BULGARIAN ACADEMY OF SCIENCES

CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **23**, No 4
Sofia • 2023 Print ISSN: 1311-9702; Online ISSN: 1314-4081
DOI: 10.2478/cait-2023-0045

Spatial and Temporal Variations on Air Quality Prediction Using Deep Learning Techniques

S. Vandhana, J. Anuradha

School of Computer Science and Engineering, Vellore Institute of Technology, Vellore 632014, Tamil Nadu, India
E-mails: januradha@vit.ac.in svandhana2012@gmail.com

***Abstract*:** *Air Pollution is constantly causing a severe effect on the environment and public health. Prediction of air quality is widespread and has become a challenging issue owing to the enormous environmental data with time-space nonlinearity and multi-dimensional feature interaction. There is a need to bring out the spatial and temporal factors that are influencing the prediction. The present study concentrates on the correlation prediction of spatial and temporal relations. A Deep learning technique has been proposed for forecasting the accurate prediction. The proposed Bi_ST model is evaluated for 17 cities in India and China. The predicted results are evaluated with the performance metrics of RMSE, MAE, and MAPE. Experimental results demonstrate that our method Bi_ST accredits more accurate forecasts than all baseline RNN and LSTM models by reducing the error rate. The accuracy of the model obtained is 94%.*

***Keywords*:** *Bi-LSTM, Air pollution, Spatial-temporal analysis, Prediction.*

1. Introduction

The primary environmental risk factor is exposure to ambient air pollution, which severely affects public health. Air pollution is considered as a time series data and its application is spread in wide variety. One such is the prediction of the future value of an object based on past value. Particles ($PM_{2.5}$ and PM_{10}, that is, particles smaller than 2.5 or 10 μm in size) and nitric oxide (Nitrogen Oxides) are the dominant components of air pollution associated with Urban development [1-3]. Epidemiological studies have shown that some air particles such as $PM_{2.5}$, $PM_{10,}$ and NO_x are the major pollutant contributors associated with different diseases [4]. Various countries recently developed air quality standards and guidelines for atmospheric environment assessment and public health protection. The concentration of air pollutants such as carbon monoxide and sulfur oxides that exceed air quality guidelines can cause chronic health problems with short-term exposure [5].

Therefore, an accurate and reliable model for assessing the status of air pollution is indispensable.

There is a need to manage urban air quality and information systems to forecast future air pollution levels over several days and to provide precise measures and control strategies in the outdoor environment. A variety of studies have recently been proposed to predict air pollutant concentrations like statistical models [6], Artificial Neural Network (ANN) [6], Weather Forecasting models [6], and other hybrid models [6]. These methods do a good job of forecasting air pollutant concentrations. Using these methods, ANN has demonstrated non-linear mapping, robustness, and self-adaptation in predictive fields. Unlike other modeling techniques, ANN makes no assumptions regarding the distribution of data. J i a n g, L i and Z h a n g [7] have developed a hybrid model using spectrum analysis technology to divide the original time series into lower layers and a harmony search algorithm to improve neural network parameters with backward propagation. F e n g et al. [8] used a genetic algorithm to match BPNN parameters for ozone prediction, and Support Vector Machine (SVM) has been used accurately to classify data into appropriate classification and prediction. Multilayer perceptron techniques and radial function, as well as a principle component-based regression analysis model, have been used to predict hourly particle concentrations in air at a volume less than 10 um (PM_{10}) on Cyprus Island [9]. A regression neural network is used to predict PM_{10} pollutants, which perform well in predicting PM_{10} at the country level [10]. The improved Elman network where both the outputs of hidden and output layers are allowed to feedback themselves through a context layer, has been used to forecast PM_{10} and has provided good forecasting compared to standard models [11]. P a i et al. [12] have used the Adaptive Neuro Fuzzy Inference System (ANFIS), which consists of ANN, fuzzy logic expression, and if-then rules to predict 24-hour oxidation concentration. It can overcome the limitations of conventional neural networks and improve prediction performance. Although these studies have obtained preferred prediction results, it is still necessary to further improve forecasting performance.

The following are the challenges involved in predicting the air quality for any region in a country.

1. First, there is a pattern connected with weather factors. Therefore, these weather patterns cause different variations according to the season and climate. So the pattern of the air quality changes. Several models are needed and taken into account for understanding the different weather patterns.

2. Second, Air Pollution is spread across different locations. The relationship between different locations is complex because they can be affected by geography, natural processes, weather phenomena, or other factors.

3. Third, air quality is a time series dependency that generates sequential properties. The periodic patterns may vary over time due to phenomena such as development, human activity, or emissions from the factory. Therefore, long-term dependencies should be explicitly modeled.

The relationship between different locations and time series dependency is addressed in this study. Geographical locations are included for the prediction of future and past values. The extent to which the timely factors that influence the

prediction are evaluated. The rest of the section is followed by the Section 2 (related work on air quality and health and deep learning techniques), Section 3 (methodology of Bidirectional long short term memory and Bidirectional with spatial and temporal interactions), Section 4 (Dataset description), Section 5 (Experimental work), Section 6 (Results), and Section 7 (Conclusion).

2. Related work

2.1. The air quality and health

Air pollution has not only caused effects on human health, it also affected the IQ level of 6-8 year children in India who have been exposed to household air pollution in the first three years of their life [13]. Air pollution impacts death and high life expectancy across the states of India and India has one of the highest average annual environmental conditions $PM_{2.5}$ particle exposure levels worldwide [14] and on the other side, it has caused critical revolution in the fields of plant growth as a cause of climate change [15]. It has affected diversity of species and the population density of forest birds [16]. Among the air pollutants, the Particulate $Matter_{2.5}$ (the size of the particle diameter less than 2.5 μm) can cause adverse effect on public health. Indoor and outdoor air quality is measured using Internet of Things air quality monitoring system. This has achieved Quality of Service in accuracy, sensing throughput and power consumption optimization [17]. Seven air pollutants O_3, NO_x, NO, NO_2, CO, SO_2, PM_{10} have been forecasted on hourly bases using Random Forest (RF) initially. Later the model has been analyzed with relation to its meteorological variables and the subsequent modelling has been done with its residuals using Auto Regressive Integrated Moving Average (ARIMA) methodology [18].

2.2. Deep learning techniques

Many researchers have applied the traditional statistical methods. S h a r m a et al., [19] have used time series regression model to forecast the air pollution, [20] has examined the relationship between meteorological fields that affect the concentration of $PM_{2.5}$ using multiple regression approach. Author has also performed various seasonal analyses under different meteorological conditions. Authors of [21] and [22] have used Autoregressive Integrated Moving Average (ARIMA) model along with Support Vector Machine and Back Propagation respectively. These hybrid models analyze the time series data and forecast the concentration of air pollution. The above are the conventional models, which underline the linear assumption of data. The underlying fact is that the real-world data has non-linear characteristics and this non-linear data limits the performance of linear models. The concept of ANN has been applied successfully in environment modeling in [23] and air pollution in [24, 25].

