

---

# Deep-learning Spatio-temporal prediction framework for PM under dynamic monitoring

Transportation Research Record  
2022, Vol. XX(X) 1–18  
©National Academy of Sciences:  
Transportation Research Board 2022  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/ToBeAssigned  
journals.sagepub.com/home/trr

SAGE

Vishesh Mittal<sup>1</sup> and Suhas Sasetty<sup>2</sup> and Rashmi Choudhary<sup>3</sup> and Amit Agarwal<sup>3</sup>

DOI: 10.1177/03611981221082589

## Abstract

With increasing air pollution and its harmful effect on the residents of developing countries, the prediction and analysis of pollutants have become an important research aspect. This study focuses on the spatio-temporal prediction of hourly particulate matter with different deep learning modeling techniques for Delhi, India. The secondary data of particulate matter concentrations and the meteorological parameters for the four static monitors in the area are collected from Central Pollution Control Board (CPCB) for dates between January 2019 to April 2021. The study area of South Delhi is divided into hexagonal grids. The datasets at the centroid of each grid are formulated with the spatial interpolation method of inverse distance weighting and Kriging. The hexagonal grids are required to collate the data coming from dynamic monitors. Three models with CNN, LSTM, and CNN-LSTM are developed for a total of fifteen cells. To evaluate developed models, mean absolute error and root mean square error are used. The results from prediction models show that CNN-LSTM models outperform the other two models. The predictions are accurate for the CNN-LSTM model compared with the values obtained from the static monitor. Also, compared to the existing and individual models, the proposed hybrid CNN-LSTM model results performed better for most of the cells. The prediction models can also provide the pollutant concentration on various routes, which can assist residents in making travel choices based on the air pollution prediction information. Planners and practitioners can replicate the developed models in other regions.

**Keywords:** Air Quality prediction, Particulate matter (PM<sub>2.5</sub>), Deep learning, CNN, LSTM, CNN-LSTM

## Introduction

With increasing motorization, industries, and urbanization, the air quality is deteriorating in the urban agglomerations of developing countries. Urban air quality indicates key concerns to the health of urban residents (1). In cities, people face high exposure while traveling for different activities and performing various activities throughout the day. People need to know the information of air pollution exposure to make choices to escape from negative externalities. The exposure to pollution is heterogeneous and dynamic. Particulate matter pollution contributes to the majority of air pollution in Indian cities. Air pollution makes people face various health problems such as throat and lung infections, heart and respiratory diseases, etc. (2, 3). The particulate matter concentrations vary swiftly with distance from sources of pollutants, wind speed, wind direction, etc.

The travel behavior of residents has been studied in literature exploring the influence of air pollution on travel choices. It is observed that along with the choice of travel modes such as active transportation, public transit, and private vehicles, route choices also vary with exposure to pollution (4). The accurate concentrations at street

level are required for providing alternate routes with information of pollution exposure on each alternative. The pollution monitoring of different types of pollutants and their individual effects on human health can assist urban residents in understanding the impacts of varying ambient air quality levels and plan preventive measures. To prevent the alarming scenarios due to degrading air quality, comprehending the situation in advance with the help of monitoring and prediction is advantageous. Accurate prediction models and forecasting are essential for making informed travel choices through which air pollution can be reduced. A dense network of monitors is essential for adequate data availability and good prediction models. The static monitors are costly

---

<sup>1</sup> Department of Computer Engineering, Vivekanand Education Society's Institute of Technology, India

<sup>2</sup> Department of Computer Science Engineering, SRM Institute of Science and Technology, India

<sup>3</sup> Department of Civil Engineering, Indian Institute of Technology Roorkee, India

**Corresponding author:**

Amit Agarwal, amitfce@iitr.ac.in

and cannot be densely placed in large cities. Therefore, spatial interpolation methods are employed for estimating the pollutant concentration at high spatial resolution. However, determining the pollutant concentrations at a later date and time, i.e., temporal predictions, can help people plan their future travel choices. Various models were used in the literature for spatial, temporal, and spatio-temporal predictions. Inverse distance weighting (IDW) method for spatial was used to study the prediction of daily PM<sub>2.5</sub> concentration, for which data from eight monitors in Delhi was used excluding the meteorological parameter (5).

The daily PM<sub>2.5</sub> was predicted at target locations, and the profiles were plotted using spatial interpolation techniques (Kriging and IDW) over data collected from 17 monitoring stations in Delhi. The observed error (difference of predicted and actual value) of Kriging and IDW was found out to be 22% and 24%, respectively (6). The trans-Gaussian spatial prediction was used to generate interpolation maps that helped in visualization of spatial variation for particulate matter (PM<sub>2.5</sub>) and gaseous pollutants (NO<sub>2</sub> and SO<sub>2</sub>) in Egypt (7).

Initially, for forecasting pollutant concentrations, regression-based models were used. Various regression-based models were used to study the air quality monitoring in Taiwan using the hourly pollutant data from 2012 to 2017, which was later converted into daily and monthly data for modeling (8). The authors considered and compared linear regression, Lasso regression, Ridge regression, random forest regression, K-neighbors regression, multi-layer perceptron regression, and decision tree regression models. A multivariate linear regression model was proposed to achieve short period prediction of PM<sub>2.5</sub> in Beijing, China. The data for aerosol optical depth obtained using remote sensing, meteorological factors (wind velocity, temperature, and relative humidity) acquired from ground monitoring, and other gaseous pollutants were considered for model development (9). The regression models are sensitive to outliers and are limited to only low-complexity features.

Along with the emerging usage of multiple regression methods, machine learning models became popular for prediction modeling. Air quality prediction was performed using linear regression, Support Vector Machine (SVM), decision tree, and Lasso regression methods, and the models were evaluated using R<sup>2</sup> values (10).

The prediction modeling of PM<sub>2.5</sub> for neighboring cities of Delhi, India, with data of six and a half months for the year 2019, was performed. A stacked regression model was proposed, which provided better results than other regression models (11).

Further, the basic machine learning techniques were used to develop a Geographically Weighted Predictor (GWP). The models were developed using data from only four locations. Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and neural network were used to investigate

atmospheric pollutant concentration prediction (12). SVM has a better ability to learn complex features, but it was not found suitable for studies with large datasets and lacks transparency in results (13). To overcome limitations of SVM online scalable SVM ensemble learning method was proposed (14). Similar to SVM, the random forest machine learning method also has an excellent ability to comprehend complex features, but real-time calculation speed and effectiveness are low. Prediction modeling and forecasting are performed through deep-learning methods for obtaining better prediction results spatiotemporally. Techniques such as Convolutional Neural Network (CNN), Artificial Neural Network (ANN), and Recurrent Neural Network (RNN) are used in deep learning. CNN is a widely used network model and image processing technique proposed by (15), and long short-term memory (LSTM) is a network model proposed by (16) to solve the longstanding problems of gradient disappearance in Recurrent neural networks (RNN), and used widely in time-series forecasting (17). The use of multi-output and multi-index of supervised learning based on LSTM was shown to predict the PM<sub>2.5</sub> based on 35 monitoring stations' data (18). It was found in the literature that the ANN worked better than conventional methods such as regression, fuzzy logic, and principal component analysis for pollutant prediction (19, 20).

With developing methodologies and the need for accurate results, hybrid models were explored by researchers. (21) has combined different models of various monitors (industrial, external source, and local monitors) into an aggregated-LSTM model to predict PM<sub>2.5</sub>. The results of different regression models based on SVM, gradient boosted tree regression, LSTM, etc., were compared and observed improved results in the aggregated model. APNet, a CNN-LSTM model, was developed formulated on 24 hours cumulated wind speed, duration of rain (in hours), to predict PM<sub>2.5</sub> of the next hour (22). (23) developed a CNN-LSTM model based on a time-distributed, one-dimensional CNN layer for forecasting the next 24 hours PM<sub>2.5</sub> concentration in a study conducted for Beijing, China. The use of two-dimensional CNN layers and batch normalization could improve the model further. An integrated dual LSTM with sequence to sequence technology was used to establish a single-factor prediction model to obtain the predicted value of each component in air quality data including the data of neighboring stations and weather parameters (24). After which, the attention mechanism-LSTM model was used as the multi-factor prediction model. XGBoosting tree was used to integrate these two models.

In India, Delhi is severely affected by air pollution, and with increasing motorization, the situation is aggravated. Lack of consistent particulate matter database at high spatial and temporal resolution obstructs the environmental assessments, accurate predictions, and modeling (25). The daily, seasonal, and regional variations also need to be

explored along with predictions at different locations. The use of numerical methods can be time-consuming, and therefore, deep-learning or hybrid models which overcome the limitations of other models should be developed for quick and efficient predictions. With increasing air pollution, efficient and accurate prediction models are required to lessen the impact by taking suitable measures against increasing air pollution.

In previous studies, the prediction models were either built for only one monitoring station or based on the data from all the monitoring stations in the vicinity combined in one model, limiting the results to a particular area. Notably, the data used for prediction is primarily from the static monitors. Given the advances in the technology of low-cost, portable air pollution monitors, the dynamic, real-time air pollution and meteorological information can be obtained by installing them on buses/ trams (26, 27). Selection of optimum routes for placement of mobile monitors for covering the need of high-density of monitoring stations was explored for high spatio-temporal resolution, which can additionally increase the dataset for prediction algorithms development (27). For real-time air pollution monitoring, bus routes were selected based on route coverage image analysis based on the non-satellite image of the study area (28). The existing prediction methodologies do not apply to such dynamic data or dynamic monitoring networks. Dynamic monitoring network refers to deploying devices/sensors (e.g., portable air pollution monitors) on traversing vehicles/ UAVs. The devices will send the data at different locations as the vehicle moves and at different time bins. Clearly, data does not belong to a fixed location, and thus, appropriate methods are required to handle/ process the data coming from dynamic networks. The current study focuses on providing a prediction framework that can also consider the dynamic data and provide high precision spatio-temporal predictions for PM<sub>2.5</sub>. The use of hexagonal grid formations and model development for each cell can help overcome the limitations of previous studies, as the pollutant concentration depends highly on local parameters. This study aims to develop a prediction model framework for high spatial and temporal resolution considering the historical particulate matter pollutant concentrations and meteorological data from the dynamic monitoring network. The current study explores the prediction modeling using three different deep learning methods and compares the results to provide the most suitable model.

