RESEARCH

Open Access

Air-pollution prediction in smart city, deep learning approach

Abdellatif Bekkar[1]* , Badr Hssina[1], Samira Douzi[2] and Khadija Douzi[1]

*Correspondence:
abd.bekkar@gmail.com
[1] FSTM, University Hassan II,
Casablanca, Morocco
Full list of author information
is available at the end of the
article

Abstract

Over the past few decades, due to human activities, industrialization, and urbanization, air pollution has become a life-threatening factor in many countries around the world. Among air pollutants, Particulate Matter with a diameter of less than $2.5\mu m$ ($PM_{2.5}$) is a serious health problem. It causes various illnesses such as respiratory tract and cardio-vascular diseases. Hence, it is necessary to accurately predict the $PM_{2.5}$ concentrations in order to prevent the citizens from the dangerous impact of air pollution beforehand. The variation of $PM_{2.5}$ depends on a variety of factors, such as meteorology and the concentration of other pollutants in urban areas. In this paper, we implemented a deep learning solution to predict the hourly forecast of $PM_{2.5}$ concentration in Beijing, China, based on CNN-LSTM, with a spatial-temporal feature by combining historical data of pollutants, meteorological data, and $PM_{2.5}$ concentration in the adjacent stations. We examined the difference in performances among Deep learning algorithms such as LSTM, Bi-LSTM, GRU, Bi-GRU, CNN, and a hybrid CNN-LSTM model. Experimental results indicate that our method "hybrid CNN-LSTM multivariate" enables more accurate predictions than all the listed traditional models and performs better in predictive performance.

Keywords: Air-pollution, $PM_{2.5}$, Deep learning, Forecasting, LSTM, GRU, CNN-LSTM

Introduction

The increase in the percentage of the urban population in the world shows that people more and more are moving to cities. According to United Nations (UN), the urban population as of 2020 is about 56.15% [1]. And it is expected that it will become 68% of the world's population will live in urban cities by 2050 [2]. The growth of urbanization and industrialization causes several problems logistics, health care, and air quality. In order to resolve these issues, and improve the quality of its citizens' lives, The smart city concept was created by integrating Information and Communication Technology (ICT), and fixed/mobile sensors. These last are installed within the city to observe real human practice. This concept become an endless source of urban data.

In the last decades, the frequent occurrence of smog caused by the increase in industrialization has harshly brought environmental pollution to its serious peak. That is, it becomes more severe than ever before. One of the hazardous pollutants is a fine particulate matter whose size is $2.5\,\mu m$ or less, also known as $PM_{2.5}$. Such particle results

Bekkar *et al. Journal of Big Data*   (2021) 8:161

Page 2 of 21

in serious health damage. According to the World Health Organization WHO, almost 90% of people breathe polluted air that exceeds the limits of WHO guidelines in terms of air quality [3], bringing about respiratory problems [4, 5], moreover, even a few hours to weeks of short-term exposure to $PM_{2.5}$ can trigger cardiovascular disease-related mortality and events [6]. The Global Burden of Diseases GBD identified that Exposure to $PM_{2.5}$ contributed to 4.2 million deaths and 115.1 million disability-adjusted life years (DALYs) globally in 2015 [7] with an increase in 2017 (4.58 million deaths and 142.52 million DALYs) [8]. This poor air quality not only threatens the health and lives of individuals but the economies as well. The report carried out by the Organization for Economic Cooperation and Development OECD has shown that air pollution could cost 1% of world Gross domestic product GDP [9].

Due to the coronavirus pandemic (COVID-19), the epidemic center in China is the first to announce a lockdown on January 23, 2020. after, other countries did the same to reduce the spread of the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). Universally the COVID-19 lockdown creates a unique and precious opportunity to evaluate and to understand human activities, and the factors affecting air pollutants. Many studies have reported the environmental effects of lockdown policies on $PM_{2.5}$ concentration in different regions due to the COVID-19 pandemic [10–12] . Moreover, various hypotheses have been studied on the relationship between $PM_{2.5}$ and covid-19. Among them, searchers are found that $PM_{2.5}$ has been an important vector in the acceleration of the spread of the COVID-19 [13]. Another paper identifies that a significant relationship between air pollution and COVID-19 infection [14]. The exposures in the long term of $PM_{2.5}$ are positively associated with higher county-level city COVID-19 mortality rates after accounting for many area-level confounders [15]. Based on the hypothesis that there is a relationship between the spread of the virus and the presence of $PM_{2.5}$ in the air, the researchers propose an innovative metric to predict COVID-19 with the machine learning model quote mirri2020covid.

An effective system for monitoring and predicting air pollution in advance has great importance for human health and government decision-making. However, the mechanism and process of $PM_{2.5}$ formation are very complex due to the complexity of its properties, such as non-linear properties in time and space [16], which have a significant impact on the accuracy of prediction. It thus requires an examining consideration. Furthermore, the air quality data is closely related to time, which means that it belongs to time series and has an apparent periodicity. Due to the data's timeliness, time predictions have become essential topics that undoubtedly require meticulous attention by academics and scholars. So doing showcases that Time series analysis plays a paramount role in many different applications, including economics, medicine, astronomy, geology, and others.

Traditional statistical methods have been widely used to process air quality forecasting problems. These methods are significantly based on the approach of using historical data for learning. Some of the notable statistical methods that have been used for air quality forecasting are Autoregressive Moving Average ARMA [17], and Autoregressive Integrated Moving Average ARIMA [18]. With the increase in the amount and the complexity of the data obtained, however, these methods can no longer meet the actual demand because of training-length time.

With artificial intelligence and big data evolution, prediction methods based on machine learning technologies are becoming more and more common. Because these types of models do not require an understanding of atmospheric pollutants' physical or chemical properties. The most popular machine learning algorithms are Multiple Linear Regression (MLR), Random Forest (RF) [19], Support Vector Regression (SVR) [20], Artificial Neural Networks (ANN) [21] that incorporate complex nonlinear relationships between the concentration of air pollutants and meteorological variables. Various ANN structures have been developed to predict air pollution over different study areas, such as neuro-fuzzy neural network (NFNN) [22], and Bayesian neural network [23], An Ensemble Approach which incorporated several different machine learning algorithms, has shown to be a robust and accurate measure of pollution levels in the Greater London area [24].

With the popularity of Artificial Intelligence, many deep learning algorithms have been developed respectively, such as Recurrent Neural Networks (RNN) and their variants. Long short-term memory (LSTM) is the most widely used model in air quality forecasting [25, 26] because it considers the temporal dependencies of a typical phenomenon observed in the $PM_{2.5}$ concentration series. Due to the complexity of $PM_{2.5}$ formation, the high accuracy and demand for predictive efficiency are essential in developing an effective model for predicting $PM_{2.5}$ concentration. We accordingly suggest comparing multivariate deep learning models based on several metrics (Average absolute error MAE,Root mean square error RMSE,The coeffcient of determination R^2).

To this end, this paper seeks to undergo a research study on the application of deep learning (LSTM, Bi-LSTM, GRU, CNN, CNN-LSTM,CNN-GRU). Hence, the study aims to unearth a comparison between the results obtained with these techniques to learn more about their efficient use in predicting $PM_{2.5}$ concentration. Moreover, our research aims to provide a $PM_{2.5}$ forecasting model with good accuracy with meteorological data and the concentration of adjacent stations.

In this study, we designed a system for the Prediction of $PM_{2.5}$ by utilizing advanced deep neural networks. We, therefore, proposed a hybrid CNN-LSTM forecasting model. Seven baseline predictive deep learning models were also built in this study for comparison with our proposed model. The key contributions of this study are :

1. This study combines the pollutant components, meteorological data, and adjacents stations in different time periods into the input variables. After preprocessing data by filling the missing values, encoding, normalizing data and analyzing the correlation between features and $PM_{2.5}$ concentration as a features selection. Spatial and temporal correlations are complex and comprehensive. In our study, historical data from the target station and adjacent stations are integrated with other features and entered into the model. From the results, the proposed combination is found more effective in extracting spatiotemporal features and performs $PM_{2.5}$ prediction accuracy more than others.

2. Through the proposed model, the Spatio-temporal characteristics of the data are extracted. By combining the advantages of the Convolutional Neural Network CNN model, which is effective at filtering out the spatial characteristics include the characteristics of the data between pollutant components and Weather and between differ-

ent adjacent stations. At the same time, an LSTM network is used for the extraction of temporal features.

3. Comparing the performances of seven popular deep learning methods in the air pollution prediction problem, we validated the practicality and feasibility of the proposed model in $PM_{2.5}$ concentration prediction by comparing the Metrics in different batch sizes, and lags. Moreover, the results achieved in this work are comparable to other state-of-the-art deep learning approaches reported in the literature.

This paper is organized as follows: "Related works" section briefly reviews the related work. "Deep learning models" section defines the basic concepts of the deep learning models, namely LSTM, Bi-LSTM, GRU, CNN, CNN-LSTM, and CNN-GRU. "Material and methods" section describes the detailed methodology of the proposed approach, including the implementation and experimental results, whereas "Results and discussions" section covers the paper's conclusion.

