



Prediction and assessment of the impact of COVID-19 lockdown on air quality over Kolkata: a deep transfer learning approach

Debashree Dutta · Sankar K. Pal

Received: 25 March 2022 / Accepted: 12 November 2022
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

Abstract The present study focuses on the prediction and assessment of the impact of lockdown because of coronavirus pandemic on the air quality during three different phases, viz., normal periods (1 January 2018–23 March 2020), complete lockdown (24 March 2020–31 May 2020), and partial lockdown (1 June 2020–30 September 2020). We identify the most important air pollutants influencing the air quality of Kolkata during three different periods using Random Forest, a tree-based machine learning (ML) algorithm. It is found that the ambient air quality of Kolkata is mainly affected with the aid of particulate matter or PM (PM_{10} and $PM_{2.5}$). However, the effect of the lockdown is most prominent on $PM_{2.5}$ which spreads in the air of Kolkata due to diesel-driven vehicles, domestic and commercial combustion activities, road dust, and open burning. To predict urban $PM_{2.5}$ and PM_{10} concentrations 24 h in advance, we use a deep learning (DL) model, namely, stacked-bidirectional long short-term memory (stacked-BDLSTM). The model is trained during the normal periods, and it shows the superiority over some supervised ML models, like support vector machine, K-nearest neighbor classifier, multilayer perceptron, long short-term memory, and statistical time series forecasting model autoregressive integrated moving

average. This pre-trained stacked-BDLSTM is applied to predict the concentrations of $PM_{2.5}$ and PM_{10} during the pandemic situation of two cases, viz., complete lockdown and partial lockdown using a deep model-based transfer learning (TL) approach (TLS-BDLSTM). Transfer learning aims to utilize the information gained from one problem to improve the predictive performance of a learning model for a different but related problem. Our work helps to demonstrate how TL is useful when there is a scarcity of data during the COVID-19 pandemic regarding the drastic change in concentration of pollutants. The results reveal the best prediction performance of TLS-BDLSTM with a lead time of 24 h as compared to some well-known traditional ML and statistical models and the pre-trained stacked-BDLSTM. The prediction is then validated using the real-time data obtained during the complete lockdown due to COVID second wave (16 May–15 June 2021) with different time steps, e.g., 24 h, 48 h, 72 h, and 96–120 h. TLS-BDLSTM involving transfer learning is seen to outperform the said comparing methods in modeling the long-term temporal dependency of multivariate time series data and boost the forecast efficiency not only in single step, but also in multiple steps. The proposed methodologies are effective, consistent, and can be used by operational organizations to utilize in monitoring and management of air quality.

D. Dutta (✉) · S. K. Pal
Center for Soft Computing Research, Indian Statistical
Institute, Kolkata 700108, India
e-mail: debashree.120@gmail.com

S. K. Pal
e-mail: sankarpal@yahoo.com

Keywords Air quality · Bidirectional LSTM · COVID-19 · Deep learning · Lockdown · $PM_{2.5}$ · PM_{10} · Random forest · Transfer learning

Introduction

The problem of urban air pollution has become increasingly significant with the rapid growth of urbanization and the advancement of industrialization which not only speeds up climate change but also affects human health. For more than 1 year, the human population of the world confronted a challenging new pandemic COVID-19 caused by the new coronavirus SARS-CoV-2. The coronavirus disease (COVID-19) spread rapidly around the world in early 2020, changing anthropogenic activities permanently (Kharroubi & Saleh, 2020). India's first coronavirus case was found in Kerala on 30 January 2020. To control the COVID-19 outbreak in India, the Government of India (GoI) declared a complete lockdown, and social distancing measures were initiated from 24 March 2020. The complete lockdown measures included the suspension of all activities in the non-essential industrial production, transportation sectors, supplies trading except a few essential industries (viz., manufacturing units of drugs, pharmaceuticals, petroleum refineries, fertilizers, food-related industries, power plants) and essential transportation. Academic, cultural, administrative, political, and religious gatherings remain prohibited. The State Government of Kolkata, West Bengal, extended complete lockdown until 5 May 2020 and extended it up to 31 May 2020. After complete lockdown, the government declares partial lockdown measures targeting non-essential activities amid restrictions and a lesser number of public vehicles. This is the stage of the step-by-step unlock period with maintained proper COVID guidelines extended up to 30 September 2020 over Kolkata. From October 2020 to mid-February 2021, India experienced a decrease in new confirmed COVID-19 infections. But India was in the grips of a devastating second wave of the virus during mid-April due to the highly infectious double mutant variant of SARS-CoV-2 (B.1.617 lineage), and cities were facing fresh lockdown 2.0. In view of the rising COVID-19 cases, the State Government of Kolkata, West Bengal government imposed a complete lockdown from 16 May 2021 to 15 June 2021.

Pollution is the leading environmental cause of morbidity and mortality worldwide. It is triggered by human activities that disrupt environmental balance. Heavy metals are critical components of air pollution because they can be poisonous and deadly even at low concentration, and even necessary elements for organisms can be harmful at high concentrations (Dutta &

Pal, 2022b; Jo et al., 2020). Sevik et al. (2020) determined the variation of Pb and Mg accumulation depending on plant species, plant organ, washing status, and traffic density in some landscape plants grown in the city center of Kastamonu. Ba is one of the most hazardous heavy metals, and it was found that Ba concentrations increased with traffic density in nearly all plant organs (Bayraktar et al., 2022; Cetin & Jawed, 2022). Cetin et al. (2022) identified and mapped the Pb and Cr pollution in the city center of Ankara. Therefore, monitoring the alteration in heavy metal concentrations is critical for the health of both human and environment. One of the most reliable methods is the use of annual rings of trees. Cesur et al. (2022) illustrated the utility of annual rings of *Cupressus arizonica* as a biomonitor in order to indicate the change of Li, Fe, and Cr concentrations. Air pollution is commonly recognized to be a matter of concern to both public wellness and economic development. Cetin and Sevik (2016) observed the change of the air quality at various points in the Kastamonu city center in terms of particulate matter and CO₂ amount. Cetin et al. (2019) assessed the temporal and regional variation of some air pollution parameters in different areas of Bursa's city center. It has been founded that indoor air quality has a direct impact on students' working capacity and, as a consequence, academic performance (Cetin, 2016; Stabile et al., 2017). Elsunousi et al. (2021) evaluated the regional and periodic change of CO₂ and particulate matter pollution in the city of Misurata, one of the major cities of Libya.

The most serious challenge in the twenty-first century is the degradation of air quality around the world due to different sorts of anthropogenic interventions (Motesaddi et al., 2017). The global response to alleviate the pandemic of COVID-19 has led to an unexpected turndown of most of the emission sources. Several researchers focusing on monitoring urban air quality during the COVID-19 lockdown have recorded significant reductions in air pollutant concentrations around the world, which have been linked to lower emissions from anthropogenic activities (Wang & Su, 2020; Zambrano-Monserrate et al., 2020; Bauwens et al., 2020; Ding et al., 2020; Shi & Brasseur, 2020; Wang et al., 2020; Goldberg et al., 2020; Berman & Ebisu, 2020). In India, studies on air quality have shown how the lockdown has improved the ambient air quality and reduced the rate of selective air pollutants during the lockdown period (Gautam, 2020; Mahato et al., 2020;

Navinya et al., 2020; Sarkar et al., 2021). According to the Central Pollution Control Board (CPCB), after the first 4 days of lockdown, air pollution levels in 88 Indian cities decreased drastically (Sharma et al., 2020). The degree of air pollutants reduction may vary from city to city because of local factors such as the restriction of the lockdown, distribution of emission sources, meteorological fluctuations, and trends in pollutant emissions (Kumar et al., 2020). Though most of the latest research have studied the change of air quality standard, effect on the environment, and causes of reduction in pollution concentrations during lockdown compared to the pre-lockdown phase, air quality forecasting during pandemic remains a challenge for researchers.

Due to the dynamic nature, volatility, and high variability in the time and space of pollutants, predicting air quality is a difficult task (Castelli et al., 2020). The traditional deterministic methods were usually designed under specific assumptions, and they cannot fit other real-world conditions (Ma et al., 2019a). Statistical approaches, as opposed to deterministic approaches, can be used in a wider range of situations as long as enough data is provided. Traditional statistical approaches, on the other hand, are based on the linear assumption, which is incompatible with the non-linear properties of the real world, thereby limiting their performance. Recently, deep learning (DL), being an advanced form of machine learning (ML), has dramatically improved the state of the art in big data analysis, such as in video analytics, speech analysis, bioinformatics, and remote sensing data. Deep learning means learning in depth in different stages. Here, the learning performance depends on the number of samples (or previous experiences). The larger the number is, the better the learning performance. Deep learning (DL) technique has been applied for a wide variety of problems such as object detection, weather prediction, motion modeling, image classification, natural language processing, speech recognition, and synthesis (Chakraborty & Pal, 2021; Dutta & Pal, 2022a; Hinton et al., 2012; Krizhevsky et al., 2017; LeCun et al., 2015; Pal et al., 2021; Young et al., 2018; Zhang et al., 2013). Since the air pollution data is big in nature, the use of a DL-based data-driven model in conjunction with advanced artificial intelligence (AI) tools for accurate representation and prediction of the air quality depending on weather and several factors appears to be logical and appropriate.

Artificial intelligence (AI) deals with simulation of human intelligence processes by a computer or a

machine. AI has provided some essential tools for local environmental protection agencies to use in making informed decisions about air pollution mitigation measures in order to reduce the public exposure risk (Jerrett et al., 2005; Masood & Ahmad, 2021). For example, many researchers used artificial neural networks (ANNs) for prediction of air quality that include multilayer perceptron (MLP), back-propagation neural network (BPNN), and adaptive neuro-fuzzy inference systems (ANFIS) (Prasad et al., 2016; Tatavarti et al., 2018; Turias et al., 2008). Pollutant and particulate concentrations are predicted using support vector regression (SVR) models (Lei et al., 2022; Liu et al., 2017). Dutta and Pal (2022b) have recently used a rough set theory-based decision support system to predict the different categories of AQI and developed a Z-number-based novel quantification measure of semantic information of AQI to assess the reliability of the outcomes.

Air pollution can have a short- or long-term impact on future conditions that may persist for hours, days, or even weeks. As a result, when forecasting the air quality, the time transit needs to be taken into account. However, most of the artificial neural network (ANN)-based approaches fail to improve the temporal lag of air pollution or develop long-term dependencies. Some researchers used advanced deep learning techniques such as recurrent neural network (RNN), long short-term memory (LSTM), long short-term memory-fully connected neural network (LSTM-FC), K-nearest neighbor, and long short-term memory (KNN-LSTM) to model the time series data to address this challenge (Ong et al., 2015; Qi et al., 2018; Tao et al., 2019; Zhao et al., 2019; Qin et al., 2019; Heydari et al., 2021).