The problem of non-linear characteristics in real world data is addressed by many researchers and started to adopt non-linear machine learning methods. ANN is one of the most widely used machine learning methods for prediction and forecasting. For example, Self-Adaptive Neuro-Fuzzy Weighted Extreme Learning Machine (ANFIS-WELM) based on the Weighted Extreme Learning Machine (WELM) and

the Adaptive Neuro Fuzzy Inference System (ANFIS) has been used for predicting the concentration of air pollutants [26]. Authors Alimissis et al. [27] have performed an estimation of urban pollution using ANN and multiple linear regression. The model is used on air pollution time series and has the advantage of both spatial and temporal correlations with high degree of accuracy. A new hybrid model MCSDE-CEEMD-ENN, Modified Cuckoo Search and Differential Evolution Algorithm Complementary Ensemble Empirical Mode Decomposition and Elman Neural Network is proposed to improve the accuracy of prediction of air pollutant concentration [28]. The hourly concentration of nitrogen dioxide at different traffic stations has been forecasted by feed-forward neural network and a parallel Genetic Algorithm has been used for selecting the inputs to the architecture of the multi-layer perceptron model [29].

Air pollution at any time can cause short or long-term impacts. The effect can continue hours, days, or even weeks. However, most of the above ANN methods cannot extend the prediction of time-delay air pollution or identify long-term dependencies. This somehow limits their performance. In order to overcome the limit of ANN, some studies have used advanced deep learning techniques to model time-series data such as Recurrent Neural Network (RNN), Gated Recurrent Unit (GRU), and Long Short-Term Memory (LSTM). These methods have been reported to perform well in various fields. Fu, Zhang and Li [30] use LSTM to predict the traffic flow in a short time and have adopted GRU Neural Network (NN) for forecasting the traffic air pollution. The results show that the RNN along with LSTM and GRU have shown better results than ARIMA. This is the first study to use GRU for predicting traffic flow. A study has shown that Convolution LSTM (ConvLSTM) outperforms Fully Connected LSTM (FC-LSTM) [31]. Authors predict the rainfall intensity of a local region in a short time period. A Long Short-Term Memory-Fully Connected (LSTM-FC) to a neural network is compared with ANN to predict $PM_{2.5}$ contamination of a specific air quality monitoring station over 48 h [32]. The results show that LSTM-FC neural network model provides better prediction performance. The results have shown that the stronger correlation is for temporal than spatial component. Deep Air Learning (DAL) model is proposed to predict the air quality [34]. The model also has embedded feature selection process along with semi-supervised learning in different layers of network. It is a spatiotemporal framework for interpolation, feature analysis and prediction. The process of LSTM takes information only from the past without considering any information from the future. The solution to LSTM is given by Bi-directional LSTM (Bi-LSTM) method. In bi-directional LSTM the hidden states are separated such that it captures the past and future information and sends it as a sequence of forward and backward, respectively. Bi-LSTM units are highly effective in deep learning processes, such as natural language processing [35] and speech recognition [36], due to their ability to learn long-term dependencies. Virendra Barot and Viral Kapadia [37] have used bidirectional and stacking technique in LSTM for fine tuning $PM_{2.5}$ and carbon monoxide in air quality parameters. The model performance is optimized by tuning hyperparameters, regularization to overcome overfitting and various merging options in bidirectional LSTM. An ensemble of deep learning techniques MultiLayer

Perceptron (MLP), BackPropagation Network (BPN) and LSTM are stacked and it is used in intrusion detection systems. The result of each deep learning model is integrated using XGBoost to achieve higher performance [38].

From the literature study, the existing models have used RNN, LSTM and Bi-LSTM in which authors have not included the influence of the geographical locations. That is, how the prediction varies when the spatial locations are also taken into consideration for forecasting the future values. The same is also applicable for past values. When the geography of the air stations has an influence in prediction, the next is to check for the time series analysis. The study has taken its inputs combined with the past values of the same location for a time series predictions. The number of previous days that is included is denoted by *t*, and the experiment is carried out for different days of *t*.

3. Methodology

The present study mainly concentrates on the spatial temporal patterns of $PM_{2.5}$. The input air pollution parameters are $PM_{2.5}$, NO_2, SO_2, CO, O_3, temperature, relative humidity, wind speed, wind direction and rainfall.

Table 0

Pollutant	Unit
PM2.5	$\mu g/m^3$
NO_2	ppb
SO_2	ppb
CO	ppb
O_3	ppb
Temperature	◦C
Relative humidity	NA
Wind speed	NA
Wind direction	NA
Rainfall	NA

The variation to the model of Bi-LSTM is included with the spatial and temporal inputs that are affecting the near prediction of $PM_{2.5}$. The proposed model is an extension of Recurrent Neural Network called as Bidirectional Long Short Term Memory with Spatial and Temporal interactions (Bi_ST).

3.1. LSTM and Bi-directional LSTM

Recurrent Neural Network (RNN) has the disadvantage of capturing the long-term dependencies from the sequence of input. RNN do not seem to have the capacity to learn the dependencies in output. Nevertheless, for a time series application,

considering the recent information alone would affect the model accuracy. This in case reduces the performance in time series applications. The problem of vanishing gradient is overcome by the method of LSTM [39]. LSTM is a cell unit with self-connected path, which allows a value with both forward pass and backward pass. The value that is passed to the unit has to be saved and then called in the required time step. The connection between the cell state helps maintaining the information in memory for a long period of time. However, in practice, the working network of LSTM is still not perfect as it supports only the forward learning. On situations like this the backward processing of the LSTM model would be useful for accurate predictions.

Graves and Schmidhuber [40] have designed a Bi-directional LSTM (Bi-LSTM) model. The model has pioneered in forecasting air quality. For example, the prediction of $PM_{2.5}$ with the combination of auto-encoders and Bi-LSTM network [41]. The model has predicted the strongest correlation between the meteorological variable rainfall and the concentration of $PM_{2.5}$. With the motivation of the related works in LSTM and Bi-LSTM, a Bi_ST is constructed for spatial and temporal conditions. The proposed framework functions on the correlations of spatial and temporal inputs being represented as a neighboring stations and memory of LSTM. Using this and combining the Tobler's First Law of Geography [42], stating all places are related, but nearby places are more related than distant places. This can be spatially illustrated by calculating the distance between two stations. Using Haversine formula the distance between two stations is calculated based on latitude and longitude of each station. The formula uses sine function in order to maintain a valid number in case of tiny distance. We also assume that the current air quality is highly linked to previous air pollution levels and also the days of the future. The distance between the two stations is calculated based on latitude and longitude of each station. The following Haversine formula is suggested to find the distance. Sine function maintains the valid number in case of tiny distance,

$$\text{haver}\sin\left(\frac{d}{R}\right) = \text{haver}\sin(\varphi_2 - \varphi_1) + \cos(\varphi_2)\text{haver}\sin(\Delta\lambda), \tag{1}$$

$$\text{haver}\sin(\theta) = \sin^2\left(\frac{\theta}{2}\right) + \cos(\varphi_2). \tag{2}$$

Here R is the radius of the earth, haversin(.) is the distance between the stations, φ_1 and φ_2 is the latitude between the points, $\Delta\lambda$ is the difference between two points.