Related works

Since the topic of $PM_{2.5}$ air pollution in cities needs urgently to be solved, $PM_{2.5}$ forecasting is absolutely a vital topic for the development of smart cities. The difficulty of prediction can be seen in the fact that $PM_{2.5}$ propagation is impacted by variations in meteorological variables, e.g. Wind speed and direction. Wind speed and direction data have a high degree of randomness and constantly change over different periods [27, 28].

Several $PM_{2.5}$ prediction methods are developed by researchers based on statistical models and machine learning techniques. Recently, the academic community has begun using deep neural networks for pollutant concentration prediction. Deep learning may solve problems by using more layers and more extensive data sets and processing all layers simultaneously to obtain more accurate results [29]. These favorable properties of deep learning make it suitable for modeling and predicting air pollution.

A wide variety of models can be used for this purpose. Authors in article [30] analyse and study the prediction $PM_{2.5}$ levels on 12 stations in Beijing using four models ARIMA, FBProphet (Facebook prophet), LSTM, and CNN. With historical air quality data, meteorological data, weather forecast data. Most of the stations showed that LSTM performed better than all other models MAE = 13.2 and RMSE = 20.8. In this study [31], the authors propose a predictive model of PM concentration at the 25 monitoring stations in Seoul, South Korea, historical PM2.5 concentration, and meteorological data is used for comparing LSTM, and DAE (Denoising AutoEncoders). The comparison showed that the LSTM prediction model was more accurate than the DAE model.

In article [32] , the authors develop a bidirectional long short-term memory (Bi-LSTM) model to predicted $PM_{2.5}$ Concentration in China. The $PM_{2.5}$ Concentration and weather from the hourly data of the US Embassy, recorded for Beijing city as input. The proposed model achieved accuracy as follows MAE = 7.53, RMSE = 9.86, and SMAPE = 0.1664. Other researchers have been interested to Predict the $PM_{2.5}$ contamination of stations in Beijing by using long short-term memory-fully connected (LSTM-FC), LSTM, and an artificial neural network (ANN) with historical air quality data, meteorological data, weather forecast data, and the day of the week data. They showed that the LSTM-FC model outperforms LSTM and the ANN, with MAE = 23.97 and RMSE =

35.82 over 1-6 h [33]. However, none of these models can make use of pollutant concentration information in neighboring areas. Changes in pollutants are related not just to time but also to space. Because a pollutant in one place may travel to other regions, spatial information must be considered.

A CNN is consisting of a series of convolutional layers used to extract the spatial features of neural networks. CNN achieved remarkable results in multi-dimensional spatial arrays Which makes it a good topic for researchers to know the environmental situation through digital images. In the article [34], the authors propose an ensemble of deep neural networks to estimate $PM_{2.5}$ concentrations from outdoor images. Three convolutional neural networks, VGG-16, Inception-v3, and Resnet50, are used as the base learners. The experimental results demonstrated that the proposed ensemble can provide a more accurate $PM_{2.5}$ estimation than all three individual deep learning networks used. CNN has proven to be powerful in spatial data processing. This method has also been used to estimate the concentration of pollutants in urban areas, usually by analyzing satellite images [35, 36]. However, sometimes there is no image data but only abstracted monitoring data, e.g., wind direction, temperature, and location.

To solve the problem of Air Pollution in Seoul city in Korea, the researchers proposed the usage of the Convolutional Long Short-Term Memory (ConvLSTM), a combination of Convolutional Neural Networks and Long Short-Term Memory, which automatically manipulates both the spatial and temporal features of the data [37]. In this paper, this Spatio-temporal model includes air pollution data, meteorological data, traffic volume, average driving speed, and air pollution indicators in outdoor areas. The proposed model has proven its superiority over the various models. In another paper [38], the authors verified the feasibility and practicability of CNN-LSTM to estimate $PM_{2.5}$ concentration in Beijing for the next hour, cumulated wind speed, and cumulated hours of rain over the last 24 h. They showed that the CNN-LSTM model outperforms other models with MAE = 14.6344 and RMSE = 24.22874.

Deep learning models

In this work, our goal is to investigate the performances of several deep learning models to forecast the concentration of $PM_{2.5}$. Thus, we decided to use the LSTM, Bi-LSTM, GRU, CNN, CNN-LSTM previously mentioned. Next, we briefly describe each network:

LSTM

LSTM is a type of recurrent neural network (RNN) that was developed in 1980 [39, 40]. RNNs are a powerful type of artificial neural network and are most likely used for time-series forecasting problems. RNN can internally maintain memory to remember things from past occurrences that can predict future events. However, RNNs frequently suffer from vanishing and exploding gradients, which leads the model learning to become too slow or stopped altogether. LSTMs were created in 1997 [41] to solve these problems. LSTMs have longer memories and can learn from inputs that are separated from each other by long time lags.

An LSTM has three gates: an input gate that determines whether or not to let the new input in, a forget gate that deletes information that is not important and an output gate that decides what information to output. These three gates are analogical gates based on the

Bekkar *et al. Journal of Big Data*   (2021) 8:161

Page 6 of 21

sigmoid function, which works on the range between 0 and 1. These three sigmoid gates can be seen in Fig. 1 below. A horizontal line that can be seen running through the cell represents the cell state.

LSTM formulas are listed below:

$$Inputgate \quad : \quad i_t = \sigma(W^{(it)}\overline{x}_t + W^{(it)}h_{t-1}) \tag{1}$$

$$Forgetgate \quad : \quad f_t = \sigma(W^{(if)}\overline{x}_t + W^{(hf)}h_{t-1}) \tag{2}$$

$$Outputgate \quad : \quad i_t = \sigma(W^{(io)}\overline{x}_t + W^{(ho)}h_{t-1}) \tag{3}$$

$$ProcessInput \quad : \widetilde{C}_t = tanh(W^{(i\widetilde{c})}\overline{x}_t + W^{(h\widetilde{c})}h_{t-1}) \tag{4}$$

$$Cellupdate \quad : \quad C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{5}$$

$$Output \quad : \quad y_t = h_t = O_t + i_t * tanh(C_t) \tag{6}$$

GRU

GRU, Gated recurrent unit is an advancement of the standard RNN [33] is included in RNN, and it is similar to an LSTM unit. The GRU unit consists of the reset and updates gate. Figure 2 shows the GRU architecture. The reset gate is designed to forget the previous state between the prior activation and the next candidate activation, whereas the update gate is used to select the number of the candidate activation that updates the cell state.

GRU formulas are listed below:

$$Resetgate \quad : \quad r_t = \sigma(W^{(ir)}\overline{x}_t + W^{(hr)}h_{t-1}) \tag{7}$$

$$Updategate \quad : \quad z_t = \sigma(W^{(iz)}\overline{x}_t + W^{(hz)}h_{t-1}) \tag{8}$$
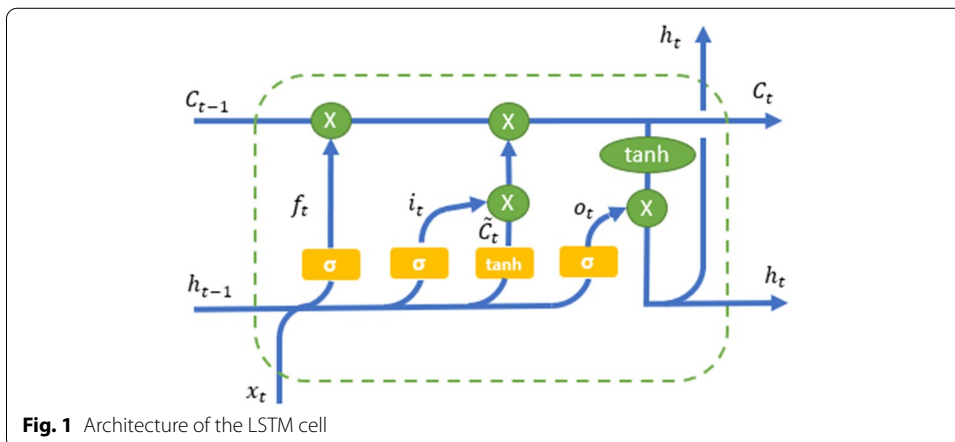
Fig. 1 Architecture of the LSTM cell