One of the recent advances in AI and deep learning is transfer learning (TL). TL means reusing a model already developed (i.e., knowledge gained) for a task as the starting point (say, previous or source task) to a different but related task (say, target task); thereby, improving the learning and prediction processes in the target task. For pattern classification, this implies an enrichment of the generalization-capability of the latter. TL differs from traditional machine learning algorithms which attempt to learn each task from scratch. TL method has been useful in transferring knowledge from previous tasks to a target task when the latter has fewer training data (Pan & Yang, 2010). TL is widely used in the domains of image classification (Dai et al., 2007), speech and natural language processing (Bel et al., 2003; Blitzer et al., 2006; Ling et al., 2008),

building utilization (Arief-Ang et al., 2018), and neurophysiological studies (Atyabi et al., 2013; Tu & Sun, 2012). To enhance the forecast accuracy, TL has also been used in atmospheric research (Tariq et al., 2021). It is widely used in prediction of air pollutants, especially in case of scarcity of data, where transferring the knowledge learned on the pre-trained model improves the prediction accuracy of the model. For example, deep learning-based stacked-bidirectional long short-term memory was applied to transfer the knowledge learned from smaller temporal resolutions to larger temporal resolution for the prediction of air quality in China where the performance of the model improves in greater time-based resolutions (Ma et al., 2019b). Ma et al. (2020) proposed a TL-based stacked-bidirectional long short-term memory (TLS-BLSTM) network that helps to improve the forecast accuracy by transferring knowledge from existing air quality stations to new stations. Fong et al. (2020) used LSTM recurrent neural networks (RNNs) to forecast the future concentration of air pollutants in Macau, having less observed data in terms of quantity and type in some air quality monitoring stations. In order to mitigate the issue of scarcity of data, Dhole et al. (2021) proposed an ensemble approach for multi-source transfer learning. This provides a cumulative prediction by transferring the knowledge learned from multiple source stations to a given target station, allowing better utilization of the data easily available from neighboring stations to improve the performance of the prediction. Gilik et al. (2022) developed a supervised model for predicting the air quality by utilizing real sensor data and transferring the model between cities.

The present study demonstrates the effectiveness of a deep model-based transfer learning approach in air quality modeling. During COVID-19 pandemic, the concentration of pollutants changed drastically as compared to the normal period. The nature and size of the data thus obtained during these two periods are different, the size being much larger in the case of normal. Therefore, to analyze the data during the pandemic, it may be logical to adopt the concept of transfer learning, whereby the knowledge gained from the normal period through training a deep network can be reused in designing an architecture appropriate for the data obtained due to pandemic situation. We propose a simple method that comprises the application of machine learning (viz., Random Forest), deep learning (viz., stacked-bidirectional LSTM), and transfer learning. The study period considered for air

quality prediction consists of three phases: normal periods (1 January 2018–23 March 2020), complete lockdown (24 March–31 May 2020), and partial lockdown (1 June–30 September 2020). We first identify the most important air pollutants (among nitrogen dioxide (NO₂), sulfur dioxide (SO₂), particulate matter or PM (PM₁₀ and PM_{2.5}), carbon monoxide (CO), ozone (O₃), and benzene) influencing the air quality of Kolkata during normal periods, complete lockdown, and partial lockdown using Random Forest algorithm. Then the prediction of concentration of the aforesaid major air pollutants is done. Here, we use a deep learning model, namely, stacked-bidirectional LSTM, which is learned from the normal periods. Afterwards, we adopt a transfer learning approach on the pre-trained deep model, stacked-bidirectional LSTM (stacked-BDLSTM), to predict the concentration during the pandemic situation, where the knowledge gathered from the normal period is transferred to build the final deep transfer learning model (TLS-BDLSTM). Finally, the prediction capability of the model, thus developed, is validated using the real-time data obtained during the lockdown due to COVID second wave (16 May–15 June 2021). The performance and feasibility of the proposed model are evaluated by comparing its output with those of stacked-bidirectional LSTM and supervised ML and statistical models in the prediction of air quality.

The novelty of the research is that it focuses on the following:

- Providing a new transfer learning-based framework that uses deep learning (DL) model, which is learned from the normal periods, and then incorporates the noble transfer learning (TL) approach to predict the concentrations of most dominant air pollutant in the pandemic situation with a small and complex dataset.
- Identification of most dominant air pollutant and assessment of effect of COVID-19 lockdown on it using Random Forest, a tree-based machine learning algorithm.
- Finally, validation of the proposed framework with real-time data of complete lockdown due to the COVID second wave 2021 and assessment of its performance with multi-step prediction.

Thus, the identification, assessment, and prediction of air quality over Kolkata were achieved using meteorological and air quality data, a user-friendly computational technique, and a deep transfer learning approach.

The remainder of the paper is structured as follows: The “**Materials and methods**” section presents an overview of the study area, data collection, detailed architectures, and algorithms of the proposed methods. It also includes the way of execution/implementation of algorithms with performance measures. In the “**Results and discussion**” section, experimental results are analyzed and discussed. Finally, the “**Conclusions**” section illustrates the concluding remarks of this investigation and proposes suggestions for future research.

Materials and methods

Study area

Kolkata (formerly Calcutta) is one of India’s major metropolitan cities. It is located in eastern India in the Ganges Delta at 22°33’N and 88°20’E, along the east bank of the Hooghly River at an elevation of about 9 m. Kolkata has a tropical wet and dry climate with summer monsoon, and wind movement in Kolkata is from SW direction; however, in winter, the calm wind is blowing from N and NE directions. Kolkata is a densely populated city with a population of 14.4 million. It is enlisted in the World’s twenty five most polluted cities along with ten worse polluted cities in India (Bera et al., 2020) facing rising air pollution and multi-pollutant crisis. The city’s economic and industrial expansion and different industries such as paper and pulp, organic and inorganic chemical industries, rubber, iron, plastic, textile and food, vehicular emissions, dust from construction sites, solid waste burning, and wind-blown dust from open lands besides thermal power plants are significantly contributing to air pollution over Kolkata. A study conducted at major traffic junctions of Kolkata showed that the key pollutants like lead, NO_x, PM₁₀, SO₂, and CO are in excess of permitted levels (Ghose et al., 2004).

Data collection

This study has been carried out with 4 years of data and observations from 2018 to 2021 over Kolkata, India. The data of daily air quality of Kolkata city are collected from the West Bengal Pollution Control Board (WBPCB). Seven air pollutants consisting of particulate matter or PM (PM_{2.5}, PM₁₀), CO, O₃, NO₂, benzene, and SO₂ are considered. Meteorological parameters used in the present study are air temperature (°C), maximum

and minimum air temperature (°C), dew point temperature (°C), precipitation (mm), relative humidity (%), visibility (km), wind speed (kmph), and air pressure (hPa). These parameter values are obtained on daily basis from the Regional Meteorological Centre, Kolkata. Satellite-based data on daily observations of aerosol optical depth (AOD) is collected from both MODIS Terra (MOD08_D3) and Aqua (MYD08_D3) collection 6.1 level 3 product for the period from 2018 to 2021. Here we have used a combined Dark Target (DT) Deep Blue (DB) AOD at 550 nm which is in better agreement with ground-based observations than the individual dark target and deep blue retrievals (Mhawish et al., 2017).

Data pre-processing and normalization

After collection of data, pre-processing of data is executed. Data pre-processing is the data mining technique so that we can improve data quality. Missing data is a common problem for most of the air pollution monitoring stations due to instrument failure, data entry error, maintenance, and other unmanageable factors. Though there are a few missing values in our data record, to fill the empty values, a linear interpolation technique is used (Plaia & Bondi, 2006). To convert the data on the same scale and make the data a pure dimensionless quantity or value, we have used max–min normalization as follows:

$$z(x) = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{1}$$

where min and max are the minimum and maximum values in x , given its range, and $z(x)$ is the normalized value of x .

Methods

Our objective is to build a framework of a deep model-based transfer learning approach in forecasting the concentration of pollutants during the pandemic situation. In this section, we will discuss the principles of different algorithms and tools used to design that framework in building the model TLS-BDLSTM, as well as some traditional ML and statistical methods used for comparison.

Random Forest algorithm

Random Forest (Breiman, 2001) is a multifunctional machine learning algorithm, widely used for classification

and regression predictive modeling problems in atmospheric science (Lepioufle et al., 2021; Yu et al., 2016). Here, we have used this algorithm to study the variable importance which is a part of designing our proposed deep TL model. Random Forest is a tree-based machine learning algorithm that makes decisions by combing the outcomes of several decision trees (DTs). Each DT relies on a random bootstrap dataset. Due to the simplification and nonparametric behaviors, classification and regression tree (CART) is generally used as a DT within the Random Forest. Thus, Random Forest is a collection of independent individual CART classifiers and can be represented as

$$\{h(x, \Theta_k), \text{ where } k = 1, 2, \dots, i, \dots, n\} \quad (2)$$

where h denotes the Random Forest classifier, x stands for input variable, and $\{\Theta_k\}$ are independent identically distributed random predictor variables, which are used for generating each CART tree.

The main purpose of the training stage of Random Forest is to construct a large number of decorrelated DTs (Liu et al., 2021). An overlap sampling solution known as “bagging” or bootstrap aggregation is used to reduce the variance associated with classification in Random Forest algorithm. It extracts observations with replacement to generate the independent bootstrap sample from the training dataset. Each DT can be trained from various bootstrap samples, resulting in more tree diversity. In certain Random Forest formulations, a subset of features is also taken for each tree. Thus, DTs within RF can be grown without pruning, resulting in a relatively small computational burden. Furthermore, while employing different bootstrap samples and node features, the noise immunity of Random Forest can be improved by averaging various decorrelated DTs. The fundamental advantage of bootstrap aggregation is that it improves model performance by decreasing the variance of the model, while keeping the bias low. It reduces the required training time and helps to avoid over-fitting. It gives a high level of predictive accuracy and maintains the correctness of the generalization even if a large part of the data is missing. The architecture of the Random Forest classification model is shown in Fig. 1. Each DT is constructed from a bootstrap sample chosen with replacement from the original dataset, and the predictions of all trees are finally combined through majority voting (Boulesteix et al., 2012).

The function of Random Forest is characterized by two basic features:

- a) Out-of-bag error (OOBE): The out-of-bag error, also called generalization error, is used for cross-validation to evaluate the performance of Random Forest. For each DT within a Random Forest, due to the bagging solution, some training data would be repeatedly used as the bootstrap sample, thereby some other observations not being selected to fit this DT. These observations are out-of-bag (OOB) instances. Nearly one-third of training data is OOB instances, which are not used in the RF training process. In other words, we can also use these OOB instances to evaluate the classification performance of by averaging the evaluation of OOB instances of DTs of Random Forest model.
- b) Variable importance (VI): The Random Forests can rank the features (variables) based on their importance in a classification problem. So, it enables reducing the dimension of the data by deleting the less important features. More elaborate variable importance measures incorporate a (weighted) mean of the individual tree’s improvement in the splitting criterion provided by each variable (Friedman, 2001). An example for such a measure in classification is “Gini importance” that represents the improvement in the “Gini gain” splitting criterion (Strobl et al., 2009). Another advanced variable importance measure available in Random Forests is the “permutation importance” measure which is directly based on the accuracy of the Random Forest model. It is obtained by permuting a feature and averaging the difference in OOBE (out-of-bag error) before and after the permutation over all trees. The underlying concept is that permuting an important feature is expected to decrease the classification accuracy more strongly than permuting a relatively unimportant feature. This operation can also be performed by selecting a subset of features, instead of permuting a feature, to determine the importance of the subset as a whole.

