Environmental Pollution 306 (2022) 119348

Contents lists available at ScienceDirect

Environmental Pollution

journal homepage: www.elsevier.com/locate/envpol

Deep neural networks for spatiotemporal $PM_{2.5}$ forecasts based on atmospheric chemical transport model output and monitoring data☆

Pu-Yun Kow [a], Li-Chiu Chang [b], Chuan-Yao Lin [c], Charles C.-K. Chou [c], Fi-John Chang [a,*]

[a] Department of Bioenvironmental Systems Engineering, National Taiwan University, Taipei, 10617, Taiwan
[b] Department of Water Resources and Environmental Engineering, Tamkang University, New Taipei City, 25137, Taiwan
[c] Research Center for Environmental Changes, Academia Sinica, Taipei, 11529, Taiwan

ARTICLE INFO

Keywords:
$PM_{2.5}$ forecast
Deep neural network (DNN)
Convolutional neural network (CNN)
Atmospheric chemical transport model
Regional air quality forecast

ABSTRACT

Reliable long-horizon $PM_{2.5}$ forecasts are crucial and beneficial for health protection through early warning against air pollution. However, the dynamic nature of air quality makes $PM_{2.5}$ forecasts at long horizons very challenging. This study proposed a novel machine learning-based model (MCNN-BP) that fused multiple convolutional neural networks (MCNN) with a back-propagation neural network (BPNN) for making spatiotemporal $PM_{2.5}$ forecasts for the next 72 h at 74 stations covering the whole Taiwan simultaneously. Model configuration involved an ensemble of massive hourly air quality and meteorological monitoring datasets and the existing publicly-available $PM_{2.5}$ simulated (forecasted) datasets from an atmospheric chemical transport (ACT) model. The proposed methodology collaboratively constructed two CNNs to mine the observed data (the past) and the forecasted data from ACT (the future) separately. The results showed that the MCNN-BP model could significantly improve the accuracy of spatiotemporal $PM_{2.5}$ forecasts and substantially reduce the forecast biases of the ACT model. We demonstrated that the proposed MCNN-BP model with effective feature extraction and good denoising ability could overcome the curse of dimensionality and offer satisfactory regional long-horizon $PM_{2.5}$ forecasts. Moreover, the MCNN-BP model has considerably shorter computational time (5 min) and lower computational load than the compute-intensive ACT model. The proposed approach hits a milestone in multi-site and multi-horizon forecasting, which significantly contributes to early warning against regional air pollution.

1. Introduction

In the era of the booming economy, the rising economic activities in developing countries have inevitably triggered a notable increase in air pollutant emissions. Despite the economic takeoff, people's quality of life has not been improved much because of air pollution (Tagaris et al., 2009). In some areas suffering from severe air pollution, people can no longer carry out human activities at will, which has seriously affected their lives. Particulate matters and nitrogen oxide are the dominant components of ambient air pollution-induced by industrialization and urbanization (Keulers et al., 2022; Trusz et al., 2020). Long-term exposure to $PM_{2.5}$ causes adverse health effects such as respiratory infections and lung cancer (Guo et al., 2022; Wang et al., 2021). Air pollution induced-health hazards raise an imperative need for $PM_{2.5}$ forecasts in support of early warning for taking air pollution control measures.

$PM_{2.5}$ forecasting faces a high degree of uncertainty because of its dynamic and complex generation mechanism. In Taiwan, the emission of $PM_{2.5}$ relates mostly to transportation, industrial production, and physical and chemical processes (Hsu and Cheng, 2016; EPA, 2016). Besides, the emission sources and compounds of $PM_{2.5}$ are dissimilar in different regions. For example, $PM_{2.5}$ emissions in northern Taiwan come mainly from automobiles, while those in southern Taiwan are associated mainly with industrial production (Chen et al., 2001; Shen et al., 2018). Researchers have been seeking novel methods in diverse perspectives to predict $PM_{2.5}$ concentrations with observational data from ground-based monitoring facilities and/or Internet of Things devices for moving toward clean and advanced smart cities (Ho et al., 2020). In general, air quality forecast models can be classified into atmospheric chemical transport (ACT) models and machine learning models (He et al., 2022). The ACT model numerically constructs nonlinear ordinary differential equations based on physical and chemical mechanisms as well as performs dynamic simulations based on

☆ This paper has been recommended for acceptance by Prof. Pavlos Kassomenos.
* Corresponding author.
E-mail address: changfj@ntu.edu.tw (F.-J. Chang).

https://doi.org/10.1016/j.envpol.2022.119348
Received 8 March 2022; Received in revised form 13 April 2022; Accepted 20 April 2022
Available online 26 April 2022

air-contaminated objects and their precursors from the past to the future. The Weather Research and Forecasting-Community Multi-Scale Air Quality Model (WRF-CMAQ) and the integration of the Weather Research and Forecasting model with Chemistry model (WRF-Chem) are ACT models (Hong et al., 2020). These models have been the mainstream approaches to modeling $PM_{2.5}$ since the last decades. Although several ACT models have the ability to simulate the future trend of $PM_{2.5}$, there is room for improvement in model accuracy, e.g., $R^2 = 0.23$ by the WRF-Chem model (Goldberg et al., 2019). Besides, the notorious complexity and high uncertainty of $PM_{2.5}$ concentration makes the development of ACT models devoted to $PM_{2.5}$ forecasting challenging (Kim et al., 2019; Lee et al., 2020).

Machine learning (ML) models are popular predictors for $PM_{2.5}$ nowadays and offer a rapid depiction of the interdependence between $PM_{2.5}$ and affecting factors (Shogrkhodaei et al., 2021; Zhou et al., 2019). Among ML techniques, artificial neural networks (ANNs) and deep neural networks (DNNs) have been widely applied to air quality prediction (Yan et al., 2020 & 2021; Zhou et al., 2021). A new research trend involving the integration of several ML techniques to improve prediction reliability and accuracy of $PM_{2.5}$ concentration has arisen lately (e.g., Jain et al., 2021; Lu et al., 2021). For instance, Kow et al. (2020) proposed to fuse a convolutional neural network (CNN) with a back-propagation neural network (BPNN) for producing regional 10-h-ahead $PM_{2.5}$ forecasts. Wang et al. (2017) configured multiple ML techniques to capture the high-level features of $PM_{2.5}$ and provided one-day-ahead $PM_{2.5}$ forecasts. Nevertheless, ML models usually encounter a common drawback: they could precisely make short-term forecasts but could not reliably predict the occurrence of long-term incidents in time because of the weak persistent behavior of the underlying data especially on long-horizons.

High spatio-temporal variability in $PM_{2.5}$ concentration poses a huge challenge in regional long-term $PM_{2.5}$ forecasting. Facilitating the advantages of both ACT and machine learning models would bear great potential to improve $PM_{2.5}$ forecast accuracy, and therefore would benefit the issuance of warnings to prevent public exposure to air pollution. This motivates us to explore a novel hybridization approach grounding on machine learning techniques driven by ensemble inputs to tackle the high nonlinearity of regional $PM_{2.5}$ forecasts at horizons extending from 1 h up to 72 h (3 days ahead). We aim to improve forecast accuracy and reliability by effectively learning the spatio-temporal features of inputs over the period spanning from the past to the future based on an ensemble of two datasets: the monitoring data; and the forecasts from the ACT model. The proposed approach (MCNN-BP) is configured with two CNNs responsible for extracting important information separately from the observed and forecasted datasets as well as one BPNN responsible for $PM_{2.5}$ forecasting. The CNN-LSTM-BP model (configured with CNN, long short term memory neural network (LSTM), and BPNN) and the ACT model are the benchmarks adopted in this study. 74 air quality monitoring stations spreading over the whole Taiwan constitute the case study. The Kriging method is used to display regional $PM_{2.5}$ forecasts in a two-dimensional (2D) map. The proposed MCNN-BP model makes a big attempt to offer $PM_{2.5}$ forecasts with multiple site/horizon attributes simultaneously. The study area, materials, and methodology are introduced in Sections 2 and 3, respectively. Section 4 presents the statistics of input datasets and forecast results. Concluding remarks are drawn in Section 5.

2. Study area and materials

A total of 74 air quality monitoring stations in Taiwan are used. We first partition Taiwan into four regions (R1, R2, R3, and R4 representing northern, central, southern, and eastern regions, respectively, Fig. 1) based on human activities. The northern region (R1), especially Taipei City, is Taiwan's economic center. Since R1 is the most populated region with heavy traffic loads, the government has set up 26 air quality monitoring stations in R1. Likewise, there are 23 and 17 monitoring

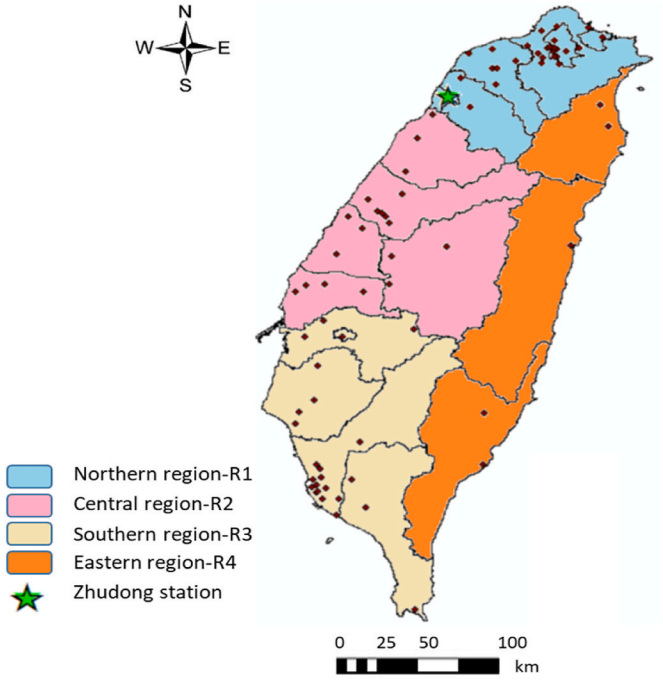

Fig. 1. Illustration of the study area and four regions (R1-R4) in Taiwan.

stations set up in southern (R3) and central (R2) regions, respectively. It is noted that Kaohsiung City in R3 is the second-largest city in Taiwan and embraces a heavy industrial area, where AQI often soars above 100. The air pollution control policy should be better implemented with strict environmental monitoring activities in response to industrial emissions. Eastern Taiwan (R4) is the least populous region famous for beautiful scenes. The main human activities in R4 are tourism. The concentration of each pollutant is obviously the lowest in Taiwan (Table 1).

There are two types of datasets for use in this study: the observation dataset (denoted as the EPA dataset) from the Environmental Protection Administration (TW EPA) and the simulated (forecasted) dataset (denoted as the AS dataset) from the ACT model (WRF-Chem) of the Academia Sinica in Taiwan (TW AS). The EPA dataset is composed of air quality and meteorological data extracted from the open data platform of the TW EPA (EPA, 2021). The AS dataset provides a huge dataset of three-dimensional regional air quality (e.g., $PM_{2.5}$) for the next 72 h. Eight input factors related to $PM_{2.5}$ formations were collected from the TW EPA in the period spanning between January 1, 2019 and June 30, 2021, including particulate matter ($PM_{2.5}$ and PM_{10}), O_3, nitrogen dioxide (NO_2), sulfur dioxide (SO_2), carbon monoxide (CO), relative humidity, and ambient temperature. It is worth noting that model inputs at time step t consist of observations (t-23, …, and t) extracted from the EPA dataset and the forecasts (t+1, …, and t+72) extracted from the AS dataset. A total of 64,861 hourly datasets (=2.5 years × 365 days x 74 stations-2664 missing datasets) were collected, where 41,511 datasets (64%) and 10,378 datasets (16%) associated with the period of January 1, 2019 and December 31, 2020 were shuffled and used in model training and validation, respectively, while 12,972 datasets (20%, a time series from January to June in 2021) were used for testing.

3. Methodology

3.1. Motivations

This study intends to increase the accuracy of 3-day-ahead regional $PM_{2.5}$ forecasts by utilizing advanced ML techniques upon ensemble inputs of the antecedent EPA monitoring dataset (representative of the past) and 3-day-ahead WRF-Chem forecasts (representative of the

Table 1
Results of statistical analyses on air quality and meteorological data (January 1, 2019−June 30, 2021).

Dataset	Factor	Indicator	R1 (North)	R2 (Central)	R3 (South)	R4 (East)	Whole Taiwan
EPA dataset	$PM_{2.5}$	Q3[a]	18.0	25.0	32.0	12.0	24.0
	($\mu g/m^3$)	Mean	13.9	18.3	23.0	9.2	17.5
		Standard deviation	9.2	13.3	16.1	5.6	13.3
		Q1[b]	8.0	8.0	10.0	5.0	8.0
	PM_{10}	Q3	35.0	45.0	60.0	27.0	45.0
	($\mu g/m^3$)	Mean	27.7	34.6	43.8	21.0	34.0
		Standard deviation	18.4	22.7	26.1	12.4	23.0
		Q1	16.0	19.0	23.0	13.0	18.0
	NO_2	Q3	22	15	18.9	8.1	18
	(ppb)	Mean	16	11.4	13.6	6.4	13.3
		Standard deviation	10.1	6.7	8.9	4.1	9.1
		Q1	8.2	6.5	6.9	3.5	6.5
	SO_2	Q3	3	3	3.7	1.9	3.1
	(ppb)	Mean	2.5	2.5	3	1.6	2.6
		Standard deviation	1.5	1.4	2	0.8	1.6
		Q1	1.6	1.7	1.8	1.2	1.6
	O_3	Q3	36	31.2	31.9	31	33
	(ppb)	Mean	26.7	24.6	24.7	24.1	25.6
		Standard deviation	13.4	11.4	12	11	12.6
		Q1	16.7	16.5	16	15.9	16
	CO	Q3	0.7	0.5	0.6	0.4	0.6
	(ppm)	Mean	0.5	0.4	0.4	0.3	0.5
		Standard deviation	0.4	0.3	0.3	0.1	0.3
		Q1	0.3	0.2	0.3	0.2	0.3
	Relative	Q3	84	83	83	83	84
	Humidity	Mean	75.2	75.5	76.5	75.2	75.8
	(%)	Standard deviation	12	10.9	9.5	10.6	10.9
		Q1	66	68	70	67	68
	Ambient	Q3	28.6	28.4	29.2	28.6	28.8
	temperature	Mean	23.4	23.7	25.1	24.1	24
	(°C)	Standard deviation	5.7	5.6	4.9	5.1	5.4
		Q1	18.9	19.3	21.4	20	19.7
AS dataset	$PM_{2.5}$ ($\mu g/m^3$)	Q3	19.3	24.2	24.2	11.7	23.1
		Mean	15.0	18.2	19.7	8.7	17.4
		Standard deviation	12.6	13.9	16.3	7.4	16.3
		Q1	6.8	8.3	8.3	3.6	6.6

[a] Third Quartile.
[b] First Quartile.

future). This study explores a hybridization approach (MCNN-BP) driven by two deep learning CNNs coupled with BPNN for producing reliable and suitable regional long-horizon $PM_{2.5}$ forecasts. The deep learning model (CNN-LSTM-BP) and the WRF-CHEM model constitute two benchmarks for comparison purpose.

Fig. 2 illustrates the proposed deep learning MCNN-BP framework for producing multi-station and multi-horizon $PM_{2.5}$ forecasts. MCNN-BP is divided into two phases: feature extraction and forecasting phases. Fig. 2(a) shows that CNN is capable of learning the spatial pattern and extracting the temporal feature of each sample. Fig. 2(b) shows that BPNN is responsible for learning the nonlinear structure from the extracted features of the corresponding time-series data. The model flowchart is illustrated in Fig. 2(c). The methods adopted are described in the following sections.

3.2. Convolutional neural network (CNN)

The CNN has become a popular feed-forward DNN in recent years because of breakthroughs in hardware. With the great capability of high-level feature extraction, the CNN can effectively distinguish one sample from the others through weighting each sample. Besides, its feature extraction can be carried out without preprocessing noise removal from datasets, which is crucial to improve model performance. The CNN structure is mainly formed by convolutional layers (feature extraction layers), pooling layers (dimension reduction layers), and fully connected layers (flatten layers). However, pooling layers may remove important information when tackling the time series problem. As a result, this study does not incorporate any pooling layer into CNN. Instead, CNN acts as a filter to extract important information while filtering out noise from samples (Ghorbani and Behzadan, 2021; Yurtsever and Yurtsever, 2019). Therefore, the CNN can stack more layers to extract more information and conduct training to prevent overfitting. Because the CNN can effectively capture the spatio-temporal dependencies in each sample and distinguish dominant features from low-level ones in samples, it has been widely used in different fields (Cao et al., 2021; Garajeh et al., 2021; Pelletier et al., 2019). The implementation of the CNN in this study is briefly introduced below.

In this study, each of the 74 stations has 913 samples (2.5 years * 365 days), where each sample is composed of 24-h antecedent data of 8 input variables (6 air-quality and 2 meteorological variables) extracted from the EPA dataset and 72-h-ahead $PM_{2.5}$ concentration extracted from the AS dataset. The EPA dataset and the AS dataset are separately fed into two CNN models, where the numbers of CNN filters are set to be 24 and 36, respectively (Table 2, Fig. 2). This is mainly because the contents and dimensions of both datasets are very different. For instance, the EPA dataset (representing the past) has 8 variables ($PM_{2.5}$, PM_{10}, O_3, NO_2, SO_2, CO, relative humidity and ambient temperature) with a dimension of 192 (=24 time steps (t-23, …, and t) x 8 variables); and the AS dataset (representing the future) has 1 variable ($PM_{2.5}$) with a dimension of 72 (=72 time steps (t+1, …, and t+72) x 1 variable).

For each dataset, the filtering process of its associated CNN is conducted on each sample, where small matrices (weights) of CNN windows are “screening” over each sample. In other words, the filtering process carries out the multiplication of each sample with the weights of windows (Kim, 2017; O’Shea & Nash, 2015). It is worth mentioning that the core of the CNN is "weight sharing," which allows the weights of each CNN filter to remain unchanged during feature extraction from each sample (Abdel-Hamid et al., 2014). That is to say, the weights for each

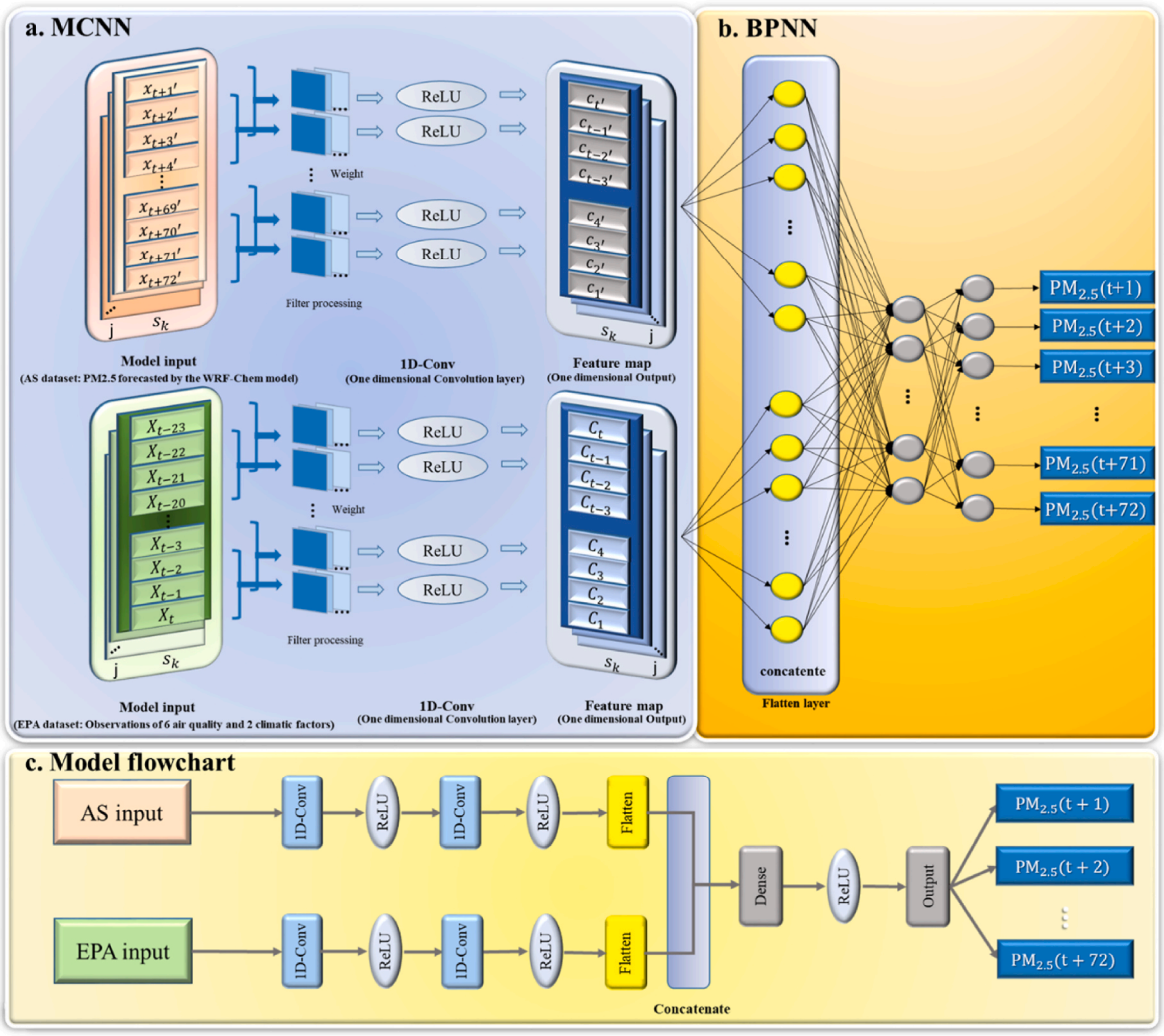

Fig. 2. The MCNN-BP deep learning approach (fusing the convolutional neural network (CNN) with a regression classifier (BPNN)). (a) Architecture of MCNN. (b) Architecture of BPNN. (c) MCNN-BP flowchart.

Table 2
Parameter settings for DNN models.

Model	Parameters							
	Epochs	Filter/Neuron	Hidden layer	Learning rate	Batch size	Kernel size	Patience (early stopping)	Optimizer
MCNN-BP	50	36 filters/ 24 filters 36 neurons 72 neurons	2 Conv. [a] layers/ 2 Conv. layers 1 FC [b] layer 1 FC layer	0.0001	64	3	15	Adam
CNN-LSTM-BP	50	36 filters/ 24 neurons 36 neurons 72 neurons	2 Conv. layers/ 2 LSTM [c] layers 1 FC layer 1 FC layer	0.0001	64	3	15	Adam

[a] Convolutional.
[b] Fully-connected.
[c] Long short term memory layer.

sample remains the same during each filtering process of the CNN, which is distinct from the behaviors other ANNs whose layers have different weights at different time steps for each sample.

A filter window determines how many time lags to screen at a time. Because this study utilizes one-dimensional CNNs (Fig. 2(a)), the length of the filter window is the only parameter to set. The number of connections between the input layer and the convolutional layer conforms to the length of the filter window, and the number of connections between the convolutional layer and the output layer is 1 (feature map).

CNN is a supervised neural network and undergoes two phases to train the model, i.e. the forward propagation and the backward propagation. In this study, the one-dimensional CNN is applied because the model requires time features only, implying CNN filters screen each sample along one direction (time step). The formulas for the one-

dimensional CNN are expressed as follows.

Forward propagation (Indolia et al., 2018):

If w_j connects the input layer and the hidden layer,

$$C_i^\ell = \sigma\left(X_i^\ell\right) \quad \text{with} \quad X_i^\ell := \sum_{j=1}^{m} w_j^\ell X_{i+j}^{\ell-1} \tag{1}$$

else if w_j connects two hidden layers,

$$C_i^\ell = \sigma\left(X_i^\ell\right) \quad \text{with} \quad X_i^\ell := \sum_{j=1}^{m} w_j^\ell C_{i+j}^{\ell-1} \tag{2}$$

where σ is the activation function (i.e. ReLU in this study), $i = 1, 2, \cdots, t$, $X_{i+j}^{\ell-1} = (X_{1+j}^{\ell-1}, X_{2+j}^{\ell-1}, \ldots X_{t+j}^{\ell-1}) \in \mathbb{R}^n$, $C_{i+j}^{\ell-1} = (C_{1+j}^{\ell-1}, C_{2+j}^{\ell-1} \cdots, C_{t+j}^{\ell-1}) \in \mathbb{R}^n$, $X_i^\ell = (X_1^\ell, X_2^\ell \cdots, X_t^\ell) \in \mathbb{R}^n$

Backward propagation (Wagaa & Kallel, 2020; Zhou, 2018):

The formulas of the backward propagation can be simplied as follows.

$$\frac{\partial E^\ell}{\partial w_i^\ell} = \sum_{j=1}^{n} \delta_i^\ell C_{i+j}^{\ell-1} \tag{4}$$

where E is the error function, δ is the partial derivative, $C_{i+j}^{\ell-1} = C_{i+j-n}^{\ell-1}$ if $i + j > n$

The formula of δ is expressed as follows.

$$\delta_i^{\ell-1} = \sum_{j=1}^{n} \delta_{i-j}^\ell w_j^\ell \tag{5}$$

where $\delta_{i-j}^\ell = \delta_{i-j+n}^\ell$ if $i - j \le 0$, $i = 1, \cdots, n$.

$$\delta_{i-j}^\ell = \delta_{i-j}^\ell \sigma'\left(X_i^\ell\right) \tag{6}$$

where $i = 1, \cdots, n$.

Replacing X_i^ℓ and C_j^ℓ with X_j^ℓ and $C_j^{\ell-1}$ in Eqs. (4) and (5), respectively,

$$\frac{\partial E^{\ell+1}}{\partial w_i} = \sum_{j=1}^{n} \delta_j^{\ell+1} \cdot \sigma'\left(X_j^\ell\right) \cdot C_{i+j}^{\ell-1} \text{ and } \delta_i^\ell = \sum_{j=1}^{n} \delta_j^{\ell+1} w_{i-j}^{\ell+1} \sigma'\left(X_j^\ell\right). \tag{7}$$

In CNN, the weight and the error term are arrays. The error term is expressed as follows.

$$\delta_i^\ell = conv\left(\delta_i^{\ell+1}, rot180\left(w_i^{\ell+1}\right)\right).\sigma'\left(X_j^\ell\right). \tag{8}$$

where the conv(.) is the convolution operation, and rot180(.) denotes the 180-degree rotation to make the convolution function perform cross-correlation (Bouvrie, 2006).

More details of CNN can be found in Albawi et al. (2017).

3.3. Long short term memory neural network (LSTM)

Recurrent neural networks (RNNs) have been widely used in time series forecasting. The LSTM is one of the famous RNNs in the deep learning field. The LSTM passes the cell state and the hidden state at the current horizon to the next horizon to preserve long-term memory and short-term memory, respectively (Wang et al., 2021). The LSTM is commonly composed of a cell, an input gate, an output gate and a forget gate to learn the continuous information of samples during model training. More details of the LSTM can be found in Yokoo et al. (2022).

3.4. Hybrid of multiple CNNs and BPNN (MCNN-BP)

The proposed MCNN-BP model seamlessly fuses two CNNs with a BPNN to make long-term regional air quality forecasts. The configuration of MCNN-BP can be divided into two phases: the feature extraction phase using CNNs; and the forecasting phase using BPNN. The model utilizes two types of input datasets (i.e. EPA and AS datasets) and individually handles the two datasets by two CNNs. Each output of the CNN contains multiple time attributes. The merits of the model are that similarity in patterns among samples is considered to be an auxiliary to model training while the noise in the dataset can be removed to improve forecast accuracy. CNNs can effectively extract the spatio-temporal features from the sample (Fig. 2(a)). The features extracted by the CNNs of the first layers are low-level features, and the superposition of low-level features form high-level features in the following layers. In the forecasting phase, the output of each CNN is flattened and concatenated in the flatten layer that is configured with two one-dimensional parallel convolutional layers determined by trial and error procedures. Then, the BPNN is responsible for making forecasts (Fig. 2(b) and (c)). 72 neurons are set up in the output layer to generate 72-h-ahead $PM_{2.5}$ forecasting.

To improve long-term forecast accuracy, the proposed model is fed with high-dimensional datasets, including the EPA dataset (horizons t-24, …, and t) and the AS dataset (horizons t+1, …, and t+72). The reason for developing the model to tackle this highly nonlinear problem is to avoid gradient vanishing and gradient explosion for aggregating information from the past to the future. As a result, the MCNN-BP model not only can overcome the bottlenecks encountered by traditional neural networks that depend solely on observed datasets but also can extend the forecast horizon to t+72 based on the past (the EPA dataset) and the future (the AS dataset) information.

3.5. Hybrid of CNN, LSTM, and BPNN (CNN-LSTM-BP)

The CNN-LSTM-BP model has three phases, which are the enhanced learning phase, the feature extraction phase, and the forecasting phase. It differs from the MCNN-BP model by replacing one CNN with an LSTM to form the enhanced learning phase for handling the EPA dataset, where the LSTM enhances model learning ability by passing current information to the next few time steps. The feature extraction phase is responsible for capturing important features from the AS dataset. Then, both the LSTM output and the flattened CNN output are concatenated to feed into the BPNN for simultaneously producing hourly $PM_{2.5}$ forecasts of 74 monitoring stations at multiple horizons.

3.6. The Weather Research and Forecasting–chemistry (WRF-Chem) model

The WRF-Chem model (version 3.9) that integrates weather (WRF) and air quality (Chem) forecast systems is a three-dimension regional air quality modeling system (Grell et al., 2005). The WRF module provides a comprehensive simulation of the movement of weather systems, representing the physical movement of air mass subject to temperature and pressure (Hafeez et al., 2021). The Chem module simulates the emission and chemical reaction of air pollutants, representing the fate and budget of air pollutants in the atmosphere (Ding et al., 2021). The hybridization of WRF and Chem can forecast multiple pollutant concentrations under different meteorological scenarios (Casazza et al., 2019). Zong et al. (2020) indicated that the simulation of $PM_{2.5}$ obtained from the WRF-Chem model produced good results. The WRF-Chem model has been used for the source apportionment and long-term forecasting of $PM_{2.5}$ (Liu et al., 2018; Reátegui-Romero et al., 2018). In this study, the emission, transport, and transformation of air pollutants within East Asia and Taiwan were simulated with a grid resolution of 9 km and 3 km, respectively. The global weather forecast data from the Operational Global Forecast System of the National Center for Environmental Prediction (NCEP) were deployed to initiate the regional model. The concentrations of $PM_{2.5}$, PM_{10}, O_3, and other critical air pollutants in Taiwan for the next 72 h were forecasted and reported on a daily basis during the study period.

3.7. Evaluation indicators

This study adopts the Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and the coefficient of determination (R^2) to evaluate forecast results.Their formulation can be found in Kow et al. (2020) and Pal (2016). RMSE represents the error between forecasted values and observed data, which magnifies the peak value of forecast error. MAE refers to the average magnitude of absolute errors between forecasted values and observed data. R^2 indicates the correlation between forecasted values and observed data, and its value ranges from 0 to 1. A model with R^2 value closer to 1 implied it offers more accurate prediction.

4. Results and discussion

The constructed MCNN-BP model can deeply learn the observed data extracted from the EPA dataset and the $PM_{2.5}$ forecasts extracted from the AS dataset. Another DNN model (i.e., CNN-LSTM-BP) and the WRF-Chem model are selected for comparison purpose. The results and findings are presented in the order of data preprocessing, model construction, and model comparison, and are shown as follows.

4.1. Data preprocessing

Since there are various heterogeneous data, data normalization is necessary before building the model. Data normalization scales the feature data proportionally to adjust each data into a common scale (e.g., 0 to 1). Then the normalized samples are randomly allocated into training (64%), validation (16%), and test (20%) datasets.

4.2. Preliminary analysis of each pollutant factor

Table 1 shows the statistical results of air quality and meteorological data used in this study. R1 has the large standard deviation values of NO_2, SO_2, and CO, which indicate the concentrations of these three substances have drastic changes. This provides evidence that the main emission sources are steam and locomotives in R1. The highest mean values of $PM_{2.5}$ and PM_{10} concentrations can be observed in R3, which was due to industrial emissions. R2 has the second high mean values of $PM_{2.5}$ and PM_{10} concentrations because Central Taiwan has been developed rapidly in recent years. Besides, R2 and R3 have relatively large standard deviations of $PM_{2.5}$ concentrations. R1 has a relatively low mean and standard deviation of $PM_{2.5}$ but the highest mean and standard deviation of O_3. R4 has the least industrial and commercial activities, and therefore its mean and standard deviation of $PM_{2.5}$, PM_{10}, and O_3 concentrations is the lowest.

According to Liu et al. (2009), meteorological factors are also important factors affecting $PM_{2.5}$ concentration. The deposition process of $PM_{2.5}$ has a high correlation with relative humidity because moisture attaches to fine particles and accumulates to bigger sizes. The relative humidity of the whole Taiwan is very high, with an annual mean value exceeding 70%, implying relative humidity is a crucial factor affecting $PM_{2.5}$ (Cao et al., 2021; Zalakeviciute et al., 2018). The temperature is also an important factor affecting $PM_{2.5}$ (Chen et al., 2021). The comparison of the mean values of $PM_{2.5}$ associated with EPA and AS datasets reflects that there exists bias of $PM_{2.5}$ in R1 and R3 since the mean value in the AS dataset is significantly higher in R1 but lower in R3. Besides, the standard deviations of the AS dataset are higher than those of the EPA dataset in all regions (R1-R4). In addition, the errors of the AS dataset can be observed, which also reflects there exist bias in the AS dataset.

One of the characteristics of ANN is fault-tolerant. That is, a few extreme values do not significantly affect model training (Duddu et al., 2020). In this study, we define extreme values to be either smaller than mean-3*standard deviation or larger than mean + 3*standard deviation (Jawlik, 2016). The proportions of outliers in the entire dataset are 1.3%, 0.16%, 2.3%, 0.5%, 1.4%, 0.2%, 1.2% and 1.1% for SO_2, ambient temperature, CO, O_3, $PM_{2.5}$, relative humidity, PM_{10} and NO_2, respectively, which are small. In this study, we delete most of the outliers while reserving some noise to test the denoise ability of the proposed model.

4.3. Parameter setting

Table 2 presents the parameter settings for the two DNN models. The most suitable values of parameters are determined by validation errors after executing a large number of trial and error procedures. The kernel size of each CNN filter is set to be 3, leading to the filter length and width being 3. The filter and the neuron number are set to be the time step of each sample divided by an integer (e.g., 1,2,3,..) to avoid overfitting. Furthermore, patience (early stopping) is set to be 15, indicating the training process will terminate earlier if there is no further reduction in the error after 15 consecutive iterations. The Adam optimizer that has the ability of both Adagrad (an optimizer that adjusts the learning rate by the gradient) and momentum is selected.

Moreover, in order to avoid gradient vanishing problems caused by the complex structure of the hybrid DNN model, the numbers of hidden layers in the convolution layer and the fully-connected layer, filters and neurons are set to be small values (a shallow architecture) in this study (Hochreiter, 1998; Kolbusz et al., 2017).

4.4. Comparison between DNN models and the WRF-Chem model for $PM_{2.5}$ forecasting

4.4.1. Performance of DNN forecasts

We first show the reasons why the ensemble of inputs is important for the deep learning model. To fully investigate the forecast effect on consecutive forecast horizons, an air pollution incident at the QianZhen air quality monitoring station in R1 (Fig. 1) was selected because of its high variation in $PM_{2.5}$ concentrations. Fig. 3 gives the $PM_{2.5}$ forecast results obtained from WRF-Chem and MCNN-BP models at consecutive horizons from 2021/01/27 to 2021/01/31.

Fig. 3(a) gives consecutive 72-h (3-day) $PM_{2.5}$ forecasts starting from 2021/01/27 8am. The MCNN-BP model performs satisfactorily especially when $PM_{2.5}$ concentration exceeds 35 μg/m3. This result indicates that the MCNN-BP model not only deeply mines the EPA dataset that strengthens the short-term forecast effect but also learns the future $PM_{2.5}$ trend from the AS dataset that produces good long-term forecast performance.

Fig. 3(b) gives consecutive 72-h $PM_{2.5}$ forecasts starting from 2021/01/28 8am. In this air pollution incident, the MCNN-BP and the WRF-Chem models behave quite differently. The MCNN-BP model performs satisfactorily even though there is a great variation of $PM_{2.5}$ concentration (>45 μg/m3) in this incident. The result indicates that the MCNN-BP model has a robust short-term forecasting ability for sudden incidents. For instance, when the concentration of an air quality factor in the EPA dataset suddenly increases, there will be a high probability for the MCNN-BP model to make a forecast of high $PM_{2.5}$ concentrations. Besides, the MCNN-BP model also inherits the long-term forecasting ability from the WRF-Chem model. Therefore, the MCNN-BP model can mimic the long-term $PM_{2.5}$ trend through the air pollution diffusion mechanism. The reason why the MCNN-BP model can better foresee sudden incidents in both short-term and long-term forecasting is attributed to its ability that learns in-depth features from observed (the past) and forecasted (the future) datasets. The forecasts of the WRF-Chem model evidently played an important role in improving the accuracy of the proposed MCNN-BP model.

Fig. 3(c) gives 72 h $PM_{2.5}$ forecasts starting from 2021/01/29 8am. In this case, both the MCNN-BP and the WRF-Chem models successfully forecasted the occurrence of a pollution incident in the future. The results of the MCNN-BP model fluctuate less because it can deeply excavate historical data to correct the forecast results of the AS dataset and make its forecasts more reasonable.

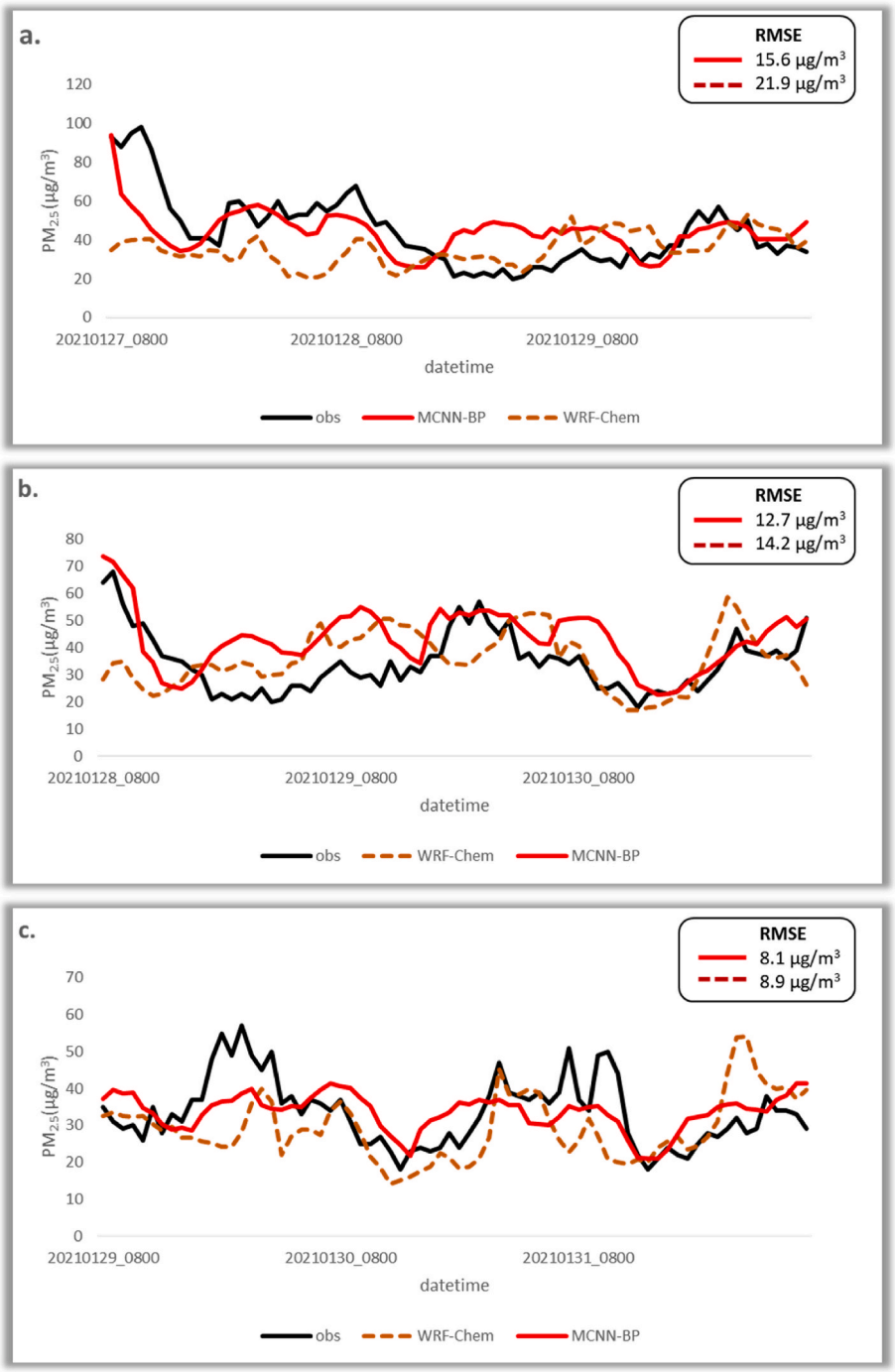

Fig. 3. $PM_{2.5}$ forecast results of WRF-Chem and MCNN-BP at consecutive horizons (72-h) starting from 2021/01/27 to 2021/01/29 at the QianZhen station, with initial forecast times at: (a) 2021/01/27 08:00:00; (b) 2021/01/28 08:00:00; and (c) 2021/01/29 08:00:00.

In brief, the three $PM_{2.5}$ incidents shown in Fig. 3 imply that an ensemble of inputs representative of the past and the future has high potential to improve the performance of the MCNN-BP model. Although the CNN-BP model occasionally underestimates $PM_{2.5}$, it still can suitably forecast $PM_{2.5}$ trends by learning $PM_{2.5}$ features in the past, which improves the outcome of the WRF-Chem model. The forecast accuracy made by the MCNN-BP is satisfactory because of its capability to remove noises from datasets and to extract useful information from two input datasets (EPA and AS datasets).

4.4.2. Performance of different DNN models for $PM_{2.5}$ with ensemble inputs

We further investigate the two DNN models for $PM_{2.5}$ forecasting. The hourly performance of both DNN models and the benchmark model (WRF-Chem) at t+24, t+48, and t+72 are shown in Table 3. The results show that MCNN-BP outperforms CNN-LSTM-BP in training, validation and testing phases. It is worth mentioning that the MAE and RMSE values of both models are relatively larger in the testing phase than in the training and validation phases. This is because some data in the test dataset are comparatively higher than those in training and validation datasets.

It is noted from Table 3 that the uncertainty of the AS dataset is noticeable and increases with the time horizon because the linearity between the WRF-Chem forecasts (the AS dataset) and observations (the EPA dataset) declines as the time horizon increases in term of R^2 (0.36,

Table 3
Hourly performance (R^2, RMSE and MAE) of the DNN models for $PM_{2.5}$ in the training, validation, and testing phases and WRF-Chem in the testing phase at t+24, t+48 and t+72.

Indicator	Horizon	Training		Validation		Testing		
		CNN-LSTM-BP	MCNN-BP [a]	CNN-LSTM-BP	MCNN-BP	CNN-LSTM-BP	MCNN-BP	WRF-Chem[b]
R^2	t+24	0.60	0.63 (5.00)	0.55	0.56 (1.82)	0.54	0.55 (1.85)	0.36
	t+48	0.57	0.6 (5.77)	0.50	0.52 (4.55)	0.49	0.5 (2.04)	0.31
	t+72	0.52	0.55 (5.77)	0.44	0.46 (1.00)	0.47	0.47 (0.00)	0.31
RMSE	t+24	7.75	7.48 (3.48)	8.00	7.92 (1.00)	11.62	11.37 (2.15)	14.43
($\mu g/m^3$)	t+48	8.20	7.79 (5.00)	8.61	8.35 (3.02)	12.49	12.03 (3.68)	14.77
	t+72	8.59	8.36 (2.68)	9.15	9.07 (0.87)	12.43	12.34 (0.72)	14.65
MAE	t+24	3.62	3.09 (14.64)	3.76	3.3 (12.23)	4.88	3.92 (19.67)	9.19
($\mu g/m^3$)	t+48	5.46	5.27 (3.48)	5.72	5.66 (1.05)	7.46	7.36 (1.34)	9.36
	t+72	5.68	5.48 (3.52)	6.13	6.02 (1.79)	7.99	7.85 (1.75)	9.61

[a] Value in parenthesis denotes the improvement of MCNN-BP over CNN-LSTM-BP (%).
[b] Atmospheric Chemical Transport (ACT) model.

0.31 and 0.31 for horizons T+24, T+48, and T+72, respectively, Table 3). If we use one CNN to handle both datasets, the AS dataset with larger uncertainty will be assigned smaller weights due to the low R^2 values between the AS dataset and the EPA dataset, compared to the EPA dataset. As a result, the importance of the AS dataset to the MCNN-BP model will be significantly reduced. In other words, two separate CNN models (i.e. one multivariate CNN model and one univariate CNN model) are implemented for deeply extracting features from the historical observation of the EPA dataset and learning the future $PM_{2.5}$ trend and corresponding uncertainty from the AS dataset separately.

We further calculate the daily performance of $PM_{2.5}$ forecasts by averaging the hourly forecasts over 24 h, as shown in Table 4. The results show that MCNN-BP outperforms CNN-LSTM-BP in all three phases. It is worth mentioning that forecast performance summarized on a daily basis is superior to hourly forecast performance in the terms of higher R^2, lower RMSE and lower MAE values. This result proves that the DNN model can be suitable and reliable to make 3-day-ahead $PM_{2.5}$ forecasts.

The MCNN-BP model is a static model, which means it neither stores nor passes the previous state to the next time step. Therefore, model performance would decrease in the testing phase when the data structure of the testing phase is inconsistent with those of the training and validation phases. As for model efficiency, the MCNN-BP model has a shorter (28.6% less) computational time than the CNN-LSTM-BP model. According to forecast performance, especially in training and validating phases, as well as computational time, the MCNN-BP model is superior to the CNN-LSTM-BP model. The main reason for the MCNN-BP model to make such achievement is its superior ability of the convolution layer for feature extraction and noise removal from EPA and AS datasets. In contrast, the CNN-LSTM-BP model has a weak denoise ability, despite its ability to transmit the information to the next few time steps to enhance model forecasting.

It is also crucial to further validate the capability and reliability of the proposed model. The comparative results demonstrate that the MCNN-BP model could significantly improve $PM_{2.5}$ forecasts and produce much higher accuracy than the WRF-Chem model. The reason is that the model effectively extracts the observed data (the past) and the forecasted data from WRF-Chem (the future), which can improve model ability through learning the future trend (the next 72 h' $PM_{2.5}$ forecasts). In other words, the MCNN-BP model is capable of deeply extracting the time feature from high-dimensional datasets, removing the noise from the multiple convolution layers, and converting the feature into useful information in the stacked fully connected layer.

It is noted that the WRF-Chem model operated on a server consumed about 8 h to produce $PM_{2.5}$ forecasts and the MCNN-BP model operated on a desktop computer consumed only about 5 min to produce $PM_{2.5}$ forecasts. Therefore, the overall computation time of the operational forecasting model integrating WRF-Chem and MCNN-BP will exceed 8 h because the MCNN-BP model requires the output of the WRF-Chem model to carry out $PM_{2.5}$ forecasting.

4.5. Regional multi-step-ahead $PM_{2.5}$ forecasting

To evaluate the forecasts performances from different perspectives, we chose two severe pollution incidents for further investigation. Fig. 4 gives the regional multi-step-ahead $PM_{2.5}$ forecasts (t+24 and t+72) obtained from the MCNN-BP model over the whole Taiwan at 8am on January 27, 2021 and February 04, 2021. As shown, the proposed model, in general, could provide suitable and reliable long-term forecasts, where the MAE values for the whole Taiwan are fairly small (less than 5 μg/m3 mostly). Fig. 4(a) gives 3-day-ahead $PM_{2.5}$ forecasts starting from 2021/01/27 8am. The immense MAE value occurs in the southern region (R3) at t+24 because the DNN model obviously overestimated $PM_{2.5}$. At t+72, although $PM_{2.5}$ is overestimated in R2, the

Table 4
Daily (average over 24 h) performance (R^2, RMSE and MAE) of the DNN forecasting models for $PM_{2.5}$ in the training, validation, and testing phases at horizons T+1, T+2 and T+3.

Indicator	Horizon (days)	Training		Validation		Testing		
		CNN-LSTM-BP	MCNN-BP	CNN-LSTM-BP	MCNN-BP	CNN-LSTM-BP	MCNN-BP	WRF-Chem
R^2	T+1	0.79	0.82 (3.80)	0.75	0.77 (2.67)	0.70	0.72 (2.86)	0.44
	T+2	0.72	0.74 (2.78)	0.66	0.67 (1.52)	0.61	0.62 (1.64)	0.41
	T+3	0.68	0.71 (4.41)	0.59	0.62 (5.08)	0.58	0.58 (0.00)	0.40
RMSE	T+1	4.59	4.30 (6.32)	4.85	4.67 (3.71)	7.12	6.75 (5.20)	10.03
($\mu g/m^3$)	T+2	5.31	5.06 (4.71)	5.73	5.60 (2.27)	8.26	8.05 (2.54)	10.23
	T+3	5.73	5.46 (4.71)	6.44	6.26 (2.80)	8.53	8.50 (0.35)	10.20
MAE	t+24	3.18	2.99 (5.97)	3.42	3.31 (3.22)	4.59	4.46 (2.83)	7.02
($\mu g/m^3$)	t+48	3.70	3.52 (4.86)	4.04	3.95 (2.23)	5.47	5.38 (1.65)	7.19
	t+72	4.02	3.78 (5.97)	4.50	4.32 (4.00)	5.74	5.74 (0.00)	7.22

[a] Value in parenthesis denotes the improvement of MCNN-BP over CNN-LSTM-BP (%).
[b] Atmospheric Chemical Transport (ACT) model.

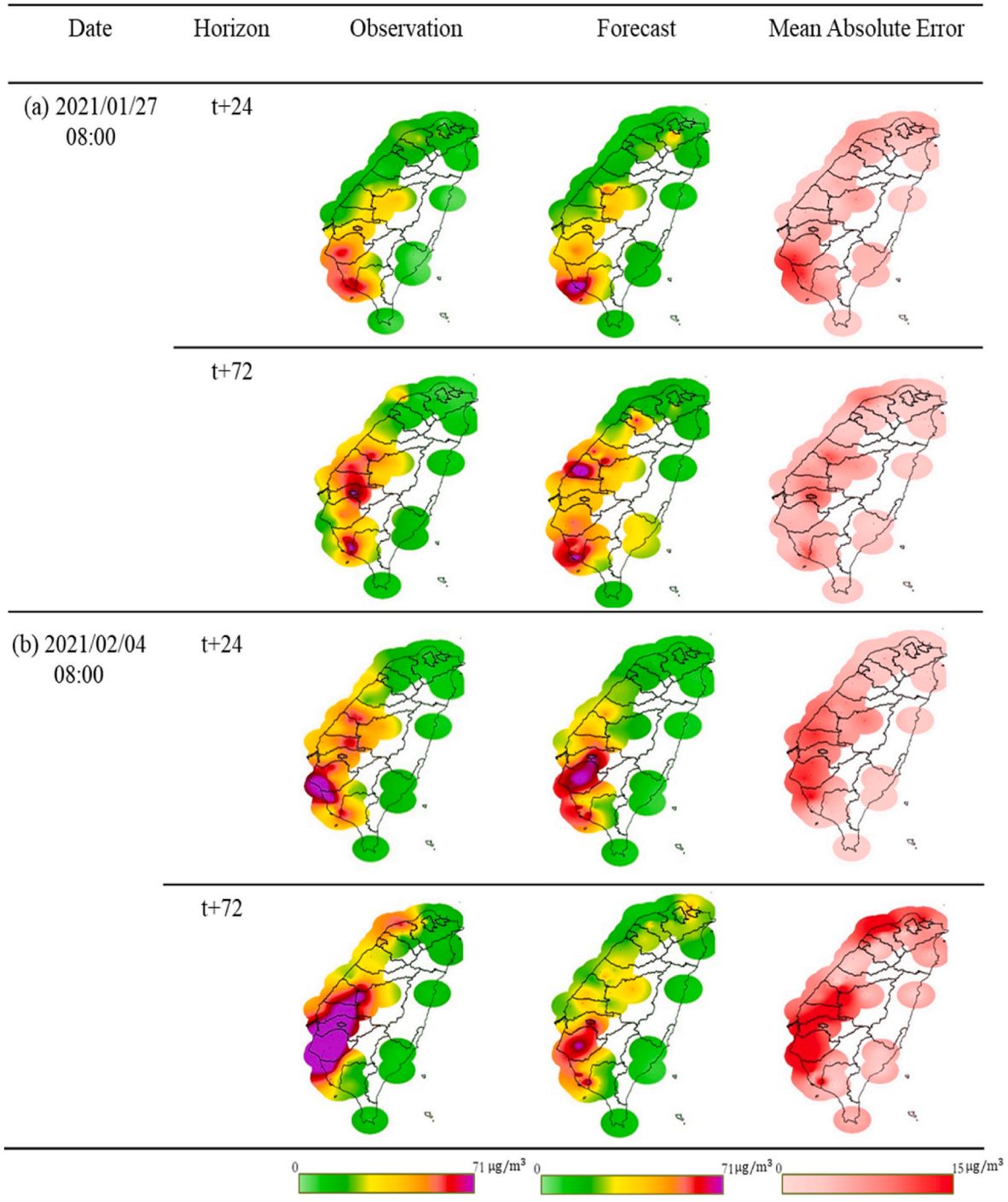

Fig. 4. $PM_{2.5}$ forecast results obtained from the MCNN-BP model at t+24 and t+72 for the whole Taiwan at (a) 2021/01/27 8am and (b) 2021/02/04 8am.

model still well forecasted the pollution incident occurring in R2 and R3 (red and yellow, respectively). The MAE values at t+24 and t+72 are considered very low because there was a serious pollution event in Taichung County (R2) and Kaohsiung County (R3). The main emission sources causing severe air pollution in Taichung and Kaohsiung are thermal power plants and heavy industrial plants, respectively (She et al., 2020; Wu et al., 2021). Furthermore, R2 and R3 are located on the leeward sides of mountain ranges, which is not conducive to atmospheric diffusion (EPA, 2022).

Fig. 4(b) gives 3-day-ahead $PM_{2.5}$ forecasts starting from 2021/02/04 8am. We notice that the proposed model overestimated $PM_{2.5}$ and produced a large MAE value in the southern region (R3) at t+24. There are also immense MAE values occurs in the northern region (R1), central region (R2) and southern region (R3) at t+72. In contrast, the proposed model underestimated $PM_{2.5}$ at t+72, especially in the western coast of Taiwan. This is mainly because transboundary pollution brought by the northeast monsoon makes the model underestimate $PM_{2.5}$ (EPA, 2022; Li et al., 2021). The results further demonstrate that the proposed hybrid deep neural network hits a milestone in suitably predicting $PM_{2.5}$ behavior and holistically addressing the interactive mechanism between air quality and meteorological factors.

5. Conclusion

The forecast of fine particulate matter concentration becomes a complex spatio-temporal problem due to a great variety of emission sources. Therefore, long-term multi-step-ahead air pollution forecasting is highly challenging. This study proposed a deep learning approach (MCNN-BP) to extract high-dimensional features from ensemble inputs and make multi-step-ahead (1- to 72-h) $PM_{2.5}$ forecasts at 74 sites concurrently. The input sources consisted of the dataset (EPA dataset) collected at air quality monitoring stations and the forecasts (AS dataset) given by the WRF-Chem model. The main contributions of the proposed MCNN-BP approach based on ensemble inputs are fourfold.

Firstly, the model can tackle the curse of dimensionality to produce reliable and suitable regional long-horizon forecasts $PM_{2.5}$. It has the merits to extract similarity in patterns among high dimensional samples to facilitate model training while removing the noise from datasets to improve forecast accuracy. Secondly, the model can foresee the occurrence of air pollution incidents through learning future information from the AS dataset and antecedent information from the EPA dataset. Thirdly, the model makes a breakthrough in extending the forecast horizon from a few hours to several days (3 days in our case). Fourthly, the comparative results demonstrated the MCNN-BP model with shorter computation time outperformed the CNN-LSTM-BP model and the WRF-Chem model.

In light of methodological transferability, future research can integrate the MCNN-BP with other methodologies such as the self-organizing map (SOM), a powerful clustering tool (Chang et al., 2020&, 2021). SOM can be implemented to reduce data's dimensionality and then increase the learning efficiency of subsequent MCNN-BP for improving model accuracy. Furthermore, real-time high temporal resolution satellite imagery can improve interpretability and predictive accuracy of air quality (Yan et al., 2021) and therefore can be a good auxiliary input to the MCNN-BP model for regional long-term $PM_{2.5}$ forecasting.

Credit author statement

Pu-Yun Kow: Conceptualization, Formal analysis, Methodology, Resources, Software, Validation, Visualization, Writing – original draft. Li-Chiu Chang: Conceptualization, Methodology, Formal analysis, Resources, Supervision. Fi-John Chang: Conceptualization, Methodology, Funding acquisition, Investigation, Project administration, Writing – review & editing. Chuan-Yao Lin: Methodology, Resources, Project administration. Charles C.-K. Chou: Methodology, Resources, Supervision, Project administration.

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.

Acknowledgments

This study is supported by the Ministry of Science and Technology, Taiwan (MOST: 109-2119-M-002-015-A; 110-2625-M-002-018-). The datasets provided by the Environmental Protection Administration and the Academia Sinica in Taiwan and the support of computation resources from the National Center for High Performance Computing in Taiwan are acknowledged. The authors would like to thank the Editors and anonymous Reviewers for their constructive comments that greatly contribute to enriching the manuscript.

References

Abdel-Hamid, O., Mohamed, A.R., Jiang, H., Deng, L., Penn, G., Yu, D., 2014. Convolutional neural networks for speech recognition. IEEE/ACM Transactions on audio, speech, and language processing 22 (10), 1533–1545.

Albawi, S., Mohammed, T.A., Al-Zawi, S., 2017, August. Understanding of a convolutional neural network. In: 2017 International Conference on Engineering and Technology (ICET). Ieee, pp. 1–6. https://doi.org/10.1109/ICEngTechnol.2017.8308186.

Bouvrie, J., 2006. Notes on Convolutional Neural Networks.

Cao, K., Zhang, Z., Li, Y., Zheng, W., Xie, M., 2021. Ship fuel sulfur content prediction based on convolutional neural network and ultraviolet camera images. Environ. Pollut. 273, 116501. https://doi.org/10.1016/j.envpol.2021.116501.

Casazza, M., Lega, M., Jannelli, E., Minutillo, M., Jaffe, D., Severino, V., Ulgiati, S., 2019. 3D monitoring and modelling of air quality for sustainable urban port planning: review and perspectives. J. Clean. Prod. 231, 1342–1352. https://doi.org/10.1016/j.jclepro.2019.05.257.

Chang, F.J., Chang, L.C., Kang, C.C., Wang, Y.S., Huang, A., 2020. Explore spatio-temporal PM2. 5 features in northern Taiwan using machine learning techniques. Sci. Total Environ. 736, 139656.

Chang, L.C., Wang, W.H., Chang, F.J., 2021. Explore training self-organizing map methods for clustering high-dimensional flood inundation maps. J. Hydrol. 595, 125655.

Chen, J.L., Lee, C.T., Chang, S.Y., Chou, C.C., 2001. The elemental contents and fractal dimensions of $PM_{2.5}$ in Taipei City. Aerosol Air Qual. Res. 1 (1), 9–20. https://doi.org/10.4209/aaqr.2001.06.0002.

Chen, H.S., Chen, K.S., Chen, C.Y., Hung, C.C., Meng, P.J., Chen, M.H., 2021. Spatiotemporal distribution of shrimp assemblages in the western coastal waters off Taiwan at the Tropic of cancer, western Pacific ocean. Estuarine. Coast. Shelf Sci. 255, 107356. https://doi.org/10.1016/j.ecss.2021.107356.

Ding, H., Cai, M., Lin, X., Chen, T., Li, L., Liu, Y., 2021. RTVEMVS: real-time modeling and visualization system for vehicle emissions on an urban road network. J. Clean. Prod. 309, 127166. https://doi.org/10.1016/j.jclepro.2021.127166.

Duddu, V., Rajesh Pillai, N., Rao, D.V., Balas, V.E., 2020. Fault tolerance of neural networks in adversarial settings. J. Intell. Fuzzy Syst. 38 (5), 5897–5907. https://doi.org/10.3233/JIFS-179677.

EPA, 2016. https://www.ey.gov.tw/File/70F71D7236BFF9AF?A=C.

EPA, 2021. https://data.epa.gov.tw/dataset/aqx_p_13.

EPA, 2022. https://airtw.epa.gov.tw/CHT/Encyclopedia/pedia08/pedia8_5.aspx.

Garajeh, M.K., Malakyar, F., Weng, Q., Feizizadeh, B., Blaschke, T., Lakes, T., 2021. An automated deep learning convolutional neural network algorithm applied for soil salinity distribution mapping in Lake Urmia, Iran. Sci. Total Environ. 778, 146253. https://doi.org/10.1016/j.scitotenv.2021.146253.

Ghorbani, Z., Behzadan, A.H., 2021. Monitoring offshore oil pollution using multi-class convolutional neural networks. Environ. Pollut. 289, 117884. https://doi.org/10.1016/j.envpol.2021.117884Get rights and content.

Goldberg, D.L., Gupta, P., Wang, K., Jena, C., Zhang, Y., Lu, Z., Streets, D.G., 2019. Using gap-filled MAIAC AOD and WRF-Chem to estimate daily $PM_{2.5}$ concentrations at 1 km resolution in the Eastern United States. Atmos. Environ. 199, 443–452. https://doi.org/10.1016/j.atmosenv.2018.11.049.

Grell, G.A., Peckham, S.E., Schmitz, R., McKeen, S.A., Frost, G., Skamarock, W.C., Eder, B., 2005. Fully coupled "online" chemistry within the WRF model. Atmos. Environ. 39, 6957–6975. https://doi.org/10.1016/j.atmosenv.2005.04.027.

Guo, Q., Wang, Y., Zhang, Y., Yi, M., Zhang, T., 2022. Environmental migration effects of air pollution: micro-level evidence from China. Environ. Pollut. 292, 118263. https://doi.org/10.1016/j.envpol.2021.118263.

Hafeez, M.A., Nakamura, Y., Suzuki, T., Inoue, T., Matsuzaki, Y., Wang, K., Moiz, A., 2021. Integration of Weather Research and Forecasting (WRF) model with regional coastal ecosystem model to simulate the hypoxic conditions. Sci. Total Environ. 771, 145290. https://doi.org/10.1016/j.scitotenv.2021.145290.

He, Z., Liu, P., Zhao, X., He, X., Liu, J., Mu, Y., 2022. Responses of surface O3 and PM2. 5 trends to changes of anthropogenic emissions in summer over Beijing during 2014–2019: a study based on multiple linear regression and WRF-Chem. Sci. Total Environ. 807, 150792. https://doi.org/10.1016/j.scitotenv.2021.150792.

Ho, C.C., Chen, L.J., Hwang, J.S., 2020. Estimating ground-level PM2. 5 levels in Taiwan using data from air quality monitoring stations and high coverage of microsensors. Environ. Pollut. 264, 114810. https://doi.org/10.1016/j.envpol.2020.114810.

Hochreiter, S., 1998. Recurrent neural net learning and vanishing gradient. Int. J. Uncertain. Fuzziness Knowledge-Based Syst. 6 (2), 107–116.

Hong, J., Mao, F., Min, Q., Pan, Z., Wang, W., Zhang, T., Gong, W., 2020. Improved PM2. 5 predictions of WRF-Chem via the integration of Himawari-8 satellite data and ground observations. Environ. Pollut. 263, 114451. https://doi.org/10.1016/j.envpol.2020.114451.

Hsu, C.H., Cheng, F.Y., 2016. Classification of weather patterns to study the influence of meteorological characteristics on $PM_{2.5}$ concentrations in Yunlin County, Taiwan. Atmos. Environ. 144, 397–408. https://doi.org/10.1016/j.atmosenv.2016.09.001.

Indolia, S., Goswami, A.K., Mishra, S.P., Asopa, P., 2018. Conceptual understanding of convolutional neural network-a deep learning approach. Procedia Comput. Sci. 132, 679–688. https://doi.org/10.1016/j.procs.2018.05.069.

Jain, S., Presto, A.A., Zimmerman, N., 2021. Spatial Modeling of Daily PM2. 5, NO2, and CO Concentrations Measured by a Low-Cost Sensor Network: Comparison of Linear, Machine Learning, and Hybrid Land Use Models. Environmental Science & Technology. https://doi.org/10.1021/acs.est.1c02653.

Jawlik, A.A., 2016. Statistics from A to Z: Confusing Concepts Clarified. John Wiley & Sons. https://doi.org/10.1108/RR-01-2017-0006.

Keulers, L., Dehghani, A., Knippels, L., Garssen, J., Papadopoulos, N., Folkerts, G., et al., 2022. Probiotics, Prebiotics, and Synbiotics to Prevent or Combat Air Pollution Consequences: the Gut-Lung axis. Environmental Pollution, p. 119066. https://doi.org/10.1016/j.envpol.2022.119066.

Kim, P., 2017. Convolutional neural network. In: MATLAB Deep Learning. Apress, Berkeley, CA, pp. 121–147. https://doi.org/10.1007/978-1-4842-2845-6_6.

Kim, H.S., Park, I., Song, C.H., Lee, K., Yun, J.W., Kim, H.K., et al., 2019. Development of a daily PM $_{10}$ and PM $_{2.5}$ prediction system using a deep long short-term memory neural network model. Atmos. Chem. Phys. 19 (20), 12935–12951. https://doi.org/10.5194/acp-19-12935-2019.

Kolbusz, J., Rozycki, P., Wilamowski, B.M., 2017. June). The study of architecture MLP with linear neurons in order to eliminate the "vanishing gradient" problem. In: International Conference on Artificial Intelligence and Soft Computing. Springer, Cham, pp. 97–106. https://doi.org/10.1007/978-3-319-59063-9_9.

Kow, P.Y., Wang, Y.S., Zhou, Y., Kao, I.F., Issermann, M., Chang, L.C., Chang, F.J., 2020. Seamless integration of convolutional and back-propagation neural networks for regional multi-step-ahead $PM_{2.5}$ forecasting. J. Clean. Prod. 261, 121285. https://doi.org/10.1016/j.jclepro.2020.121285.

Lee, M., Lin, L., Chen, C.Y., Tsao, Y., Yao, T.H., Fei, M.H., Fang, S.H., 2020. Forecasting air quality in Taiwan by using machine learning. Sci. Rep. 10 (1), 1–13. https://doi.org/10.1038/s41598-020-61151-7.

Li, Z., Chen, W.T., Chang, I.C., Hung, C.C., 2021. Dynamic relationship between air pollution and economic growth in Taiwan deduced from mathematical models. Clean 49 (10), 2100081. https://doi.org/10.1002/clen.202100081.

Liu, Y., Paciorek, C.J., Koutrakis, P., 2009. Estimating regional spatial and temporal variability of PM2.5 concentrations using satellite data, meteorology, and land use information. Environ. Health Perspect. 117 (6), 886–892. https://doi.org/10.1289/ehp.0800123.

Liu, S., Hua, S., Wang, K., Qiu, P., Liu, H., Wu, B., et al., 2018. Spatial-temporal variation characteristics of air pollution in Henan of China: localized emission inventory, WRF/Chem simulations and potential source contribution analysis. Sci. Total Environ. 624, 396–406. https://doi.org/10.1016/j.scitotenv.2017.12.102.

Lu, T., Marshall, J.D., Zhang, W., Hystad, P., Kim, S.Y., Bechle, M.J., et al., 2021. National Empirical Models of Air Pollution Using Microscale Measures of the Urban Environment. Environmental Science & Technology. https://doi.org/10.1021/acs.est.1c04047.

O'Shea, K., Nash, R., 2015. An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458. https://doi.org/10.48550/arXiv.1511.08458.

Pal, R., 2016. Predictive Modeling of Drug Sensitivity. Academic Press. https://doi.org/10.1016/B978-0-12-805274-7.00004-X.

Pelletier, C., Webb, G.I., Petitjean, F., 2019. Temporal convolutional neural network for the classification of satellite image time series. Rem. Sens. 11 (5), 523. https://doi.org/10.3390/rs11050523.

Reátegui-Romero, W., Sánchez-Ccoyllo, O.R., de Fatima Andrade, M., Moya-Alvarez, A., 2018. $PM_{2.5}$ estimation with the WRF/Chem model, produced by vehicular flow in the lima metropolitan area. Open J. Air Pollut. 7 (3), 215–243. https://doi.org/10.4236/ojap.2018.73011.

She, H., Cheng, P.H., Yuan, C.S., Yang, Z.M., Ie, I.R., 2020. Chemical characteristics, spatiotemporal distribution, and source apportionment of PM2. 5 surrounding industrial complexes in Southern Kaohsiung. Aerosol Air Qual. Res. 20 (3), 557–575. https://doi.org/10.4209/aaqr.2020.01.0007.

Shen, Y.L., Liu, W.T., Lee, K.Y., Chuang, H.C., Chen, H.W., Chuang, K.J., 2018. Association of PM2. 5 with sleep-disordered breathing from a population-based study in Northern Taiwan urban areas. Environ. Pollut. 233, 109–113. https://doi.org/10.1016/j.envpol.2017.10.052.

Shogrkhodaei, S.Z., Razavi-Termeh, S.V., Fathnia, A., 2021. Spatio-temporal modeling of pm2. 5 risk mapping using three machine learning algorithms. Environ. Pollut. 289, 117859. https://doi.org/10.1016/j.envpol.2021.117859.

Tagaris, E., Liao, K.J., DeLucia, A.J., Deck, L., Amar, P., Russell, A.G., 2009. Potential impact of climate change on air pollution-related human health effects. Environ. Sci. Technol. 43 (13), 4979–4988. https://doi.org/10.1021/es803650w.

Trusz, A., Ghazal, H., Piekarska, K., 2020. Seasonal variability of chemical composition and mutagenic effect of organic PM2. 5 pollutants collected in the urban area of Wrocław (Poland). Sci. Total Environ. 733, 138911. https://doi.org/10.1016/j.scitotenv.2020.138911.

Wagaa, N., Kallel, H., 2020. October). Vector-based back propagation algorithm of supervised convolution neural network. In: 2020 International Conference on Control, Automation and Diagnosis (ICCAD). IEEE, pp. 1–6.

Wang, D., Wei, S., Luo, H., Yue, C., Grunder, O., 2017. A novel hybrid model for air quality index forecasting based on two-phase decomposition technique and modified extreme learning machine. Sci. Total Environ. 580, 719–733. https://doi.org/10.1016/j.scitotenv.2016.12.018.

Wang, B., Yuan, Q., Yang, Q., Zhu, L., Li, T., Zhang, L., 2021. Estimate hourly PM2. 5 concentrations from Himawari-8 TOA reflectance directly using geo-intelligent long short-term memory network. Environ. Pollut. 271, 116327. https://doi.org/10.1016/j.envpol.2020.116327.

Wang, Y.S., Chang, L.C., Chang, F.J., 2021. Explore regional $PM_{2.5}$ features and compositions causing health effects in Taiwan. Environ. Manag. 67 (1), 176–191. https://doi.org/10.1007/s00267-020-01391-5.

Wu, Y.S., Tsai, C.Y., Chang, K.H., Chiang, C.F., 2021. Impact of air pollutants emitted by Taichung power plant on atmospheric PM2. 5 in central Taiwan. Aerosol Air Qual. Res. 21 (4), 200358. https://doi.org/10.4209/aaqr.2020.06.0358.

Yan, X., Zang, Z., Luo, N., Jiang, Y., Li, Z., 2020. New interpretable deep learning model to monitor real-time PM2. 5 concentrations from satellite data. Environ. Int. 144, 106060.

Yan, X., Zang, Z., Jiang, Y., Shi, W., Guo, Y., Li, D., et al., 2021a. A Spatial-Temporal Interpretable Deep Learning Model for improving interpretability and predictive accuracy of satellite-based PM2. 5. Environ. Pollut. 273, 116459.

Yan, X., Zang, Z., Zhao, C., Husi, L., 2021b. Understanding global changes in fine-mode aerosols during 2008–2017 using statistical methods and deep learning approach. Environ. Int. 149, 106392.

Yokoo, K., Ishida, K., Ercan, A., Tu, T., Nagasato, T., Kiyama, M., Amagasaki, M., 2022. Capabilities of Deep Learning Models on Learning Physical Relationships: Case of Rainfall-Runoff Modeling with LSTM, 802. Science of The Total Environment, p. 149876.

Yurtsever, M., Yurtsever, U., 2019. Use of a convolutional neural network for the classification of microbeads in urban wastewater. Chemosphere 216, 271–280. https://doi.org/10.1016/j.chemosphere.2018.10.084.

Zalakeviciute, R., López-Villada, J., Rybarczyk, Y., 2018. Contrasted effects of relative humidity and precipitation on urban $PM_{2.5}$ pollution in high elevation urban areas. Sustainability 10 (6), 2064. https://doi.org/10.3390/su10062064.

Zhou, X., 2018, April. Understanding the convolutional neural networks with gradient descent and backpropagation. In: Journal of Physics: Conference Series, 1004. IOP Publishing, p. 12028. https://doi.org/10.1088/1742-6596/1004/1/012028. No. 1.

Zhou, Y., Chang, F.J., Chang, L.C., Kao, I.F., Wang, Y.S., Kang, C.C., 2019. Multi-output support vector machine for regional multi-step-ahead $PM_{2.5}$ forecasting. Sci. Total Environ. 651, 230–240. https://doi.org/10.1016/j.scitotenv.2018.09.111.

Zhou, H., Zhang, F., Du, Z., Liu, R., 2021. Forecasting PM2. 5 using hybrid graph convolution-based model considering dynamic wind-field to offer the benefit of spatial interpretability. Environ. Pollut. 273, 116473. https://doi.org/10.1016/j.envpol.2021.116473.

Zong, P., Zhu, Y., Wang, H., Liu, D., 2020. WRF-chem simulation of winter visibility in Jiangsu, China, and the application of a neural network algorithm. Atmosphere 11 (5), 520. https://doi.org/10.3390/atmos11050520.