## Data & Methods

For the modeling based on historical data, secondary data in the form of pollutant concentrations and meteorological parameters are collected from Central Pollution Control Board (CPCB) (29). Hourly data from multiple air pollution monitoring stations are collected for January 2019 to April 2021. **There were no monitors in South Delhi before the year**

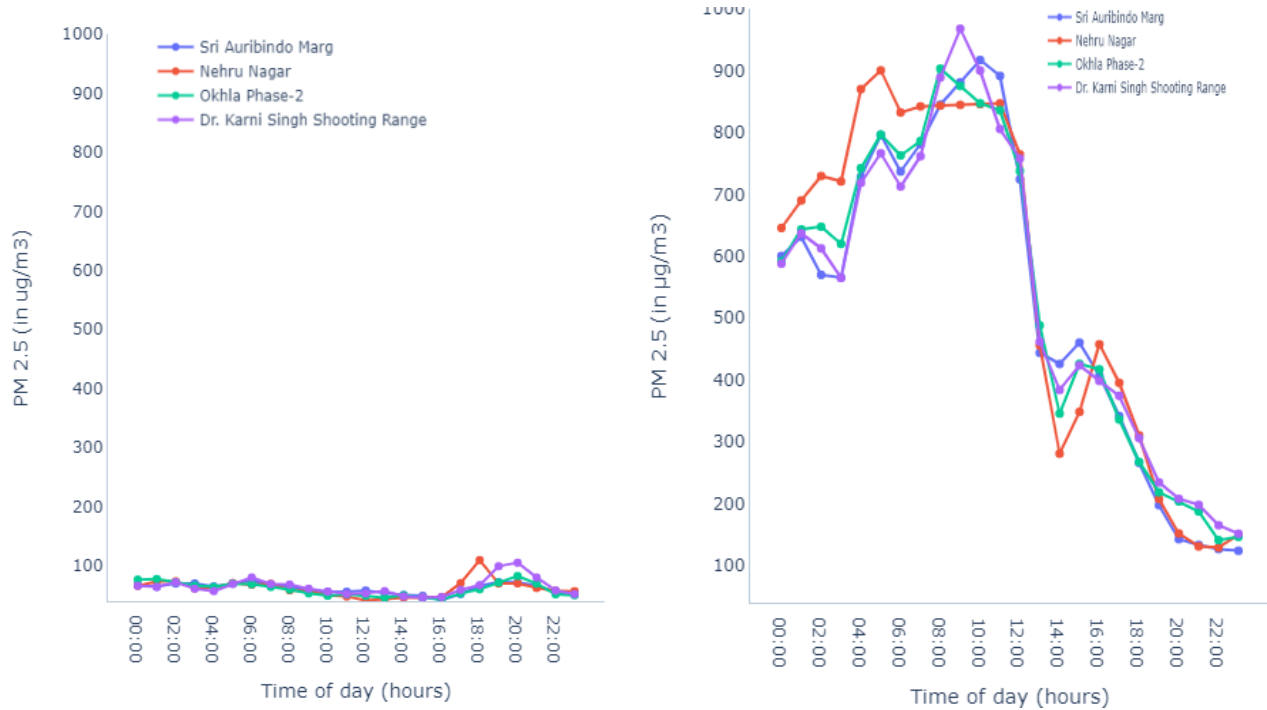
**2019, and therefore, historical data for 2019-2021 is considered in the present study (29).** Meteorological parameters are included in prediction models as these parameters are likely to influence the pollutant concentrations significantly (30). Particulate matters, PM<sub>2.5</sub> and PM<sub>10</sub> (in  $\mu\text{g}/\text{m}^3$ ) and the meteorological parameters, such as, temperature (in  $^{\circ}\text{C}$ ), barometric pressure (in  $\text{mm Hg}$ ), humidity (in %), wind speed ( $\text{m}/\text{s}$ ) and wind direction are used in this study. The clockwise angle from the north defines the wind direction.

The hourly variations of PM<sub>2.5</sub> and PM<sub>10</sub> (aerodynamic particles size less than 2.5 and 10 micrometers) concentrations of the four static monitors on two different days in different seasons are shown in Figure 1. It can be observed that PM<sub>2.5</sub> and PM<sub>10</sub> concentrations are high in November (winter season) compared to June (summer season) due to various reasons. At the same time, the value varies significantly with the time of the day. **Increased local emissions due to crop and biomass waste burning, winter inversion (low wind speed with dip in temperature leads to accumulated pollutants at lower height), and absence of photo-chemical reactions in winters assists in increased concentrations in the winter season compared to summer (31).**

## Study Area

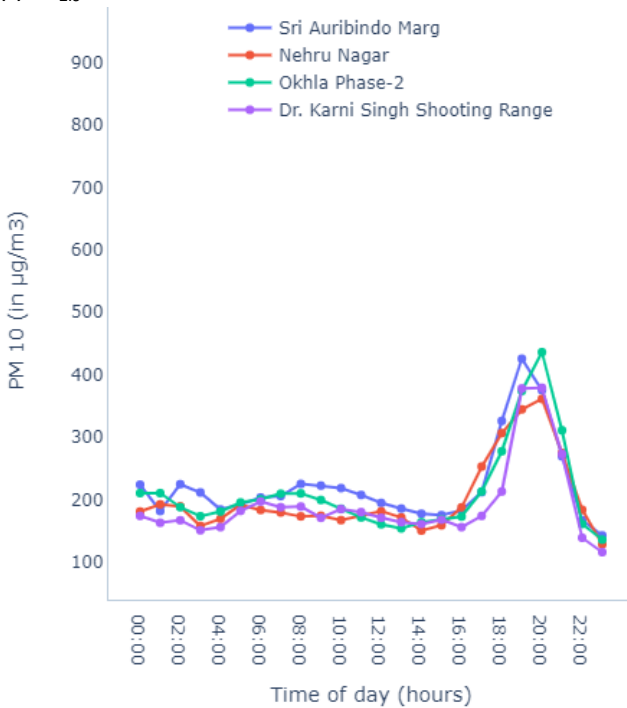
The majority of Indian cities suffer from extremely high levels of urban air pollution, particularly in the form of small-sized particulate matter (5, 32). According to the World Health Organization (WHO), the capital city of India, Delhi is one of the most polluted cities in the world (5). Pollution in Delhi usually spikes during the winters. For instance, it exceeded almost 40 times the WHO's ambient concentration limits in the first week of November 2019 for PM<sub>2.5</sub>, i.e., a concentration value of around  $950 \mu\text{g}/\text{m}^3$  (33). Delhi is divided into 11 districts, and out of which the South Delhi district has the highest number of static monitors concerning its area. There are 4 monitors in the South Delhi district for which the data is collected from the CPCB. The available static monitors in the area located at Dr. Karni Singh shooting range, Nehru Nagar, Okhla Phase- 2, and Sri Aurobindo Marg, all of which are deployed by Delhi Pollution Control Committee (DPCC) authority (see Figure 2).

The data should belong to monitors closer to the location where prediction is to be made for having better prediction. Ideally, this would mean having a dense monitoring network. A model trained with the more extensive urban agglomeration area data is less likely to be accurate than multiple models trained with the data from localized, multiple, smaller zones. Further, in a dynamic monitoring network, having data from the same point is not possible, **i.e., data will be distributed over space and time.** Therefore, the cells are created, and models are developed for each cell by converging the data points at the centroid of the cell. In spatial analysis, grids of points or polygons are frequently used to sample, index, or partition an area. The



(a) PM<sub>2.5</sub>, June 05, 2019.

(b) PM<sub>2.5</sub>, Nov. 03, 2019.



(c) PM<sub>10</sub>, June 05, 2019.

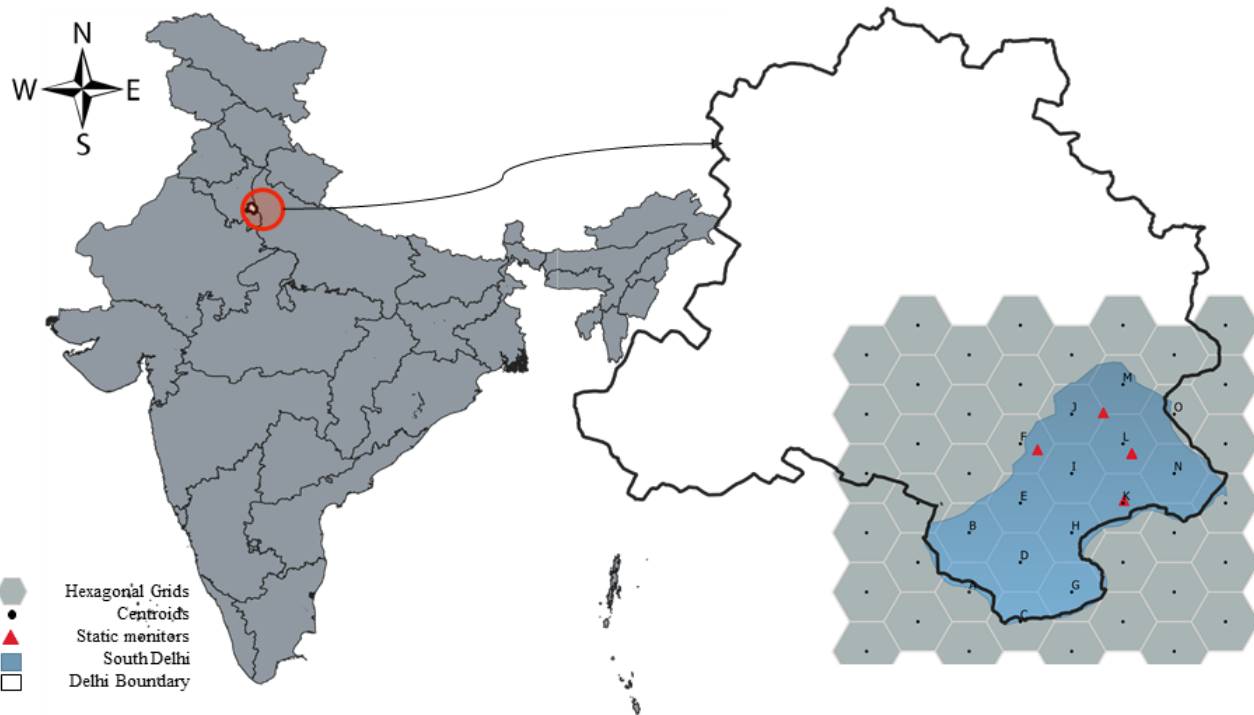


(d) PM<sub>10</sub>, Nov. 03, 2019.

**Figure 1.** Seasonal variation of PM<sub>2.5</sub> and PM<sub>10</sub> for June 05, 2019 (left) and Nov. 03, 2019 (right).

square and triangular shapes are commonly employed in the grid designs because of their simplicity in the definition, data storage, and ease of re-sampling to different spatial scales. Although hexagonal grids are not very common, these are

becoming more popular, keeping the fact that hexagonal grids are a better approximation of the human vision grid (34). In the past, many studies used hexagonal grids for spatial visualization, and analysis (35). Regular hexagons feature



**Figure 2.** Hexagonal grids and static monitors in South Delhi

additional symmetries and are the closest form of the circle, so they are more appropriate for spatial data sets that suit this study perfectly. Hexagons are better at fitting the earth's curvature than squares when dealing with a large spatial area where the globe's curvature becomes essential. Therefore, hexagonal grids of 5 km side-to-side spacing covering the whole area of South Delhi are opted for the current study, as shown in Figure 2. With this, there are a total of 15 cells in South Delhi (see Figure 2), for which the prediction models are developed.