Fig. 2 Architecture of the GRU cell

Fig. 3 Architecture of the Bi-LSTM model
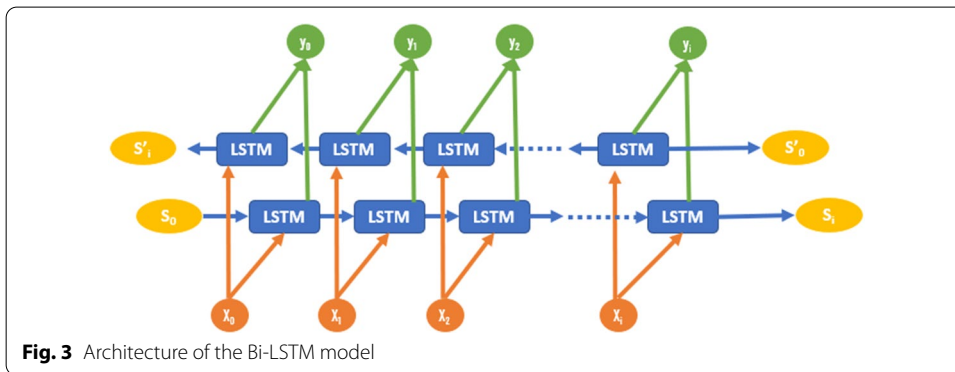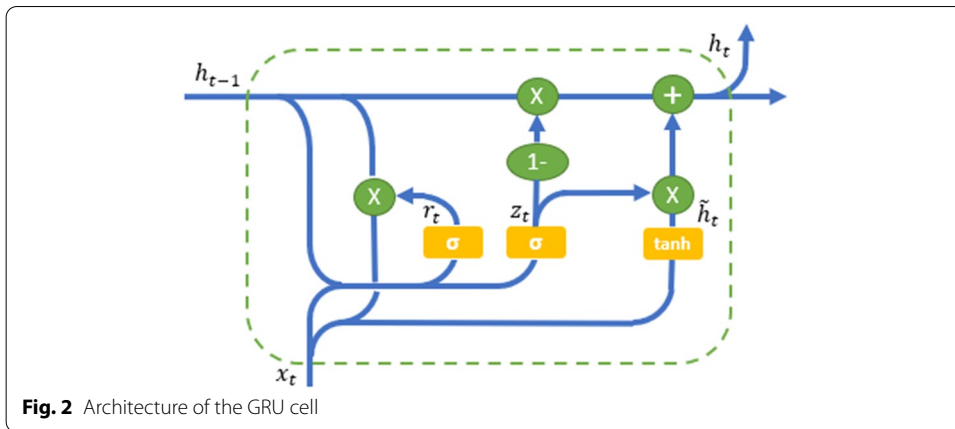
$$ProcessInput \quad : \quad \widetilde{h}_t = tanh(W^{(\widetilde{ih})}\overline{x}_t + W^{(h\widetilde{h})}h_{t-1}) \tag{9}$$

$$Hiddenstateupdate \quad : \quad h_t = (1 - z_t) * h_{t-1} + z_t * \widetilde{h}_t \tag{10}$$

$$Output \quad : \quad y_t = h_t \tag{11}$$

Bi-LSTM

Standard RNN and LSTM often ignore future information in time-processing, while Bi-LSTM can take advantage of future information. The basic structural idea of Bi-LSTM is that the front and back layers of each training sequence are two LSTM networks, respectively. Moreover, the LSTM networks are both connected to one input and one output layer. The output layer can obtain past information of each point from the input sequence and get future information from each point through this structure. as shown in Fig. 3.

CNN

CNN has been successfully applied to computer vision and medical image analysis [42]. Moreover, in this paper auteurs proposes a multiscale fully convolutional neural network (MFCN) for change detection in high-resolution remote sensing images [43].

In our model, the convolutional layers are constructed using one-dimensional kernels that move through the sequence (unlike images where 2D convolutions are utilized). These kernels act as filters that are learned during training. As in many CNN architectures, the deeper the layers get, the higher the number of filters. The architecture of CNN is shown in Fig. 4.

CNN-LSTM

The use of classical CNN architecture is the best choice when input networks are 2-D or 3-D tensors like images or videos [44]. Since LSTMs architectures are more adapted for 1-D Data, a new variant of LSTM called Convolutional LSTM or ConvLSTM [45] has been designed. In this architecture, the LSTM cell, which contains a convolution operation and input dimension of data, is kept in the output layer instead of just a 1-D vector. A convolution operation replaces matrix multiplication at each gate of classical LSTM. We can say that ConvLSTM architecture merges the capabilities of CNN and LSTM Network. It was normally developed for 2-D Spatio-temporal data such as satellite images.

Another approach to working with Spatio-temporal data is to combine CNN and LSTM layers, one block after another. Such architecture is called Convolutional-LSTM (CNN-LSTM) and was initially named Long-term Recurrent Convolutional Network or LRCN model. In the first part of this model, convolutional layers extract essential features of input data, and the results are flattened in a 1-D tensor so that they can be used as input for the second part of the model (LSTM). Finally, before passing data in the last hidden layer, information has to be reshaped in the original form of input data. The architecture of CNN-LSTM is shown in Fig. 5.
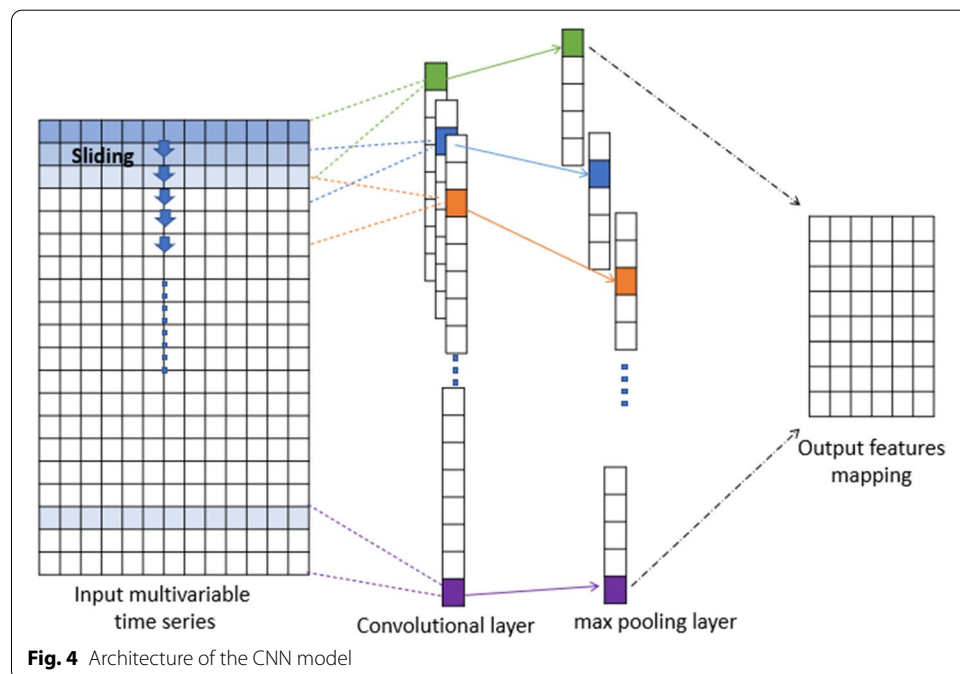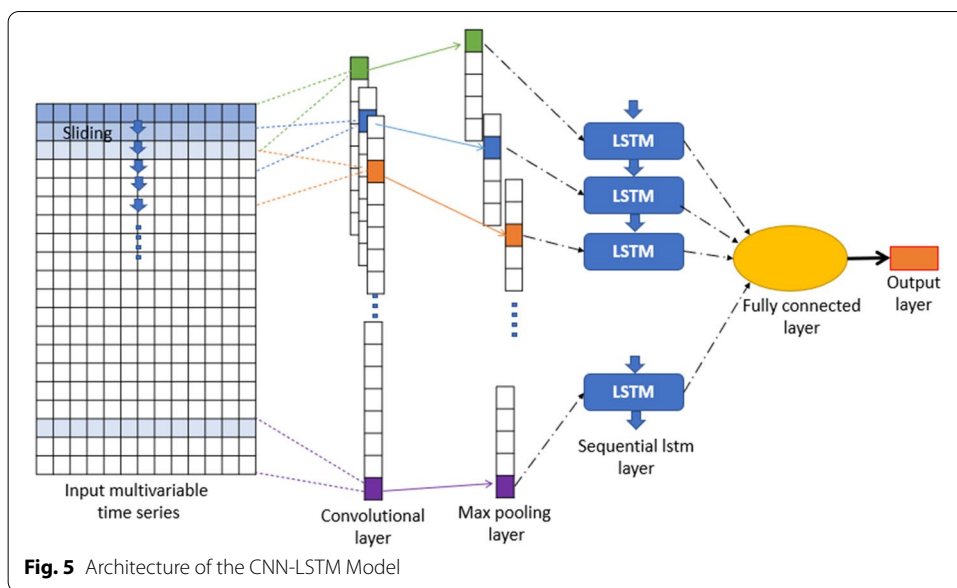

Fig. 4 Architecture of the CNN model

Fig. 5 Architecture of the CNN-LSTM Model

Material and methods

Dataset

The dataset chosen in this article (420768 instances and 18 attributes) comes from the UCI Machine Learning Repository [46]. this dataset shows the concentration of air pollutants and air quality at 12 sites. The air quality data comes from the Beijing Municipal Environmental Monitoring Center. The meteorological data indicating the air quality for each site is matched with the nearest meteorological station of the China Meteorological Administration. as shown in Fig. 6.

Data pre-processing

Missing values

This dataset includes 35064 records with multi-features in each station. The period of recording is from March 1st, 2013, to February 28th, 2017. The data are composed of: date, the concentration of $PM_{2.5}$, PM_{10}, Sulfur dioxide SO_2, Nitrogen dioxide NO_2, carbon monoxide CO, ozone O_3, dew point, temperature, atmospheric pressure, combined wind direction, cumulated wind speed, cumulated hours of snow, and rain. However, Air quality and meteorological monitoring equipment will cause leakage in data collection due to machine failure, due to some uncontrollable reasons. The existence of such missing values will have some impact on data mining.