LSTM, bidirectional LSTM, and stacked-LSTM

Long short-term memory (LSTM) This is an advanced recurrent neural network (RNN), a sequential network, designed to prevent the neural network output for a given input from either decaying or exploding as it cycles through the feedback loops. It provides better performance compared to other RNN architectures because

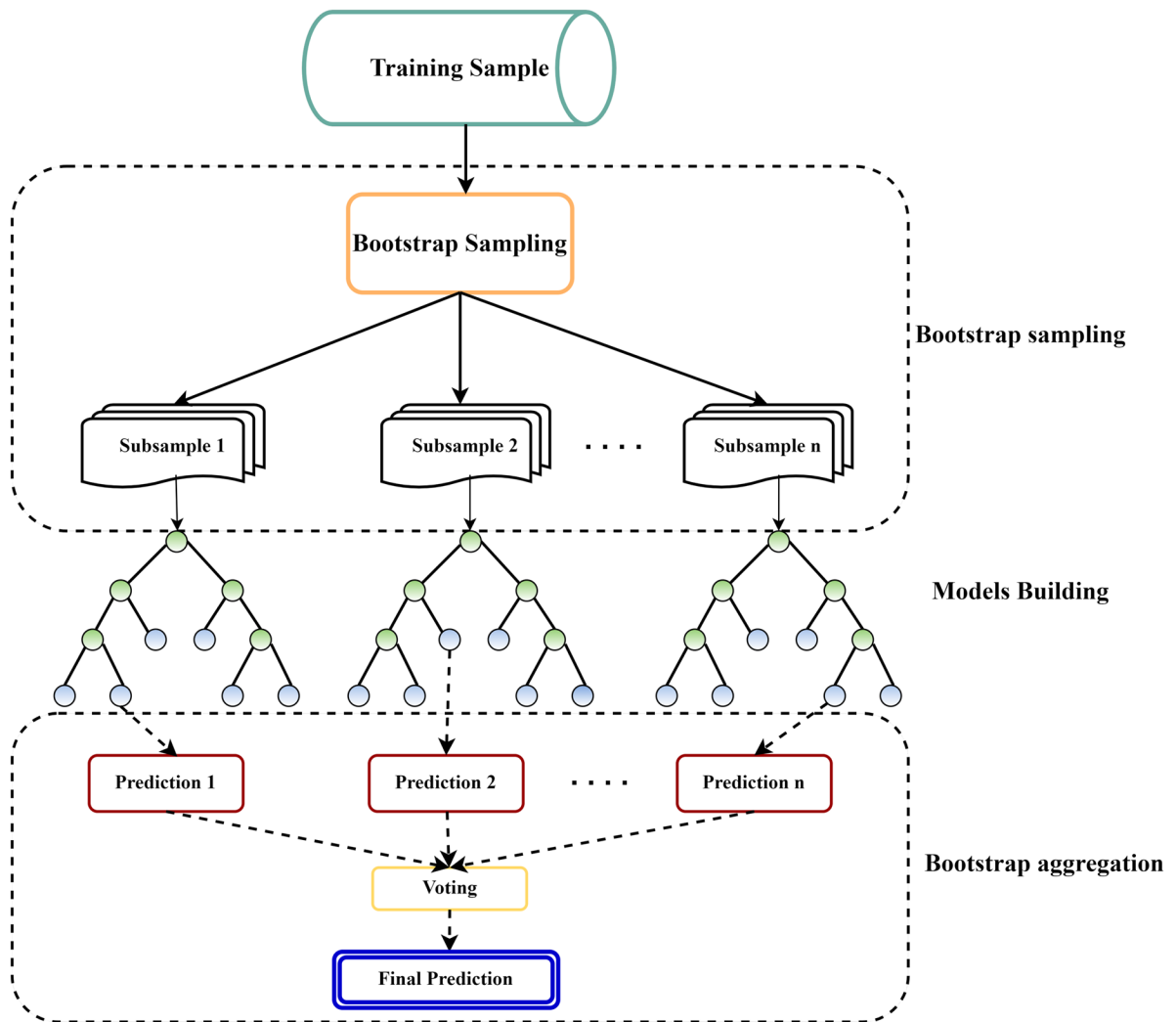


Fig. 1 Structure of Random Forest classification model

it can solve the problem of the disappearing gradient of RNN in dealing with long-term dependence. LSTMs are effectively applicable to several sequence learning problems including sequence prediction, speech recognition, language modeling and translation, and weather forecasting. LSTM is a type of supervised deep neural network formed based on artificial neural networks (ANN) and recurrent neural networks (RNN). Artificial neural networks (ANN) are composed of a large number of highly interconnected processing elements (neurons) working together to solve a problem. ANN usually consists of three layers—one input layer, one or more hidden layer, and one output layer. It is also known as a Feed-Forward Neural

network because inputs are processed only in the forward direction. However, the prediction accuracy is limited to the degree of stability of the time series and can fail when it presents complex dynamic behavior. To overcome this problem, recurrent neural network (RNN) is introduced that improves prediction accuracy. The recurrence feature in RNN allows forward and backward connections (recurrent or feedback), forming cycles within the network architecture, which uses previous states as a basis for the current state, and facilitates the learning of dynamic relationships (Andrea Sánchez-Sánchez et al., 2020). However, regular RNNs suffer from the vanishing or blowing up gradient during the back propagation (BP) process, and thus, it is

incapable of learning from long time lags and unable to capture the long-term dependencies in the input sequences (Bengio et al., 1994; Gers et al., 1999).

LSTM was introduced by Hochreiter and Schmidhuber (1997) as a modified version of RNN, which is well-suited to classify process and predict time series, given time lags of unknown duration. Due to the gated structure, LSTM architecture can deal with long-term dependencies to allow useful information to pass along the LSTM network. The memory cell of LSTM comprises three gates: namely, forget gate, input gate, and output gate that control the system's state. Gates are basically a way to selectively allow information through.

Figure 2a presents the architecture of a memory cell of LSTM. C_t and C_{t-1} are the current and previous cell states, respectively. h_t and h_{t-1} are respectively the current and previous hidden states. x_t is the input to the current cell. “ σ ” in rectangular box and “ \tanh ” in rectangular box represent Sigmoid-activated NN (neural networks) and Tanh (hyperbolic tangent)-activated NN, respectively, while “ \tanh ” in elliptical box denotes element-wise Tanh operator (not a NN). “ \times ” denotes element-wise multiplication, and “+” denotes element-wise summation. Note that each line in Fig. 2a carries an entire vector of equal number of components.

The first step in LSTM involves the identification of information that is not required in next steps and needs to be forgotten before passing to the cell state. This decision is made by a Sigmoid-activated NN layer, called the forget gate layer, whose output is vector f_t . This gate performs concatenation of the two vectors h_{t-1} and x_t , and this combination is introduced into the element-wise Sigmoid function (Eq. (3)) operator

$$\sigma(t) = \frac{1}{1 + e^{-t}}, \quad (3)$$

which outputs a number between 0 and 1 corresponding to each number in the cell state C_{t-1} . Here 1 means “completely keep this” (i.e., completely remembering), whereas 0 indicates “completely forgetting” (i.e., get rid of this).

The output f_t vector can be represented as

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (4)$$

where W_f and b_f are the parameters of the weight matrix and the bias vector, respectively, which are learned during training of the Sigmoid-activated NN. Note that in Eq. (4), the element within (.) is a vector, and σ (.)

means the point-wise Sigmoid function being applied on each (scalar) component of that vector, thereby resulting in a series of numbers (lying between 0 and 1) corresponding to those in the cell state C_{t-1} . That means, f_t represents a series of such numbers between 0 and 1. Then, the element-wise multiplication of f_t and C_{t-1} (Fig. 2a) decides to which extent the existing information in the cell is forgotten.

The next step is to decide what new information to store in the cell state. This is done by the input gate which has two parts. In one part, a Sigmoid-activated NN layer is used to generate an output vector (i_t) for that gate to decide which values to be updated. In another part, a Tanh-activated NN layer is used to generate a vector of new candidate values, \tilde{C}_t , that should be added to the state. Their expressions involving concatenation of the vectors h_{t-1} and x_t are shown in Eqs. (5) and (6).

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (5)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (6)$$

Here, W_i and b_i and W_c and b_c are the weight matrices and the bias vector parameters of the Sigmoid-activated NN layer and Tanh-activated NN layer, respectively. These parameters are learned during their training. i_t and C_t are then combined through element-wise multiplication to create a cell state update. Equation (7) represents the new cell state C_t which is updated from the previous cell state, C_{t-1} .

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (7)$$

Finally, the output gate decides the information of the cell state C_t that is to be provided to the new hidden state h_t . This output is a filtered version of C_t . To obtain this, a sigmoid-activated NN layer is first used to generate an output vector (o_t) that decides which parts of the cell state C_t are going to contribute to the output. Then, we pass the cell state C_t through \tanh operator to transform its element values to lie between -1 and $+1$ and finally multiply (element-wise) the transformed values by the Sigmoid gate output (o_t) that represents a set of values in the range (0, 1). This means that the multiplication operator only output the parts that we have decided to contribute. Expressions of o_t and the new hidden state h_t , thus obtained, are given in Eqs. (8) and (9), respectively.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (8)$$