Long Short Term Network (LSTM) is capable of learning long-term dependencies. LSTM have been explicitly developed to avoid a long-term dependency problem. The structure of LSTM is a chain like structure with four neural network layers, which interacts in a special way. A single LSTM unit is given in Fig. 1. LSTM consist of cell unit, forget gate layer, activation layer and output layer.

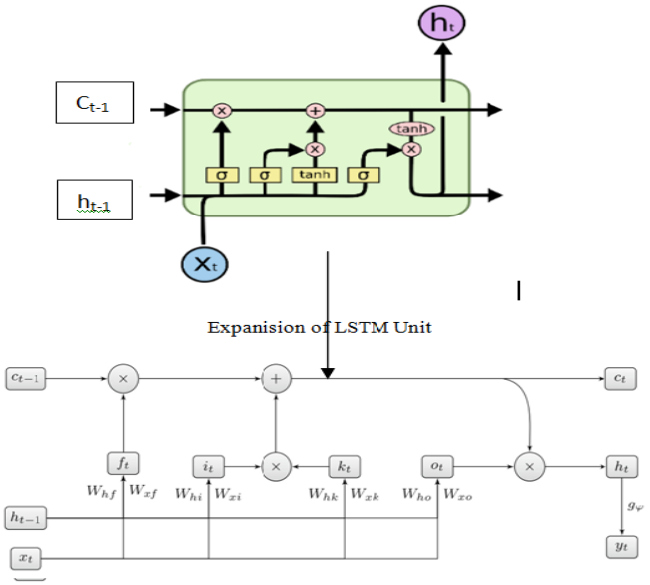

Fig. 1. Single LSTM unit.
Source: **www.inblog.in**

The first layer in LSTM is forget gate, which decides what information should be removed from the cell state. It shuts the old memory from further processing. The inputs given to the cell state are h_{t-1} and x_t, which outputs a number between 0 and 1 for each of the output from the cell state.

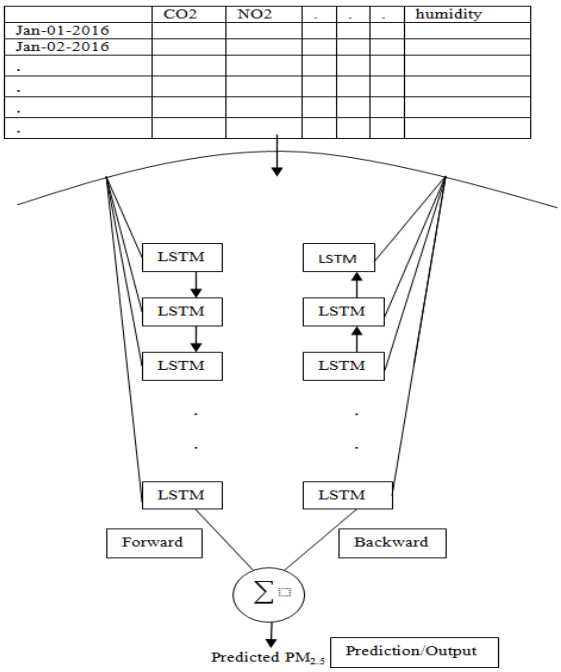

Fig. 2. Architecture of the model

(3) $$f_t(x_t, h_{t-1}) = \sigma(W_{xf} x_t + W_{hf} h_{t-1} + b_f) \begin{cases} 1 & \text{completelyremembers} \\ 0 & \text{completelyforgets} \end{cases},$$

where: W_{xf} are weight matrix associated with input gate *f* at time *t*; W_{hf} are weight matrix associated with hidden state for gate *f*; b_f are bias vector for the forget gate.

The next step is the input gate layer and tanh layer. The input gate layer decides the value to be updated and tanh creates the vector of new candidate values, k_t, which is added to the state. The two layers are combined and updated:

(4) $$i_t = \sigma(W_{xi} x_t + W_{hi} h_{t-1} + W_{ci} h_{t-1} + b_k),$$

where: W_{xi} are weight matrix associated with input gate; W_{hi} are weight matrix of hidden state at the input gate; W_{ci} are weight matrix associated with the cell state; b_k are bias to the forget gate at index *k* or particular time.

(5) $$k_t = \tanh(W_{ij} x_t + W_{hj} h_{t-1} + b_k).$$

where: W_{ij} are weight matrix associated with the input at time step *j* and output at time step *i*; W_{hj} are weight matrix associated with the hidden state at time step *j* and output at time step *i*; W_{ci} are weight matrix associated with the cell state at time step *i*.

The value to the new state is created by the following:

(6) $$c_t = c_{t-1} \times f_t + i_t \times k_t.$$

The output of the output gate layer is,

(7) $$o_t = \sigma(W_{xo} x_t + W_{ho} h_{t-1} + b_o),$$

where: W_{xo} are weight matrix associated with the input for the output gate; W_{ho} are weight matrix associated with the hidden state for the output gate.

(8) $$h_t = \tanh(c_t) \times o_t.$$

The architecture of the model is given in Fig. 2. Each input is given to the different LSTM unit, where the processing is done based on time series input along with the spatial and temporal neighbors as explained in Algorithm 1. Backward LSTM works in the same way. Prediction results from forward and backward LSTM are merged using the sum function. Finally, the merged results are given as prediction output. Input data contains the air pollution parameters and meteorological data fetched into the LSTM unit.

3.2. Bidirectional LSTM with Spatial-Temporal prediction Algorithm

The developed Bi_ST – Bidirectional LSTM with Spatial-Temporal (ST) prediction algorithm is given below.