In time-independent (non-chronological) data to replace missing field values, the most popular approaches are the mean or median value. However, in the case of a time series, this is not the case. To resolve incomplete data problems, many imputation techniques are adopted. A study has shown that the linear interpolation method is the best method to estimate hourly monitoring data for PM_{10} for all percentages of simulated missing values [47].

Fig. 6 The distribution of monitoring stations in Beijing
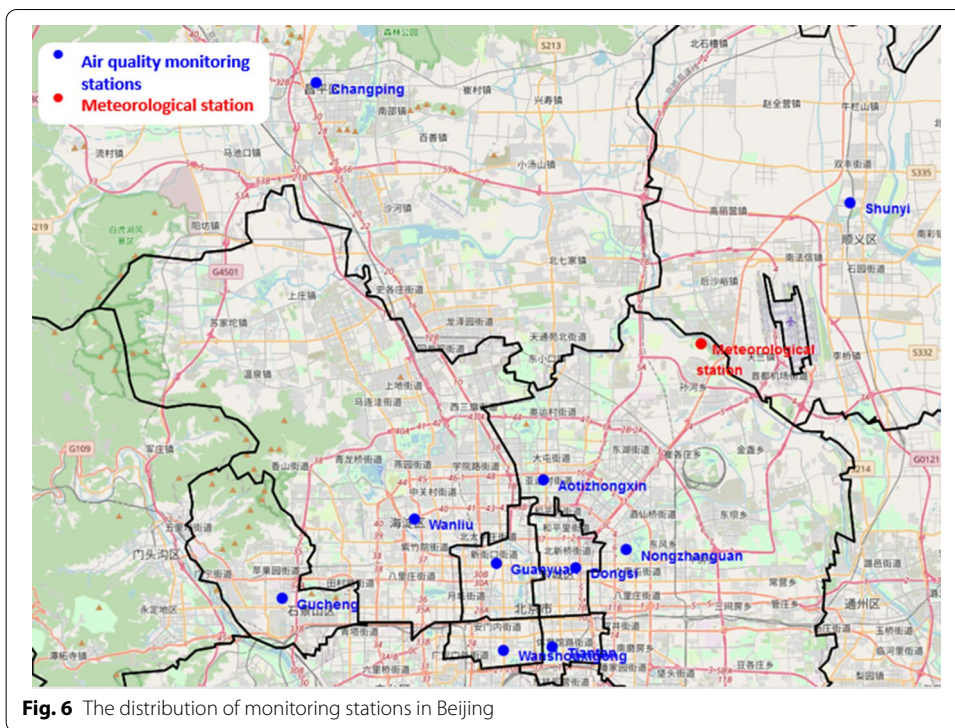
The processed data set contains less than 4% missing values, these missing values were addressed by linear spline imputation. The $SL(x)$ equation can adapt to local anomalies without affecting the interpolation values at other points.

The equation of the spline linear interpolation function is:

$$SL(x) = f(x_{i-1})\frac{x - x_i}{x_{i-1} - x_i} + f(x_i)\frac{x - x_{i-1}}{x_i - x_{i-1}} \quad x \in [x_{i-1}, x_i], i = 1, 2, 3, ..., n \qquad (12)$$

where x is the independent variable, $x_0, x_1, ... x_n$ are known values of the spline and $SL(x)$ the linear spline that interpolates f at these points.

Encoding categorical variables

In this analysis, the wind factor is an essential indicator of atmospheric activity. The pollutant concentration is affected by the wind speed [27], and the wind direction is crucial in determining the concentrations of $PM_{2.5}$ [28]. The wind direction attribute is categorical data, admits 16 values: N, NNE, NE, ENE, E, ESE, SE, SSE, S, SSW, SW, WSW, W, WNW, NW, and NNW. To convert each cardinal wind direction to a value of degrees azimuth. We have divided the compass into 16 sectors of 22.5 degrees each. North was given a value of zero and with clockwise displacement, the value increase by 22.5. The direction of each segment is 22.5 degrees. as shown in Fig. 7.

Normalization

In order to improve the prediction accuracy, we normalize the values of $PM_{2.5}$ concentration using the Min-Max normalization, the method is given in the equation 13:

Fig. 7 Wind Direction and Degree Values
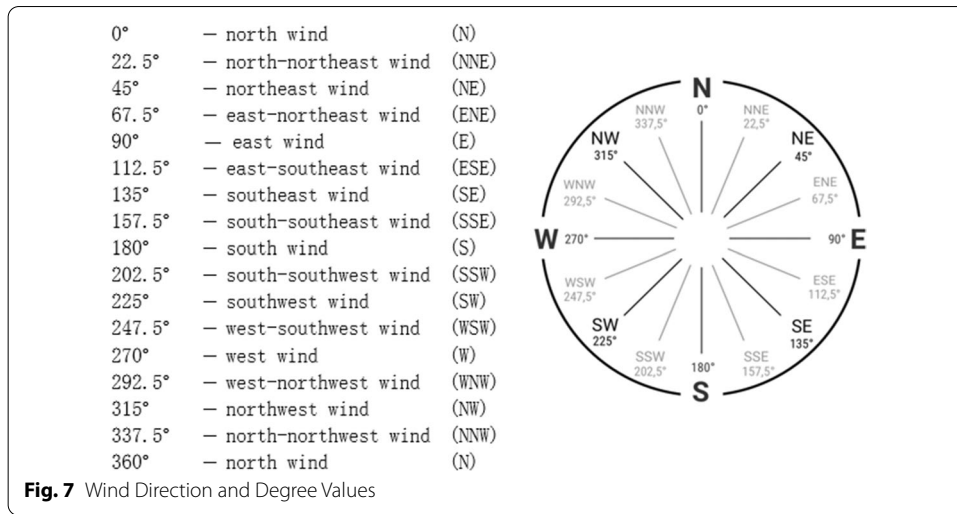
$$x = \frac{x - min}{max - min} \tag{13}$$

Feature selection

In machine learning applications, features selection is an essential step that can be done in several ways. Most of the previous work has applied a mathematical correlation to find the relationship between the input and output variables [48–51]. When there are many features to enter the network for training, finding the correlation between the target output value and those features reduces the complexity of training and improves performance [48].

The Pearson correlation is the most popular method used to find the correlation between two variables. The following equation can calculate its coefficient *r*:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{14}$$
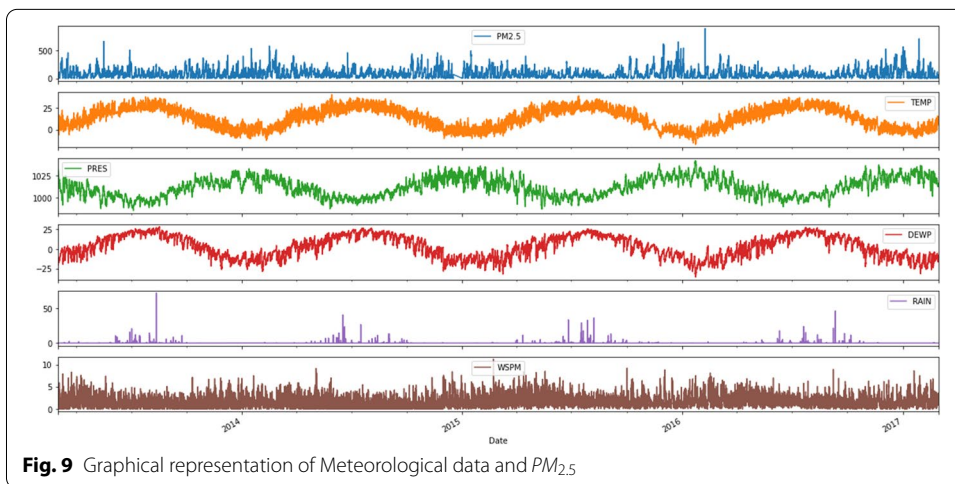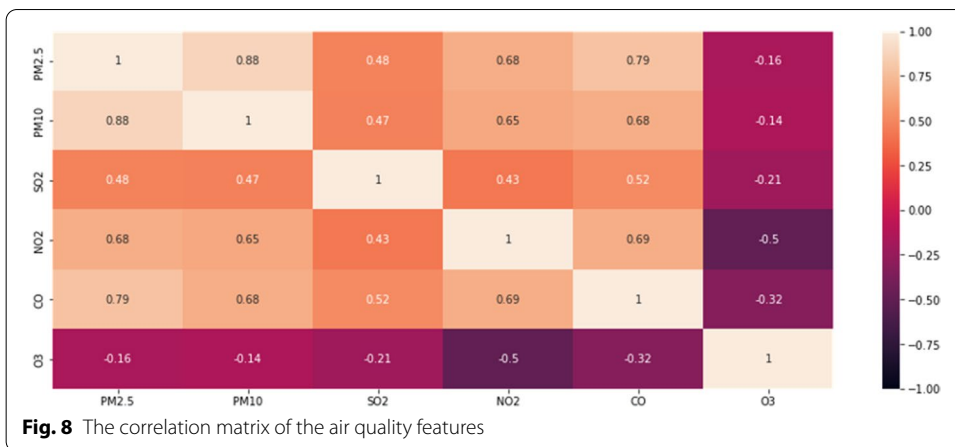
where x and y represent variables, and $\bar{x}$ and $\bar{y}$ represent the mean of the variables.

Air quality feature

In the atmosphere, we detect different pollutants, the increase of their concentrations negatively affects the quality of the air. We calculated the correlations between the features, of the air quality and we found a high correlation value between $PM_{2.5}$, PM_{10}, and CO as shown in Fig. 8.

Meteorological feature

Weather parameters (atmospheric temperature, atmospheric pressure, wind speed, wind direction, and relative humidity) affect air quality. For example, high wind speed will reduce the concentration of $PM_{2.5}$, high humidity generally worsens air pollution, and high air pressure generally results in good air quality [50, 51]. Therefore, meteorological parameters are of prime importance for the task of forecasting air quality (Fig. 9).

Fig. 8 The correlation matrix of the air quality features

Fig. 9 Graphical representation of Meteorological data and $PM_{2.5}$

Spatial analysis

We performed the spatial correlation between Aotizhongxin station (target) and other adjacent stations. We used Pearson correlation to select the correlated $PM_{2.5}$ monitoring stations around the target.

The results are shown in Fig. 10. All correlation values are above 0.80 indicate that there is a strong spatial correlation between the selected stations.

The data set has been split into two, a training set and a test set. 80% (28,052 h) of the dataset was taken as a training set. The remaining 20% (7012 h) becomes the test set used to test the model and analyze its accuracy.

Evaluation index of the models

Once the structure of the model is determined, the training set is used to train the network until convergence. In order to assess the efficiency of the model, three indicators are used in this article, including the mean absolute error (MAE), the mean squared error (RMSE), and the coefficient of determination (R^2).

Fig. 10 The Spatiotemporal Correlation Analysis
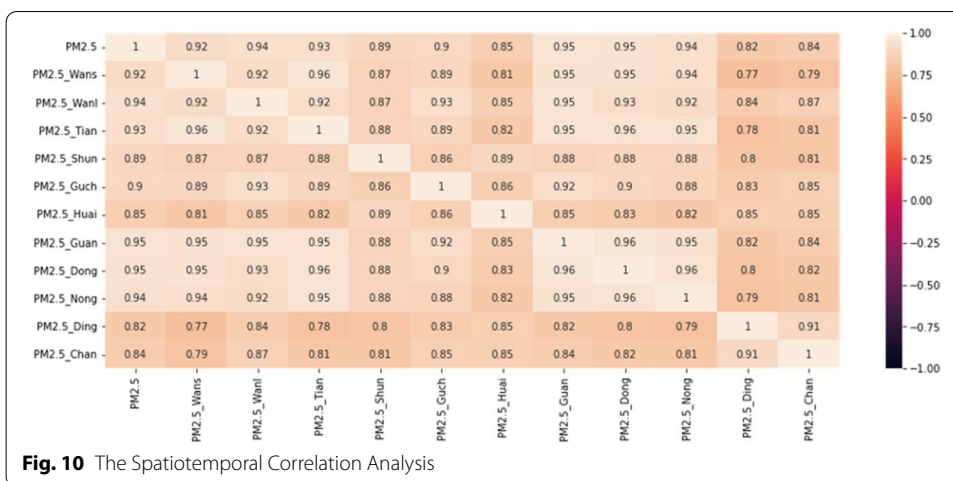
MAE

MAE (Mean Absolute Error) is the arithmetic mean of the absolute values of the deviations between the true value and the model prediction value of all samples, which can better reflect the real prediction error situation. The calculation formula is as follows:

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

RMSE

RMSE (Root Mean Square Error) is the square root of the mean of the square of all of the error. It may well reflect the accuracy of the prediction error. The calculation formula is shown below:

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

R^2

The coefficient of determination reflects the proportion of all variations of the dependent variable that can be explained by the independent variable through the regression relationship. The closer the value of R^2 is to 1 becomes, the better the independent variable can explain the dependent variable. See the calculation formula below:

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

In these three equations, n is the sample size, y_i and $\hat{y}_i$ represent the real value and predicted value at time, respectively; $\bar{y}_i$ denotes the mean of all real values.

Results and discussions

We designed our models with various Python packages, including Scikit-Learn, Keras, and native TensorFlow. For hardware, We ran our heavier workloads on Google Colab, which housed NVIDIA's Tesla T4 GPU.

In this research, the prediction of the concentration of $PM_{2.5}$ was simulated using various deep-learning models. In this section, the historical observation $PM_{2.5}$ data are compared with the computed $PM_{2.5}$ from artificial neural networks such as LSTM, GRU, Bi-LSTM, Bi-GRU, CNN, CNN-LSTM, and CNN-GRU tested in one and seven lag days. Figure 11 shows the workflow for predicting $PM_{2.5}$ concentrations.

Each network attempts to predict the results as accurately as possible. The value of the accuracy in the network is achieved by the cost function trying to punish the network when it fails. The optimal output is the lowest cost.

In this study, for all networks, we applied MSE (Mean Squared Error) as a cost function. A repetition step in training generally works with a division of training data named a batch size. The number of samples for each batch is a hyperparameter, generally obtained by trial and error.

The value of this parameter in all models is 24, 32, 64, and 128, respectively, as this study has shown. In each repetition step, the cost function is computed as the mean MSE of these observed and predicted $PM_{2.5}$ concentration samples. The number of iteration steps for neural networks is named an epoch; in each epoch, the streamflow time series is simulated by the network once. Like other networks, neurons or network layers can be selected arbitrarily in recurrent networks. In our study for comparing models with each other, the structures of all recurrent network models are created identically.

- In LSTM, GRU, BI-LSTM, and BI-GRU each network, four hidden layers are used, 200 units in the first layer, then 100 in the second layer, and 50 units in the last two layers. The last layer output of the network is linked to a dense layer with a single output neuron. Between the layers, a dropout equal to 10% is used. In all networks, the ReLU [52] activation function is applied for the hidden layers.

Fig. 11 Workflow for predicting $PM_{2.5}$ concentrations
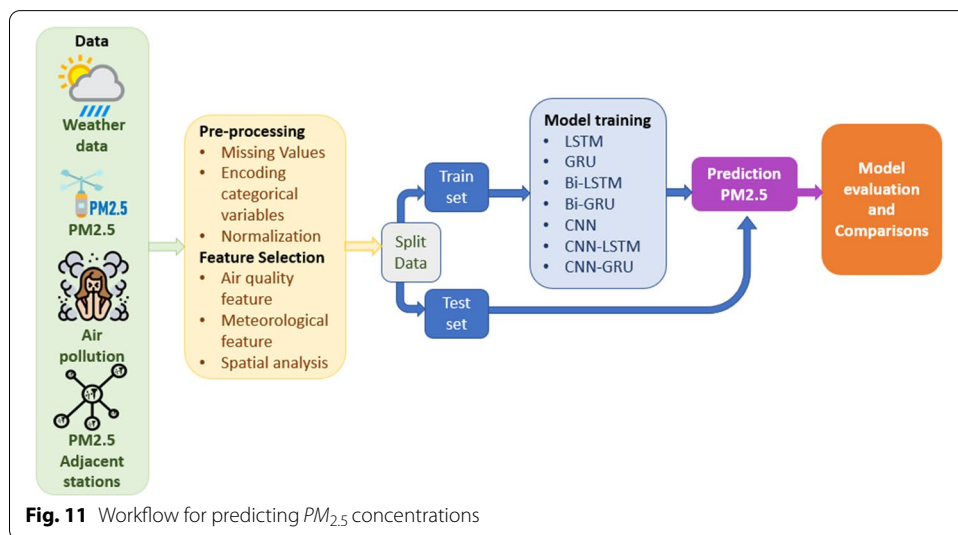
- In CNN-LSTM and CNN-GRU each network of them contains 1D CNN, which Contains three convolutional layers, with 64, 64, and 32 feature detectors successively, the length of the convolution window is 3 with causal padding. Between the three convolutional layers, the BatchNormalization layer is used. All is followed by a MaxPooling1D layer with a pool size of 3. This last is linked to LSTM/GRU, which Contains two layers with 100 and 50 units per layer, then a dense layer with a single output neuron. An overview of the proposed CNN–LSTM models architecture is depicted in Fig. 12.

The main advantage of using ReLU is that there is a fixed derivative for all inputs greater than 0. This constant derivative speeds up network learning. Each method is run with 200 epoch, and a $EarlyStopping(min\_delta = 1e - 3, patience = 50)$. All models are run with different Batch sizes. As seen in Table 1, the Batch size as one of the influential parameters plays a primary role. We used Adam as an optimizer with the learning rate (0,001) and learning rate decay (0.0001). As showcased in Table 1, three different evaluation criteria compare seven different prediction methods.

Table 1 summarizes the MAE, RMSE, and R^2 values for the concentration of $PM_{2.5}$ in air generated by the model prediction models. In 65 models, the RMSE values for the 1-day lags were the smallest. However, the results show that the CNN-LSTM performed best in a one-hour forecast compared to other models under the same conditions, and different batches sizes. Moreover, these results show that the CNN-LSTM with 32 in batch size is more accurate in the different lags, with an advantage in 1-day lag.