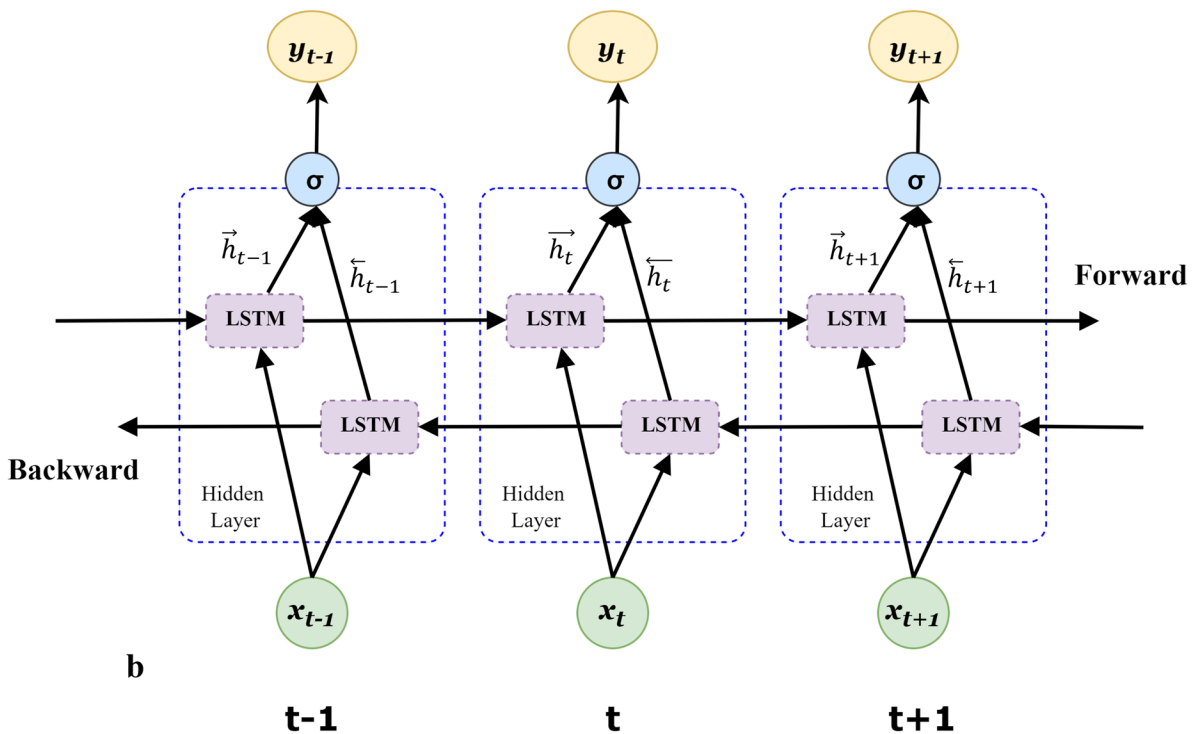
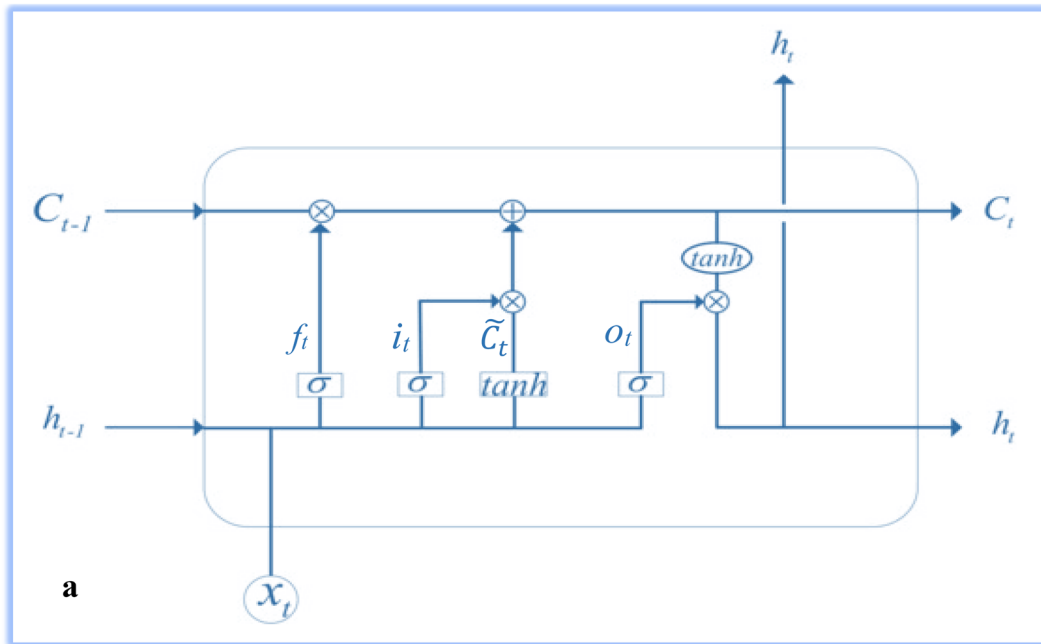


Fig. 2 The architecture of **a** long short-term memory (LSTM) and **b** bidirectional LSTM (BDLSTM)

where W_o and b_o are a set of learning parameters defined for the output gate.

$$h_t = o_t * \tanh(C_t) \tag{9}$$

where \tanh is an element-wise hyperbolic tangent function defined as

$$\tanh(t) = \frac{e^t - e^{-t}}{e^t + e^{-t}} \tag{10}$$

Bidirectional LSTM (BDLSTM) This is an extension of traditional LSTMs that can enhance model performance on sequence classification problems. The traditional LSTM (unidirectional) stores information of the past because the only inputs it has seen from the past. But the structure of BDLSTM allows the networks to have both backward and forward information about the sequence at every time step and using the two hidden states combined, and it is possible at any point in time to preserve information from both past and future. There are two hidden layers: one is forward LSTM layer, and the other is the backward LSTM layer. Both hidden layers are connected to the same output layer. The forward layer output sequence \vec{h}_t is iteratively calculated using inputs in a positive sequence from time $t=1$ to time $t=T$, while the backward layer output sequence, \overleftarrow{h}_t , is calculated using the reversed inputs from time $t=T$ to $t=1$. Both outputs are then concatenated using a Sigmoid function and fed to the final output vector

(Fraivan & Alkhodari, 2020) which is denoted by y_t and can be represented as (Yu et al., 2015):

$$y_t = W_{hy}^+ \vec{h}_t + W_{hy}^- \overleftarrow{h}_t + b_y \tag{11}$$

W_{hy}^+ is the weight from the forward LSTM layer to the output layer, and W_{hy}^- is the weight from the backward LSTM layer to the output layer. b_y is the output bias.

Figure 2b shows the architecture of the BDLSTM network.

Stacked-bidirectional LSTM (stacked-BDLSTM)

This architecture comprises multiple bidirectional LSTM layers (Fig. 3) where the output of the non-last layer is fed into the next layer. Each layer includes two LSTM cells with opposite directions, which are used to capture the forward and backward dependency, respectively.

In the following, we describe briefly some classification/regression ML models, namely, multilayer perceptron, support vector machine, K-nearest neighbor classifier, and statistical model autoregressive integrated moving average which are used for comparative study.

Multilayer perceptron (MLP)

Artificial neural network (ANN) is a mathematical model or information processing paradigm that is based on the way biological nervous systems process

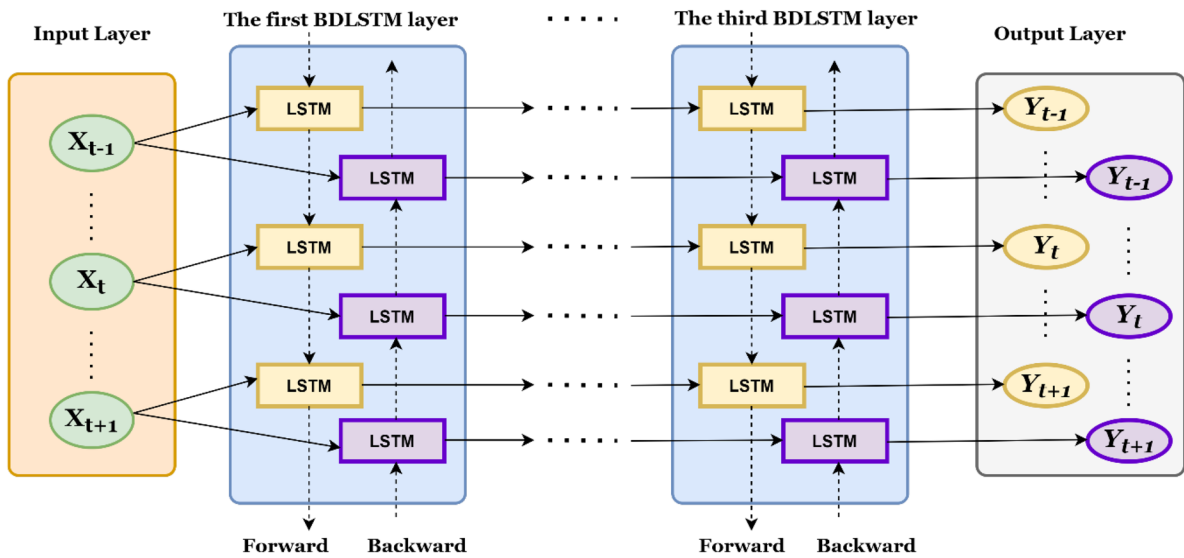


Fig. 3 The structure of stacked-bidirectional LSTM (stacked-BDLSTM)

information. It is made up of a network of highly interconnected processing elements, or artificial neurons, that work together to solve challenges. As mentioned in the sub-section “LSTM, bidirectional LSTM, and stacked-LSTM,” neural networks come in a number of different forms, but the feed-forward network architecture is the most popular. Different numeric inputs to the network are transferred forward from an input layer, through one or more hidden layers, to an output layer. The data is passed through a network and modeled based on the “weights” assigned to each connected link. By adjusting the weights binding its neurons together, a network can be trained to respond to different inputs in different ways; therefore, a functional mapping to a set of predictands from a set of predictors is generated.

The internal activity of the neuron can be explained as follows: let each neuron k receive incoming signals from every neuron j in the previous layer. Each incoming signal (x_j) is associated with a weight (w_k). The net input, v_k , to neuron k is a sum of the incoming signal times the weight. It is expressed as

$$V_k = \sum_{j=1}^n w_{kj} x_j \tag{12}$$

That means, the net input, v_k , equals the sum of the weight times the input signal for all the inputs to the neuron k from neuron j , starting at output of neuron $j=1$ and ending at $j=n$.

The activation function acts as a squashing function, such that the output of a neuron in a network lies between certain values (usually 0 and 1 or -1 and 1) (Pal & Mitra, 1999). In general, there are three types of activation functions (ϕ). Firstly, the threshold function which takes on a value of 0, if the summed input is less than a certain threshold value (v), and the value 1 if the summed input is greater than or equal to the threshold value. That is,

$$\phi(v) = \begin{cases} 1 & \text{if } v \geq 0 \\ 0 & \text{if } v < 0 \end{cases} \tag{13}$$

The multilayer perceptron (MLP) is perhaps the most used neural network architecture. Due to the dynamic nature of the atmosphere, accurate weather parameter prediction is a complex task. Multilayer perceptron is applied to prediction of atmospheric parameters, like temperature, wind speed, rainfall, fog, and pollution events (Dutta & Chaudhuri, 2015; Esteves et al., 2019; Cifuentes et al., 2020). It is a

feed-forward network of interconnected neurons usually trained by using the error back-propagation (BP) algorithm. The BP algorithm works by iteratively changing the interconnecting weights of the network in order to lower the overall error that exists between the observed values and modeled network outputs. The output of a neuron can be expressed as,

$$y = \phi \left(\sum_{i=1}^n w_i x_i + b \right) = \phi(w^T x + b) \tag{14}$$

where w is the weight vector, x denotes the input vector, b is the bias, and ϕ is the activation function Eq. (13).

It may be mentioned that artificial neural networks (ANNs) have the ability to learn the relation, whatever complicated, between input and output from examples, and therefore are considered a good candidate for machine learning. ANNs enjoy the characteristics, like adaptivity, speed, robustness/ruggedness, and optimality that any designer would love to have in her/his systems.

Support vector machine (SVM)

A support vector machine (SVM) is a supervised learning algorithm and is considered an effective tool for the prediction and analysis of air quality (García Nieto et al., 2013; Wang et al., 2021). The primary idea underlying SVM is to map the original datasets to a higher-dimensional space of features and construct an optimal separating plane (SP), from which the distance to all the data points is minimal (Ji et al., 2017). It was first developed to solve classification problems and later generalized to handle regression problems. This method is called support vector regression (SVR). SVR usually first trains the model according to the existing data to predict the time series.