Algorithm 1. Bi_ST Algorithm

Step 1. Input: Bi_ST (City C, Date d)

Step 2. */* Split data in 10 fold cross validation */*

Step 3. Data=KFold (n_splits=10)

Step 4. */* Generate data for training and testing */*

Step 5. */* Fit an LSTM model to training data */*
Step 6. Procedure fit_lstm (train, epoch, neurons) model= Sequential()
Step 7. Model.add (LSTM(neurons), stateful=True)
Step 8. Model.compile (loss='mean_squared_error', optimizer='adam')
Step 9. For I in range (epoch) **do**
    Step 9a. Model.fit (X, y, epochs=50, shuffle=False)
    Step 9b. Model.reset_states()
Step 10. End
Step 11. Procedure forecast_lstm (model, X)
Step 12. Predict = model.predict (X)
Step 13. Return predict
Step 14. */* Fit lstm model */*
Step 15. Lstm_model=fit.lstm(train,epoch,neurons)
Step 16. */* Forecast training dataset */*
Step 17. Lstm_model.predict(train)
Step 18. For *i* in range (length(test)) **do**
    Step 18a. X=test[*i*]
    Step 18b. Predict=forecast_lstm(lstm_model, X)
    Step 18c. Predictions.append (predict)
    Step 18d. Expected=test[*i*]
Step 19. end
Step 20. MSE=mean_square_error(expected, predictions)
Step 21. RMSE=sqrt(MSE)
    Step 21a. / * Bi_ST function */
Step 22. Bi_ST (City C, date d)
Step 23. Input: $P=(P_1, P_2, \ldots, P_m)$, air pollution and meteorological parameters
Step 24. $P_i \in (s_i, t_i)$ s is a spatial neighbor and *t* is a temporal value
Step 25. Output: Predicted $PM_{2.5}$
Step 26. Begin
Step 27. Let N_1 be a subset of $S=\{n_{11}, n_{12},\ldots\}$ where n_1 and n_2 are spatial neighbors of C
Step 28. Let N_2 be a subset of $T=\{n_{21}, n_{22},\ldots\}$ where n_{2i} are temporal info of N_1
Step 29. Initialize model with N_1 and N_2
Step 30. Train Bi_ST model with N_1 and N_2
Step 31. Return(predicted value for C)
Step 32. end

The model is first trained with model parameters and then it is is tested for the given spatial and temporal inputs. The final results are evaluated in Section 6 (Results).

4. Dataset description

The data is collected from Central Pollution Control Board. The parameters chosen are Meteorological parameters (temperature, relative humidity) and Air Pollution parameters ($PM_{2.5}$, NO_2, SO_2, CO, O_3). The data collected from are the year 2015 to

2019 (five years) in the study area. The air quality data in this study have been extracted from Central Pollution Control Board CPCB for India and for China from Environmental Protection Administration and Central Weather Bureau in Taiwan. Table 1 gives the air monitoring sites in the dataset.

Table 1. Dataset

Locations in India	Locations in China
Tirupathi, Gaya, Velachery, Haldia, Agra, Kanpur, Hyderabad, Mumbai, Nagpur, BTM Layout, Kadabesanaalli, Anandh Vihar	Beijing, Chengdu, Guangzhou, Shanghai, Shenyang

5. Experimental work

To explore the efficiency of the proposed method, the input parameter of the neuron is taken in temporal and spatial relations. Different pollutants are given as an input to neurons. The two combinations of input such as temporal input and spatial input are given to a neuron. The prediction of $PM_{2.5}$ is the output of neuron. The model is trained with the number of spatial neighbors (s) as 2 and the temporal connections (t) as 4 for India and China.

The training of the neural network is done with the training set, which is partitioned into fixed size batches, and each batch is processed as a sequence in one training epoch. The geographical and the temporal neighbors are taken as s and t, respectively. The value of s is chosen according to the air stations location. In our experiment, $s \in \{1, 2\}$ and $t \in \{1, 2, 3, 4\}$, i.e., the numbers of spatial neighbors chosen are two nearby places and the numbers of temporal relations chosen are previous four days. Consider, the current day is taken as t and the remaining days are ($t - 1$, $t - 2$, $t - 3$, $t - 4$). From the experiment it is verified that, when the value of influential days is increased the performance of Bi_ST is reduced (when $t > 4$). Experimental results for the same are provided in table A3. The past information that is far from a particular day is influencing the model performance. Therefore, it is concluded that the influence of historical information diminishes over time. In our experiment, memory unit in LSTM, Bi-LSTM and Bi_ST Algorithm is given the input of past four days for training the algorithm in case of temporal relation and the number of neighbors 2(s1 and s2) in case of spatial relation. The number of cities taken as spatial neighbors depends on the dataset and also on the location of air stations. Rectification nonlinearity (ReLu) activation is applied to hidden layers. The model also uses dropout regularization. The use of activation function is that it normalizes the output and provides a continuous first order derivative. This is used in the back propagation process. The ReLu function crushes a large input space between 0 and ∞. Therefore, a significant change in the input causes a slight change in the output. Adam optimizer is used to train the model by minimizing the Root Mean Squared Error between the predicted and the ground truth-values.

Table 2. Training details of Bi_ST Algorithm

Parameters	India	China
Instances	17500	7300
Batch size	20	10
Drop out	0.3	0.3
Training epoch	50	30
Number of neighbors	2	2
Number of influential days	4	4
Number of input neurons	8	8
Number of hidden layers	3	3
Activation function	ReLu	ReLu
Number of output	1	1

6. Results

The results discussed in this section are examined to assess the predictability of the Bi_ST model. The stability of the model with respect to spatial and temporal aspects is explained. As 10-fold cross validation is used, the dataset is shuffled randomly such that the data is divided into 10 groups. In this approach the entire data is taken as training sample as well as for testing sample. This results in increasing the skill of training sample.

The further experiment is carried out for the neighbors and the temporal factors associated with the input model. Table A1 and table A2 show the results of the learning model with neighbors and temporal series respectively. The evaluation of the model has been done with the performance measures such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE).

6.1. Bi_ST model performance in case of spatial neighbors

The number of neighbors included for forecasting has influenced the accuracy rate. When we consider the spatial neighbors, it is important to note the location of the monitoring station also. According to dataset, spatial neighbors are chosen as 2. It may increase when the air monitoring stations are located nearby. In case of India there are only few cities, which have two or three stations nearby. However, in China, air monitoring stations are not located in a close distance. So, considering for both India and China and based on the geography of locations, neighbors are chosen as 2. If the distance is too far, then there is no meaning in including that station for prediction. Moreover, in some places the concentrations of $PM_{2.5}$ are three times higher than the prescribed amount of air pollution parameters by Central Pollution Control Board (CPCB). For example, the average concentration of $PM_{2.5}$ for the month of April 2019 in Anandh Vihar, Delhi is 272.38, which indicate that it has a hazardous air quality. The prescribed level if $PM_{2.5}$ in air is 60. Based on Pollution Control Board in India, the range of $PM_{2.5}$ falling under the different category is given in Table 3. Spatial patterns of air pollution are difficult to generalize for the entire city as the neighboring locations influence the pattern. The maximum of two neighbors are included for a prediction. Table 3 depicts the estimated measures for the cities in India and China. The regions of air pollution monitoring stations in China

are in a distance location, prediction is done based on the nearest two locations according to the dataset. The same is applicable for Indian regions where the RMSE and MAE values are reduced when compared with Bi-LSTM model. Error rate given by MAPE for Bi_ST model is reduced to 3-5% comparatively.

