A Hybrid Time Series Model for the Spatio-Temporal Analysis of Air Pollution Prediction Based on $PM_{2.5}$

Naushad Ahmad and Vipin Kumar(✉)

Computer Science and Information Technology,
Mahatma Gandhi Central University, Bihar 845401, India
rt.vipink@gmail.com

Abstract. The issue of air pollution represents a formidable challenge for the global community. This challenge is further compounded by the increasing levels of air pollution witnessed daily. The base pollutants which are $PM_{2.5}$, PM_{10}, NO_2, SO_2, NO, CO, and O_3. $PM_{2.5}$ air pollution has been associated with various health complications, including respiratory diseases, heart diseases, cancer, and premature death. This attention has been brought about by the prevalence of air pollutants and their impact on human health. The author has proposed a solution by utilizing a hybrid model incorporating spatiotemporal correlation analysis. Specifically, the author suggested a model that merges CNN and LSTM. The outcomes of this model have been remarkable, with an enhancement of 64.26% (MSLR), 67.75% (MSE), 24.46% (MAPE), 35.20% (MAE), and 29.52% (RMSE) when compared to models that lack spatiotemporal analysis. Recently, there has been growing attention towards the detrimental effects of air pollution from both the government and the public. Despite the numerous efforts to curb air pollution, it continues to pose a significant threat to human health and the environment. Therefore, there is a need to continue exploring innovative solutions, such as the proposed hybrid model, to mitigate the impact of air pollution.

Keywords: Air pollution · Deep Learning · $PM_{2.5}$ · spatio-temporal

1 Introduction

The globe will have a challenging time dealing with air pollution. The amount of air pollution is rapidly rising every day. A single person or nation cannot stop these global issues alone. It is accountable for all people and takes preventative measures on your end. The public and the government have been paying close attention lately to the worsening of air quality, the frequent occurrences of air pollutants, and the health effects they have Kumar and Ahmad (2022). Significant pollutants include PM, CO, O_3, NO_2, and SO_2. Heart attacks, coughing fits, and breathing difficulties can all co-occur while breathing in particles with a diameter of less than or equal to 10 μm. According to estimations from the

A. Verma et al. (Eds.): ANTIC 2023, CCIS 2093, pp. 62–81, 2024.
https://doi.org/10.1007/978-3-031-64067-4_5

World Health Organization (WHO), 92% of people worldwide are exposed to pollutants at levels deemed harmful to their health Lee et al. (2014).

Numerous nations worldwide must constantly work to measure, research, avoid, and regulate these. Air pollution remains a persistent and significant concern among the G20 countries. Under the India Presidency, the G20 countries convened to address climate and environmental challenges. The G20 theme,"One Earth, One Family, One Future, emphasizes the interconnectedness of all living beings on Earth and in the universe. The main goals of G20 are to protect the planet from degradation and promote sustainable and balanced growth. Air quality, the allowable threshold for concentrations of the principal pollutants, measurements, and standards are all governed by official documents in the European Union Kutlar Joss et al. (2017). Each nation employs a distinct Air-Quality Index (AQI) based on its geographical location to ensure its populace's well-being. Within India, regulatory bodies such as the Central Pollution Control Boards (CPCB), State Pollution Control Boards (SPCB), Pollution Control Committees (PCCs), and the National Environmental Engineering Research Institute (NEERI) operate in major urban centres, striving to address the issue Ganguly (2019). In a collective effort, countries convened at the Sharm el-Sheikh Climate Change Conference (COP 27) to embark on actions that align with the global climate objectives outlined in the Paris Agreement and the Convention Gao et al. (2017). The European Union has established a notable collection of legislation, supplemented by significant international agreements like the Paris Agreements, to set strict legal limits and target thresholds for the concentrations of critical atmospheric pollutants to be achieved by the Member States in the upcoming decadesLeptien et al. (2019). The quantity of missing parameter values significantly influences the efficacy of the model. Consequently, one of the prevalent inquiries in data science pertains to developing and implementing an appropriate and efficient data supplementation technique to reduce data interference Sinharay et al. (2001). Only through proper categorization of the data can the impact of air pollution on human health be mitigated. The issue of class imbalance arises in various categorization challenges. Researchers need help with attempting to glean insights from imbalanced data Ketu and Mishra (2021).

In light of the continuous advancement of AI technology, the (Berral et al., 2018) utilization of statistical modelling, particularly models that integrate machine learning (ML) and deep learning (DL) approaches, has generated interest among a broad range of scientists and researchers across various sectors Berral-García (2018). Deep learning, currently the most prevalent data-driven methodology, can automatically extract and comprehend the fundamental characteristics of diverse air quality data. Deep learning has substantially progressed in image processing, natural language comprehension, and air pollution prognostication Liao et al. (2020). In a broader context, deterministic and statistical techniques are viable approaches for anticipating air pollution concentrations Gautam et al. (2019). Consequently, statistical and time series modelling methodologies are predominantly employed in short-term air pollution forecasting, catering to operational requirements Espinosa et al. (2021). In comparison to the feedforward neural network (FNN), the recurrent neural network (RNN) has

exhibited remarkable proficiency in capturing temporal dependencies and has been effectively employed in addressing diverse real-world challenges Ye et al. (2018). Devising strategies for integrating different methodologies to enhance air quality prediction accuracy is essential. Furthermore, distinct approaches necessitate varying computational settings and resources. For instance, conventional machine learning techniques rely on extensive data platforms to make predictions, whereas deep learning methods yield superior results when implemented on graphics processing units (GPUs) Zaini et al. (2022).

The novelty of the research work is listed below:

- One of the most polluted city (i.e., Muzaffarpur, India) data set have be utilized for analysis of the proposed model which has information of spatial and temporal of $PM_{2.5}$
- A CNN-LSTM model has been proposed to capture the spatio-temporal information to enhance the performance.
- The effectiveness of the proposed model has been compared with traditional deep learning model based on with and without spatio-temporal information.

The paper is structured: The first section commences with an Introduction, wherein the subsequent subsections discuss the research issues and motivation. The second section comprises a comprehensive review of the most up-to-date literature. The third section begins with Material. Fourth and fifth section contains the spatio-temporal correlation analysis and deep learning model of time series analysis. Section 6 contains the results and discussion, the last section contains the description of the conclusion.