For a training dataset $\{(x_i, y_i), i = 1, \dots, n\}$, $x \in R^m, y \in R$ where n is the total number of data patterns, x is the input vector of m components, and y is the corresponding output value, the SVM regression function can be expressed as follows:

$$f(x) = w \cdot \phi(x) + b \tag{15}$$

where w is the weight vector, b is the bias, and $\phi(x)$ indicates the non-linear transfer function (Zhao et al., 2021). The parameters w and b , which define the location of SP,

can be determined by minimizing the following regularized risk function:

$$\text{Minimize} : \frac{1}{2} \|w\|^2 + c \sum_{i=1}^n (\xi_i + \xi_i^*) \tag{16}$$

Subject to

$$y_i - w \cdot \phi(x) - b \leq \varepsilon + \xi_i \quad w \cdot \phi(x) + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i \geq 0, \xi_i^* \geq 0$$

where C is the regularization parameter and ξ and ξ^* are slack variables. Equation (16) is solved in a dual form utilizing the Lagrangian multipliers. That is,

$$\text{Maximize} : -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (a_i - a_i^*)(a_j - a_j^*)K(x_i, x_j) \\ - \sum_{i=1}^n (a_i - a_i^*) + \sum_{i=1}^n (a_i - a_i^*)y_i \tag{17}$$

Subject to $\sum_{i=1}^n (a_i - a_i^*) 0 \leq a_i \leq C 0 \leq a_i^* \leq C$

Where $K(x_i, x)$ is the kernel function.

By imposing the Karush–Kuhn–Tucker (KKT) optimality condition, w^* is obtained, that is,

$$w^* = \sum_{i=1}^n (a_i - a_i^*) \cdot K(x_j, x) \tag{18}$$

Finally, the SVM is expressed as follows:

$$f(x) = \sum_{i=1}^n (a_i - a_i^*) \cdot K(x_i, x) + b \tag{19}$$

The kernel function for SVM can be of four types: polynomial, sigmoid, linear, and radial basis function (RBF).

K-nearest neighbor (KNN) classifier

It is a simple and widely used machine learning model for classification and regression. This classical data mining tool has been applied for the prediction of air quality (Dragomir, 2010; Tella & Balogun, 2021). The KNN classification algorithm can be extended with regression for forecasting purposes. It is based on the assumption that objects that are close together will have similar characteristics (i.e., similarity in terms of proximity). When $K=1$, the nearest neighbor rule is the simplest form of KNN classifier (also called, multi-prototype minimum distance classifier). The KNN classifier provides predictions based on results from the K -neighbors closest to

that point. As a result, to make predictions with KNN classifier, we need to first identify the metric that will be used to calculate the distance between the request point and the case reference point in the sample. The Euclidean distance is one of the most frequently used distance metrics for determining this similarity. The Euclidean distance $d(X, Y)$ between two points (vectors) X and Y in Euclidean n -space is expressed as,

$$d(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{20}$$

where x_i and y_i are the i^{th} components of X and Y respectively. One may note that the Euclidean distance is only valid for continuous variables. After determining the value of K , we can make predictions based on K -nearest neighbor (KNN) examples. For classification of an unknown point (pattern), we decide the class to which majority of K -nearest neighbors belong.

The KNN classifier is reputed to be able to generate piecewise linear decision boundaries and, thereby, is quite efficient in handling non-convex and linearly nonseparable pattern classes. However, the performance of KNN depends on the value of K . The value of K is chosen to be “odd” to avoid tie in a decision. Further, the larger the values of K , the more is the computational time.

Autoregressive integrated moving average (ARIMA)

The ARIMA procedure uses the autoregressive integrated moving average (ARIMA) or autoregressive moving average (ARMA) model to evaluate and forecast equally spaced univariate time series data, transfer function data, and intervention data (Chaudhuri & Dutta, 2014). An ARIMA model predicts a value in a response time series as a linear combination of its own past values, past errors (also known as shocks or innovations), and present and past values of other time series. ARIMA represents a general class of time series models that combines several time series techniques such as differencing, autoregressive models, and moving average models (Montgomery & Johnson, 1976). An autoregressive (AR) model of the order p is the one in which the current observation X_t is regressed on previous observations $X_{t-1} + X_{t-2} + \dots + X_{t-p}$ of the same time series. This may be expressed as

$$X_t = \xi + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + e_t \quad (21)$$

where $\phi_1, \phi_2, \dots, \phi_p$ are the regression coefficients, ξ is a constant term, and e_t denotes random error.

If the time series is non-stationary, then it is first transformed into a stationary time series by a process, called differencing, in which the previous observation X_{t-1} is subtracted from the current observation X_t . The differencing can be performed with the help of a difference operator ∇ , defined as $\nabla X_t = X_t - X_{t-1}$ where differencing order is 1, and $\nabla^2 X_t = \nabla(\nabla X_t)$ where differencing order is 2, and so on.

An ARIMA model is generated by combining the three techniques, viz., autoregressive models, moving average models, and differencing. A general ARIMA model of the order (p, d, q) may be expressed as,

$$\phi(B)\nabla^d X_t = \theta(B)e_t \quad (22)$$

In the context of the present study, X_t and e_t represent pollutant and random error terms at time t , respectively, while d represents the order of differencing.

In the following, we provide some definitions characterizing transfer learning. The way of its implementation in building the proposed deep TL model is explained in sub-section “[Implementation of methodology](#).”

Transfer learning: some definitions

Transfer learning provides an improvement in learning in a new task through the transfer of knowledge from a related task that has already been learned (Torrey & Shavlik, 2010). The gain from such transfer of knowledge is particularly apparent whenever the data are ample in the source domain but scarce in the target domain. In this sub-section, some definitions characterizing transfer learning are given as follows:

Definition 1 (Lu et al., 2015): (Domain) A domain, which is denoted by $D = \{\chi, P(X)\}$ is composed of two components:

- 1) Feature space χ
- 2) Marginal probability distribution $P(X)$, where $X = \{x_1, \dots, x_n\} \in \chi$

Definition 2 (Lu et al., 2015): (Task) A task, which is denoted by $T = \{Y, f(\cdot)\}$ consists of two components:

- 1) A label space $Y = \{y_1, \dots, y_m\}$
- 2) An objective predictive function $f(\cdot)$ which is not observed and can be learned by pairs $\{x_i, y_i\}$ where $x_i \in X$ and $y_i \in Y$

The function $f(\cdot)$ can be used to predict the corresponding label, $f(x_i)$, of a new instance x_i . From a probabilistic viewpoint, $f(x_i)$ can be represented as $P(y_i|x_i)$.

Definition 3 (Lu et al., 2015): (Transfer learning) With notations of domain and task, the transfer learning (TL) is defined as follows:

Given a source domain D_s and learning task T_s , and a target domain D_t and learning task T_t , TL aims to help improve the learning of the target predictive function $f_t(\cdot)$ in D_t using the knowledge in D_s and T_s where $D_s \neq D_t$ or $T_s \neq T_t$.

In the above definition, the condition $D_s \neq D_t$ refers to either $\chi_s \neq \chi_t$ or $P_s(X) \neq P_t(X)$, i.e., the source and target domains have different feature spaces or marginal probability distributions. On the other hand, the condition $T_s \neq T_t$ means either $Y_s \neq Y_t$ or $P(Y_s|X_s) \neq P(Y_t|X_t)$, i.e., the source and target domains have different label spaces or conditional probability distributions. If the TL improves the overall performance by using entirely D_t and T_t , the outcome is referred to a positive transfer. Otherwise, negative transfer occurs when the information learned from a source domain has a detrimental effect on a target learner.

Implementation of methodology

At the outset, the collected data during the period from 2018 to 2021 are separated into three parts. These are the data from 1 January 2018 to 23 March 2020 as normal periods, the data from 24 March 2020 to 31 May 2020 as complete lockdown periods, and the data from 1 June to 30 September 2020 as partial lockdown periods. Validation is done using the real-time data obtained during the lockdown due to COVID second wave (16 May–15 June 2021).

Proposed research framework

Prediction of air quality is a major challenge in early warning and control of urban air pollution. COVID-19 lockdown has led to the short-term beneficial effects

on the environment where the concentration of pollutants marks down significantly associated with reduced emissions from anthropogenic activities. Thus, the air quality data obtained during the pandemic situation is quite different with respect to both nature and size, with the size of the data being substantially smaller. The objective of the present study is to predict concentrations of air pollutants during the pandemic situation. Transfer learning makes use of the knowledge learned while solving one problem and applying it to a different but related problem, so it does not need to learn from scratch. As a result, TL can address the most significant issue even when there is a lack of well-labeled training data. So, in the said pandemic situation, it may be rational to employ the concept of transfer learning.

The overall research framework is illustrated in block diagram (Fig. 4). To address our research objectives, we have first identified the important pollutants (among nitrogen dioxide (NO_2), sulfur dioxide (SO_2), particulate matter or PM (PM_{10} and $\text{PM}_{2.5}$), carbon monoxide (CO), ozone (O_3), and benzene) affecting the air quality of Kolkata during the complete lockdown, partial lockdown, and normal periods. This is done through ranking of their importance using Random Forest classifier. The dataset has seven columns corresponding to above mentioned pollutants and several rows representing the number of observations during complete lockdown, partial lockdown, and normal periods. The numbers of rows during complete lockdown, partial lockdown, and normal periods are 69, 122, and 813, respectively. Table 1 depicts the statistical characteristics of pollutants during the complete lockdown, partial lockdown, and normal periods. The coefficient of variation (CV) is the ratio of the standard deviation to the mean of a dataset and is a measure of precision. Thus, the lower the CV, the more precise the values (Table 1).

After obtaining the most effective pollutants to assess the air quality during the mentioned time periods, we propose a new air quality forecasting framework that can learn the temporal dependencies of multivariate air quality-related time series data and forecast the concentration of that pollutant. The input matrices are prepared with the data of ten parameters including air temperature ($^{\circ}\text{C}$), maximum and minimum air temperature ($^{\circ}\text{C}$), dew point temperature ($^{\circ}\text{C}$), daily precipitation (mm), relative humidity (%), visibility(km), wind speed (kmph), air pressure (hPa), and AOD. So, there are four types of datasets for four different periods, viz., normal periods, complete lockdown, partial lockdown, and validation. The columns

of the dataset represent the above-mentioned input parameters (total ten columns), and the rows represent the number of observations during the four periods. The numbers of rows during normal periods, complete lockdown, partial lockdown, and validation are 813, 69, 122, and 31, respectively. It is clear from the dataset that the number of observations during complete lockdown, partial lockdown, or validation is smaller than that of the normal periods. Our objective is to predict the concentrations of most effective pollutant as the output class.