Figures 13, 14, 15 shows the MAE, RMSE, and R^2 in 1 and 7-days lags of the seven models in 32 Batch size. These values are between the predicted and true values of $PM_{2.5}$ concentration.
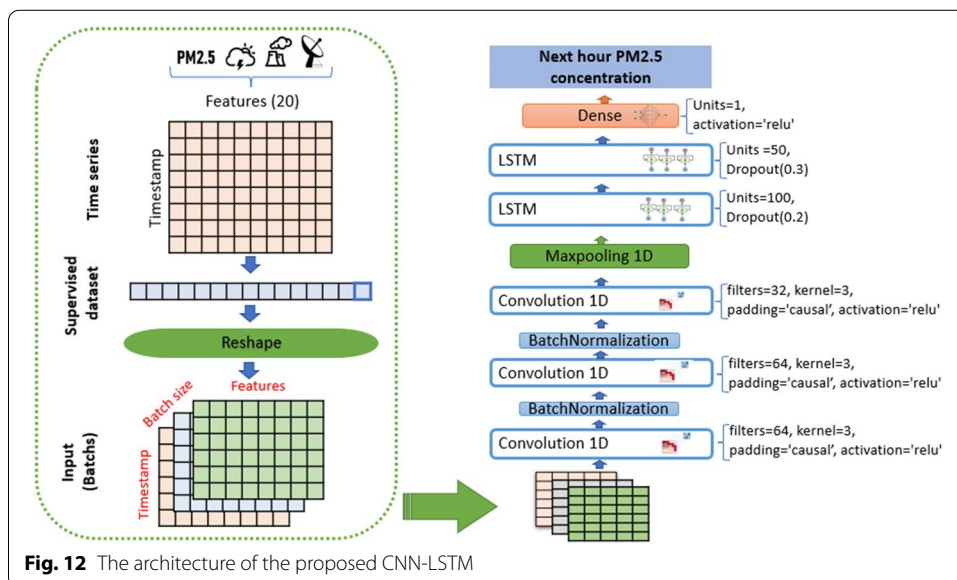
This Figure indicates that :

Fig. 12 The architecture of the proposed CNN-LSTM

Table 1 Results of different models in 1 and 7-days Lags (Bold indicates the best results.)

| Model	Batch	1 Day			7 Day		
		MAE	RMSE	R^2	MAE	RMSE	R^2
GRU	24	9.960	16.571	0.979	12.868	21.526	0.953
	32	9.541	16.609	0.978	11.055	18.122	0.974
	64	10.021	17.220	0.977	12.258	19.797	0.965
	128	9.842	16.904	0.978	11.899	19.560	0.970
LSTM	24	9.102	15.999	0.980	11.916	19.255	0.969
	32	9.503	16.217	0.980	11.623	19.154	0.972
	64	9.252	16.044	0.980	11.551	19.155	0.970
	128	9.725	16.997	0.978	11.823	19.227	0.971
Bi-LSTM	24	8.947	15.710	0.982	12.204	20.050	0.966
	32	8.868	15.597	0.982	11.253	18.488	0.974
	64	9.561	16.380	0.980	12.055	19.323	0.969
	128	9.488	16.456	0.979	11.753	18.113	0.971
Bi-GRU	24	9.907	16.859	0.978	11.488	18.854	0.969
	32	9.692	16.712	0.978	11.984	19.294	0.970
	64	9.192	16.196	0.981	11.631	19.289	0.969
	128	9.230	16.046	0.981	11.553	19.113	0.970
CNN	24	9.663	17.062	0.978	10.693	17.962	0.973
	32	9.591	16.981	0.979	11.150	18.606	0.975
	64	9.261	16.667	0.979	10.621	18.431	0.975
	128	9.974	17.636	0.977	10.674	18.636	0.974
CNN-LSTM	24	9.198	16.523	0.981	9.353	16.724	0.978
	32	**6.742**	**12.921**	**0.989**	**9.034**	**16.625**	**0.979**
	64	7.869	15.757	0.982	9.885	18.373	0.976
	128	8.940	16.337	0.980	9.037	16.524	0.979
CNN-GRU	24	9.812	17.554	0.977	9.912	18.564	0.965
	32	9.459	17.700	0.977	9.459	18.700	0.967
	64	9.433	16.836	0.979	9.653	17.856	0.976
	128	9.499	17.285	0.979	9.949	17.885	0.976
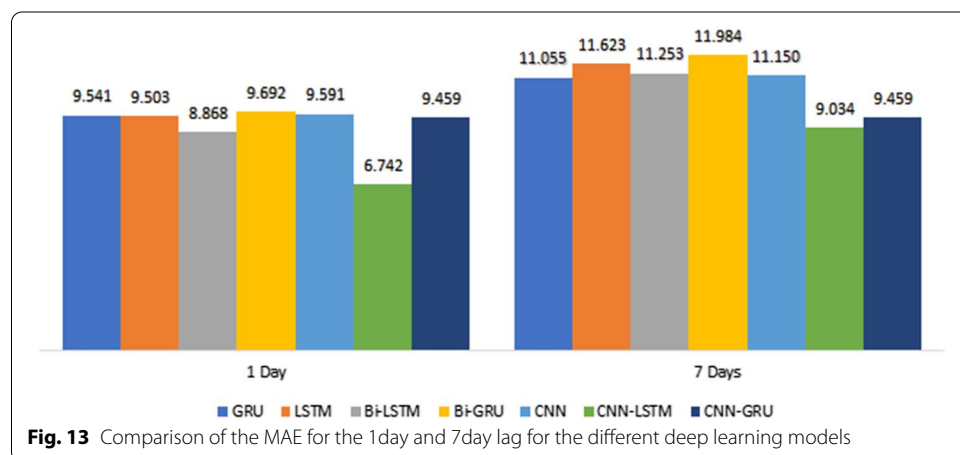
Fig. 13 Comparison of the MAE for the 1day and 7day lag for the different deep learning models
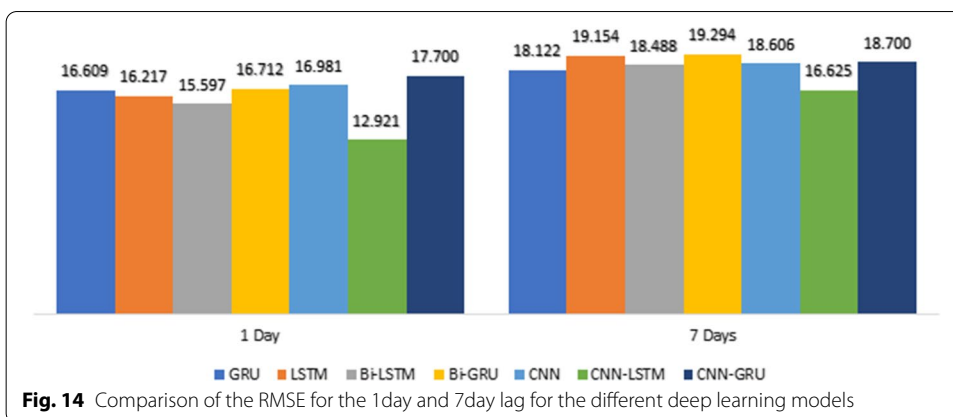
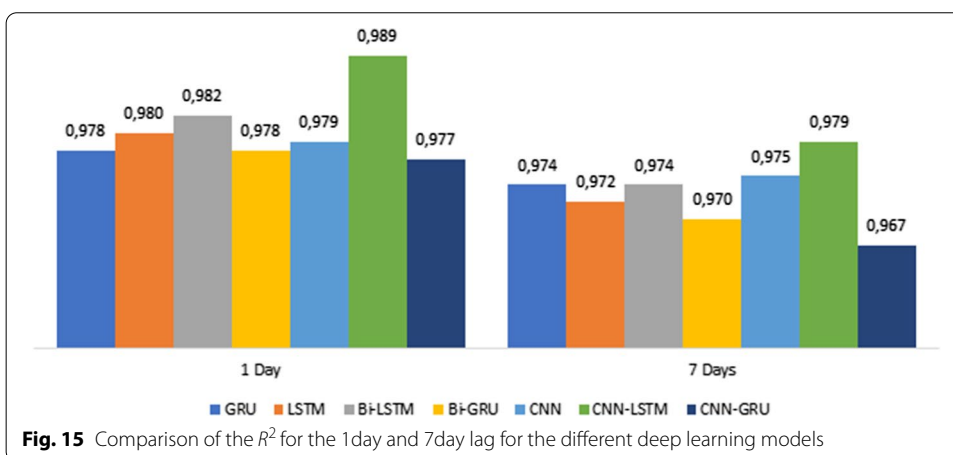**Fig. 14** Comparison of the RMSE for the 1day and 7day lag for the different deep learning models

Fig. 15 Comparison of the R^2 for the 1day and 7day lag for the different deep learning models

- By comparing CNN with LSTM in 1-day lags, the MAE and RMSE of LSTM decrease, R^2 increases, MAE decreases from 9.591 to 9.503, and RMSE decreases from 16.981 to 16.217, so LSTM was better than CNN. However, the error measurement indexes MAE and RMSE of CNN-LSTM are the smallest, and the maximum R^2 is close to 1.
- By Comparing CNN-LSTM with LSTM based on metrics MAE and RMSE. The proposed model in this paper has the smallest value of MAE and RMSE than those LSTM without the CNN layer, R^2 has a certain improvement, MAE decreases from 9.503 to 6.742, RMSE decreases from 16.217 to 12.921, and R^2 increases close to 1.
- By comparing CNN-LSTM in 1-days lags with 7-days lags, the MAE and RMSE increases, R^2 decreases, MAE increases from 6.742 to 9.034, and RMSE increases from 12.921 to 16.625 (Fig. 16).