2 Literature Review

The present study Tan et al. (2022a) proposes a model that exhibits superior performance compared to 25 benchmark models, thereby highlighting its significant application potential and value in predicting levels of inhalable particulate matter ($PM_{2.5}$) in both spatial and temporal domains. In this paper Fang et al. (2022), a novel framework called DESA has been developed to forecast $PM_{2.5}$ levels, which are associated with various spatiotemporal components that make the prediction process challenging. The research paper Liu et al. (2022) investigates the spatiotemporal variations and sources of $PM_{2.5}$ concentrations in the largest region of central China, which is grappling with a severe air pollution issue. The present paper Zhang et al. (2023b) posits an adaptive spatiotemporal prediction network (ASTP-NET) as a means of achieving precise $PM_{2.5}$ concentration forecasting. The network comprises an encoder that extracts input data features, a decoder that maps the features to a future time series prediction, and an objective function that gauges the model's multi-step prediction fluctuations. The paper Wang et al. (2022) suggested that urban dwellers' physiological, psychological, and behavioural aspects are adversely affected by elevated pollution levels in urban settings. The current Rincon et al. (2023) study employed a range of analytical techniques, including bivariate analysis, multivariate linear regression analysis, and logistic regression analysis, to examine the complex relationship between

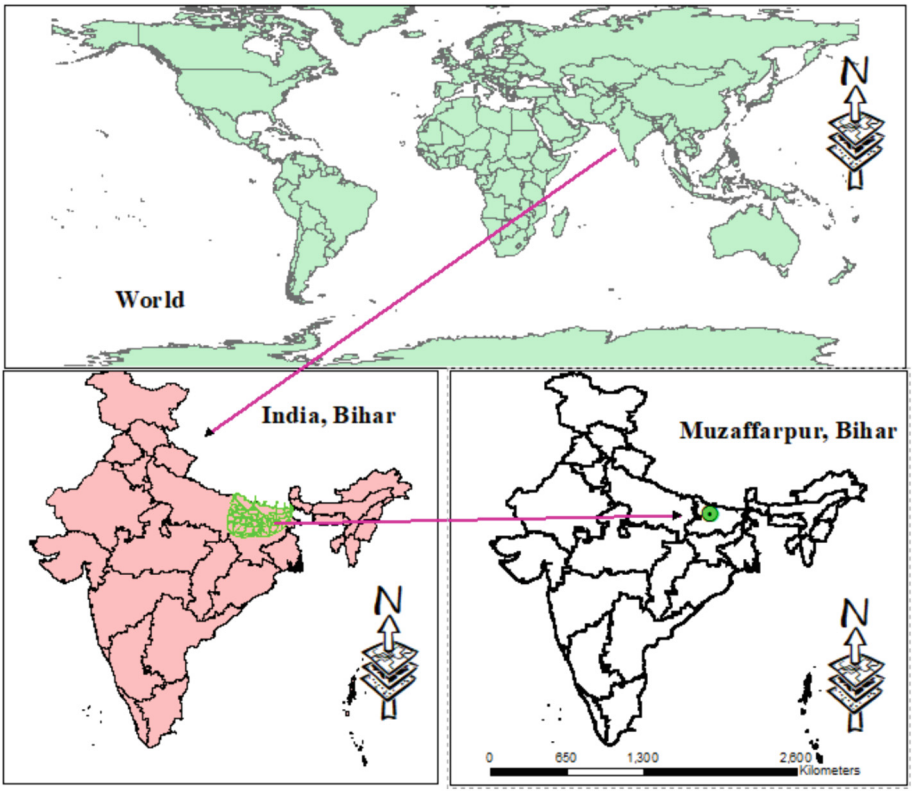


Fig. 1. Location of Studies of $PM_{2.5}$ concentration

PM concentrations and various meteorological variables, as well as the occurrence of events and the characteristics of land use.

The analysis conducted by Yang et al. (2022) focuses on examining long-term measurements of ambient antimony sb $PM_{2.5}$ in Beijing. The primary objective of this investigation is to reveal the spatial and temporal distributions of antimony, identify its principal sources, and provide predictions regarding potential future trends Yang et al. (2022). In the research conducted by Cai et al. (2022), attention is given to the examination of the spatial and temporal characteristics of nitrogen dioxide (NO_2), delicate particulate matter ($PM_{2.5}$), and ozone (O_3) in Fujian Province, which is located in Southeast China, throughout the year 2015. Additionally, this study explores the impact of landscape metrics that define green spaces on air pollutants through Pearson correlation analysis at various spatial extents and seasons Cai et al. (2022). The investigation carried out by Chicas et al. (2023) focuses on the cartographic and mathematical representation of the spatiotemporal dispersion of $PM_{2.5}$ in the central region of Bangladesh from 2002 to 2019, with a specific emphasis on identifying areas with high concentrations in urban areas Chicas et al. (2023).

Table 1. Recent list of literature with the details of hybrid model, type of analysis, evaluation and area of study

Reference	Model Name	Analysis	Error Parameters	Area of study
Liu et al. (2023a)	GCN-CNN	Spatio-temporal	MAE, RMSE	Beijing, China
Fang et al. (2022)	DESA	spatiotemporal attention	RMSE	China
Tan et al. (2022b)	SES-IDW and LSTM	Spatiotemporal	RMSE, R2	Worldwide
Liu et al. (2023b)	TWR-XGBoost	Spatiotemporal Distribution	R2, RMSE, and MAE	Chengdu-Chongqing, China
He et al. (2023)	CLSTM-GPR	spatial-temporal	R2, RMSE, and MAE	Jiangmen and Huizhou, China
Chen et al. (2023)	CNN-RF Ensemble	spatiotemporal factors	RMSE, and MAE	Fuxing and Fongshan, Taiwan
Ma (2023)	Diffusion model	Spatiotemporal dynamic	RMSE, and MAE	Beijing, China
Ma et al. (2023)	ST-Informer	SpatioTemporal-Informer	R2, RMSE, and MAE	Beijing-Tianjin-Hebei, China
Zhang et al. (2023a)	CCA-LSTM	spatiotemporal dependence	RMSE, and MAE	Xi'an, China

The Li and Huo (2023) paper posits a composite framework, namely CNN-BiGRU. The framework employs the convolution operation to extract salient features from single-dimensional data and join it with the circular neural network BiGRU to formulate and prognosticate the $PM_{2.5}$ concentration. Regrettably, this study did not obtain any particular endowment from funding organizations in public, commercial, or not-for-profit sectors. The machine learning model Wu et al. (2023), which was stacked and employed in the investigation, exhibited superior performance compared to other models in its ability to simulate the correlation between the concentrations of $PM_{2.5}$ and the influencing factors.

3 Materials

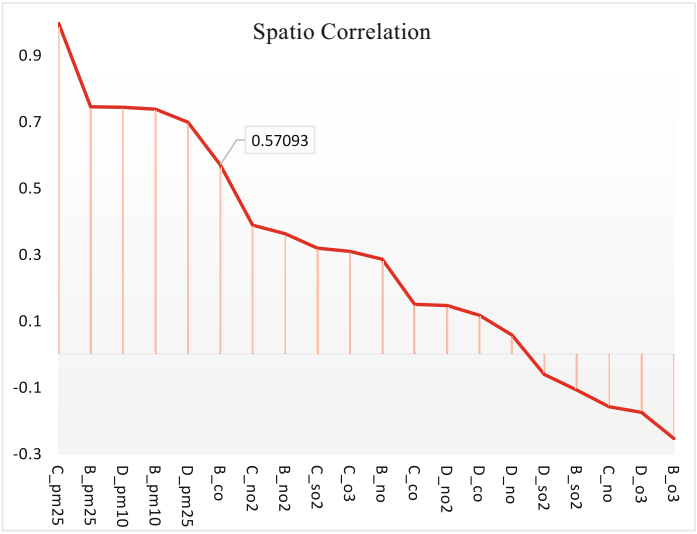


Fig. 2. The correlation coefficients of different features from surrounding stations