The learning process for prediction (forecasting) of the concentration of air pollutants involves the following steps:

- 1) First, train the deep learning model, i.e., stacked-BDLSTM, with the data of normal periods, which is a large source dataset of 2 years.
- 2) This pre-trained model is then applied to the target datasets by changing the layers of stacked-BDLSTM (Fig. 9) and then fine-tuning the model. The target dataset comprises the data during complete lockdown of about 2 months (case study 1) and partial lockdown of about 4 months (case study 2) in 2020 for COVID-19 pandemic. The size of these datasets is comparatively much smaller than those of the source data of the normal periods. The resulting model thus obtained is called transfer learning-based stacked-bidirectional long and short-term memory (TLS-BDLSTM) model. This is practically a supervised model where a deep learning architecture is used to demonstrate the effectiveness of transfer learning. This model may be used for forecasting.
- 3) Finally, the TLS-BDLSTM is validated with real-time data with the multi-step prediction of air pollutants during validation time which is the complete lockdown period due to the COVID second wave (16 May–15 June 2021).

Implementation of different ML and statistical models

As mentioned before, the performance of the proposed TLS-BDLSTM model has been compared with different popular ML methods, viz., SVM, KNN, MLP, LSTM, and statistical model ARIMA in a part of the investigation. For implementing these ML and statistical models, their architectures are chosen through trial and error during the training over normal periods. We have used the architecture that produces the

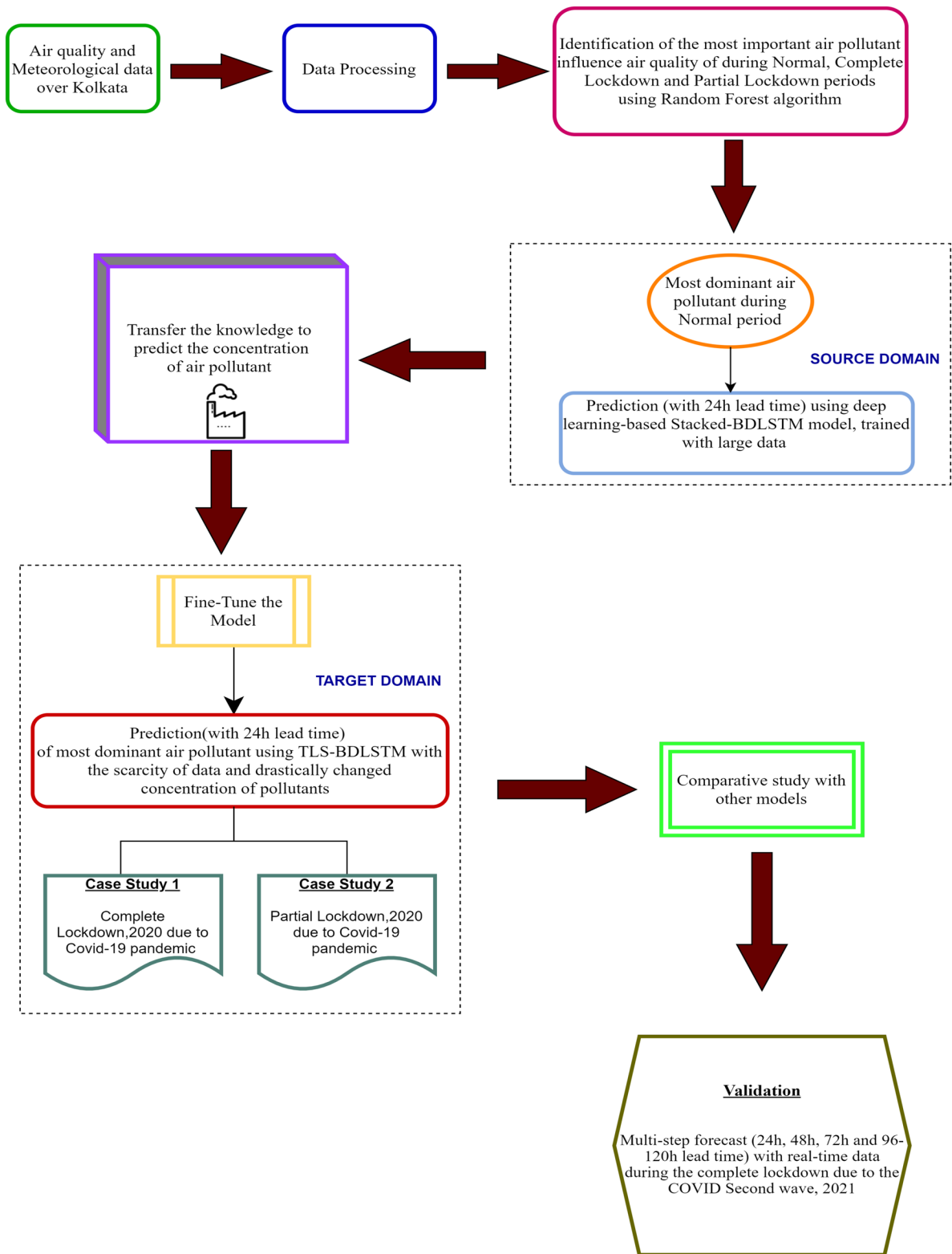


Fig. 4 Block diagram of proposed research framework with transfer learning for prediction of the concentration of air pollutants over Kolkata

Table 1 Statistical characteristics of pollutants during the complete lockdown, partial lockdown, and normal periods

Pollutants	Complete lockdown			Partial lockdown			Normal periods		
	Average	Standard deviation	Coefficient of variation (CV)	Average	Standard deviation	Coefficient of variation (CV)	Average	Standard deviation	Coefficient of variation (CV)
PM _{2.5} (µg/m ³)	23.05	13.15	57.03	18.69	6.86	36.71	92.85	22.93	24.69
PM ₁₀ (µg/m ³)	45.41	18.30	40.29	50.64	11.81	23.32	174.05	38.85	22.32
NO ₂ (µg/m ³)	10.32	3.67	35.64	28.80	9.63	33.45	57.18	12.94	22.64
SO ₂ (µg/m ³)	6.83	2.35	34.43	5.89	1.59	27.10	12.08	2.26	18.75
Benzene (µg/m ³)	1.67	0.88	52.69	3.32	5.37	161.57	9.03	4.12	45.66
CO (mg/m ³)	0.25	0.07	31.01	0.32	0.39	120.18	0.89	0.33	37.62
O ₃ (µg/m ³)	36.44	16.95	46.52	35.69	16.07	45.03	35.85	10.98	30.63

best outcomes in terms of error (RMSE/MAE) in this investigation. We have applied these optimum models throughout our study periods, i.e., during the complete lockdown, partial lockdown, normal periods, and during the time of validation. For example, the structure of MLP model which is found best here (Table 2) is 10: 10–6–1:1, i.e., 10 inputs with a single hidden layer having 6 nodes and a single output. There are many guidelines for determining the number of hidden layers and the correct number of neurons to use in the hidden layers (Chaudhuri et al., 2015; Mitra et al., 1996). The number of hidden layers in practical problems does not exceed two, and the number of hidden neurons should be between the sizes of the input and the output layers.

For KNN, we experimented with $k=1, 3, 5, 7$, and \sqrt{n} (where n denotes the number of training samples) (Pal et al., 2012), and we have found that $k=5$ is the best for prediction (Table 3). Similarly, ARIMA (2, 1, 2) model, i.e., ARIMA (p, d, q) model with $p=2$, the number of autoregressive terms, $d=1$, the degree of differencing and $q=2$, and the number of moving average terms, shows the optimum result, and therefore, it is taken for our study (Table 4).

For SVM model, RBF (radial basis function) was adopted as the kernel function because it can map input vectors into a high-dimensional feature space in a non-linear fashion. Hence, it has the ability to model complicated non-linear relationships. During SVM model selection, the determination of the optimal combination of C and gamma is highly important in order to build high-performance regression models. C is the regularization parameter that controls the degree of empirical error in optimization problems, and gamma is the RBF kernel parameter that significantly affects the generalization ability of SVM. During the development of the SVM model, the optimal pairwise C and gamma were identified using the fivefold cross-validation and grid search methods. We have examined C value with $C=0.1, 1, 10, 100$, and 1000 and gamma value with 1, 0.1, 0.01, 0.001, and 0.0001 during investigation. The optimal values found of these two parameters are $C=10$ and $\text{gamma}=0.0001$. Table 5 depicts that for LSTM model, the number of hidden layers considered was one with 64 hidden neurons. The learning rate was set to 0.001, while the epoch was set to 200 using Adam Optimizer. For stacked-BDLSTM model, we have chosen

Table 2 MLP model performances for predicting PM₁₀ and PM_{2.5} concentrations during training (optimal MAE and RMSE values are indicated in bold)

MLP	PM _{2.5}			PM ₁₀		
	MAE	RMSE	Time (min)	MAE	RMSE	Time (min)
MLP (10: 10–4–1:1)	17.01	22.97	02:36	32.76	41.94	02:36
MLP (10: 10–6–1:1)	15.58	20.30	02:36	31.72	40.11	02:37
MLP (10: 10–4–4–1:1)	18.97	24.92	02:38	32.03	40.89	02:38
MLP (10: 10–6–6–1:1)	17.26	23.08	02:38	36.66	45.92	02:38
MLP (10: 10–4–4–4–1:1)	33.67	40.29	02:39	50.62	58.02	02:39

Table 3 KNN model performances for predicting PM₁₀ and PM_{2.5} concentrations during training (optimal MAE and RMSE values are indicated in bold)

KNN	PM _{2.5}			PM ₁₀		
	MAE	RMSE	Time (min)	MAE	RMSE	Time (min)
KNN (K = 1)	23.19	29.02	02:55	36.79	48.54	02:55
KNN (K = 3)	22.78	27.89	02:55	35.04	47.92	02:55
KNN (K = 5)	21.68	26.69	02:55	33.91	45.05	02:55
KNN (K = 7)	22.59	28.93	02:55	35.23	48.39	02:56
KNN (K = √n)	25.62	31.02	02:58	39.53	51.90	02:59

three bidirectional LSTM layers, each having 64 hidden neurons. The learning rate was set to 0.001, while the epoch was set to 300 using Adam Optimizer (Table 5).

Three phases of comparison of different models

To further assess the effectiveness and performance of the TLS-BDLSTM model, a detailed comparative study is made in its three junctures (points) with some supervised conventional ML models, like support vector machine (SVM), K-nearest neighbor classifier (KNN), multilayer perceptron (MLP), long short-term memory (LSTM), and statistical model autoregressive integrated moving average (ARIMA). These three junctures are (1) output of stacked-BDLSTM (normal periods), (2) output of TLS-BDLSTM (case studies 1 and 2 in the complete and partial lockdowns), and (3) output of TLS-BDLSTM (validation period). It may be mentioned that during case studies 1 and 2, and validation, the output of TLS-BDLSTM is also compared with pre-trained stacked-BDLSTM. The objective is to demonstrate the effectiveness of both deep learning and transfer learning separately with respect to conventional machine learning and statistical techniques and the relevance in using TL over DL to improve the performance of the latter.

While comparing, the input variables are kept the same for all ML models during the three above-mentioned periods, viz., normal periods, complete and partial lockdowns, and during validation with real-time data. It is mentioned in the sub-section “Autoregressive integrated

moving average (ARIMA)” that the statistical ARIMA model predicts the current value of the time series using a linear combination of past values and innovations. It is applied similarly in the above-mentioned time periods. Furthermore, all the models have the same training and validation sets during different study periods.