Overall, observations from Table 1 and Figures 13, 14, and 15 show that the performance of CNN-LSTM in 1-day lags is the best among the Seven models. In terms of forecasting accuracy, MAE is 6.742 and RMSE is 12.921, which is the smallest among the seven forecasting models and has high forecasting accuracy, in terms of forecasting performance, and the R^2 of CNN-LSTM is 0.989, Therefore, the CNN-LSTM proposed in
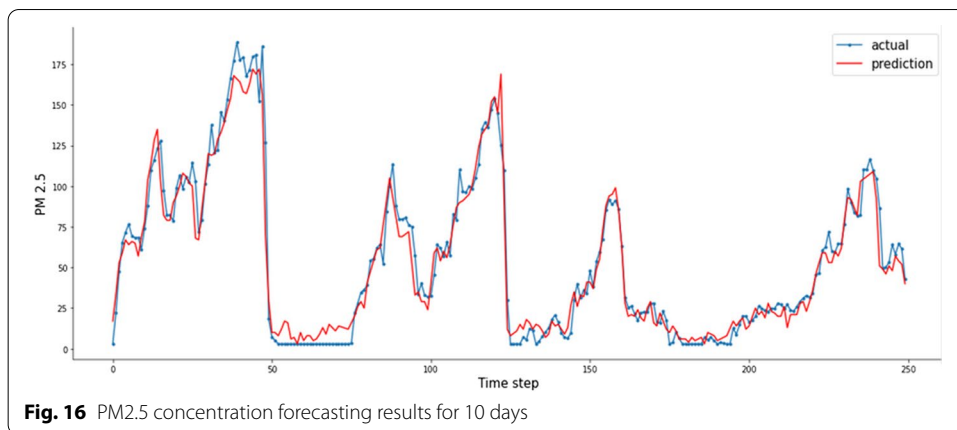
Bekkar *et al. Journal of Big Data*   (2021) 8:161

Page 18 of 21

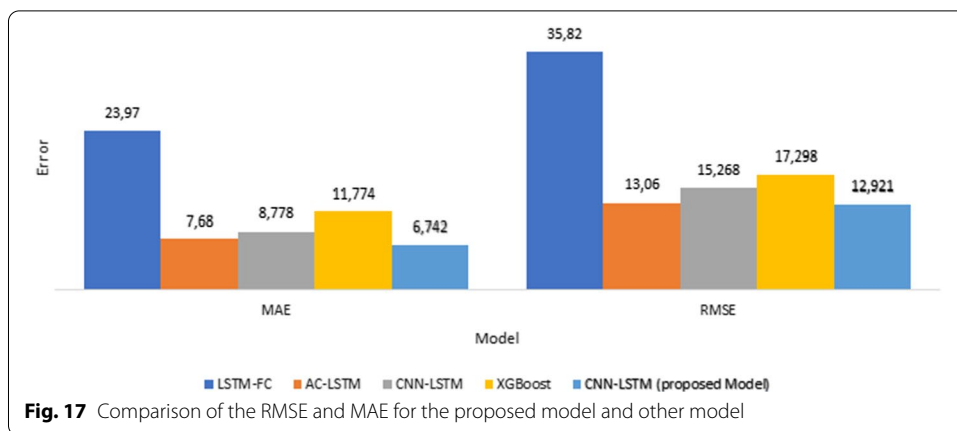**Fig. 16** PM2.5 concentration forecasting results for 10 days

Fig. 17 Comparison of the RMSE and MAE for the proposed model and other model

this paper is superior to the other comparative models, so the predicted value has a good explanation for the true value.

Comparison with recent work

Four recently published models are suggested, such as AC-LSTM by [53], LSTM-FC by [33], XGBoost by [54], and CNN-LSTM by [55], which are evaluated for comparing the performance of the proposed model. Those four models were also used to forecast pollutant particles $PM_{2.5}$. The comparison investigation was using the same two metrics, MAE, RMSE.

A comparative examination of MAE and RMSE, as shown in Fig. 17, shows that not only the lowest mean absolute error but also the lowest root mean square error occurs in the suggested model.

In this study, we developed a CNN-LSTM, which can effectively perform Spatio-temporal prediction, and used it to predict air quality in Beijing. The data of $PM_{2.5}$ concentration, concentrations of air pollutants highly correlated with $PM_{2.5}$, meteorological data, and $PM_{2.5}$ concentrations were collected from several locations of adjacent monitoring stations. The $PM_{2.5}$ prediction model showed high predictive

accuracy and explanatory power, as well as the potential for future improvement by introducing a long-term prediction model.

- First, the CNN-LSTM prediction model can be expected to produce high $PM_{2.5}$ prediction accuracy by learning Spatio-temporal information from big data. In the case of previous prediction models, it is difficult to effectively learn Spatio-temporal information. The CNN-LSTM prediction model directly manages space-time information from adjacent stations.
- Second, we can learn effectively with the CNN-LSTM model by using data from adjacent monitoring stations. Existing air quality monitoring models have shown limitations in measuring and predicting particulate matter, due to ignorance of the effects of pollution in places not covered by the monitoring station. However, the prediction model proposed in this article can support the effects of uncovered areas.
- Third, our model was only applied in the city of Beijing in China due to the limitation of hourly open access data. In the future, the proposed model can be comprehensively evaluated by applying it to other study areas or to other time periods once the greatest amount of data is available.

However, our study has a limit. The concentrations of pollutants of foreign origin affecting Beijing were not taken into account in this study. For example, the air pollution caused by other Chinese cities is carried by the wind.

Conclusion

In this paper, we proposed a hybrid model based on CNN and LSTM, which was used to predict the $PM_{2.5}$ of air pollutants in the urban area of Beijing. First of all, the historical data of the stations were analyzed for correlation. After experimental comparison, a feature with a higher correlation coefficient with the $PM_{2.5}$ was selected, weather data, and correlation between other stations. Secondly, based on the proposed hybrid model, we also used CNN to effectively extract the spatial characteristics of and the internal characteristics between different attributes; simultaneously, LSTM was used to obtain the time features and obtain a more accurate and stable prediction effect. Through performance evaluation and comparison of results, the main findings of this paper are as follows: this model can effectively extract the temporal and spatial features of the data through CNN and LSTM, and it also has high accuracy and stability. Due to the periodicity of the air quality data, a 24h was chosen for the input values.

Declarations

Ethics approval and consent to participate
The author confirms the sole responsibility for this manuscript. The author read and approved the final manuscript

Author details
[1]FSTM, University Hassan II, Casablanca, Morocco. [2]FMPR, University Mohammed V, Rabat, Morocco.