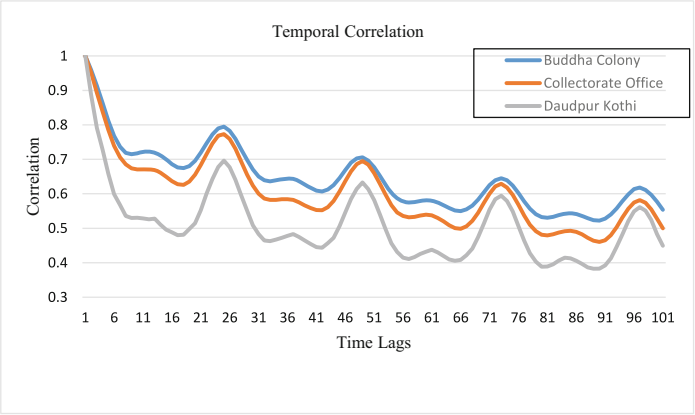


Fig. 3. The autocorrelation coefficients of the different time lags

In this paper, the area of study is Muzaffarpur Collectorate Office (C) and its surrounding monitoring stations, including Buddha Colony (B) and Daudpur Kothi (D) Muzaffarpur, Bihar, India. (Fig. 1) Researchers point to the area of studies of $PM_{2.5}$ concentration under the world map. It is well known that Muzaffarpur city and its areas suffer from severe $PM_{2.5}$ pollution. Despite being geographically near to one another, these places have recently seen significant urbanization and industrialization. The Pearson correlation of all surrounding stations for all features (see Fig. 6). This study uses historical data from Muzaffarpur, which includes information on pollutant concentrations for things like $PM_{2.5}$, NO_2, SO_2, and ozone. The autocorrelation analysis indicates that the initial 24 lags exhibit a prominent peak in positive correlation (Fig. 5). Upon identifying this initial peak in positive correlation, we conducted data decomposition using a period of 24, aligning with this observed peak positive correlation (as demonstrated in Fig. 4).

Table 1 displays the hourly dataset's statistical information. Ten thousand twenty-four (10024) samples total from midnight on December 3, 2021, to 2:00 PM on January 24, 2023, are included in the dataset (https://airquality.cpcb.gov.in/)ccr/#/caaqm-dashboard-all/caaqm-landing). Some pollutant concentration data is gathered from an online tool for monitoring air quality (CPCB) Bhawan and Nagar (2020). Handling missing values through mean imputation and min and max normalizing features to ensure uniformity and model stability.

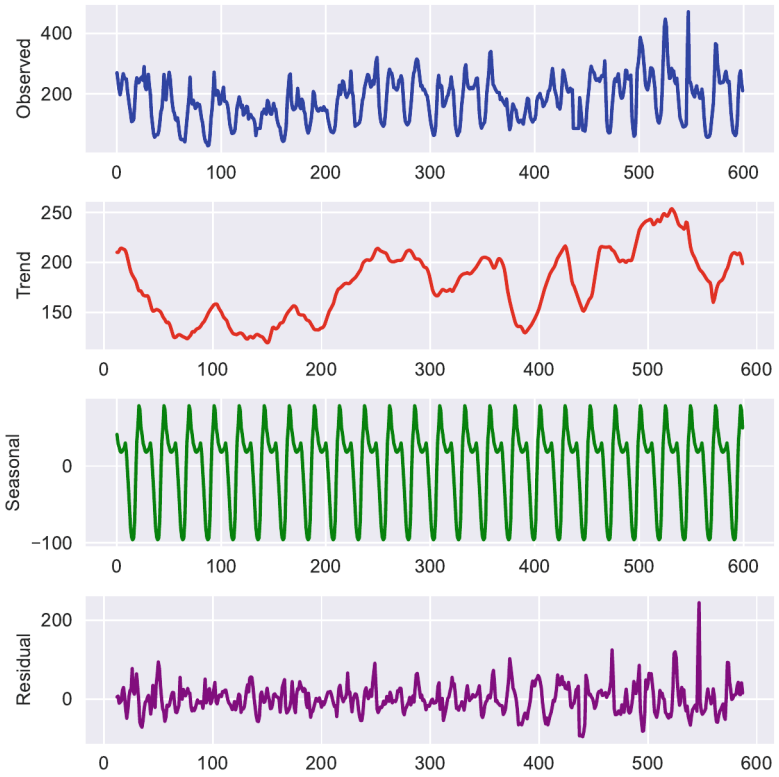


Fig. 4. Data decomposition from rows = 600 period = 24 of $PM_{2.5}$ Collectorate office (C)

These variables include $PM_{2.5}$, PM_{10}, NO_2, NO, SO_2, CO, and O_3, with their respective units provided in micrograms per cubic meter ($\mu g/m^3$) or milligrams per cubic meter (mg/m^3). The table offers insights into each variable's data type (numeric), the observed range of values, and their average values. Additionally, it introduces two calculated sub-indices, "Pollution Grade" and "AQI Grade" represented as grading or numeric values within the range of 1 to 6. While this (Table 2) provides a clear summary of the dataset's statistical characteristics, it does not delve into the specific calculation methods for the sub-indices.

Let's define the multivariate data C_d in terms of features where i^{th} features is denoted as $[C_{d_i}(t)]^T$, therefore the whole data set can be written as:

$$C_d = \left[C_{d_1}(t)^T, C_{d_2}(t)^T, \ldots C_{d_N}(t)^T\right] \tag{1}$$

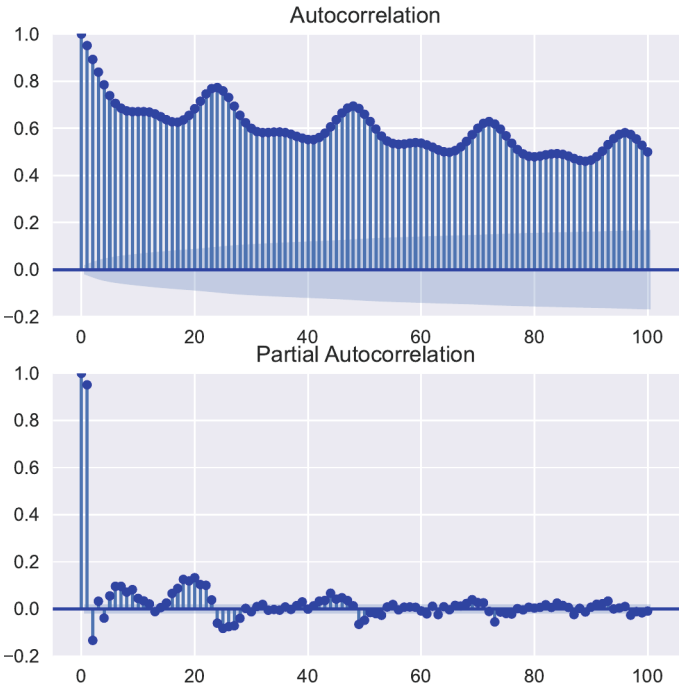


Fig. 5. Data decomposition from head rows = 600 acf and pacf plot lags = 100 of $PM_{2.5}$ Collectorate office (C)