### Dataset Formation

The detailed process of flow of this paper is shown in Figure 3. The first part is to select the study area and collect the hourly dataset of static monitors in south Delhi. The next step is to generate augmented datasets using various interpolation techniques like IDW and Kriging., followed by preprocessing the data, including interpolating the null values, normalizing the data, and converting the data into a sliding window technique. Further, the dataset is divided into training, validation, and testing sets. Then, various deep learning models, such as CNN, LSTM, and CNN-LSTM, are developed for model prediction. Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) is used for comparing the model's accuracy.

**Data Preprocessing** Preprocessing of data is performed to address the noise in the data, missing values, and other erroneous variables in the collected data to develop a

robust prediction model. In machine learning models, data preprocessing refers to converting the raw data into readable input for the machine learning model. There were missing values in the air quality data, and the reason could have been sensor device failure or network issues in data storing. Null values in a dataset can lead to inaccurate results or failure of the model itself. The null values are replaced with mean values of the previous hour and following hour values to eliminate the errors in the model that can occur due to missing values.

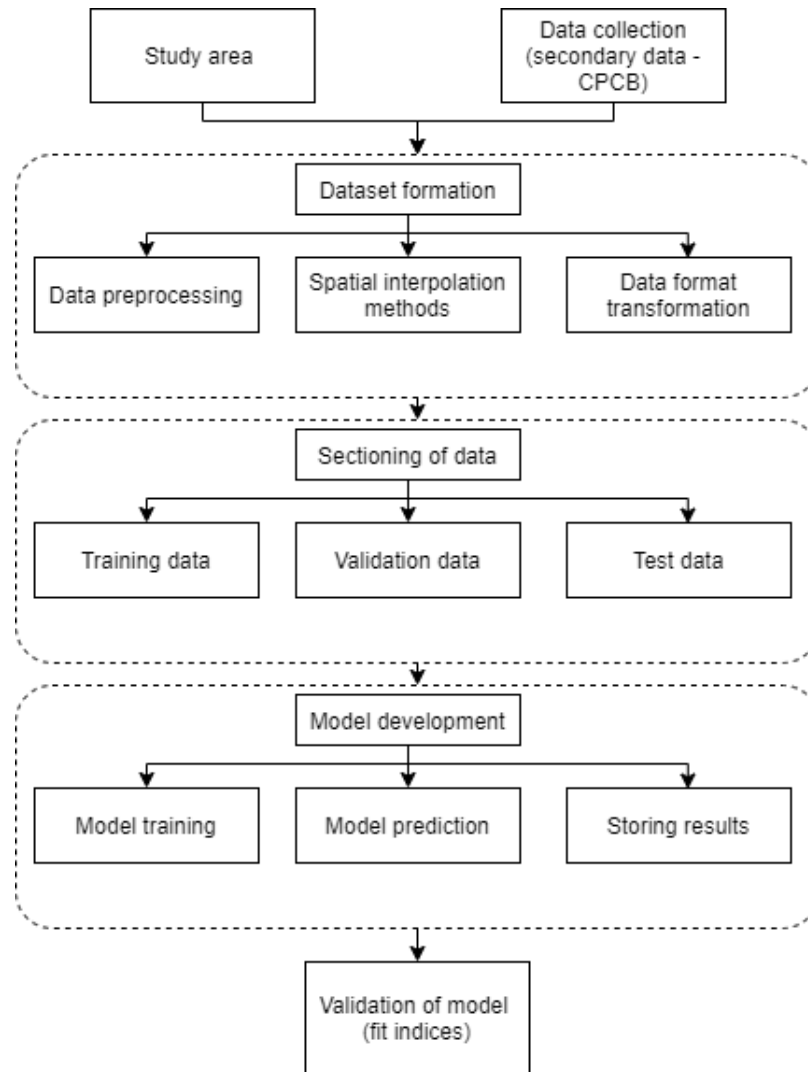
**Normalization** Normalization gives equal importance to each parameter so that no single variable control model performance in one direction just because they are in more absolute terms. Normalization of the dataset can also help in boosting the training speed and improving prediction accuracy. Also, variables assessed at different scales may not contribute equally to the model fitting and which might lead to bias, so the dataset is normalized between 0 and 1 using Equation (1).

$$X' = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad (1)$$

where,  $X_i$  is the original sample value,  $X_{max}$  and  $X_{min}$  are the maximum and minimum value of the whole parameter of original data and  $X'$  is the value after normalization.

**Correlation** Understanding the relationship between the parameters is essential while studying the machine learning





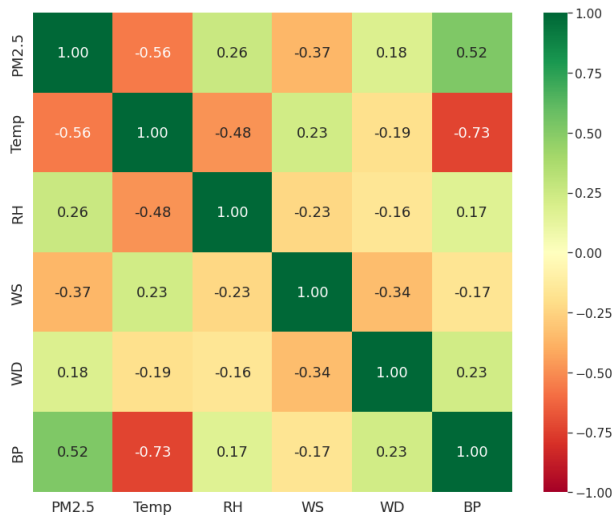
**Figure 3.** Methodology for the prediction modeling

model dataset. The Figure 4 displays Pearson's correlation of heatmap for different parameters in the dataset. The Pearson's correlation ranges from -1 to 1, where the absolute value shows the intensity of correlation (1 being highest, 0 being lowest), and the sign shows the direction (positive or negative). The correlation value close to zero indicates the independence and absence of a relationship between the parameters. In the Figure 4, higher correlation values are represented as dark green/ red panels.

For  $PM_{2.5}$ , a negative correlation is observed with temperature and wind speed, which indicates that with the increase in temperature or wind speed, the  $PM_{2.5}$  concentration will decrease. It can also be verified with Figure 1, which shows a lower concentration of  $PM_{2.5}$  in summers compared to winters. A positive correlation of  $PM_{2.5}$  with barometric pressure and relative humidity is observed, which shows that with an increase in pressure and humidity, the concentration of particulate matter also

increases. There are no parameters with exceptionally strong correlation. Therefore, all the parameters are considered for further modeling.

*Spatial Interpolation Techniques* After the hexagonal grid formation for dividing the study area into smaller spatial regions, the data point for each cell is fixed as the centroid of each hexagon, respectively. Several spatial interpolation methods were used widely in past studies, such as IDW, Kriging, spline, Thiessen polygon, kernel density estimation, etc. (36). There is no rule of thumb for the selection of interpolation methods. Two of the available spatial interpolation techniques are widely used and, thus, employed to generate the dataset at each centroid for grids. Figure 5 explains the steps followed for the creation of an augmented dataset for each cell. Initially, the raw data from four static monitors in South Delhi is collected from CPCB. Then the centroids of all the cells are located. The



**Figure 4.** Correlation between input parameters of the model (Temp: temperature, RH: relative humidity, WS: wind speed, WD: wind direction, BP: barometric pressure)

inverse distance weighting method is used to determine the pollutant concentrations ( $PM_{2.5}$  and  $PM_{10}$ ) at centroids of the hexagonal cells, and the ordinary 2D Kriging method is employed for the spatial prediction of meteorological parameters such as relative humidity and barometric pressure.

In addition to the data from static monitors, the different data points from the dynamic monitoring networks within a grid will be collated using IDW and 2D Kriging to the centroid of the grids, as explained in the following sections. The results from both methods are concatenated at the end to form a single dataset for each cell.

**Inverse Distance Weighting (IDW)** In IDW, the known points (points where data is directly measured from monitors) on the determined values (from interpolation) depend on the distance between known and unknown points. The controlling power of the influence can be externally defined for a smooth resulting surface. Power of 2 is most commonly used in the literature (5) and hence used in this study. The mathematical formula for IDW is given in Equation (2).

$$z_j = \frac{\sum_i z_i}{\sum_i \frac{1}{d_{ij}^n}} \quad (2)$$

where,  $z_i$  is value at known point,  $d_{ij}$  is distance between known and unknown point,  $z_j$  is predicted value at unknown point and  $n$  is a user selected exponent (often 1, 2 or 3).

**2D Kriging** Kriging's method of spatial interpolation is flexible to adopt different parameters and variogram forms and has a wide range of applications (36). In IDW, the weights are solely dependent on the distance, while in Kriging, the weights depend on the overall spatial arrangement of measured points and the distance between measure and prediction points. When IDW is applied for

the spatial interpolation of meteorological parameters, the interpolated values for relative humidity and pressure are observed far from the actual value. Therefore the Kriging method is employed, which is more realistic. The equation for Kriging interpolation is shown in Equation (3). Kriging is a two-step process, it includes:

- creation of variograms and co-variance to determine the spatial auto-correlation, and
- prediction of unknown values.

Different types of variograms as circular, spherical, Gaussian, exponential, and linear can be used. Ordinary Kriging is widely used, with an assumption of unknown constant means. In this study, the PyKriging toolkit is used to perform ordinary 2D Kriging interpolation on the dataset (37).

$$z_j = \sum_{i=1}^n \lambda_i z_i \quad (3)$$

where,  $z_i$  is value at known point,  $z_j$  is the predicted value at unknown point,  $\lambda_i$  is unique weight for location  $i$  and  $n$  is a user selected exponent (often 1, 2 or 3).

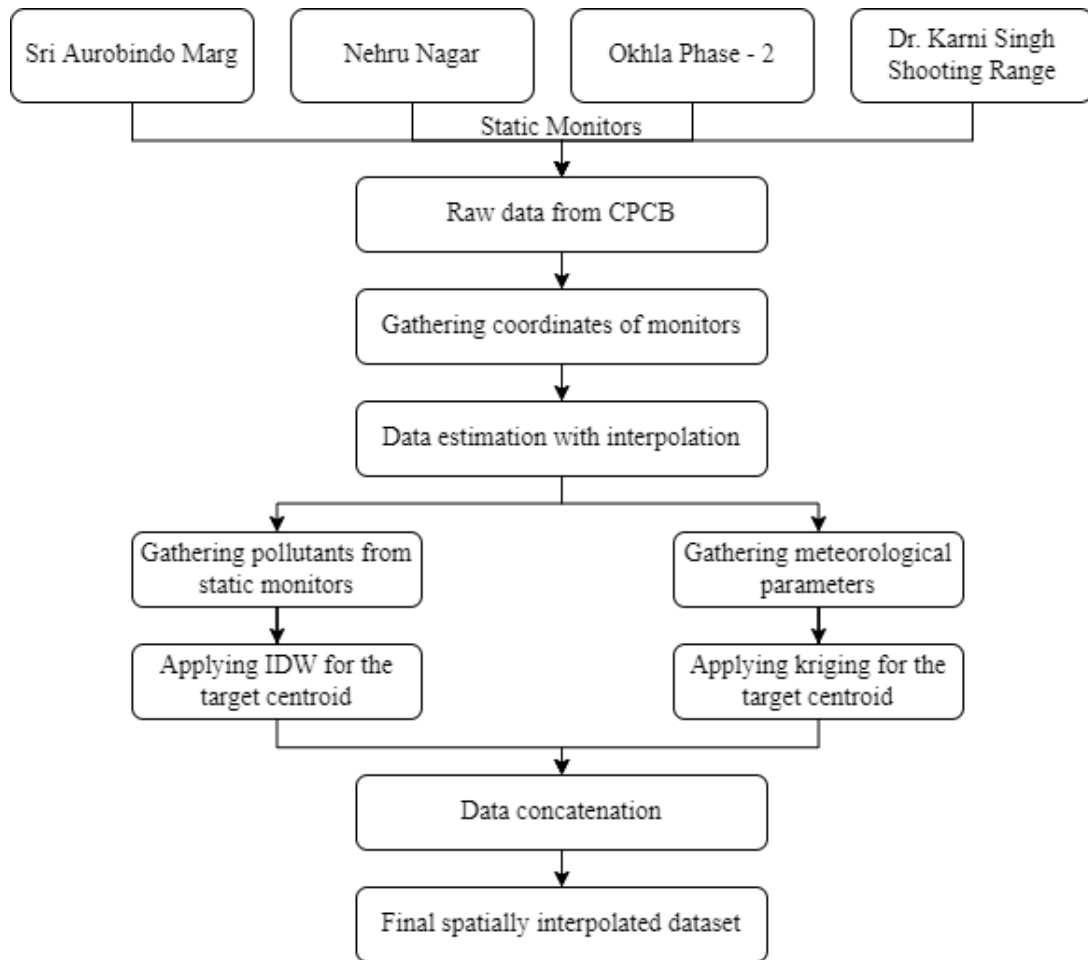
**Data transformation** Data transformation refers to the process of converting raw data into data ready for modeling by removing unused columns, changing data types, altering the timestamps, and handling the errors in data. The factors which influence variables from time to time are determined through time series models. While training the model, the results for a subsequent hour, from various sliding window sizes, are used varying from 24 to 96 hours, out of which 72 hours is found as the most optimal value during the model formulation and experimentation. For this study sliding window approach is adopted, where the dataset of duration from January 2019 to April 2021 was divided into 72-hour window and 1-hour stride.

This technique is followed for all the models to reshape the information by a fixed window so that the model is comprehensible for the complete information possible at a given period to achieve an accurate prediction.

### Data sectioning

For the development of models, based on literature, the data is divided into three sections, as follows:

- Training data: The section of data used to learn, see and train the models, 80% of the prepared dataset is used to train the models.
- Validation data: This data section is used to validate and tune the trained model by altering the hyperparameters. The models do not learn from validation data and are only used for development. 10% of the dataset is used for validating the models.
- Test data: This part of data provides an unbiased evaluation of the trained and validated models. The



**Figure 5.** Augmented dataset creation

remaining 10% of the dataset is used for testing the models.

After sectioning, machine learning models are developed on the training data, which are explained in the following section.

## Model Development

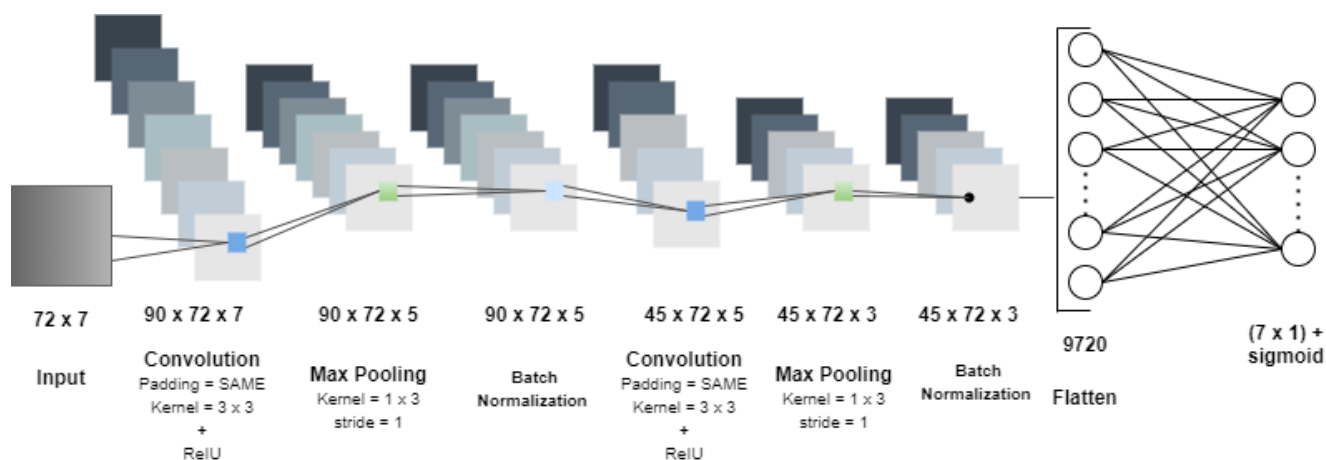
It is evident from exploratory data analysis that the pollutants and meteorological parameters highly influence one another. These parameters tend to follow specific trends and patterns that are extracted using deep neural networks specially designed for this task. In this study, three types of models are used, i.e., Convolutional Neural Network (CNN), Long-Short Term Memory (LSTM), and CNN-LSTM, as discussed in the literature review.

### Convolutional Neural Network

CNN models are mainly used for visual tasks but are not limited to these. Figure 6 shows the architecture for the

CNN model, which is applied for prediction of  $PM_{2.5}$ . CNN is used to capture various trends of pollutants and other meteorological parameters by treating the  $72 \times 7$  matrix, which is created using a sliding window approach, and a pattern is identified using CNN pattern recognition (38). Each convolution layer contains a series of filters known as kernels. A kernel is a matrix that moves over the input data, performs the dot product with the sub-region of input data, and gets the output as the matrix of dot products, and each filter tries to learn a new trend in the previous hours. Through hyperparameter tuning of the CNN architecture model, the best parameters to the architecture are shown in Figure 6. The CNN architecture shown at Figure 6 uses a  $3 \times 3$  kernel with 90 feature maps with stride set to 1 and padding set to SAME. Rectified Linear Unit (ReLU) is used as an activation function, which can solve the problem of gradient disappearance, and its calculation speed and convergence speed are faster than other activation functions (39). It is followed by a max-pooling layer of size  $1 \times 3$ , which reduces its resolution and complexity to achieve translation invariance. After that, there is a layer of Batch Normalization





**Figure 6.** Architecture for CNN model

that helps deep neural networks to standardize the inputs to a layer for each mini-batch. This is followed with the same sequence of layers with 45 filters in convolution layer followed again with a max-pooling layer of size  $1 \times 3$  and a batch normalization connecting to a dense network, using Sigmoid in the end to predict the output. As a consequence of the normalized values (range  $[0,1]$ ), the Sigmoid function at the outer layer is logical. The Sigmoid activation function is used in several previous studies for deep learning models (22, 23). At the last stage of the network, with the help of a flatten layer, the output from the Batch Normalization layer was fed into a fully connected dense layer that is reshaped into 9720 neurons ( $45 \times 72 \times 3$ ), 7 parameters are predicted, including the pollutant concentrations and the meteorological parameters.

### Long-Short Term Memory

Long short-term memory (LSTM) is a variant of the traditional neural networks that exhibit dynamic behavior. It helps the network capture the data's context for various operations, especially time series forecasting. Since LSTM models are generally used for sequential analysis and have the capability of long-term dependencies, they can be trained to predict all the parameters for air quality for the next hour by the historical data gathered for various grids. The proposed system is trained with the dataset at the centroid for each cell, shown in Figure 2.

A four-layer LSTM architecture is employed for the current study. A  $72 \times 7$  matrix data layer is used for the input, which depicts the data for 72 hours and 7 parameters. Hyperparameter tuning is performed to get good prediction model architecture on LSTM architecture by working with different LSTM cells and layers and used the best prediction architecture as shown in Figure 7. The first two LSTM layers have 80 and 50 LSTM cells, respectively, with a dropout layer of 25%. It manages to stop all LSTM cells in a layer

from synchronously optimizing their weights, thus helping the model prevent it from over-fitting conditions. Then the following layers are supported by one more LSTM layer with 25 cells followed by a fully connected layer and output neural network layer with activation function as sigmoid as shown in Figure 7.

### CNN-LSTM Model

The present study also proposes a hybrid CNN-LSTM model, which combines the time-series model of LSTM with CNN to extract effective features from data. The architecture of the proposed CNN-LSTM is shown in Figure 8. The inputs of CNN-LSTM are the tensors of the PM concentrations and meteorological parameters over the last 72 hours. The output is received in terms of prediction for all the input data parameters for the next hour. Unlike regular CNN or LSTM architectures, the first half of the architecture is CNN and is used for feature extraction. The following half of the architecture is LSTM forecasting, which analyzes the features extracted by CNN and then estimates the PM concentrations and meteorological parameters. Moreover, to improve the accuracy, additional batch normalization and dropout layers are added to the architecture of CNN-LSTM.

CNN is utilized for feature extraction; specifically, two two-dimensional convolution layers and two batch normalization layers are constructed. To prepare the data into the format required by the LSTM, a reshape layer is connected to the LSTM, it is a layer that reshapes inputs into the given shape. As the various filter maps contain the reoccurring patterns and features now, they feed to LSTM. There are two LSTM layers and in-between a dropout layer to avoid over-fitting, which is a common aspect in neural networks, and there are many solutions available. Among all, dropout is one of the efficient ones. Therefore, a dropout layer is added, whose output is connected to the LSTM layer for prediction and finally joined to a fully connected layer and

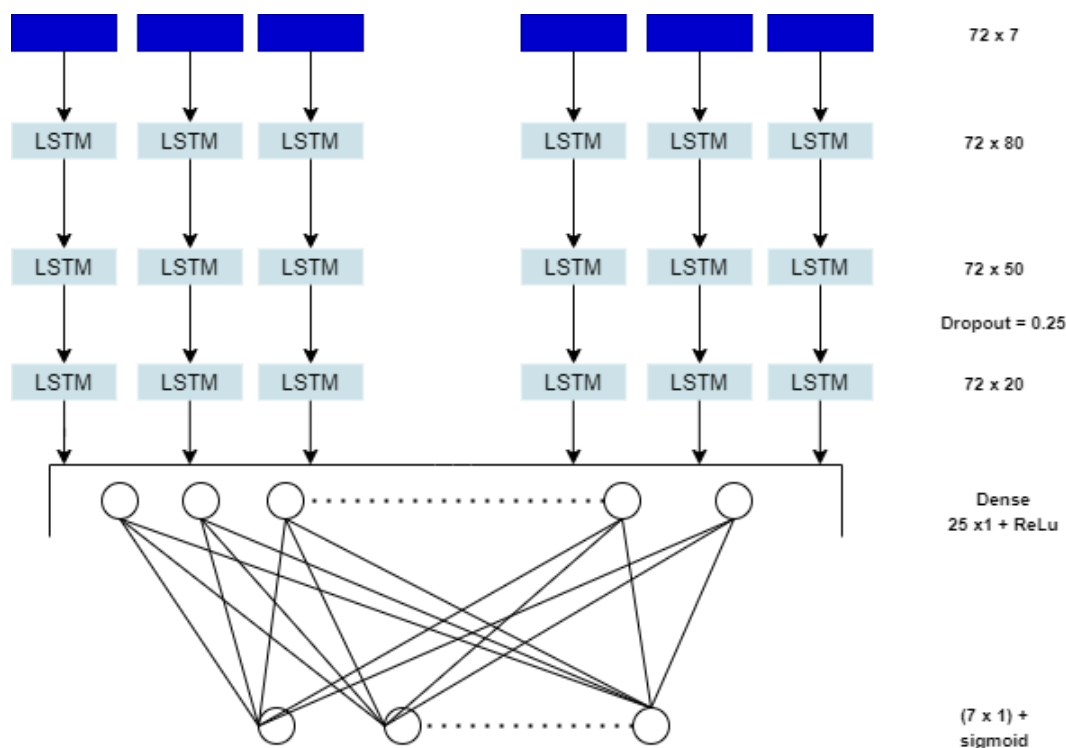


Figure 7. Architecture for LSTM model

an output layer as shown in the Figure 8. The CNN-LSTM architecture used in this study comprises two convolutional layers each of size 100 and 50, respectively, i.e., the first convolution layer has 100 unique feature maps used to extract the various trends and the second convolution layer has 50 unique feature maps. Each of the convolution layers is followed by a batch normalization layer which solves the issue of internal covariate shift (40). The output from the second batch normalized layer is reshaped to feed it as the input to the LSTM layer, and there are 50 units in the first LSTM layer, which are followed by a dropout of 0.25 that prevents neural networks from overfitting (41). Finally, one more LSTM layer is added of 25 units, followed by a fully connected dense layer that gives final output using sigmoid as an output function.

## Results & Discussions

This study explores deep learning techniques for spatio-temporal prediction of particulate matter and meteorological parameters. For this, three deep learning models are developed for the temporal prediction of  $PM_{2.5}$ , which provides an understanding of the trends and prediction for particulate matters over a selected region of Delhi. For the development of the deep learning models, Google Colab platform is used (42). Each of the models is trained with 30 epochs. The configuration included 2vCPU, 12 GB RAM, and GPU performances of 4.1 and 8.1 TFLOPS for Tesla K80

and T4 GPUs. Average training time per epoch is observed as 3 seconds for CNN, 6 seconds for LSTM, and 16 seconds for proposed CNN-LSTM model. With increasing complexity the computational burden of the model also increases.

One of the static monitors is at the centroid of cell K; therefore, the actual data is available for this cell and used for sanity testing. The actual values are the  $PM_{2.5}$  concentration data from the static monitor. In contrast, the interpolated values are obtained from applying IDW on data from all four static monitors and estimating the values at the centroid of a cell. As proposed in the present study, the data from mobile monitoring network would be aggregated to these cells. Since the dynamic monitoring network was not deployed, no historical data was available. Due to the COVID-19 pandemic, the portable devices could not be deployed and test data is also imputed from the static monitors.

To evaluate and compare the accuracy of the deep learning models for  $PM_{2.5}$  predictions using the input data, the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) values are used. Evaluation parameters such as RMSE and MAE explain the performance of deep learning models developed for forecasting. These can be used to achieve the desired accuracy of models. MAE (Equation (4)) measures the average magnitude of the errors of the predictions without considering their direction by taking modulus. It is an average over the test sample of the absolute differences between predicted and actual observations where all individual differences have equal weight. Similarly,

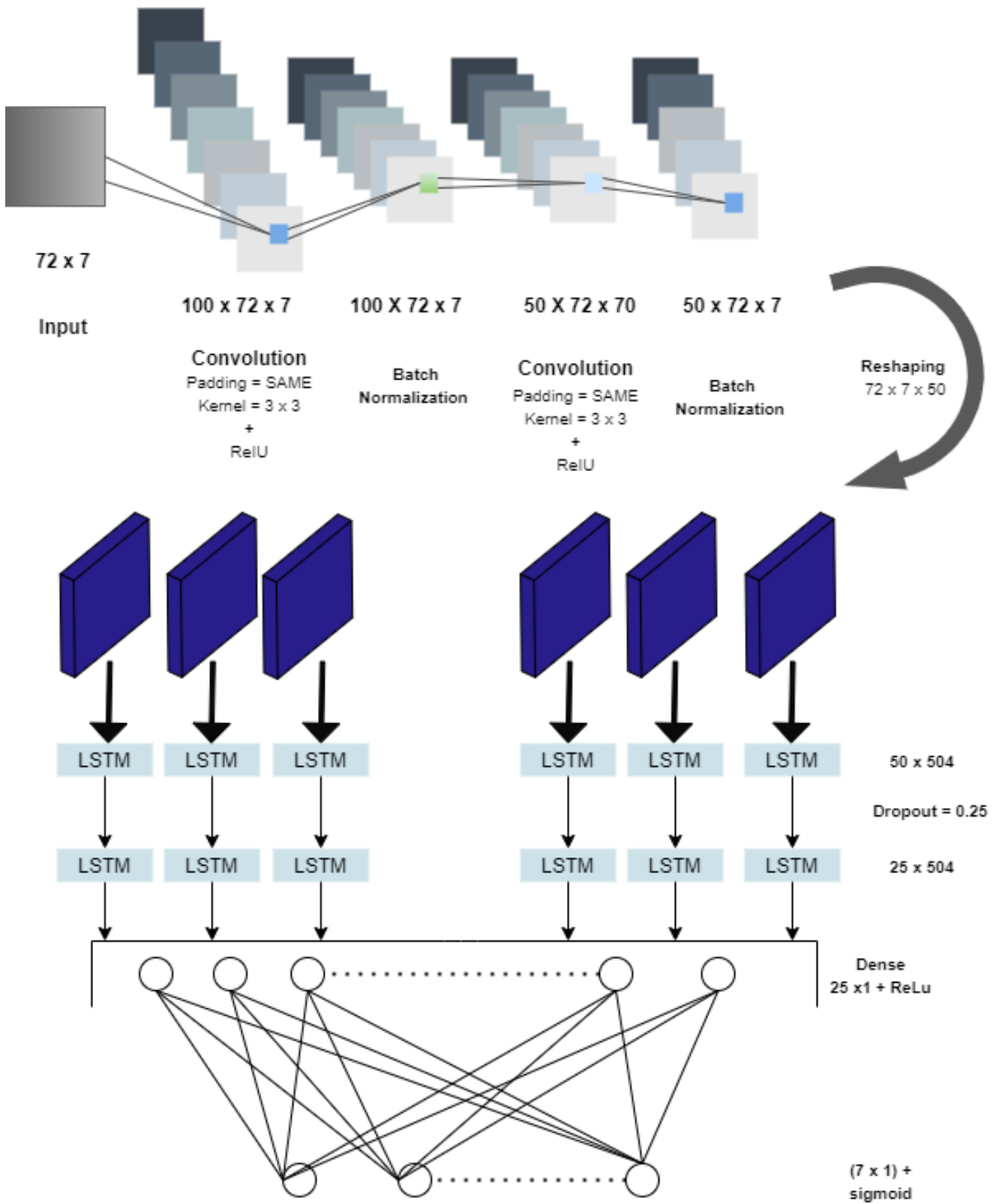


Figure 8. Architecture for CNN-LSTM model

RMSE (Equation (5)) follows the quadratic scoring rule that also measures the average magnitude of the error. It is the square root of the average of squared differences between predicted and actual observation, and it allows us to estimate the standard deviation ( $\sigma$ ) of the error for a typical single observation.

The MAE and RMSE are given by Equations (4) and (5).

$$MAE = \frac{\sum_{i=1}^n |x_i - y_i|}{n} \quad (4)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad (5)$$

Where,  $x_i$  is the estimated time series,  $y_i$  is the actual observations time series and  $n$  is the number of non-missing data points.

Figure 9 shows the mean absolute errors in each model for all cells, and it can be observed that the MAE value for the CNN-LSTM model is lesser than the MAE value for CNN and LSTM models for the majority of cells. Similar to MAE values, Figure 10 shows the root mean square error values of each model for all grids. A similar trend is followed that shows only three out of fifteen grids with RMSE value for the CNN-LSTM model has higher RMSE value for at least one of the other two models.

The comparative results and evaluation of the models confirm that the combined model of CNN-LSTM outperforms CNN and LSTM models for 10 cells out of a total of 15 cells. Further, the predicted value is compared with the actual value for all three models.

Figure 11, Figure 12, and Figure 13 show the difference between predicted and actual values of  $PM_{2.5}$  over the duration of 360 hours from 15<sup>th</sup> to 30<sup>th</sup> April, 2021, for the CNN, CNN-LSTM, LSTM model for cells H, L and N. Along with confirmation from MAE and RMSE values about CNN-LSTM being the most suitable model (Figures 9 and 10), the Figure 11c, Figure 12c, and Figure 13c also show that the curve for actual and predicted values of  $PM_{2.5}$  is with minimum deviations, i.e., for CNN-LSTM model.

For Determining the model accuracy, hourly  $PM_{2.5}$  concentrations data for two months from 1<sup>st</sup> May to 30<sup>th</sup> June 2021, is collected for four static monitors from CPCB. For the centroid of each cell, the data is interpolated using IDW and Kriging. As shown in Figure 2, one of the static monitors located at Dr. Karni Singh Shooting range lies in grid K. Therefore, to compare the actual values obtained from the static monitor and predicted values estimated from models, grid K is selected. Figure 14 compares the actual (data from static monitors) and predicted values of  $PM_{2.5}$  using three deep learning techniques CNN, LSTM, and CNN-LSTM. It confirms that the CNN-LSTM can predict the  $PM_{2.5}$  efficiently.

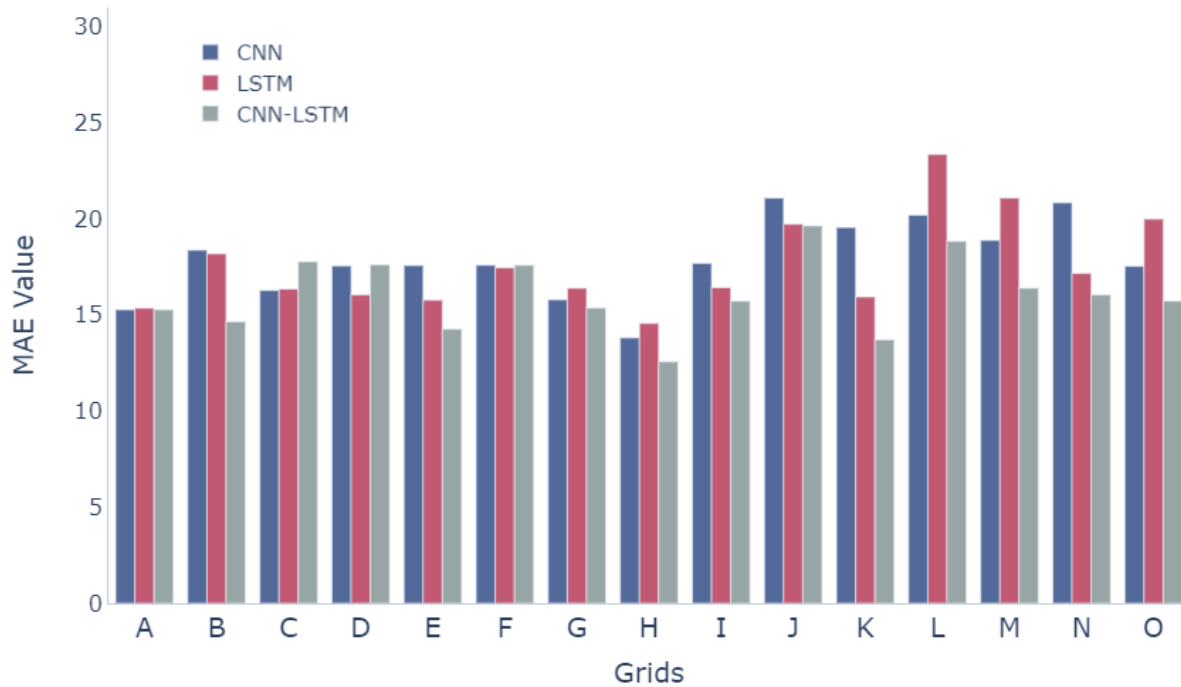
The present study shows the results for CNN-LSTM along with the individual models of CNN and LSTM, presented for 360 hours in Figures 11 to 13, are better than the results shown in the literature (21), where the results for CNN, and LSTM individually were explained for next 8 hours time span. A CNN-LSTM model was developed without batch normalization, which is a very crucial step to solve the problem of internal covariate shift, and also used one dimensional CNN layers that limit the motion of kernel to one direction only and limit the full strength of CNN-LSTM (23). On the contrary, the current study uses two-dimensional CNN layers to improve the strength of CNN-LSTM models. It is found that temperature is highly correlated to  $PM_{2.5}$  trends (Figures 1 and 4), while the influence of temperature was not considered for  $PM_{2.5}$  prediction using similar a model of CNN-LSTM (22).

The Figures 15a and 15b depicts the spatial prediction at the centroid of each cell using the CNN-LSTM model in peak (08:00) and off-peak (15:00) hours for Feb. 01, 2021 and June 28, 2021, respectively. Firstly, as expected, the winter (Figure 15a) and summer (Figure 15b) seasons have significant difference in the  $PM_{2.5}$ . Interestingly, a major difference can be observed within South Delhi depending on the cell location and time of the day, i.e., the road users may get a chance to alter the choice (e.g., route and departure time) and reduce the air pollution exposure. The accuracy of the model can be further enhanced by using the dynamic monitoring network.

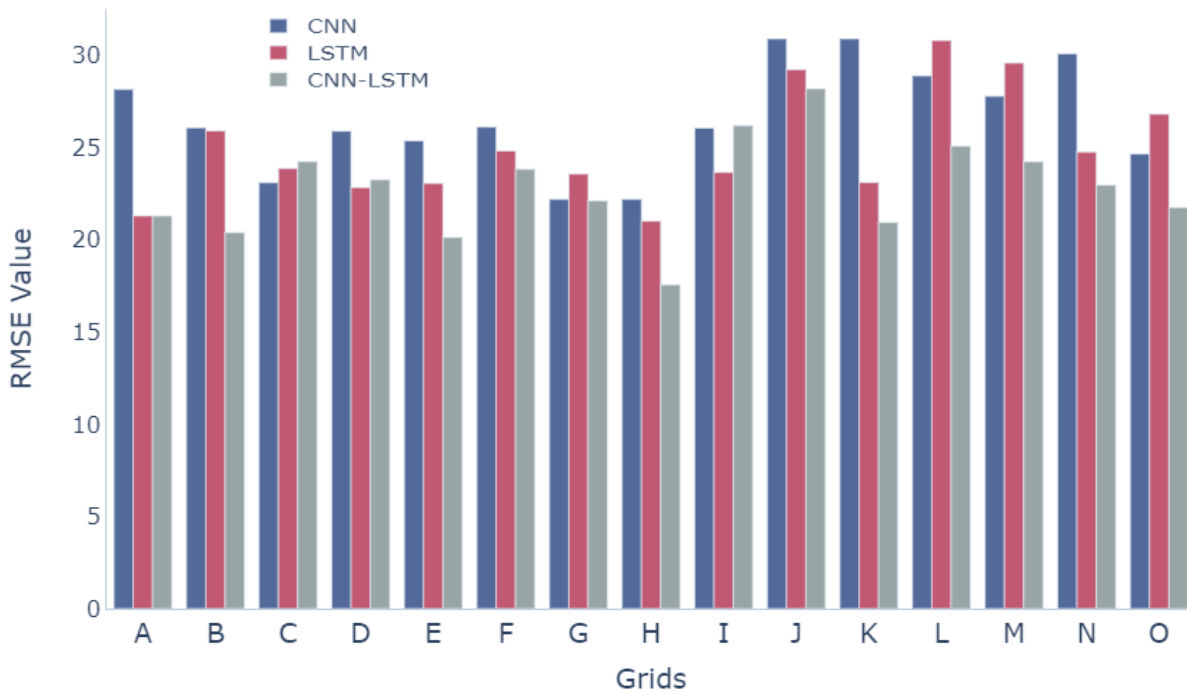
The practical applications of this study lie in estimating the accurate PM concentrations for researchers, policymakers, officials related to the environment, and the public. The commuters in Delhi are willing to update their travel choices provided air pollution and route information is provided to them (43). Commuters can consider the pollution levels in changing their travel behavior, such as changing the travel mode, changing the departure time, choosing to travel or staying at home, etc. The different routes possibilities (e.g., shortest path, greenest path, balanced path, etc.) would facilitate the commuters to reduce their air pollution exposure based on real-time congestion and air pollution patterns (44). In the long run, residence selection (e.g., location choice) can also depend on the levels of air pollution in different areas. This indicates that with the help of a prediction framework, a provision for disseminating air pollution information can be provided, which is likely to reduce the air pollution exposure of the daily commuters.

## Conclusions

To understand dynamic air pollution and push people to make informed decisions and prevent health issues due to increasing air pollution, providing information at all routes is essential. People face maximum exposure to air pollution while traveling. Therefore, the travel choices such as travel mode and route influence the exposure and effect on human



**Figure 9.** Mean absolute error in each model for all grids

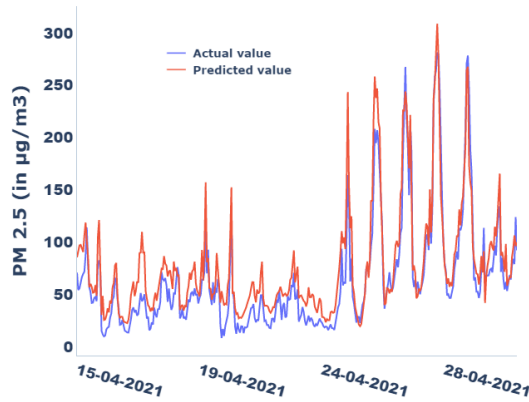


**Figure 10.** Root mean square error in each model for all grids

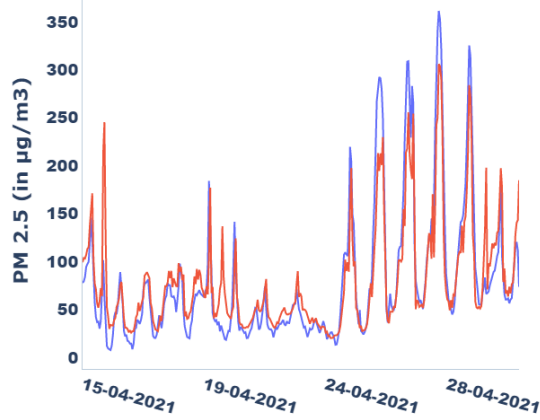
health extremely. For estimating exposure on all commuters' routes, the concentration of pollutants on each point in the vicinity is required. Due to cost factors, the monitors'

network cannot be dense enough to provide accurate values at each location. For determining pollutant concentration at all routes, high-resolution spatial prediction and accurate

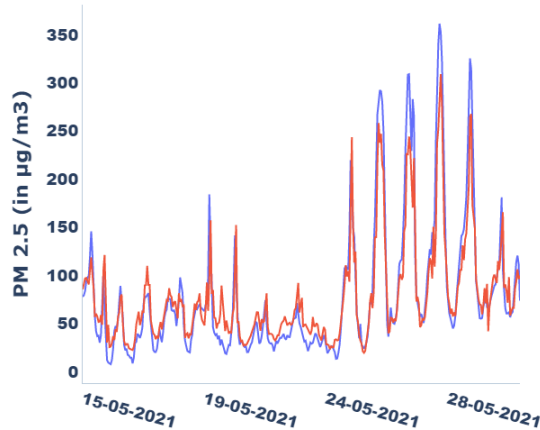




(a)



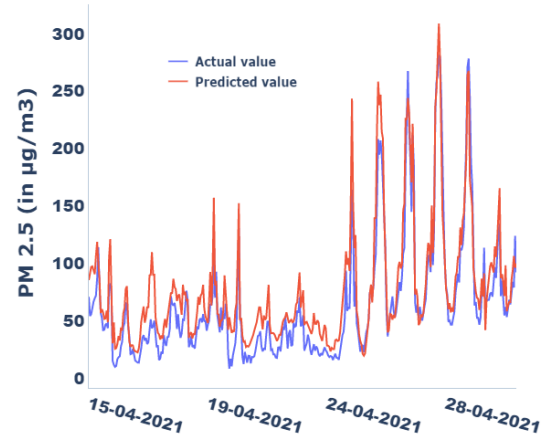
(b)



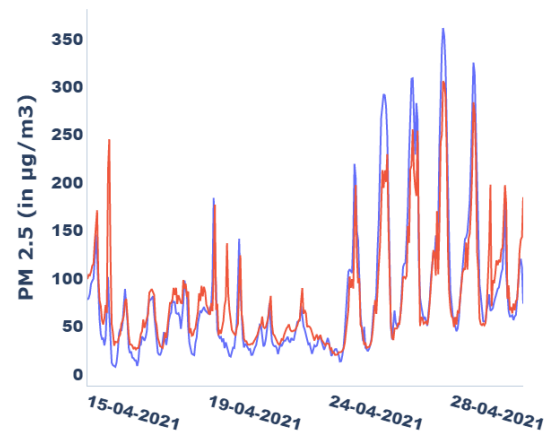
(c)

**Figure 11.** The results of CNN (a), LSTM (b), and CNN-LSTM (c) for grid H.

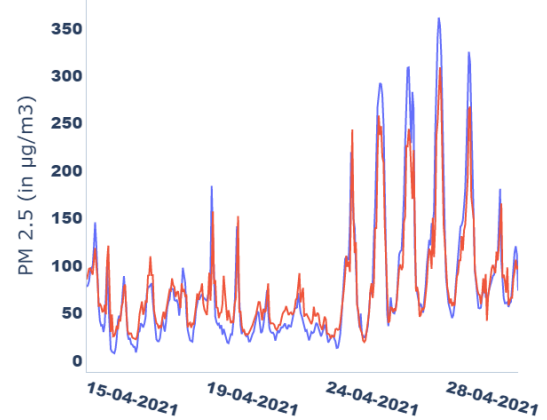
forecasting are required. This paper proposed a framework for the development of a prediction model for PM<sub>2.5</sub> pollutants in the South Delhi region of India. The model is applicable for a dynamic monitoring network. Historical data were collected from Central Pollution Control Board (CPCB)



(a)



(b)

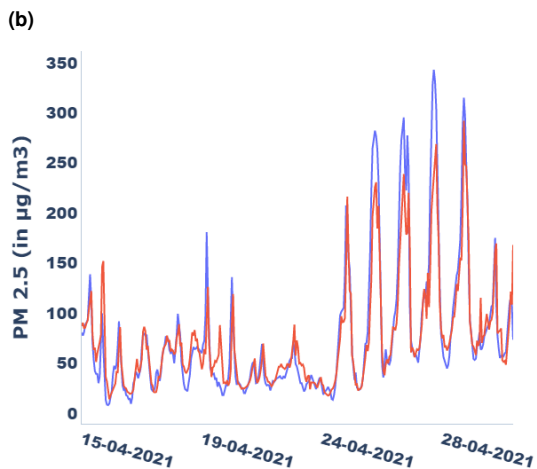
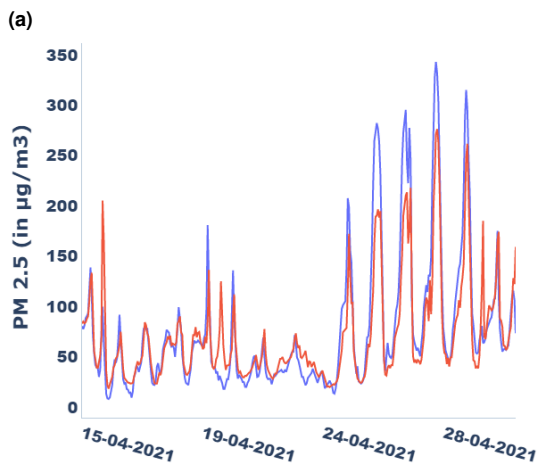
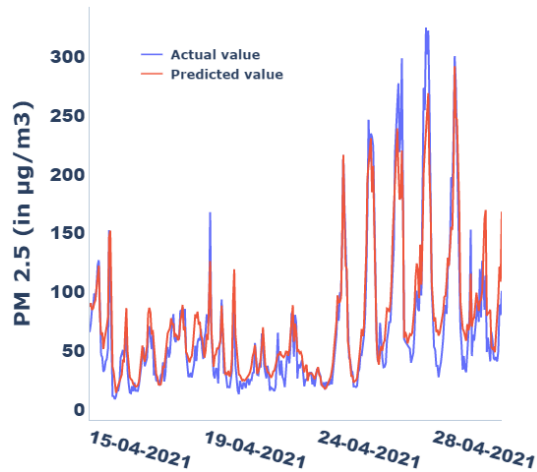


(c)

**Figure 12.** The results of CNN (a), LSTM (b), and CNN-LSTM (c) for grid L.

for all meteorological parameters and particulate matters of four static monitors from South Delhi.

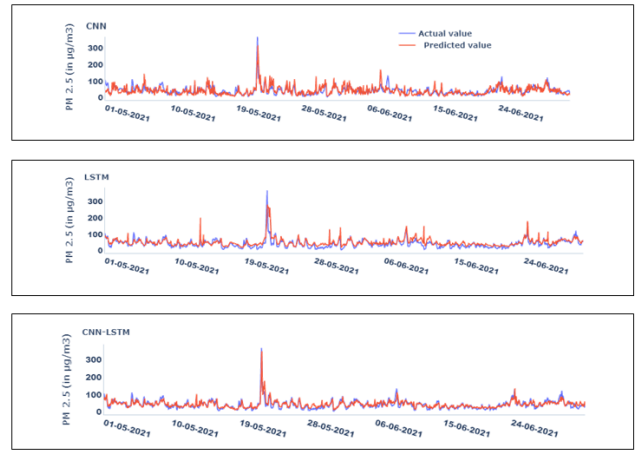
Initially, the model was developed through two techniques, i.e., CNN and LSTM, on the meteorological parameters



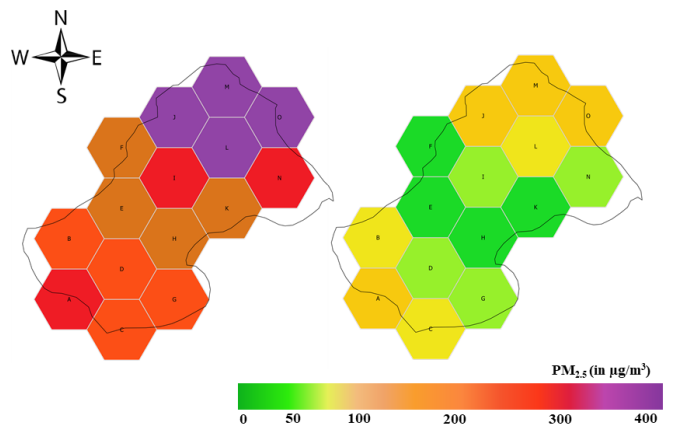
(c)

**Figure 13.** The results of (a) CNN, (b) LSTM, and (c) CNN-LSTM for grid N.

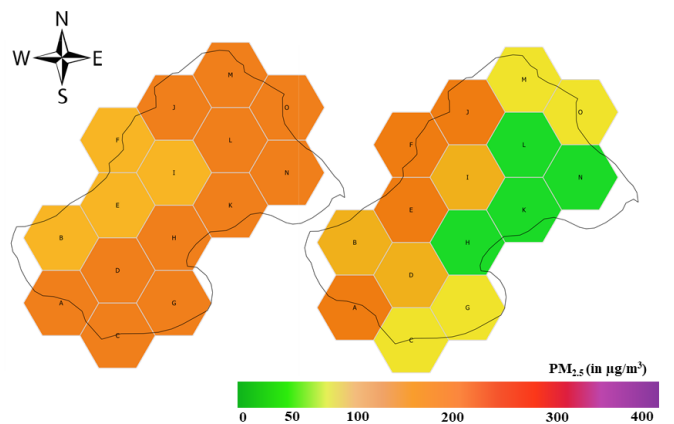
of temperature, relative humidity, pressure, wind speed and direction, and particulate matters. The prediction model inputs the previous 72 hours data and predicts the following



**Figure 14.** Results of CNN (top), LSTM (middle), and CNN-LSTM (bottom) for cell K, for time period from May 01, 2021 to June 30, 2021.



(a) Feb. 01, 2021, 08:00 (peak hour, left) and 15:00 (off peak, right)



(b) June 28, 2021, 08:00 (peak hour, left) and 15:00 (off peak, right)

**Figure 15.** Spatial prediction of PM<sub>2.5</sub> from CNN-LSTM for Feb. 01, 2021 and June 28, 2021 in peak hour (left) and off peak hour (right).

hour data of particulate matter and other meteorological parameters. A hybrid model was proposed as a combination of CNN and LSTM, where CNN extracts the effective features and LSTM manages time-series data processing to improve the accuracy of the prediction model. Through experimental results, we found that the CNN-LSTM model outperforms the remaining existing models in terms of MAE and RMSE as shown in Figure 9 and Figure 10. The proposed model may provide abrupt predictions if external environmental factors like a sudden increase in toxic gases and traffic congestion are present. As the unknown causes of change in pollution concentrations cannot be added in models, this limitation exists. Due to the pandemic situation, portable monitors cannot be deployed, and thus, the data from dynamic monitoring networks are not part of the results. However, the dynamic data from mobile monitors can be accommodated in the proposed framework, which is likely to further improve the model's accuracy. This study shows the proof of concept for providing a prediction framework for dynamic air pollution and meteorological data, in which a grid size of 5 km is assumed. A smaller grid size will increase the number of models and thus the requirement of computational efforts but may also increase the spatial accuracy (45).

In the future, the authors wish to utilize the data from the dynamic monitors for the prediction framework, exploit a way to reduce the need for a model for each grid, and implement a concept of hybrid grid sizes for the dynamic data. Since the framework is transferable, the planners and practitioners can employ the same models in different cities and use the predicted values to provide green travel routes.

## Acknowledgements

The authors would like to acknowledge the financial support by Science and Engineering Research Board (SERB) for this study (Grant number - SRG/2020/001147). The financial support by SPARK, Indian Institute of Technology Roorkee for internship of Mr. Suhas Sasetty is also acknowledged. The authors are also thankful to Central Pollution Control Board (CPCB) for making the monitored data available on website for public use.

## Author Contributions

The authors confirm contribution to the paper as follows: Conception and design of study: AA, RC; Literature review: VM, SS, RC; data collection and model development: VM, SS, RC; analysis and interpretation of results: VM, SS, RC, AA; draft manuscript preparation: VM, SS, RC, AA. All authors reviewed the results and approved final version of the manuscript.

## References

1. Castell, N., F. R. Dauge, P. Schneider, M. Vogt, U. Lerner, B. Fishbain, D. Broday, and A. Bartonova. Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? *Environment International*, Vol. 99, 2017, pp. 293–302. doi:10.1016/j.envint.2016.12.007.
2. Dey, S., L. D. Girolamo, A. van Donkelaar, S. Tripathi, T. Gupta, and M. Mohan. Variability of outdoor fine particulate (PM<sub>2.5</sub>) concentration in the Indian subcontinent: A remote sensing approach. *Remote Sensing of Environment*, Vol. 127, 2012, pp. 153–161. doi:10.1016/j.rse.2012.08.021.
3. Bell, M. L., F. Dominici, K. Ebisu, S. L. Zeger, and J. M. Samet. Spatial and Temporal Variation in PM<sub>2.5</sub> Chemical Composition in the United States for Health Effects Studies. *Environmental Health Perspectives*, Vol. 115, No. 7, 2007, pp. 989–995. doi:10.1289/ehp.9621.
4. Haddad, H. and A. de Nazelle. The role of personal air pollution sensors and smartphone technology in changing travel behaviour. *Journal of Transport & Health*, Vol. 11, 2018, pp. 230–243. doi:10.1016/j.jth.2018.08.001.
5. Gorai, A. K., P. B. Tchounwou, S. Biswal, and F. Tuluri. Spatio-temporal variation of Particulate Matter (PM<sub>2.5</sub>) concentrations and its health impacts in a mega city, Delhi in India. *Environmental Health Insights*, Vol. 12, 2018, p. 117863021879286. doi:10.1177/1178630218792861.
6. Shukla, K., P. Kumar, G. S. Mann, and M. Khare. Mapping spatial distribution of particulate matter using Kriging and Inverse Distance Weighting at supersites of megacity Delhi. *Sustainable Cities and Society*, Vol. 54, 2020, p. 101997. doi: 10.1016/j.scs.2019.101997.
7. Ahmed, S. O., R. Mazloum, and H. Abou-Ali. Spatiotemporal interpolation of air pollutants in the Greater Cairo and the Delta, Egypt. *Environmental Research*, Vol. 160, 2018, pp. 27–34. doi: 10.1016/j.envres.2017.09.005.
8. Doreswamy, H. K. S, Y. KM, and I. Gad. Forecasting Air Pollution Particulate Matter (PM<sub>2.5</sub>) Using Machine Learning Regression Models. *Procedia Computer Science*, Vol. 171, 2020, pp. 2057–2066. doi:10.1016/j.procs.2020.04.221.
9. Zhao, R., X. Gu, B. Xue, J. Zhang, and W. Ren. Short period PM<sub>2.5</sub> prediction based on multivariate linear regression model. *PLOS ONE*, Vol. 13, No. 7, 2018, p. e0201011. doi: 10.1371/journal.pone.0201011.
10. Aarhi, A., P. Gayathri, N. R. Gomathi, S. Kalaiselvi, and V. Gomathi. Air quality prediction through regression model. *International journal of scientific and technology research*, Vol. 9.
11. Sekhar, S. R. M., G. M. Siddesh, S. Jain, T. Singh, V. Biradar, and U. Faruk. Assessment and Prediction of PM<sub>2.5</sub> in Delhi in View of Stubble Burn from Border States Using Collaborative Learning Model. *Aerosol Science and Engineering*, Vol. 5, No. 1, 2020, pp. 44–55. doi:10.1007/s41810-020-00083-1.
12. Phruksahiran, N. Improvement of air quality index prediction using geographically weighted predictor methodology. *Urban Climate*, Vol. 38, 2021, p. 100890. doi:10.1016/j.uclim.2021.

- 100890.
13. Karamizadeh, S., S. M. Abdullah, M. Halimi, J. Shayan, and M. javad Rajabi. Advantage and drawback of support vector machine functionality. In *2014 International Conference on Computer, Communications, and Control Technology (I4CT)*. IEEE, 2014. doi:10.1109/i4ct.2014.6914146.
  14. Ali, S., S. S. Tirumala, and A. Sarrafzadeh. SVM aggregation modelling for spatio-temporal air pollution analysis. In *17th IEEE International Multi Topic Conference 2014*. IEEE, 2014. doi:10.1109/inmic.2014.7097346.
  15. Lecun, Y., L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, Vol. 86, No. 11, 1998, pp. 2278–2324. doi: 10.1109/5.726791.
  16. Hochreiter, S. and J. Schmidhuber. Long Short-Term Memory. *Neural Computation*, Vol. 9, No. 8, 1997, pp. 1735–1780. doi: 10.1162/neco.1997.9.8.1735.
  17. Elsworth, S. and S. Güttel. Time Series Forecasting Using LSTM Networks: A Symbolic Approach, 2020. [arXiv: 2003.05672](https://arxiv.org/abs/2003.05672).
  18. Seng, D., Q. Zhang, X. Zhang, G. Chen, and X. Chen. Spatiotemporal prediction of air quality based on LSTM neural network. *Alexandria Engineering Journal*, Vol. 60, No. 2, 2021, pp. 2021–2032. doi:10.1016/j.aej.2020.12.009.
  19. Srimuruganandam, B. and S. S. Nagendra. ANN-based PM prediction model for assessing the temporal variability of PM10, PM2.5 and PM1 concentrations at an urban roadway. *International Journal of Environmental Engineering*, Vol. 7, No. 1, 2015, p. 60. doi:10.1504/ijee.2015.069266.
  20. Bai, L., J. Wang, X. Ma, and H. Lu. Air Pollution Forecasts: An Overview. *International Journal of Environmental Research and Public Health*, Vol. 15, No. 4, 2018, p. 780. doi:10.3390/ijerph15040780.
  21. Chang, Y.-S., H.-T. Chiao, S. Abimannan, Y.-P. Huang, Y.-T. Tsai, and K.-M. Lin. An LSTM-based aggregated model for air pollution forecasting. *Atmospheric Pollution Research*, Vol. 11, No. 8, 2020, pp. 1451–1463. doi:10.1016/j.apr.2020.05.015.
  22. Huang, C.-J. and P.-H. Kuo. A Deep CNN-LSTM Model for Particulate Matter (PM2.5) Forecasting in Smart Cities. *Sensors*, Vol. 18, No. 7, 2018, p. 2220. doi:10.3390/s18072220.
  23. Li, T., M. Hua, and X. Wu. A Hybrid CNN-LSTM Model for Forecasting Particulate Matter (PM2.5). *IEEE Access*, Vol. 8, 2020, pp. 26933–26940. doi:10.1109/access.2020.2971348.
  24. Chen, H., M. Guan, and H. Li. Air quality prediction based on integrated dual LSTM model. *IEEE Access*, Vol. 9, 2021, pp. 93285–93297. doi:10.1109/access.2021.3093430.
  25. Pande, P., S. Dey, S. Chowdhury, P. Choudhary, S. Ghosh, P. Srivastava, and B. Sengupta. Seasonal transition in PM10 exposure and associated all-cause mortality risks in India. *Environmental Science & Technology*, Vol. 52, No. 15, 2018, pp. 8756–8763. doi:10.1021/acs.est.8b00318.
  26. Choudhary, R. and A. Agarwal. Optimization of bus routes to maximize coverage of air pollution monitors. In *6<sup>th</sup> Conference of Transportation Research Group of India (CTRG)*. 2021. URL <http://faculty.iitr.ac.in/~amitfce/publications.html>.
  27. Saukh, O., D. Hasenfratz, and L. Thiele. Route selection for mobile sensor nodes on public transport networks. *Journal of Ambient Intelligence and Humanized Computing*, Vol. 5, No. 3, 2012, pp. 307–321. doi:10.1007/s12652-012-0170-7.
  28. Kaivonen, S. and E. C.-H. Ngai. Real-time air pollution monitoring with sensors on city bus. *Digital Communications and Networks*, Vol. 6, No. 1, 2020, pp. 23–30. doi:10.1016/j.dcan.2019.03.003.
  29. CPCB. *Ambient air quality data of Delhi stations*. Tech. rep., Central Pollution Control Board, 2018. URL <http://www.cpcb.nic.in/>.
  30. Deswal, S. and P. Chandna. Effects of meteorological conditions on concentration of air pollutants in Delhi. *International Journal of Environment and Pollution*, Vol. 42, No. 1/2/3, 2010, p. 58. doi:10.1504/ijep.2010.034226.
  31. Cichowicz, R., G. Wielgosiński, and W. Fetter. Dispersion of atmospheric air pollution in summer and winter season. *Environmental Monitoring and Assessment*, Vol. 189, No. 12. doi:10.1007/s10661-017-6319-2.
  32. Guttikunda, S. K., K. Nishadh, and P. Jawahar. Air pollution knowledge assessments (APnA) for 20 Indian cities. *Urban Climate*, Vol. 27, 2019, pp. 124–141. doi:10.1016/j.uclim.2018.11.005.
  33. Pandey, A., M. Brauer, M. Cropper, and et al. Health and economic impact of air pollution in the states of India: the Global Burden of Disease Study 2019. *The Lancet Planetary Health*, Vol. 5, No. 1, 2021, pp. e25–e38. doi:10.1016/s2542-5196(20)30298-9.
  34. Apte, M., Y. Y. Y. Agarwadkar, S. Azmi, and A. B. Inamdar. Understanding grids and effectiveness of hexagonal grid in spatial domain. *IJCA Proceedings on International Conference on Recent Trends in Information Technology and Computer Science 2012*, Vol. ICRTITCS, No. 1, 2013, pp. 25–27.
  35. Agarwal, A. *Mitigating negative transport externalities in industrialized and industrializing countries*. Ph.D. thesis, TU Berlin, Berlin, 2017. doi:10.14279/depositonnce-5825.
  36. Wu, W., R. Gan, J. Li, X. Cao, X. Ye, J. Zhang, and H. Qu. A Spatial Interpolation of Meteorological Parameters considering Geographic Semantics. *Advances in Meteorology*, Vol. 2020, 2020, pp. 1–14. doi:10.1155/2020/9185283.
  37. Murphy, B. S. PyKrig: Kriging toolkit for python, 2021. URL <https://pypi.org/project/PyKrig/>.
  38. de Mingo López, L. F., C. M. Lucas, N. G. Blas, and K. Ivanova. Pattern Recognition with Convolutional Neural Networks: Humpback Whale Tails, 2019. doi:10.20944/preprints201902.0257.v1. Preprint.
  39. Chae, S., J. Shin, S. Kwon, S. Lee, S. Kang, and D. Lee. PM10 and PM2.5 real-time prediction models using an interpolated convolutional neural network. *Scientific Reports*, Vol. 11, No. 1. doi:10.1038/s41598-021-91253-9.
  40. Ioffe, S. and C. Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,

2015. [arXiv:1502.03167](https://arxiv.org/abs/1502.03167).
41. Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *The Journal of Machine Learning Research*, Vol. 15, No. 1, 2014, p. 1929–1958.
  42. Bisong, E. Google Colaboratory. In *Building Machine Learning and Deep Learning Models on Google Cloud Platform*. Apress, 2019, pp. 59–64. doi:10.1007/978-1-4842-4470-8\_7.
  43. Meena, K. K., V. Singh, and A. Agarwal. Perception of commuters towards air quality in Delhi. In *6<sup>th</sup> Conference of Transportation Research Group of India (CTRG)*. WP-39, 2021. URL <http://faculty.iitr.ac.in/~amitfce/publications.html>.
  44. Choudhary, R., S. Ratra, and A. Agarwal. Fusing real-time congestion and air pollution in a multi-modal routing engine. In *12<sup>th</sup> International Scientific Conference on Mobility and Transport (mobil.TUM 2022)*. WP-44, accepted, 2022. URL <http://faculty.iitr.ac.in/~amitfce/publications.html>.
  45. Choudhary, R. and A. Agarwal. *Impact of grid size on spatiotemporal prediction of fine Particulate Matter*. Tech. Rep. WP-45, Indian Institute of Technology Roorkee, 2021. URL <http://faculty.iitr.ac.in/~amitfce/publications.html>.