Performance indices

We have implemented three indicators, namely, root mean square error (RMSE), mean absolute error (MAE), and R² (or coefficient of determination) which are used to interpret the experimental results as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \tag{27}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \tag{28}$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{29}$$

The predicted, observed, and mean values of the parameter are denoted by \hat{y} , y , and \bar{y} , respectively, and n is the number of cases (observations). RMSE shows the gap between the observed value and the predicted value, and it is highly sensitive than other metrics to the occasional large error. MAE is effective and unambiguous estimation of average error and indicates the

Table 4 ARIMA model performances for predicting PM₁₀ and PM_{2.5} concentrations during training (optimal MAE and RMSE values are indicated in bold)

ARIMA (p, d, q)	PM _{2.5}		PM ₁₀	
	MAE	RMSE	MAE	RMSE
ARIMA (1, 1, 1)	22.87	30.01	35.90	44.90
ARIMA (0, 2, 2)	20.07	26.59	34.66	42.22
ARIMA (2, 1, 2)	19.13	27.31	33.01	41.95

Table 5 LSTM and stacked-BDLSTM model performances for predicting PM₁₀ and PM_{2.5} concentrations during training (optimal MAE and RMSE values are indicated in bold)

Models	Hidden neurons (in each layer)	Learning rate	Epoch	PM _{2.5}			PM ₁₀		
				MAE	RMSE	Time (min)	MAE	RMSE	Time (min)
LSTM	10	0.001	200	12.37	20.09	03:09	28.00	37.25	03:09
	32	0.001	200	11.98	18.70	03:10	26.88	34.95	03:09
	64	0.001	200	11.00	17.67	03:10	26.19	34.29	03:10
	10	0.001	300	12.50	20.01	03:10	27.02	35.35	03:11
	32	0.001	300	11.59	19.89	03:11	26.50	34.63	03:11
	64	0.001	300	13.16	21.33	03:11	29.79	38.08	03:12
Stacked-BDLSTM	10	0.001	200	13.37	20.34	05:58	20.23	30.84	06:01
	32	0.001	200	10.89	18.85	05:58	17.09	28.80	06:02
	64	0.001	200	11.55	18.07	06:00	16.72	26.49	06:02
	10	0.001	300	12.68	17.91	06:02	15.90	25.21	06:04
	32	0.001	300	12.15	17.28	06:01	16.45	25.82	06:06
	64	0.001	300	10.49	16.99	06:03	15.72	24.35	06.05

robustness of models. The coefficient of determination, R² statistic, assesses prediction performance of a statistical model. It illustrates the proportion of variance in the outcome variable that is determined by the forecast.

Results and discussion

Here we explain in detail the performance of our proposed models. This includes the selection of significant pollutants affecting the air quality through ranking, demonstrating the impact of various meteorological parameters in forecasting, prediction efficiency of the

Table 6 Spearman rank order correlation of 24 h before input values with concentration of PM_{2.5} and PM₁₀ on current day

Inputs	Output	
	PM _{2.5}	PM ₁₀
Air temperature (°C)	0.56	0.57
Maximum temperature (°C)	0.49	0.58
Minimum air temperature (°C)	0.28	0.44
Dew point temperature(°C)	0.11	0.10
Daily precipitation (mm)	−0.46	−0.31
Relative humidity (%)	0.12	−0.55
Visibility(km)	−0.77	−0.72
Wind speed (kmph)	−0.09	−0.08
Air pressure(hPa)	0.33	0.22
AOD	0.19	0.18

TLS-BDLSTM, and a comparative study with various conventional ML and statistical models.

Selection-based ranking of pollutants

Random Forest, as explained in the sub-section “[Random Forest algorithm](#),” is one of the most powerful machine learning algorithms in modern times which is built on decision trees. Decision tree is known for its ability to select the most “important” features from many features while ignoring irrelevant ones. Furthermore, decision trees provide an explicit model describing the relationship between features and predictions, which makes the model interpretation easier. As an ensemble of trees, the Random Forest inherits the capacity to select an “important” feature.

Here we present the classification-based ranking of pollutants using Random Forest classifier during normal periods, complete lockdown, and partial lockdown at an interval of 24 h. The dataset is divided into two parts, viz., 75% for training data and 25% for testing data, to see the accuracy of the classification results. In the development of Random Forest model, we have tuned hyperparameters for maximum performance. We have used grid search method from sklearn using python to tune our hyperparameters and calculate the model performance for each combination by using five-fold cross validation. We have tuned two hyperparameters, namely, n_estimators (number

of trees in the forest) and `max_depth` (max number of levels in each decision tree), during our investigation. The values of these parameters are chosen as follows: `n_estimators` (10, 100, 200, 300, 500, and 1000) and `max_depth` (10, 50, and 100). The best estimator method produces a model with the best performance parameters, which in this case is `max_depth=100` and `n_estimators=1000`. Our subsequent investigation is done with this tuned model. Figure 5 provides the pollutants' importance that highly influences the air quality of Kolkata during the three mentioned periods. PM_{10} concentration in air is ranked as the most important air pollutant throughout complete lockdown and partial lockdown; however, the concentration of $PM_{2.5}$ shows the most dominant parameter during normal periods (Fig. 6).

One may note that PM_{10} is contributed significantly by vehicles, road dust followed by secondary inorganic aerosols, domestic and commercial combustion, and construction activity. On the other hand, domestic and commercial combustion activities, cars, road dust, open burning, secondary inorganic aerosols, and agricultural and industrial activities all make a substantial contribution to $PM_{2.5}$. Thus, from Fig. 6, it is found that particulate matter or PM (PM_{10} and $PM_{2.5}$) is the effective parameter of air quality of Kolkata throughout the study period. Both PM_{10} (particulate matters with a diameter of 10 microns or less) and $PM_{2.5}$ (particles that have diameter 2.5 microns or less) can penetrate the respiratory system; however, atmospheric particulate matter $PM_{2.5}$ which is about 3% the diameter of a human hair is more harmful because it travels deep into the lungs and causes a wide range of respiratory diseases, including cancer.

$PM_{2.5}$ has been identified as the primary cause of the adverse cardiovascular effects of air pollution on human health, based on various epidemiological research and significant clinical observations (Brook et al., 2010; Pope et al., 2004; Sun & Ulintz, 2016). Commercial vehicles of more than 15 years old mainly are driven by diesel and poorly maintained still playing in Kolkata Metropolis. The solid material in diesel exhaust with less than 1 μm in diameter is a fraction of particulate matter $PM_{2.5}$. In Kolkata this fine particulate matter $PM_{2.5}$ mainly comes from vehicle (25%) and secondary aerosols (32%) (CSIR-NEERI, 2019). It is apparent from our observation that the concentration of PM_{10} is consistent during the partial lockdown and normal periods and slightly variable in the time of complete

lockdown. On the other hand, $PM_{2.5}$ concentration is highly variable during the complete lockdown period as its coefficient of variation value is very high comparative to other pollutants (Table 1).

From Fig. 6, it can be further depicted that the feature importance ranking of $PM_{2.5}$ decreases from partial lockdown to complete lockdown as compared to normal periods. So, it can be delineated that $PM_{2.5}$ is a very significant pollutant during the normal periods and the effect of the lockdown on $PM_{2.5}$ is the most significant for Kolkata. A decline was seen for $PM_{2.5}$ with concentration falling by 40% and 75% in the first and second phases of lockdown respectively, against pre-lockdown levels (CPCB, 2020). This could be attributed to the absence of non-essential vehicles and combustion activities in industrial as well as commercial areas and restrictions on construction activities along with reduced dust resuspension. On the other hand, thermoelectric power plant and factories maintained their activities during the pandemic. So, it is clear that fine particulate matter (PM), i.e., PM_{10} and $PM_{2.5}$, is of greatest concern in Kolkata. Therefore, prediction of the concentration of $PM_{2.5}$ and PM_{10} , which contribute considerably to the air quality index (AQI), is essential to sustaining acceptable air quality levels and the prevention of health risks associated with air pollution.

Impact of meteorological parameters

As stated earlier, in order to forecast the concentrations of most dominating pollutants, i.e., $PM_{2.5}$ and PM_{10} over Kolkata, the input matrices are prepared with the data of 10 meteorological parameters including air temperature ($^{\circ}\text{C}$), maximum and minimum air temperature ($^{\circ}\text{C}$), dew point temperature ($^{\circ}\text{C}$), daily precipitation (mm), relative humidity (%), visibility (km), wind speed (kmph), air pressure (hPa), and AOD, and the outputs are concentrations of PM_{10} ($\mu\text{g}/\text{m}^3$) and $PM_{2.5}$ ($\mu\text{g}/\text{m}^3$). Meteorological parameters have a significant impact on ambient air quality by influencing both directly and indirectly the emissions, transport, formation, and deposition of air pollutants. Therefore, the above data characteristics are crucial for forecasting the concentrations of $PM_{2.5}$ and PM_{10} .

Table 6 displays Spearman rank-order correlations between 24 h before input values and the concentration of $PM_{2.5}$ and PM_{10} on the current day. This reveals the potential of inputs for forecasting the target output. It is observed that the air temperature is positively correlated