Table 2. Statistical information of dataset used in this research

Variables	Units	Data Type	Range	Average
$PM_{2.5}$	$\mu g/m^3$	Numeric	[1, 545]	85.81
PM_{10}	$\mu g/m^3$	Numeric	[1, 974]	219.786
NO_2	$\mu g/m^3$	Numeric	[0.1, 68.7]	6.61
NO	$\mu g/m^3$	Numeric	[0.1, 63.7]	4.09
SO_2	$\mu g/m^3$	Numeric	[0.1, 28.63]	8.35
CO	mg/m^3	Numeric	[0, 5.63]	1.02
O_3	$\mu g/m^3$	Numeric	[0.1, 133.4]	28.05
Pollution Grade	calculating sub-index	Grading/Numeric	[1, 6]	——
AQI Grade	calculating sub-index	Grading/Numeric	[1, 6]	——

$$C_{d_{norm}}(t) = \frac{C_d(t) - C_{d_{min}}(t)}{C_{d_{max}}(t) - C_{d_{min}(t)}}, \quad (f = 1, 2, 3, \ldots, N) \tag{2}$$

$$C_n = \left[C_{n_1}(t)^T, C_{n_2}(t)^T, \ldots C_{n_M}(t)^T\right] \tag{3}$$

$$C = \left[C_h, C_{h+1}, C_{h+2}, \ldots C_T\right] \tag{4}$$

where h is the number of time series historical data used in the neural network hybrid model. The collected data set C was divided into three parts: training data, testing data, and validating data. The hybrid model training with 80% data, 10% data used for testing, and remaining for validation.

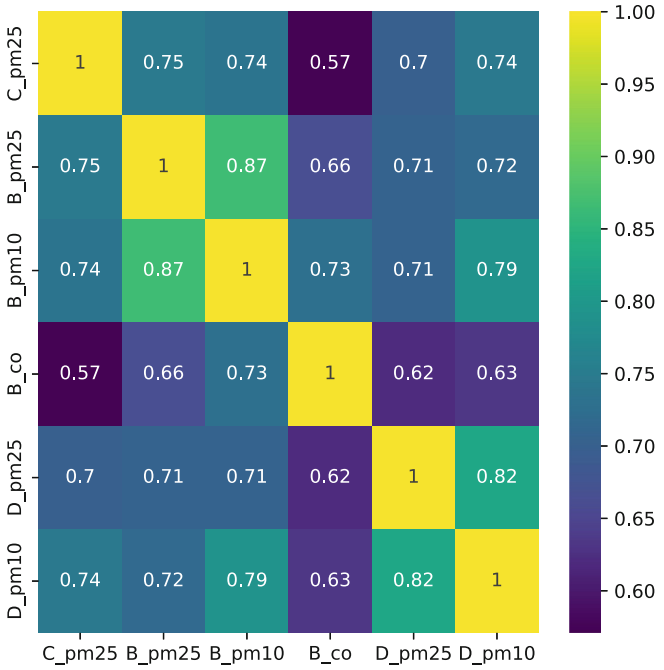


Fig. 6. Pearson correlation for all surrounding station features

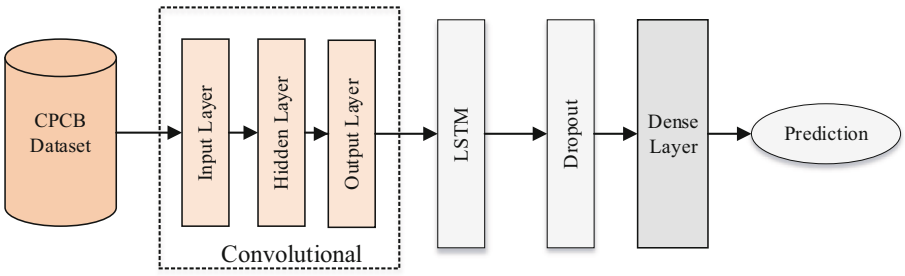


Fig. 7. The Network architecture of the CNN-LSTM hybrid model

A Hybrid model, which combines CNN and LSTM networks, has been precisely developed to accurately forecast the $PM_{2.5}$ levels in air pollution data. The innovative architectural design has shown (Fig. 7) remarkable effectiveness in capturing and understanding both spatial and temporal patterns in air quality

prediction tasks. The model can holistically comprehend the complex dynamics of air pollution by harnessing the power of historical air quality variables as input data. These crucial variables are thoughtfully organized into time series windows, typically hourly, allowing for detailed data analysis. The initial component of this sophisticated architecture, namely the CNN encoder, plays an indispensable and pivotal role in extracting intricate spatial patterns from the input data via the integration of convolution layers. Following the extraction of spatial features, the LSTM decoder seamlessly takes over, focusing its efforts on comprehending and leveraging the temporal dependencies within the data. The LSTM network, composed of multiple cells, effectively preserves and retains valuable information from previous temporal instances, thereby empowering the model to make accurate predictions at the present instance. Finally, the output layer of this model comprises one or more neurons, seamlessly generating continuous predictions of $PM_{2.5}$ levels for subsequent time instances, thus effectively contributing to the scientific understanding and forecasting of air pollution dynamics.

Algorithm 1: A hybrid time series model (CNN-LSTM) for the spatiotemporal analysis of air pollution prediction based on $PM_{2.5}$

Input: Feature set $X_f = \left\{[X_{f_i}(t)]^T\right\}_{l=1}^{m}$ // Multi-variate dataset has to follow Ergodic property where $[X_{f_i}(t)]^T = \left\{X_{f_j}(t)\right\}_{j=1}^{m}$

Output: $\hat{y}_{n+k}$ // final prediction of proposed model (CNN-LSTM)

1 Initialization: data collection $C_d = \left[\, C_{d_1}(t)^T, C_{d_2}(t)^T, \ldots C_{d_N}(t)^T \,\right]$

2 The input gate of CNN will utilize the c_j^l as : $c_j^l = \sum_i X_i^{(l-1)} \cdot W_{ij}^l + b_j^l$ and $X_j^l = \text{LeakyReLU}(c_j^l)$ // Equation 5 and 6

3 The output of CNN as the input for LSTM as: $h_t = \text{LSTM}(x_t, h_{t-1})$ // Equation 22

4 Now, X_t utilises for LSTM output X_t to Dropout(X_t) and Flatten(X_j^l) // Equation 24 and 25

5 Get the final prediction as $\hat{y}_{n+k} = \sum_{j=1}^{n} \phi_{n,j}(k) \cdot y_{n-j+1}$ // See equation 27

6 The final prediction will be measured with method (Em) as $E = E(y, \hat{y}, Em)$ // See equation 34

7 Return : $(\hat{y}_{n+k})$

8 End

4 Spatio-Temporal Correlation Analysis

We consider the spatial correlation of $PM_{2.5}$ concentration from various air quality data of monitoring stations due to the severe pollution in Muzaffarpur, Bihar, and its close geographic position. The method of calculating Pearson's correlation coefficient is frequently used to assess the degree of correlation between various variables. Correlation coefficients can be used to filter the model features. The calculated outcomes of variables from various cities are shown in Fig. 6. From

0.253483 to 0.746036, the correlation coefficient values are shown. It has been shown that the correlation coefficient for various air pollution concentrations decreases with increasing distance from the Muzaffarpur Collectorate office (C), such as Buddha Colony (B) and Daudpur Kothi (D). In addition, the threshold value of 0.5 for the correlation coefficient is chosen in this research for feature selection (Fig. 2). When the coefficient is more significant than 0.5, the variables have a strong association. We call all three stations by their short names: B, C, and D. The CO, $PM_{2.5}$, and PM10 from Buddha Colony (B), as well as the $PM_{2.5}$ and PM_{10} from Daudpur Kothi (D), appear to have a substantial correlation with the PM2.5 in the Collectorate Office (C). The spatial correlation offers substantial assistance to improve prediction performance rather than creating a distinct model for each monitoring station.