Table 3. Range of Pollutant quality

Pollutant quality (PM 2.5)	Range
Good	0-30
Satisfactory	31-60
Moderately Polluted	61-90
Poor	91-120
Severe	120 and above

6.2. Bi_ST model performance in case of temporal relations

Temporal inputs of the developed model are given in table A3. Temporal values are taken for past five days as t_1, t_2, t_3 and t_4 and t_5. The values from the past have the influence in predicting the current value. From the time series inputs given to the Bi_ST model, is noted that the error rate is increased when the memory of the LSTM is increased to t_5. It is examined that to a certain extent only, the past values can influence the current predictions. When $t > 4$ there is a deviation in results and error rate is increased. So in our model, the value of t is fixed to four.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(O_i - P_i)^2}\ , \tag{9}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\,O_i - P_i\,|, \tag{10}$$

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|\,O_i - P_i\,|}{O_i}, \tag{11}$$

where O_i is the observed air quality, P_i is the estimated air quality, and N denotes the number of data samples. The absolute error and relative error are measured by RMSE, MAE and MAPE, respectively. When MAPE is reduced, the optimal structure of the model is determined. Fig. 3 represents the predicted value for different t values. The predicted plot is for next 15 days. The different number of influential days has shown different values, in which the number of influential days with t+4 has given results that are close to the actual values. From this, it is concluded that the value of temporal days is restricted to four in this study. For the model with the predicted values, the performance measure graph is given in Fig. 4. For the various temporal values, the measure of RMSE has been reduced. The findings have concluded that, the improvement rates in RMSE notably increase from t_1 to t_4 for all the inputs. However, for t_5, there is a significant increase in RMSE and MAPE values. The increase in RMSE and MAPE is seen for both India and China. As so the time series input in fitted to t_4 and t_5 is not taken into account.

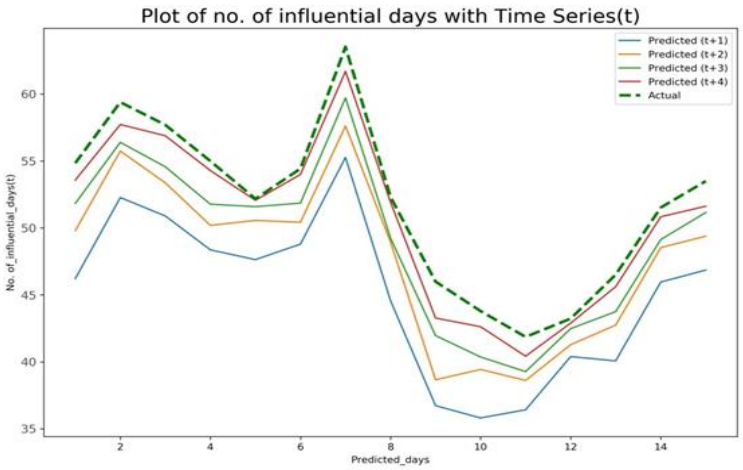

Fig. 3. Plot of predicted value of $PM_{2.5}$ with different number of influential days for the city of Agra

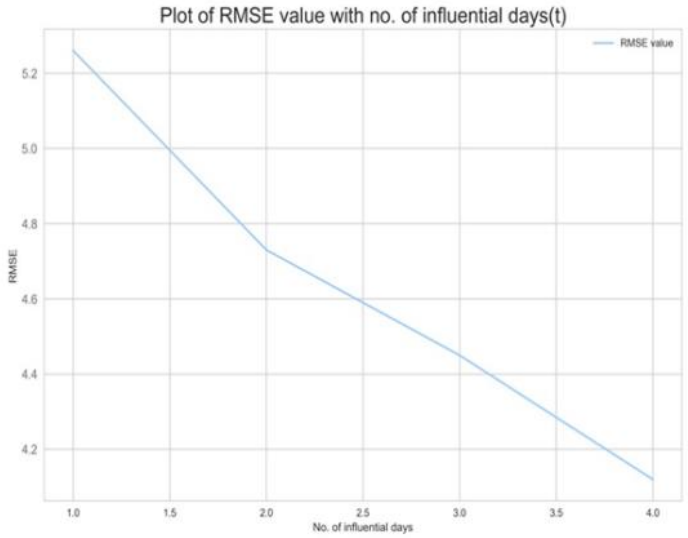

Fig. 4. Plot of RMSE values for different (*t*) days

The developed model has produced the best results in the performance of all measures with the improvement in RMSE values and the percentage of error rate measured with MAPE is also reduced. The error rate is also comparatively reduced with the inclusion of Spatial and Temporal relations in the input neuron. The method has been applied for the 17 cities included in the study. Each city has been tested with the spatial and temporal inputs and the results are recorded in table A1 and table A2. An example for the predicted $PM_{2.5}$ for the city of Agra is given in Fig. 5, Fig. 6 and Fig. 7 for RNN model, LSTM model and Bi_ST model respectively. Fig. 4 depicts the error rate in prediction and it has consecutively decreased for t+4 days of prediction. The observed and predicted $PM_{2.5}$ values for the city of Agra are plotted in Fig. 3 with different influential t values.

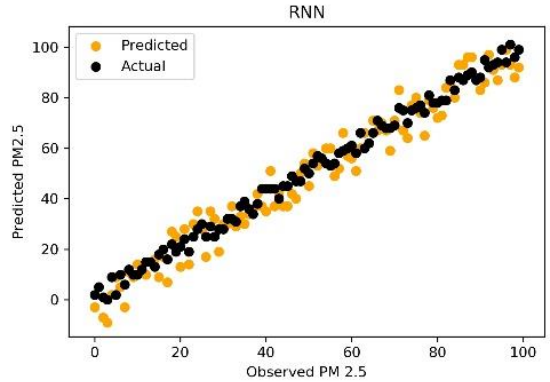

Fig. 5. Sample prediction of RNN for city Agra

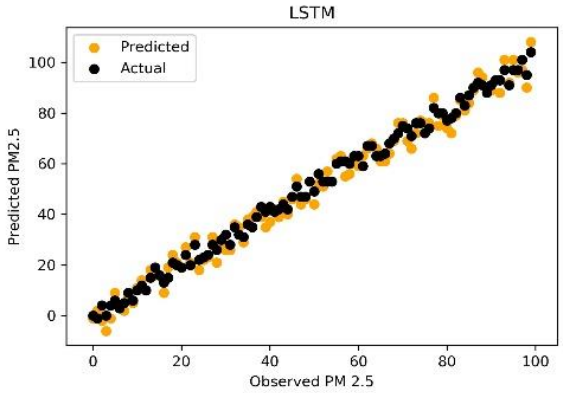

Fig. 6. Sample prediction of Bi-LSTM for city Agra

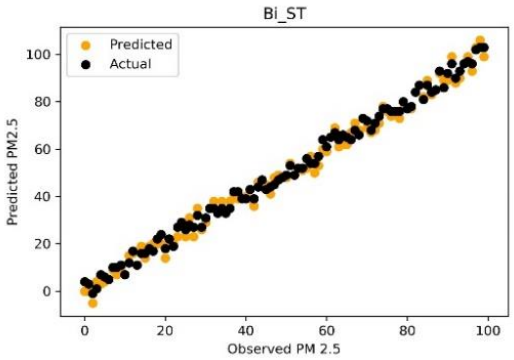

Fig. 7. Sample prediction of Bi_ST for city Agra

Fig. 8 shows the actual value recorded in dataset. The actual data recorded are for the cities Beijing, Chengdu, Guang Zhou, Shanghai and Shenyang. The city of Beijing has shown the difference in its predicted value. The predicted value from the Bi_ST model is given in Fig. 9. The predictions are almost similar to that of actual value except with the city of Beijing. Colors represent the severity of the pollution. Green, Orange and Red are labelled as good, moderate and hazardous.