References

1. Urban population (% of total population). https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS Accessed 20 Oct 2021.
2. Department of Economic and Social Affairs: Urban Population Change; 2018. https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html. Accessed 20 Oct 2021.
3. Nada Osseiran, Christian Lindmeier: 9 out of 10 people worldwide breathe polluted air, but more countries are taking action; 2018. https://www.who.int/news/item/02-05-2018-9-out-of-10-people-worldwide-breathe-polluted-air-but-more-countries-are-taking-action Accessed 20 July 2021.
4. Ailshire JA, Crimmins EM. Fine particulate matter air pollution and cognitive function among older US adults. Am J Epidemiol 2014;180(4):359–66. https://doi.org/10.1093/aje/kwu155. https://academic.oup.com/aje/article-pdf/180/4/359/8640802/kwu155.pdf.
5. Pöschl U. Atmospheric aerosols: composition, transformation, climate and health effects. Angewandte Chemie Int Ed. 2005;44(46):7520–40. https://doi.org/10.1002/anie.200501122.
6. Du Y, Xu X, Chu M, Guo Y, Wang J. Air particulate matter and cardiovascular disease: the epidemiological, biomedical and clinical evidence. J Thoracic Dis. 2016;8(1):8.
7. Cohen AJ, Brauer M, Burnett R, Anderson HR, Frostad J, Estep K, Balakrishnan K, Brunekreef B, Dandona L, Dandona R, et al. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the global burden of diseases study 2015. Lancet. 2017;389(10082):1907–18.
8. Bu X, Xie Z, Liu J, Wei L, Wang X, Chen M, Ren H. Global pm2.5-attributable health burden from,. to 2017: estimates from the global burden of disease study 2017. Environ Res. 1990;2021(197):111123.
9. OCDE. The economic consequences of outdoor air pollution; 2016, p. 116. https://doi.org/10.1787/9789264257474-en. https://www.oecd-ilibrary.org/content/publication/9789264257474-en.
10. Mo Z, Huang J, Chen Z, Zhou B, Zhu K, Liu H, Mu Y, Zhang D, Wang S. Cause analysis of pm 2.5 pollution during the covid-19 lockdown in Nanning, China. Sci Rep. 2021;11(1):1–13.
11. Rodríguez-Urrego D, Rodríguez-Urrego L. Air quality during the covid-19: Pm2.5 analysis in the 50 most polluted capital cities in the world. Environ Pollut. 2020. https://doi.org/10.1016/j.envpol.2020.115042.
12. Zoran MA, Savastru RS, Savastru DM, Tautan MN. Assessing the relationship between surface levels of pm2.5 and pm10 particulate matter impact on covid-19 in Milan, Italy. Sci Tot Environ. 2020;738:139825.
13. Md N, Wai Y, Ibrahim N, Rashid Z, Mustafa N, Hamid H, Latif M, Er S, Yik L, Alhasa K, et al. Particulate matter (pm2.5) as a potential sars-cov-2 carrier; 2020.
14. Zhu Y, Xie J, Huang F, Cao L. Association between short-term exposure to air pollution and covid-19 infection: evidence from China. Sci Tot Environ. 2020;727:138704.
15. Wu X, Nethery RC, Sabath M, Braun D, Dominici F. Air pollution and covid-19 mortality in the United States: strengths and limitations of an ecological regression analysis. Sci Adv. 2020;6(45):4049.
16. Lu D, Mao W, Xiao W, Zhang L. Non-linear response of pm2.5 pollution to land use change in China. Remote Sens. 2021;13(9):1612.
17. Bartholomew DJ. Time series analysis forecasting and control. J Oper Res Soc. 1971;22(2):199–201. https://doi.org/10.1057/jors.1971.52.
18. Kumar U, Jain V. Arima forecasting of ambient air pollutants (o 3, no, no 2 and co). Stochastic Environ Res Risk Assess. 2010;24(5):751–60.
19. Yu R, Yang Y, Yang L, Han G, Move OA. Raq-a random forest approach for predicting air quality in urban sensing systems. Sensors. 2016;16(1):86.
20. Lin K-P, Pai P-F, Yang S-L. Forecasting concentrations of air pollutants by logarithm support vector regression with immune algorithms. Appl Math Comput. 2011;217(12):5318–27.
21. Wang P, Liu Y, Qin Z, Zhang G. A novel hybrid forecasting model for pm10 and so2 daily concentrations. Sci Tot Environ. 2015;505:1202–12.
22. Mishra D, Goyal P. Neuro-fuzzy approach to forecasting ozone episodes over the urban area of Delhi, India. Environ Technol Innov. 2016;5:83–94.
23. Zaidan MA, Dada L, Alghamdi MA, Al-Jeelani H, Lihavainen H, Hyvärinen A, Hussein T. Mutual information input selector and probabilistic machine learning utilisation for air pollution proxies. Appl Sci. 2019;9(20):4475.
24. Danesh Yazdi M, Kuang Z, Dimakopoulou K, Barratt B, Suel E, Amini H, Lyapustin A, Katsouyanni K, Schwartz J. Predicting fine particulate matter (pm2. 5) in the greater London area: an ensemble approach using machine learning methods. Remote Sens. 2020;12(6):914.

Bekkar *et al. Journal of Big Data*   (2021) 8:161

Page 21 of 21

25. Salman AG, Heryadi Y, Abdurahman E, Suparta W. Single layer & multi-layer long short-term memory (lstm) model with intermediate variables for weather forecasting. Procedia Comput Sci. 2018;135:89–98.

26. Tsai Y-T, Zeng Y-R, Chang Y-S. Air pollution forecasting using rnn with lstm. In: 2018 IEEE 16th Intl Conf on dependable, autonomic and secure computing, 16th intl conf on pervasive intelligence and computing, 4th intl conf on big data intelligence and computing and cyber science and technology congress (DASC/PiCom/DataCom/CyberSciTech). IEEE; 2018, p. 1074–9.

27. Shi P, Zhang G, Kong F, Chen D, Azorin-Molina C, Guijarro JA. Variability of winter haze over the Beijing-Tianjin-Hebei region tied to wind speed in the lower troposphere and particulate sources. Atmos Res. 2019;215:1–11.

28. Pohjola MA, Kousa A, Kukkonen J, Härkönen J, Karppinen A, Aarnio P, Koskentalo T. The spatial and temporal variation of measured urban pm 10 and pm 2.5 in the Helsinki metropolitan area. Water Air Soil Pollut Focus. 2002;2(5):189–201.

29. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44.

30. Garg S, Jindal H. Evaluation of time series forecasting models for estimation of pm2.5 levels in air. In: 2021 6th international conference for convergence in technology (I2CT). IEEE; 2021, p. 1–8.

31. Xayasouk T, Lee H, Lee G. Air pollution prediction using long short-term memory (lstm) and deep autoencoder (dae) models. Sustainability. 2020;12(6):2570.

32. Jeya S, Sankari L. Air pollution prediction by deep learning model. In: 2020 4th international conference on intelligent computing and control systems (ICICCS). IEEE; 2020, p. 736–41.

33. Zhao J, Deng F, Cai Y, Chen J. Long short-term memory-fully connected (lstm-fc) neural network for pm2.5 concentration prediction. Chemosphere. 2019;220:486–92.

34. Rijal N, Gutta RT, Cao T, Lin J, Bo Q, Zhang J. Ensemble of deep neural networks for estimating particulate matter from images. In: 2018 IEEE 3rd international conference on image, vision and computing (ICIVC). IEEE; 2018, p. 733–8.

35. Zhang L, Li D, Guo Q. Deep learning from spatio-temporal data using orthogonal regularizaion residual cnn for air prediction. IEEE Access. 2020;8:66037–47.

36. Li J, Jin M, Li H. Exploring spatial influence of remotely sensed pm2.5 concentration using a developed deep convolutional neural network model. Int J Environ Res Public Health. 2019;16(3):454.

37. Le V-D, Bui T-C, Cha S-K. Spatiotemporal deep learning model for citywide air pollution interpolation and prediction. In: 2020 IEEE international conference on big data and smart computing (BigComp); 2020, p. 55–62.

38. Huang C-J, Kuo P-H. A deep cnn-lstm model for particulate matter (pm2.5) forecasting in smart cities. Sensors. 2018;18(7):2220.

39. Werbos PJ. Generalization of backpropagation with application to a recurrent gas market model. Neural Netw. 1988;1(4):339–56.

40. Robinson A, Fallside F. The utility driven dynamic error propagation network. Cambridge: University of Cambridge Department of Engineering; 1987.

41. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80.

42. Albawi S, Mohammed TA, Al-Zawi S. Understanding of a convolutional neural network. In: 2017 international conference on engineering and technology (ICET). IEEE; 2017, p. 1–6.

43. Li X, He M, Li H, Shen H. A combined loss-based multiscale fully convolutional network for high-resolution remote sensing image change detection. IEEE Geosci Remote Sens Lett. 2021. https://doi.org/10.1109/LGRS.2021.3098774.

44. Ji S, Xu W, Yang M, Yu K. 3d convolutional neural networks for human action recognition. IEEE Trans Pattern Anal Mach Intell. 2012;35(1):221–31.

45. Liu Y, Zheng H, Feng X, Chen Z. Short-term traffic flow prediction with conv-lstm. In: 2017 9th international conference on wireless communications and signal processing (WCSP). IEEE; 2017, p. 1–6.

46. Zhang S, Guo B, Dong A, He J, Xu Z, Chen SX. Cautionary tales on air-quality improvement in Beijing. The Royal Society Publishing; 2017. https://archive.ics.uci.edu/ml/datasets/Beijing+Multi-Site+Air-Quality+Data. Accessed 20 July 2021.

47. Norazian M, Al Bakri AMM, Shukri YA, Azam RN. Estimation of missing values for air pollution data using interpolation technique. Simulation. 2006;75:94.

48. Tao Q, Liu F, Li Y, Sidorov D. Air pollution forecasting using a deep learning model based on 1d convnets and bidirectional gru. IEEE Access. 2019;7:76690–8.

49. Freeman BS, Taylor G, Gharabaghi B, Thé J. Forecasting air quality time series using deep learning. J Air Waste Manage Assoc. 2018;68(8):866–86. https://doi.org/10.1080/10962247.2018.1459956 (**PMID: 29652217**).

50. Zheng Y, Yi X, Li M, Li R, Shan Z, Chang E, Li T. Forecasting fine-grained air quality based on big data. In: Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining; 2015, p. 2267–76.

51. Zheng Y, Liu F, Hsieh H-P. U-air: when urban air quality inference meets big data. In: Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining; 2013, p. 1436–44.

52. Agarap AF. Deep learning using rectified linear units (relu). arXiv preprint arXiv:1803.08375; 2018.

53. Li S, Xie G, Ren J, Guo L, Yang Y, Xu X. Urban pm2.5 concentration prediction via attention-based cnn-lstm. Appl Sci. 2020;10(6):1953.

54. Pan B. Application of xgboost algorithm in hourly pm2. 5 concentration prediction. In: IOP conference series: earth and environmental science, vol. 113. IOP publishing; 2018, p. 012127.

55. Wardana I, Gardner JW, Fahmy SA. Optimising deep learning at the edge for accurate hourly air quality prediction. Sensors. 2021;21(4):1064.

Publisher's Note