The autocorrelation coefficients of $PM_{2.5}$ from various monitoring sites are shown in (Fig. 3). With the lag time, it is clear that the curve has a falling tendency. The chart shows that the effect of the $PM_{2.5}$ concentration data on the current situation decreases with increasing time. Additionally, as the lag time lengthens, the drop rate gradually slows, with the initial descent speed being the greatest. According to the research mentioned above, it is clear that $PM_{2.5}$ at the Collectorate office (C) has a substantial spatiotemporal connection with nearby monitoring stations, which is advantageous to prediction accuracy.

5 Deep Learning Model of Time Series Analysis

The ANN is a valuable early-stage mathematical model due to its superior ability to handle nonlinear issues while simulating the structure of brain neurons. The multilayer perceptron (MLP) Taud and Mas (2018), a standard neural network structure, has been extensively used recently. The MLP comprises input, output, and hidden layers. However, the conventional MLP model with a three-layer neural structure performs poorly when data volume and feature size increase. To address this problem, well-known neural networks such as CNN Albawi et al. (2017) and LSTM Greff et al. (2016) use growing network structural complexity. This study presents the coupling of the CNN and LSTM models to address the issue of time series prediction.

5.1 Convolutional Neural Network Model

The proposed network has exhibited impressive performance in identifying handwritten fonts, which has captured the attention of scholars. A CNN comprises numerous feature maps in each layer, each consisting of multiple neurons. The convolution of the current neuron is achieved by combining the neuron's output in the higher layer with a convolutional kernel. The convolutional kernel is a predetermined weight matrix that effectively extracts characteristics from the local sensing region. Most of a convolutional neural network's architecture comprises convolutional and fully connected layers. The basic modules of a CNN include the convolutional layer and the pooling layer in the hidden layer. After the

convolutional layer has gathered local features from the data, CNN can automatically extract features from data sequences, such as text and images. The network's conventional structure includes 1D, 2D, and 3D CNNs. Since $PM_{2.5}$ data is one-dimensional, a 1D CNN was employed in this study to train features.

$$c_j^l = \sum_i X_i^{(l-1)} \cdot W_{ij}^l + b_j^l \tag{5}$$

$$X_j^l = \text{LeakyReLU}(c_j^l) \tag{6}$$

In Eq. (6) denotes the convolutional kernel.

5.2 Long Short-Term Memory Model

LSTM is a recurrent neural network architecture designed for processing and modelling sequential data, making it particularly well-suited for tasks involving time series, natural language processing, and speech recognition. Unlike traditional RNNs, LSTMs have a unique memory cell that can capture long-range dependencies in the data, preventing the vanishing gradient problem. LSTMs consist of three gates-the input gate, forget gate, and output gate-that regulate the flow of information within the network, allowing it to remember or forget information over time selectively. This architecture enables LSTMs to effectively capture and store temporal patterns in sequential data, making them a fundamental component in various deep learning applications (Fig. 8).

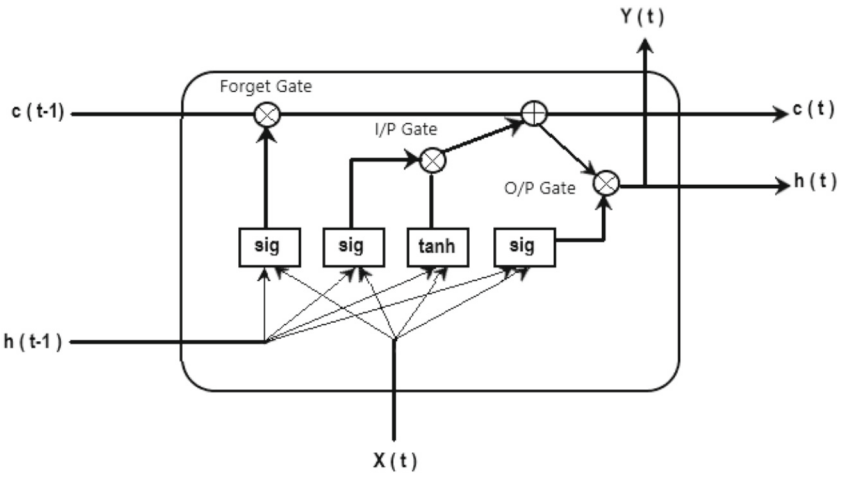


Fig. 8. The LSTM model architecture

$$h_t = \mathcal{H}\left(\mathbf{WC}_t \oplus \mathbf{Wh}h_{t-1} + \mathbf{bh}\right) \tag{7}$$

$$Y_t = \mathbf{WhY}h_t + \mathbf{bY} \tag{8}$$

$$i_t = \sigma\left(\mathbf{WC}_i \oplus \mathbf{Whi} h_{t-1} + \mathbf{bi}\right) \tag{9}$$

$$f_t = \sigma\left(\mathbf{WC}_f \oplus \mathbf{Whf} h_{t-1} + \mathbf{bf}\right) \tag{10}$$

$$o_t = \sigma\left(\mathbf{WC}_o \oplus \mathbf{Who} h_{t-1} + \mathbf{bo}\right) \tag{11}$$

$$a_t = f_t \cdot a_{t-1} + i_t \cdot \tanh\left(\mathbf{WC}_g \oplus \mathbf{bg}\right) \tag{12}$$

$$h_t = a_t \cdot \tanh(a_t) \tag{13}$$

where $\mathcal{H}$ is a hidden layer function, σ is the logistic sigmoid function, and the various $\mathbf{W}$ terms and $\mathbf{b}$ terms represent weight matrices and bias vectors.

5.3 The Hybrid CNN-LSTM Model

The hybrid paradigm of CNN-LSTM has been initially employed in both computer vision and text processing domains. For this purpose, CNN has been utilized to extract features from image, numerical, and text data. The LSTM has been utilized subsequently for the processing of these extracted features. Similarly, in this study, CNN has been employed to extract features from time series data, and LSTM has been developed for prediction based on the output of the CNN model. The detailed architecture of the CNN-LSTM model has been presented (Fig. 7). This hybrid model employs a one-dimensional convolutional layer as the fundamental layer due to time series data's distinctiveness. A flattened layer has been constructed at the end of both CNN and LSTM layers to feed the CNN output into the LSTM. A fully linked layer has also been built to decode the LSTM output. Eventually, the provided model can produce the outcomes of the predictions.

The present study exhibits the specific CNN-LSTM parameters employed therein. In lieu of other commonly employed activation functions, the LeakyReLU function was selected as the activation function due to its unique structure, which effectively mitigates the problem of gradient disappearance in neural networks. Furthermore, Nadam, a robust parameter optimizer, was utilized in this study in substitution for the gradient descent method. The Nadam parameter optimizer permits dynamic adjustment of the parameter learning rate, thus increasing the likelihood of the parameter's deviation from the local optimum.

The function $x : T \to \mathbb{R}^k, t \mapsto x_t$ is an observed time series. This can be written as:

$$x_t \mid x_t \in \mathbb{R}^k, \quad t \in T \tag{14}$$

To define a stochastic process, the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with random variables $\{X_t \mid X_t \in \mathbb{R}^k, t \in T\}$ follows:

$$\{X_t \mid X_t \in \mathbb{R}^k, \quad t \in T\}, \quad (X_t)_{t \in T} \sim \mathbb{P} \tag{15}$$

If the data set X_t has an equidistant time series (e.g., hourly), then $T \subseteq \mathbb{Z}$. For a constant mean μ:

$$(x_{t_i})_{i=1,\ldots,n} = (X_{t_i}(\omega))_{i=1,\ldots,n} \tag{16}$$

where Ω is the space of functions X, $\mathcal{F}$ is a σ-algebra on Ω, and $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$.

$$\exists \mu \in \mathbb{R}^k \text{ such that } \lim_{n\to\infty} \mathbb{E}[(\bar{x} - \mu)^2] = 0 \tag{17}$$

$$\Sigma : \mathbb{Z} \times \mathbb{Z} \to \mathbb{R} \tag{18}$$

Covariance expression: The auto-covariance function (ACF) of X_t are :

$$\Sigma(s,t) = \text{cov}(X_s, X_t) = \mathbb{E}[(X_s - \mu_s)(X_t - \mu_t)] \tag{19}$$

$$c_j^l = \sum_i X_i^{(l-1)} \cdot W_{ij}^l + b_j^l \tag{20}$$

$$X_j^l = \text{LeakyReLU}(c_j^l) \tag{21}$$

$k(x)$ typically denotes the convolutional kernel.

Given a sequence of input data x_t at time t, an LSTM model processes the data and produces an output y_t at each time step. The LSTM model can be represented as:

$$h_t = \text{LSTM}(x_t, h_{t-1}) \text{ see Eq. 7 to 13} \tag{22}$$

$$X_t = W_h \times h_t + b_h \tag{23}$$

where h_t is the hidden state at time t, h_{t-1} is the hidden state at time $t-1$ (initially set to 0), W_h and b_h are the weight and bias for the output layer, respectively.

$$X_t = \text{Dropout}(X_t) \tag{24}$$

$$X_j^l = \text{Flatten}(X_j^l) \tag{25}$$

$$y_k^{(l+1)} = \text{FC}(W_{kj}^{(l+1)} X_j^l + b_k^{(l+1)}) \tag{26}$$

$$\hat{y}_{n+k} = \sum_{j=1}^{n} \phi_{n,j}(k) \cdot y_{n-j+1} \tag{27}$$

where $\hat{y}$ is the predicted value and n is the data points.

These metrics are commonly used to evaluate the performance of regression models. Given predicted values y_{pred} and actual target values y_{true} for a set of data points, the metrics can be computed as follows:

$$\text{MSE}_n(k) = \inf_{Y \in L_{1,n}} E\left[\frac{(y_{n+k} - \hat{y}_{n+k})^2}{n}\right] \tag{28}$$

$$\text{RMSE}_n(k) = \sqrt{\inf_{Y \in L_{1,n}} E\left[\frac{(y_{n+k} - \hat{y}_{n+k})^2}{n}\right]} \tag{29}$$

$$\text{MAE}_n(k) = \inf_{Y \in L_{1,n}} E\left|\frac{y_{n+k} - \hat{y}_{n+k}}{n}\right| \tag{30}$$

$$\text{MAPE}_n(k) = \inf_{Y \in L_{1,n}} E \left| \frac{y_{n+k} - \hat{y}_{n+k}}{y_{n+k}} \right| \tag{31}$$

$$P_{L_{1,n}}(y_{n+k}) = \sum_{j=1}^{n} \phi_{n,j}(k) \cdot y_{n-j+1} \tag{32}$$

$$\hat{y}_{n+k} = \sum_{j=1}^{n} \phi_{n,j}(k) \cdot y_{n-j+1} \tag{33}$$

$$E = E(y, \hat{y}, E_m) \tag{34}$$

6 Results and Discussion

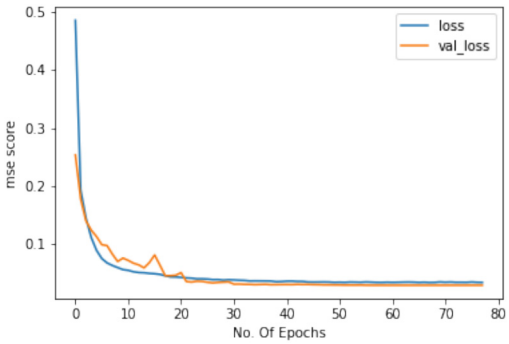


Fig. 9. Training and validation loss of proposed model (CNN-LSTM)

Table 3. Comparison of proposed model with and without spatiotemporal

Types	Model	RMSE	MAPE	MAE	MSE	MSLR
Without Spatiotemporal	CNN-LSTM	52.604385	13.293694	27.726294	2767.2212	0.04921075
With Spatiotemporal	CNN-LSTM	40.61487	10.680623	20.506994	1649.5674	0.029958937
% Improvement	CNN-LSTM	29.5200	24.4655	35.2040	67.7543	64.2623

The (Table 3) presents a comparison between two variations of a predictive model, namely CNN-LSTM, in terms of their performance metrics across various evaluation criteria. The two model variations are categorized as "Without Spatiotemporal" and "With Spatiotemporal" and their performance is measured using several metrics: Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Mean Squared Error (MSE),

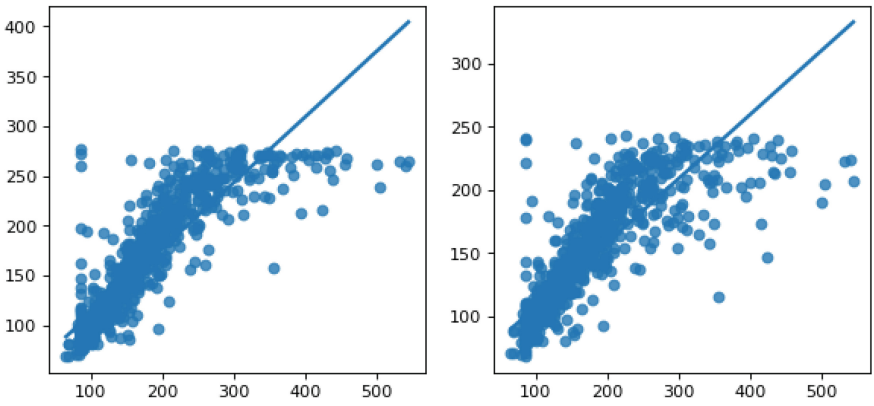
and a metric labelled as "MSLR". The results indicate that the "With Spatiotemporal" model outperforms the "Without Spatiotemporal" model across all metrics, showcasing significant improvements in RMSE, MAPE, MAE, MSE, and MSLR by approximately 29.52%, 24.47%, 35.20%, 67.75%, and 64.26%, respectively (Fig. 9).

Table 4. Comparison of the proposed model to traditional deep learning model over various performance measures

Types	Model	RMSE	MAPE	MSE
Without Spatiotemporal	MLP	114.223	57.807	13046.929
Without Spatiotemporal	LSTM	56.101	90.767	3147.320
Without Spatiotemporal	**CNN-LSTM**	**52.604**	**13.294**	**2767.221**
With Spatiotemporal	MLP	113.884	57.433	12969.455
With Spatiotemporal	LSTM	43.345	100.717	1878.789
With Spatiotemporal	**CNN-LSTM**	**40.615**	**10.681**	**1649.567**

This (Table 4) provides a comparative analysis of three different deep learning models performance, specifically "MLP," "LSTM," and "CNN-LSTM," concerning various evaluation metrics when applied to a task where spatiotemporal information is not considered ("Without Spatiotemporal"). The metrics used for Comparison include RMSE, MAPE, and MSE. Notably, the "CNN-LSTM" model stands out with the lowest values for RMSE (52.604385), MAPE (13.293694), and MSE (2767.2212) among the three models, indicating its superior predictive accuracy compared to the other traditional deep learning models, MLP and LSTM, for this particular task.

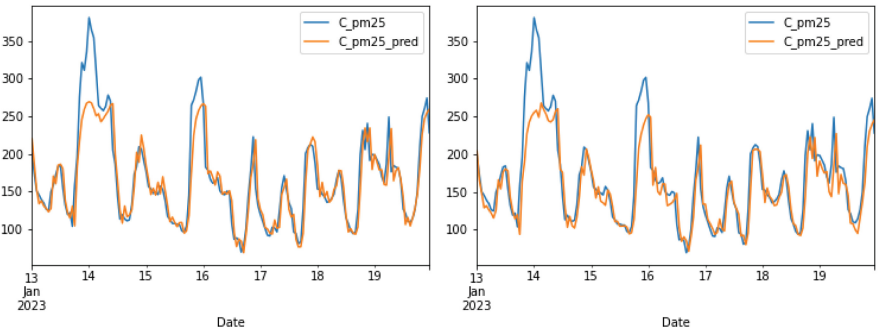
This (Table 4) presents a comparative evaluation of three different deep learning models, namely "MLP," "LSTM," and 'CNN-LSTM" in the context of a task that incorporates spatiotemporal information ("With Spatiotemporal"). The Comparison is based on several performance metrics, including RMSE, MAPE, and MSE. Notably, the "CNN-LSTM" model exhibits the best performance among the three models, with the lowest values for RMSE (40.61487), MAPE (10.680623), and MSE (1649.5674), indicating its superior predictive accuracy in capturing spatiotemporal patterns compared to the traditional deep learning models, MLP and LSTM, for this specific task.



(a) $PM_{2.5}$ prediction regression plot with statiotemporal (b) $PM_{2.5}$ prediction regression plot without statiotemporal

Fig. 10. Proposed model comparison of actual vs prediction regression plot with 45-degree Results

The proposed hybrid model with spatiotemporal (CNN-LSTM) have maximum improvement 29.52% (RMSE), 24.46% (MAPE), 35.20% (MAE), 67.75% (MSE), and 64.26% (MSLR) to without spatiotemporal (see Table 3). The proposed model outperforms the base deep learning model (MLP, LSTM). The performance measure for the proposed mode to 52.60 (RMSE), 13.29 (MAPE), and 2769.22 (MSE) for without spatiotemporal (see Table 4). The proposed model outperforms to the base deep learning model (MLP, LSTM). The performance measure for the proposed mode to 40.61 (RMSE), 10.68 (MAPE), and 1649.56 (MSE) for spatiotemporal (see Table 4). The Comparison of actual vs predicted regression plot for with and without spatiotemporal (Fig. 10). The Comparison of actual vs predicted line plot for one week of Muzaffarpur Collectorate office (C) (Fig. 11).



(a) $PM_{2.5}$ prediction results CNN-LSTM with statiotemporal (b) $PM_{2.5}$ prediction results CNN-LSTM without statiotemporal

Fig. 11. The proposed model to hourly $PM_{2.5}$ prediction for one week from 13 Jan 2023 to 20 Jan 2023

7 Conclusion

Air pollution, a widespread and enduring issue that afflicts societies worldwide, poses a significant challenge requiring immediate attention. The necessity to combat this pressing problem has prompted the author to propose an innovative and transformative hybrid model. The region adjacent to the collectrote office (C), Bihar is also contaminated, and the repercussions of this pollution are also detrimental. Both Buddha colony (B) and Daudpur kothi (D) are situated closer to the air pollution station near the collectrote office (C). This proposed model, which cleverly incorporates the principles of spatiotemporal correlation analysis, has emerged as a promising solution for predicting air quality. Remarkably, this model has exhibited unparalleled excellence, surpassing the performance of its predecessors by considerable margins. Significant advancements have been observed, with improvements of 64.26 per cent in the Mean Square Logarithmic Error (MSLR), 67.75 percent in the MSE, 24.46 percent in the MAPE, 35.20 percent in the MAE, and 29.52 percent in the RMSE. The increasing levels of concern expressed by the government and the public regarding the harmful effects of air pollutants on human health and the environment have become increasingly evident and undeniable.

Declarations

Availability of Data and Material. The data is publically available at (https://airquality.cpcb.gov.in/ccr/#/caaqm-dashboard-all/caaqm-landing).

Competing Interests. The authors have no conflicts of interest to declare.

Funding. This study did not receive explicit financial assistance from funding organizations in governmental, corporate, or philanthropic domains.

References

Albawi, S., Mohammed, T.A., Al-Zawi, S.: Understanding of a convolutional neural network. In: 2017 International Conference on Engineering and Technology (ICET), pp. 1–6. IEEE (2017)

Berral-García, J.L.: When and how to apply statistics, machine learning and deep learning techniques. In: 2018 20th International Conference on Transparent Optical Networks (ICTON). pp. 1–4. IEEE (2018)

Bhawan, P., Nagar, E.A.: Central pollution control board (2020)

Cai, L., Zhuang, M., Ren, Y.: Spatiotemporal characteristics of no2, pm2. 5 and o3 in a coastal region of southeastern china and their removal by green spaces. Int. J. Environ. Health Res. **32**, 1–17 (2022)

Chen, M.H., Chen, Y.C., Chou, T.Y., Ning, F.S.: Pm2. 5 concentration prediction model: a CNN–RF ensemble framework. Int. J. Environ. Res. Publ. Health **20**, 4077 (2023)

Chicas, S.D., Valladarez, J.G., Omine, K., Sivasankar, V., Kim, S.: Spatiotemporal distribution, trend, forecast, and influencing factors of transboundary and local air pollutants in Nagasaki prefecture, Japan. Sci. Rep. **13**, 851 (2023)

Espinosa, R., Palma, J., Jiménez, F., Kamińska, J., Sciavicco, G., Lucena-Sánchez, E.: A time series forecasting based multi-criteria methodology for air quality prediction. Appl. Soft Comput. **113**, 107850 (2021)

Fang, S., Li, Q., Karimian, H., Liu, H., Mo, Y.: Desa: a novel hybrid decomposing-ensemble and spatiotemporal attention model for pm2. 5 forecasting. Environ. Sci. Pollut. Res. **29**, 54150–54166 (2022)

Ganguly, R.: Indexing method for assessment of air quality: a case study for Dharamshala city in India. In: Global Perspectives on Air Pollution Prevention and Control System Design, pp. 68–85. IGI Global (2019)

Gao, Y., Gao, X., Zhang, X.: The 2 c global temperature target and the evolution of the long-term goal of addressing climate change-from the United Nations framework convention on climate change to the Paris agreement. Engineering **3**, 272–278 (2017)

Gautam, J., Gupta, A., Gupta, K., Tiwari, M.: Air pollution concentration calculation and prediction. In: Rathore, V.S., Worring, M., Mishra, D.K., Joshi, A., Maheshwari, S. (eds.) Emerging Trends in Expert Applications and Security. AISC, vol. 841, pp. 245–251. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-2285-3_30

Greff, K., Srivastava, R.K., Koutník, J., Steunebrink, B.R., Schmidhuber, J.: LSTM: a search space odyssey. IEEE Trans. Neural Netw. Learn. Syst. **28**, 2222–2232 (2016)

He, J., et al.: A hybrid CLSTM-GPR model for forecasting particulate matter (pm2. 5). Atmos. Pollut. Res. **14**, 101832 (2023)

Ketu, S., Mishra, P.K.: Scalable kernel-based SVM classification algorithm on imbalance air quality data for proficient healthcare. Complex Intell. Syst. **7**, 2597–2615 (2021)

Kumar, V., Ahmad, N.: Deep learning for air quality prediction after COVID-19 pandemic based on pollutant and meteorological data (2022). Available at SSRN 4292346

Kutlar Joss, M., Eeftens, M., Gintowt, E., Kappeler, R., Künzli, N.: Time to harmonize national ambient air quality standards. Int. J. Publ. Health **62**, 453–462 (2017)

Lee, B.J., Kim, B., Lee, K.: Air pollution exposure and cardiovascular disease. Toxicol. Res. **30**, 71–75 (2014)

Leptien, E., Mochalova, G., Albrecht, E.: European Union policy for sustainable development. In: Schmidt, M., Giovannucci, D., Palekhov, D., Hansmann, B. (eds.) Sustainable Global Value Chains. NRMT, vol. 2, pp. 85–106. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-14877-9_5

Li, X., Huo, H.: Prediction of pm2. 5 concentration based on CNN-BiGRU model. Acad. J. Sci. Technol. **5**, 1–8 (2023)

Liao, Q., Zhu, M., Wu, L., Pan, X., Tang, X., Wang, Z.: Deep learning for air quality forecasts: a review. Current Pollut. Rep. **6**, 399–409 (2020)

Liu, B., Wang, M., Guesgen, H.: A hybrid model for spatial–temporal prediction of pm2. 5 based on a time division method. Int. J. Environ. Sci. Technol. 1–12 (2023a)

Liu, M., Luo, X., Qi, L., Liao, X., Chen, C.: Simulation of the spatiotemporal distribution of pm2. 5 concentration based on GTWR-XGBoost two-stage model: a case study of Chengdu Chongqing economic circle. Atmosphere **14**, 115 (2023b)

Liu, X., Zhao, C., Shen, X., Jin, T.: Spatiotemporal variations and sources of pm2. 5 in the central plains urban agglomeration, china. Air Qual. Atmos. Health **15**, 1507–1521 (2022)

Ma, Y.: Spatiotemporal dynamic interpolation simulation and prediction method of fine particulate matter based on multi-source pollution model. In: E3S Web of Conferences, EDP Sciences, p. 03008 (2023)

Ma, Z., et al.: Spatial and temporal characteristics analysis and prediction model of pm2. 5 concentration based on spatiotemporal-informer model. PloS ONE **18**, e0287423 (2023)

Rincon, G., Morantes, G., Roa-López, H., Cornejo-Rodriguez, M.d.P., Jones, B., Cremades, L.V.: Spatio-temporal statistical analysis of pm1 and pm2. 5 concentrations and their key influencing factors at Guayaquil city, Ecuador. Stochastic Environ. Res. Risk Assess. **37**, 1093–1117 (2023)

Sinharay, S., Stern, H.S., Russell, D.: The use of multiple imputation for the analysis of missing data. Psychol. Methods **6**, 317 (2001)

Tan, J., Liu, H., Li, Y., Yin, S., Yu, C.: A new ensemble spatio-temporal pm2. 5 prediction method based on graph attention recursive networks and reinforcement learning. Chaos Solitons Fract. **162**, 112405 (2022a)

Tan, S., et al.: Reconstructing global pm2. 5 monitoring dataset from OpenAQ using a two-step spatio-temporal model based on SES-IDW and LSTM. Environ. Res. Lett. **17**, 034014 (2022b)

Taud, H., Mas, J.: Multilayer perceptron (MLP). In: Geomatic Approaches for Modeling Land Change Scenarios, pp. 451–455 (2018)

Wang, W., Xia, S., Zhu, Z., Wang, T., Cheng, X.: Spatiotemporal distribution of negative air ion and pm2. 5 in urban residential areas. Indoor Built Environ. **31**, 1127–1141 (2022)

Wu, Y., Du, N., Wang, L., Cai, H., Zhou, B.: Analysis of the gridded influencing factors of the pm2. 5 concentration in Sichuan province based on a stacked machine learning model. Int. J. Environ. Res. **17**, 6 (2023)

Yang, C., et al.: Spatiotemporal distributions and source apportionment of pm2. 5-bound antimony in Beijing, China. J. Geophys. Res. Atmos. **127**, e2021JD036401 (2022)

Ye, J., Wang, L., Li, G., Chen, D., Zhe, S., Chu, X., Xu, Z.: Learning compact recurrent neural networks with block-term tensor decomposition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9378–9387 (2018)

Zaini, N., Ean, L.W., Ahmed, A.N., Malek, M.A.: A systematic literature review of deep learning neural network for time series air quality forecasting. Environ. Sci. Pollut. Res. 1–33 (2022)

Zhang, X., Chen, Z., et al.: CCA-LSTM: a novel hybrid prediction network based on convolutional channel attention mechanism and LSTM for pm2. 5 concentration prediction (2023a)

Zhang, X., Li, Q., Liang, D.: An adaptive spatio-temporal neural network for pm2. 5 concentration forecasting. Artif. Intell. Rev. 1–28 (2023b)