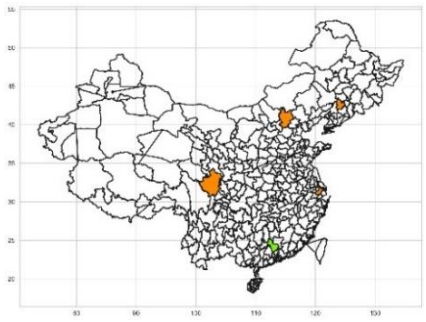

Fig. 8. Actual value for *h*+1 in China for different cities

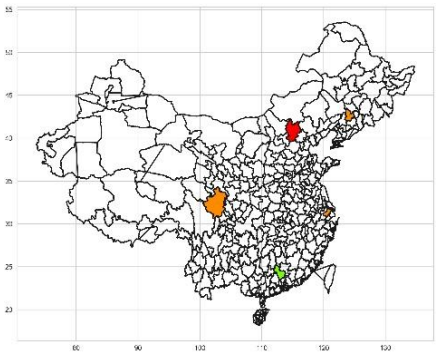

Fig. 9. Predicted value for *h*+1 in China for different cities

Fig. 10 is the actual value recorded in dataset. The actual data recorded for the cities Tirupathi, Gaya, Velachery, Haldia, Agra, Kanpur, Hyderabad, Mumbai, Nagpur, BTM Layout, Kadabesanaalli, Anandh Vihar. Few cities have shown the difference in its predicted value. The predicted value from the Bi_ST model is given in Fig. 11. The predicted values vary for BTM Layout, Agra, Haldia and Kanpur. Colors represent the severity of the pollution. Green, Orange and Red are labelled as good, moderate and hazardous.

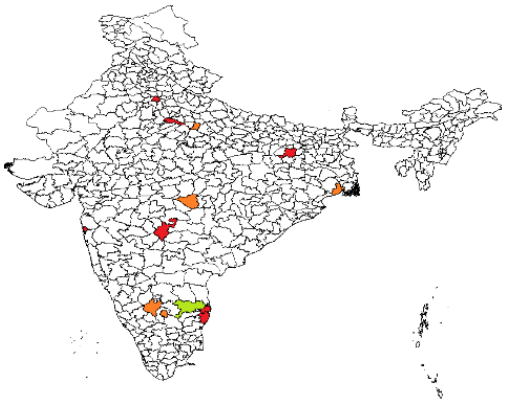

Fig. 10. Actual value for *h*+1 in India for different cities

Table 8. Model comparison with RNN, LSTM and Bi_ST

Method	Dataset	Accuracy	Precision	Recall	F1-Score
RNN	India	85.34	87.58	88.37	87.97
	China	86.49	86.34	87.24	86.76
LSTM	India	90.28	91.36	92.17	92.21
	China	89.25	90.24	89.31	89.83
Bi_ST	India	**93.76**	92.83	91.72	93.27
	China	**94.08**	91.62	90.33	90.97

Model with spatial and temporal inputs have given better accuracy of 93.76 and 94.08 for India and China, respectively. With the developed model, the prediction is given for the year 2020. Air quality for the year 2020 for India and China is mapped in Fig. 12 and Fig. 13, respectively.

The color representation says the range or severity of the air pollution in that particular place or region. Predictions are done for the future using bidirectional.

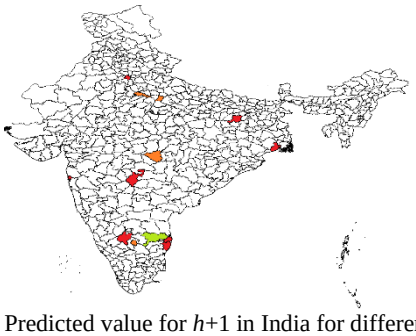

Fig. 11. Predicted value for h+1 in India for different cities

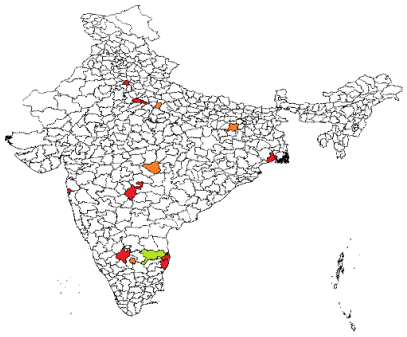

Fig. 12. Predicted air quality of India for the year 2020

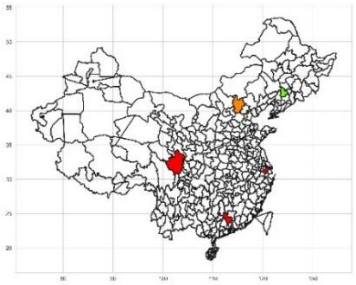

Fig. 13. Predicted air quality of China for the year 2020

Long Short Term Memory with Spatial and Temporal interactions model is given in Figs 12 and 13 for India and China, respectively. Cities with red color are in a hazardous range of air quality, orange represents moderate condition of air, and green represents the good air quality.

7. Conclusion

People living in developed cities have a greater possibility of exposure to air pollutants. From this prediction, we can conclude that exposure to air pollution from various locations can be identified. The solutions provided for environmental issues are crucial in controlling air pollution and can have significant societal and technical impacts. The method of including spatial neighbors and temporal relations for predicting the nearby values with the Bi_ST model is superior with the other two methods. The neuron, which learns both the past and future values of information and acts in a forward and backward direction, is a better way, suitable for the time series model. For India and China, the prediction rate is increased and the error rate is reduced to 3 to 5 %. The subject area chosen is much important for the Public Health Control on account for the measures on the population who is suffering from allergies and respiratory diseases.