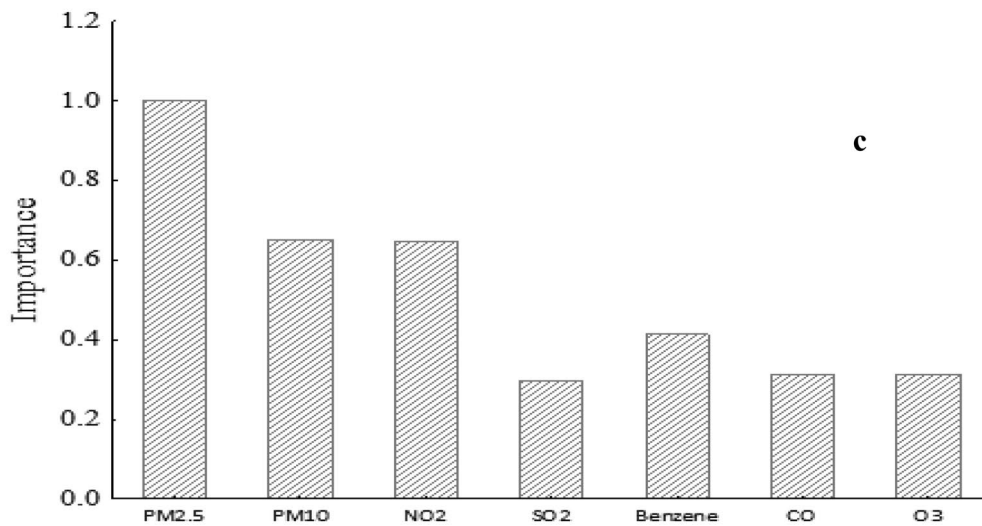
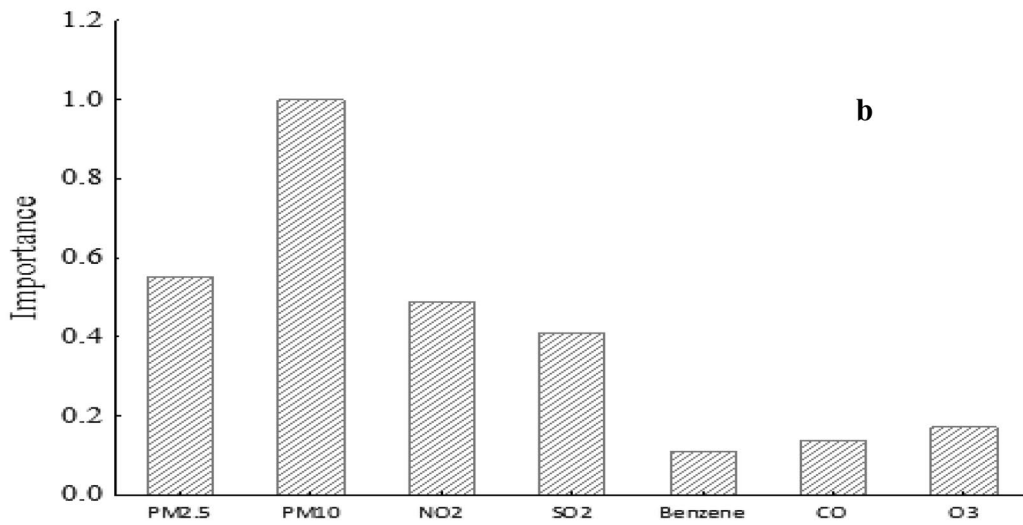
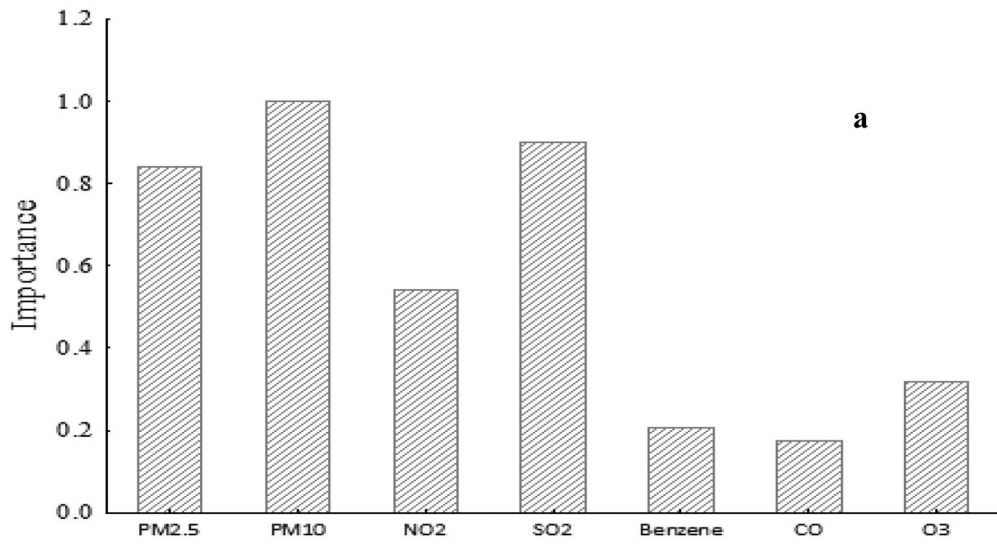


Fig. 5 Predictor importance of air quality over Kolkata during **a** complete lockdown, **b** partial lockdown, and **c** normal periods

with both $PM_{2.5}$ and PM_{10} . The photochemical reaction between precursors can be accelerated by a high temperature. This is owing to the fact that particle formation is affected by temperature. The same findings are made in the case of maximum and minimum temperature.

The airborne fine particles near the ground act as a condensation nucleus for dew. Air pollution is most serious during haze and foggy weather due to the enhancement of the fraction of fine particulate matter. There is a weakly positive correlation of 0.11 and 0.10 that was found in the case of dew point temperature with $PM_{2.5}$ and PM_{10} , respectively.

The relative humidity is positively correlated with $PM_{2.5}$ (0.12) indicating that as relative humidity increases, so does $PM_{2.5}$ concentration. Higher relative humidity causes a weaker diffusion of $PM_{2.5}$. As a result of this environment, the hygroscopicity of pollutants will increase, and the chemical transformation of pollutants will speed up, exacerbating the degree of air pollution (Li et al., 2017). But relative humidity is negatively correlated with PM_{10} (−0.55) because it affects the natural deposition process of particulate matter. Moisture particles adhere to particulate matter. As humidity rises, the size of the particulate matter also increases owing to the accumulation of particulate matter concentration. Moisture particles grow to a point where dry deposition takes place and thus reducing the concentration of PM_{10} in the atmosphere.

Both $PM_{2.5}$ and PM_{10} are negatively but very weakly correlated with wind speed. In calm weather, pollutants tend to accumulate. When the wind speed is high enough, it facilitates the migration of particulate matter and thus reduces the concentration. A positive correlation is found in atmospheric pressure with both $PM_{2.5}$ and PM_{10} because high pressure is associated with low atmospheric boundary layer height, and thus the vertical dispersion of air pollutants is restricted (Wu et al., 2017) (Table 6).

A significant strong negative correlation is found with visibility (−0.77 and −0.72, respectively) which means heavy air pollution causes lower visibility and vice versa. Precipitation effectively removes atmospheric particulate matter. Daily precipitation is seen to be negatively correlated with particulate matter $PM_{2.5}$ and PM_{10} (−0.46 and −0.31, respectively) which reveals that precipitation has a wet scavenging effect on these pollutants. $PM_{2.5}$ and PM_{10} are found to be positive though weakly correlated with AOD. Both PM concentration and AOD indicate

the turbidity of the atmosphere to some extent, and both depict part of the suspended matter in the atmosphere.

Autocorrelation of PM_{10} and $PM_{2.5}$

Due to rapid urbanization, climate change, environmental degradation, or pandemics, air pollution patterns are changing, and it is influenced by several factors. Autocorrelation is the crucial parameter that represents the linear relationship between lagged values of a time series. Figure 7 shows autocorrelation coefficients for the daily concentration of PM_{10} and $PM_{2.5}$ with time lag. When the time lag is smaller, the autocorrelation coefficient is stronger, and it shows a decreasing trend when the time lag increases. This reveals that the earlier concentrations have a weaker impact on the current concentration of PM (both $PM_{2.5}$ and PM_{10}), whereas PM concentrations closer to the current time have a stronger influence on PM concentration at the current time. When time lag is smaller than 4, autocorrelation coefficients are higher than 0.5 which depicts that PM concentrations are strongly correlated to each other. Though PM_{10} is heavier and thus tends to settle, $PM_{2.5}$ remains suspended for longer. Thus, PM concentrations in the past few days will impact the observed data in the future.

Prediction performance and comparative study

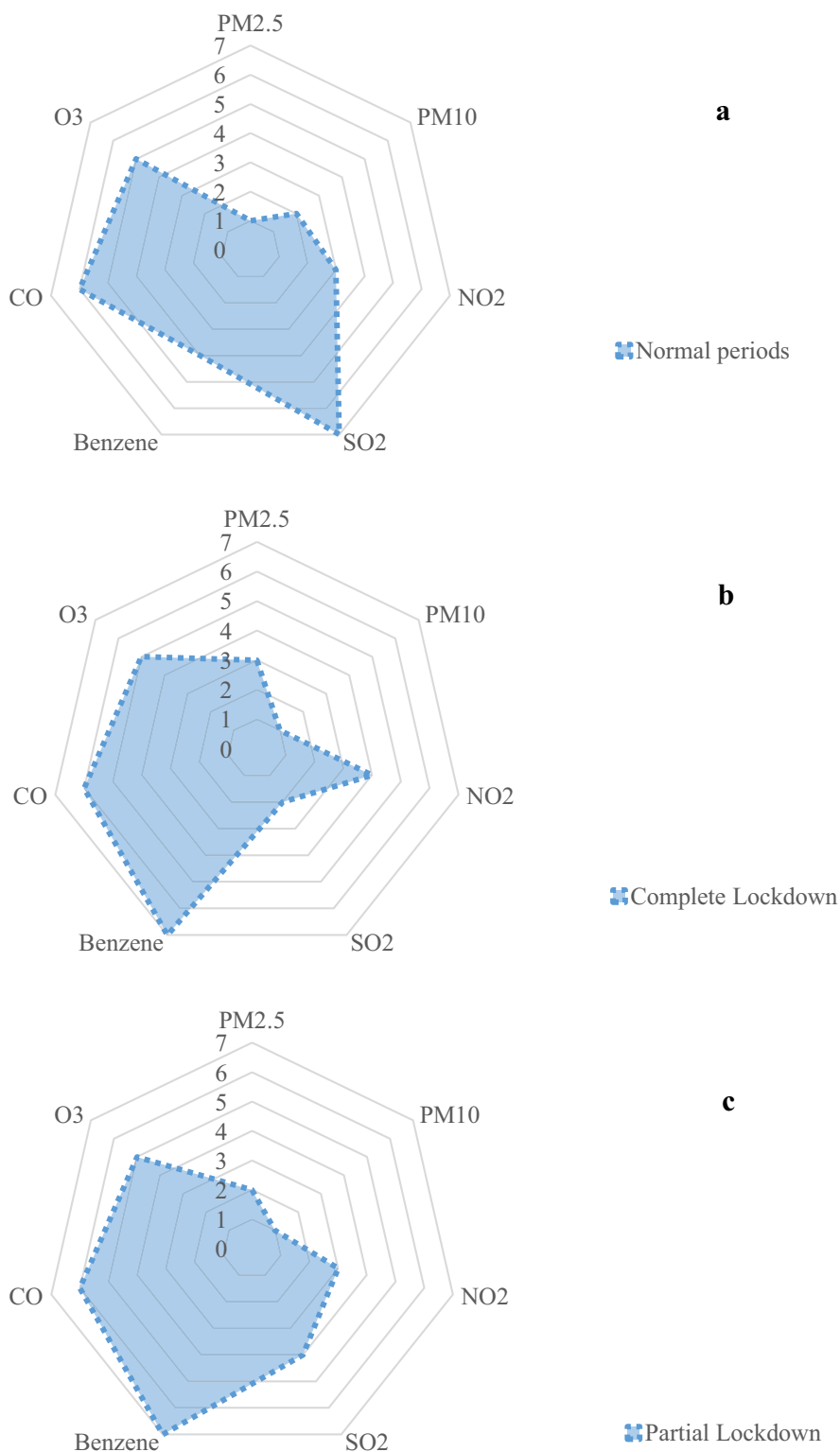
One may note that the forecasting parameters of the proposed TLS-BDLSTM model are PM_{10} and $PM_{2.5}$ which were obtained by the Random Forest algorithm as the most significant air pollutants. As further evidence towards their significance, one may consider Fig. 8 which shows a remarkable change in concentration of PM_{10} and $PM_{2.5}$ during the complete lockdown and partial lockdown phases. