEEE Access* 10 (2022) 108469–108482, <https://doi.org/10.1109/ACCESS.2022.3213926>.
- [118] B. Zhang, G. Zou, D. Qin, Q. Ni, H. Mao, M. Li, RCL-Learning: ResNet and convolutional long short-term memory-based spatiotemporal air pollutant concentration prediction model, *Expert Systems with Appl.* 207 (2022) 118017, <https://doi.org/10.1016/j.eswa.2022.118017>.
- [119] X. Cheng, W. Zhang, A. Wenzel, J. Chen, Stacked ResNet-LSTM and CORAL



- model for multi-site air quality prediction, *Neural Computing and Applicat.* 34 (2022) 13849–13866, <https://doi.org/10.1007/s00521-022-07175-8>.
- [120] G. Zhao, G. Huang, H. He, H. He, J. Ren, Regional spatiotemporal collaborative prediction model for air quality, *IEEE Access* 7 (2019) 134903–134919, <https://doi.org/10.1109/ACCESS.2019.2941732>.
- [121] Y. Qi, Q. Li, H. Karimian, D. Liu, A hybrid model for spatiotemporal forecasting of PM<sub>2.5</sub> based on graph convolutional neural network and long short-term memory, *Science of The Total Environ.* 664 (2019) 1–10, <https://doi.org/10.1016/j.scitotenv.2019.01.333>.
- [122] S. Li, G. Xie, J. Ren, L. Guo, Y. Yang, X. Xu, Urban PM<sub>2.5</sub> concentration prediction via attention-based CNN–LSTM, *Applied Sciences* 10 (2020) 1953, <https://doi.org/10.3390/app10061953>.
- [123] P.W. Soh, J.W. Chang, J.W. Huang, Adaptive deep learning-based air quality prediction model using the most relevant spatial-temporal relations, *IEEE Access* 6 (2018) 38186–38199, <https://doi.org/10.1109/ACCESS.2018.2849820>.
- [124] M. Yang, H. Fan, K. Zhao, PM<sub>2.5</sub> prediction with a novel multi-step-ahead forecasting model based on dynamic wind field distance, *Int. J. Environ. Res. Public Health* 16 (2019) 4482, <https://doi.org/10.3390/ijerph16224482>.
- [125] D. Li, J. Liu, Y. Zhao, Prediction of multi-site PM<sub>2.5</sub> concentrations in Beijing using CNN-Bi LSTM with CBAM, *Atmosphere* 13 (2022) 1719, <https://doi.org/10.3390/atmos13101719>.
- [126] A.S.A. Moursi, N. El-Fishawy, S. Djahel, M.A. Shouman, Enhancing PM<sub>2.5</sub> prediction using NARX-based combined CNN and LSTM hybrid model, *Sensors* 22 (2022) 4418, <https://doi.org/10.3390/s22124418>.
- [127] M. Zhu, J. Xie, Investigation of nearby monitoring station for hourly PM<sub>2.5</sub> forecasting using parallel multi-input 1D-CNN-biLSTM, *Expert Systems with Applic.* 211 (2023) 118707, <https://doi.org/10.1016/j.eswa.2022.118707>.
- [128] U. Pak, J. Ma, U. Ryu, K. Ryom, U. Juhyok, K. Pak, C. Pak, Deep learning-based PM<sub>2.5</sub> prediction considering the spatiotemporal correlations: a case study of Beijing, China, *Sci. Total Environ.* 699 (2020) 133561, <https://doi.org/10.1016/j.scitotenv.2019.07.367>.
- [129] S. Du, T. Li, Y. Yang, S.J. Horng, Deep air quality forecasting using hybrid deep learning framework, *IEEE Transact. Knowledge and Data Engineering* 33 (2021) 2412–2424, <https://doi.org/10.1109/TKDE.2019.2954510>.
- [130] J. Zhu, F. Deng, J. Zhao, H. Zheng, Attention-based parallel networks (APNet) for PM<sub>2.5</sub> spatiotemporal prediction, *Sci. Total Environ.* 769 (2021) 145082, <https://doi.org/10.1016/j.scitotenv.2021.145082>.
- [131] Q. Zhang, Y. Han, V.O.K. Li, J.C.K. Lam, Deep-AIR: a hybrid CNN-LSTM framework for fine-grained air pollution estimation and forecast in metropolitan cities, *IEEE Access* 10 (2022) 55818–55841, <https://doi.org/10.1109/ACCESS.2022.3174853>.
- [132] A.S. Mohan, L. Abraham, An ensemble deep learning model for forecasting hourly PM<sub>2.5</sub> concentrations, *IETE J. Res.* 69 (2023) 6832–6845, <https://doi.org/10.1080/03772063.2022.2089747>.
- [133] D. Li, J. Liu, Y. Zhao, Forecasting of PM<sub>2.5</sub> concentration in Beijing using hybrid deep learning framework based on attention mechanism, *Applied Sci.* 12 (2022) 11155, <https://doi.org/10.3390/app122111155>.
- [134] S. Gunasekar, G. Joselin Retna Kumar, Y. Dileep Kumar, Sustainable optimized LSTM-based intelligent system for air quality prediction in Chennai, *Acta Geophysica* 70 (2022) 2889–2899, <https://doi.org/10.1007/s11600-022-00796-6>.
- [135] Q. Tao, F. Liu, Y. Li, D. Sidorov, Air pollution forecasting using a deep learning model based on 1D convnets and bidirectional GRU, *IEEE Access* 7 (2019) 76690–76698, <https://doi.org/10.1109/ACCESS.2019.2921578>.
- [136] Q. Zhang, S. Wu, X. Wang, B. Sun, H. Liu, A PM<sub>2.5</sub> concentration prediction model based on multi-task deep learning for intensive air quality monitoring stations, *J. Cleaner Product.* 275 (2020) 122722, <https://doi.org/10.1016/j.jclepro.2020.122722>.
- [137] M. Faraji, S. Nadi, O. Ghaffarparand, S. Homayoni, K. Downey, An integrated 3D CNN-GRU deep learning method for short-term prediction of PM<sub>2.5</sub> concentration in urban environment, *Sci. Total Environ.* 834 (2022) 155324, <https://doi.org/10.1016/j.scitotenv.2022.155324>.
- [138] P.W. Chiang, S.J. Horng, Hybrid time-series framework for daily-based PM<sub>2.5</sub> forecasting, *IEEE Access* 9 (2021) 104162–104176, <https://doi.org/10.1109/ACCESS.2021.3099111>.
- [139] Y.-S. Mao, S.-J. Lee, C.-H. Wu, C.-L. Hou, C.-S. Ouyang, C.-F. Liu, A hybrid deep learning network for forecasting air pollutant concentrations, *Applied Intelligence* 53 (2023) 12792–12810, <https://doi.org/10.1007/s10489-022-04191-y>.