References

1. Lin, B., J. Zhu. Changes in Urban Air Quality during Urbanization in China. – J. Clean. Prod., Vol. **188**, 2018, pp. 312-321.
2. Li, L., et al. Evaluation of Future Energy Consumption on PM2.5 Emissions and Public Health Economic Loss in Beijing. – J. Clean. Prod., Vol. **187**, 2018, pp. 1115-1128.
3. Li, N., et al. Potential Impacts of Electric Vehicles on Air Quality in Taiwan. – Sci. Total Environ., Vol. **566-567**, 2016, pp. 919-928.
4. Wang, Y., M. Sun, X. Yang, X. Yuan. Public Awareness and Willingness to Pay for Tackling Smog Pollution in China: A Case Study. – J. Clean. Prod., Vol. **112**, 2016., pp. 1627-1634
5. Kurt, A., B. Gulbagci, F. Karaca, O. Alagha. An Online Air Pollution Forecasting System Using Neural Networks. – Environ. Int., Vol. **34**, 2008, No 5, pp. 592-598.
6. Moisan, S., R. Herrera, A. Clements. A Dynamic Multiple Equation Approach for Forecasting PM2.5 Pollution in Santiago, Chile. – Int. J. Forecast., Vol. **34**, 2018, No 4, pp. 566-581.
7. Jiang, P., R. Li, K. Zhang. Two Combined Forecasting Models Based on Singular Spectrum Analysis and Intelligent Optimized Algorithm for Short-Term Wind Speed. – Neural Comput. Appl., Vol. **30**, 2018, No 1.
8. Feng, Y., W. Zhang, D. Sun, L. Zhang. Ozone Concentration Forecast Method Based on Genetic Algorithm Optimized Back Propagation Neural Networks and Support Vector Machine Data Classification. – Atmos. Environ., Vol. **45**, 2011, No 11, pp. 1979-1985.
9. Paschalidou, A. K., S. Karakitsios, S. Kleanthous, P. A. Kassomenos. Forecasting Hourly PM10 Concentration in Cyprus through Artificial Neural Networks and Multiple Regression Models: Implications to Local Environmental Management. – Environ. Sci. Pollut. Res., Vol. **18**, 2011, No 2, pp. 316-327.
10. Antanasijević, D. Z., M. D. Ristić, A. A. Perić-Grujić, V. V. Pocajt. Forecasting Human Exposure to PM10 at the National Level Using an Artificial Neural Network Approach. – J. Chemom., Vol. **27**, 2013, No 6, pp. 170-177.
11. Wu, S., Q. Feng , Y. Du, X. Li. Artificial Neural Network Models for Daily PM10 Air Pollution Index Prediction in the Urban Area of Wuhan, China. – Environ. Eng. Sci., Vol. **28**, 2011, No 5, pp. 357-363.

12. P a i, T. Y., K. H a n a k i, H. C. S u, L. F. Y u. A 24-h Forecast of Oxidant Concentration in Tokyo Using Neural Network and Fuzzy Learning Approach. – Clean – Soil, Air, Water, Vol. **41**, 2013, No 8, pp. 729-736.
13. B r a b h u k u m r, A., P. M a l h i, K. R a v i n d r a, P. V. M. L a k s h m i. Exposure to Household Air Pollution during First 3 Years of Life and IQ Level Among 6-8-Year-Old Children in India – A Cross-Sectional Study. – Sci. Total Environ., Vol. **709**, 2020, p. 135110.
14. B a l a k r i s h n a n, K., et al. The Impact of Air Pollution on Deaths, Disease Burden, and Life Expectancy across the States of India: The Global Burden of Disease Study 2017. – Lancet Planet. Heal., Vol. **3**, 2019, No 1, pp. e26-e39.
15. P a n d e y, V., E. O k s a n e n, N. S i n g h, C. S h a r m a. Impacts of Air Pollution and Climate Change on Plants: Implications for India. 1st Ed. Vol. **13**. Elsevier Ltd., 2013.
16. S a h a, D. C., P. K. P a d h y. Effect of Air and Noise Pollution on Species Diversity and Population Density of Forest Birds at Lalpahari, West Bengal, India. – Sci. Total Environ., Vol. **409**, 2011, No 24, pp. 5328-5336.
17. B a r o t, V., V. K a p a d i a, S. P a n d y a. QoS Enabled IoT Based Low Cost Air Quality Monitoring System with Power Consumption Optimization. – Cybernetics and Information Technologies, Vol. **20**, 2020, No 2, pp. 122-140.
18. G o c h e v a-I l i e v a, S. G., A. V. I v a n o v, I. E. L i v i e r i s. High Performance Machine Learning Models of Large Scale Air Pollution Data in Urban Area. – Cybernetics and Information Technologies, Vol. **20**, 2020, No 6, pp. 49-60.
19. S h a r m a, N., S. T a n e j a, V. S a g a r, A. B h a t t. Forecasting Air Pollution Load in Delhi Using Data Analysis Tools. – Procedia Comput. Sci., Vol. **13**2, 2018, pp. 1077-1085.
20. G u p t a, P., S. A. C h r i s t o p h e r. Particulate Matter Air Quality Assessment Using Integrated Surface, Satellite, and Meteorological Products: Multiple Regression Approach. – J. Geophys. Res. Atmos., Vol. **114**, 2009, No 14, pp. 1-13.
21. W a n g, P., H. Z h a n g, Z. Q i n, G. Z h a n g. A Novel Hybrid-Garch Model Based on ARIMA and SVM for PM2.5 Concentrations Forecasting. – Atmos. Pollut. Res., Vol. **8**, 2017, No 5, pp. 850-860.
22. N i, X. Y., H. H u a n g, W. P. D u. Relevance Analysis and Short-Term Prediction of PM2.5 Concentrations in Beijing Based on Multi-Source Data. – Atmos. Environ., Vol. **150**, 2017, No February 2017, pp. 146-161.
23. G a r d n e r, M. W., S. R. D o r l i n g. Artificial Neural Networks (the Multilayer Perceptron) – A Review of Applications in the Atmospheric Sciences. – Atmos. Environ., Vol. **32**, 1998, No 14-15, pp. 2627-2636.
24. G r i v a s, G., A. C h a l o u l a k o u. Artificial Neural Network Models for Prediction of PM10 Hourly Concentrations, in the Greater Area of Athens, Greece. – Atmos. Environ., Vol. **40**, 2006, No 7, pp. 1216-1229.
25. I g l e s i a s-O t e r o, M. A., M. F e r n á n d e z-G o n z á l e z, D. R o d r í g u e z-C a r i d e, G. A s t r a y, J. C. M e j u t o, F. J. R o d r í g u e z-R a j o. A Model to Forecast the Risk Periods of Plantago Pollen Allergy by Using the ANN Methodology. – Aerobiologia (Bologna)., Vol. **31**, 2015, No 2, pp. 201-211.
26. L i, Y., P. J i a n g, Q. S h e, G. L i n. Research on Air Pollutant Concentration Prediction Method Based on Self-Adaptive Neuro-Fuzzy Weighted Extreme Learning Machine. – Environ. Pollut., Vol. **241**, 2018, pp. 1115-1127.
27. A l i m i s s i s, A., K. P h i l i p p o p o u l o s, C. G. T z a n i s, D. D e l i g i o r g i. Spatial Estimation of Urban Air Pollution with the Use of Artificial Neural Network Models. – Atmos. Environ., Vol. **191**, 2018, pp. 205-213.
28. Y a n g, Z., J. W a n g. A New Air Quality Monitoring and Early Warning System: Air Quality Assessment and Air Pollutant Concentration Prediction. – Environ. Res., Vol. **158**, 2017, No May, pp. 105-117.
29. N i s k a, H., T. H i l t u n e n, A. K a r p p i n e n, J. R u u s k a n e n, M. K o l e h m a i n e n. Evolving the Neural Network Model for Forecasting Air Pollution Time Series. – Eng. Appl. Artif. Intell., Vol. **17**, 2004, No 2, pp. 159-167.
30. F u, R., Z. Z h a n g, L. L i. Using LSTM and GRU Neural Network Methods for Traffic Flow Prediction. – In: Proc. of 31st Youth Acad. Annu. Conf. Chinese Assoc. Autom. (YAC'16), No December, 2017, pp. 324-328.

31. S h i, X., Z. C h e n, H. W a n g. Convolutional LSTM Network. – Nips, 2015, pp. 2-3.
32. Z h a o, J., F. D e n g, Y. C a i, J. C h e n. Long Short-Term Memory – Fully Connected (LSTM-FC) Neural Network for PM2.5 Concentration Prediction. – Chemosphere, Vol. **220**, 2019, pp. 486-492.
33. T o n g, W., L. L i, X. Z h o u, A. H a m i l t o n, K. Z h a n g. Deep Learning PM2.5 Concentrations with Bidirectional LSTM RNN. – Air Qual. Atmos. Heal., Vol. **12**, 2019, No 4, pp. 411-423.
34. Q i, Z., T. W a n g, G. S o n g, W. H u, X. L i, Z. Z h a n g. Deep Air Learning: Interpolation, Prediction, and Feature Analysis of Fine-Grained Air Quality. – IEEE Trans. Knowl. Data Eng., Vol. **30**, 2018, No 12, pp. 2285-2297.
35. X u, C., L. X i e, X. X i a o. A Bidirectional LSTM Approach with Word Embeddings for Sentence Boundary Detection. – J. Signal Process. Syst., Vol. **90**, 2018, No 7, pp. 1063-1075.
36. L i n, B. Y., F. X u, Z. L u o, K. Z h u. Multi-Channel BiLSTM-CRF Model for Emerging Named Entity Recognition in Social Media. – In: Proc. of 3rd Workshop on Noisy User-Generated Text, September 2017, pp. 160-165.
37. B a r o t, V., V. K a p a d i a. Long Short Term Memory Neural Network-Based Model Construction and Fne-Tuning for Air Quality Parameters Prediction. – Cybernetics and Information Technologies, Vol. **22**, 2022, No 1, pp. 171-189.
38. I k r a m , S. T., A. K. C h e r u k u r i, B. P o o r v a, P. S. U s h a s r e e, Y. Z h a n g, X. L i u, G. L i. Anomaly Detection Using XGBoost Ensemble of Deep Neural Network Models. – Cybernetics and Information Technologies, Vol. **21**, 2021, No 3, pp. 175-188.
39. H o c h r e i t e r, S., J. U r g e n S c h m i d h u b e r. Long Shortterm Memory. – Neural Comput., Vol. **9**, 1997, No 8, 17351780.
40. G r a v e s, A., J. S c h m i d h u b e r. Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures. – Neural Networks, Vol. **18**, 2005, No 5-6, pp. 602-610.
41. Z h a n g, B., H. Z h a n g, G. Z h a o, J. L i a n. Constructing a PM2.5 Concentration Prediction Model by Combining Auto-Encoder with Bi-LSTM Neural Networks. – Environmental Modelling & Software, Vol. **124**, 2020, p. 104600.
42. T o b l e r, A. W. R. Clark University. – Science (80), Vol. **13**, 1889, No 332, pp. 462-465.

Appendix A. First appendix

Table A1. Bi_ST Performance measure for the nearest two neighbours as S1 and S2

Region	RMSE		MAE		MAPE	
	S=1	S=2	S=1	S=2	S=1	S=2
Tirupathi	3.76	3.52	1.28	1.67	18.17%	17.69%
Gaya	4.23	3.96	2.01	1.82	17.56%	16.23%
Velachery	4.17	3.67	2.19	1.06	16.93%	18.27%
Haldia	4.68	4.27	2.58	2.86	16.06%	19.27%
Agra	5.64	5.41	3.71	3.26	19.84%	18.77%
Kanpur	4.71	3.82	2.49	2.11	19.81%	18.37%
Hyderabad	3.94	3.15	1.86	1.75	19.74%	19.51%
Mumbai	3.72	3.44	2.14	1.76	18.59%	17.73%
Nagpur	3.68	3.87	1.61	1.73	19.67%	18.37%
BTM Layout	4.19	3.88	2.71	2.19	18.78%	16.49%
Kadabesanaalli	4.94	4.77	2.18	2.06	19.78%	17.46%
Anandh Vihar	3.76	3.62	1.83	1.74	17.74%	17.27%
Beijing	3.67	3.51	1.49	1.34	18.51%	18.06%
Chengdu	3.46	3.73	1.48	1.75	18.45%	17.94%
Guangzhou	4.71	4.29	2.41	2.18	16.56%	16.42%
Shanghai	3.91	3.67	1.67	1.54	16.75%	15.84%
Shenyang	4.51	4.29	2.34	2.67	17.64%	17.24%

Table A2. Bi_ST Algorithm with different number of influential days

Region	Temporal relation	RMSE	MAE	MAPE
Tirupathi	*t*=1	3.76	1.28	19.17%
	t=2	3.52	1.67	17.69%
	t=3	3.29	1.23	18.40%
	t=4	3.16	1.17	17.22%
	***t*=5**	3.94	2.12	19.23%
Gaya	*t*=1	3.17	1.08	16.37%
	t=2	3.67	1.27	17.82%
	t=3	2.83	1.63	17.82%
	t=4	2.56	1.37	16.75%
	***t*=5**	3.59	2.46	18.49%
Agra	*t*=1	3.41	1.37	16.67%
	t=2	3.69	1.64	16.32%
	t=3	3.16	1.41	15.66%
	t=4	3.66	1.27	15.71%
	***t*=5**	4.87	2.69	20.49%
	***t*=5**	4.71	2.62	19.62%
Beijing	*t*=1	3.06	1.69	16.41%
	t=2	3.17	1.56	16.84%
	t=3	2.61	1.59	16.67%
	t=4	2.54	1.12	15.94%
	***t*=5**	5.41	3.53	19.26%
Chengdu	*t*=1	2.59	1.65	16.35%
	t=2	2.84	1.51	16.48%
	t=3	2.46	1.59	16.74%
	t=4	2.37	1.27	16.69%
	t=5	4.61	2.93	19.65%

Received: 20.03.2023; Accepted: 28.08.2023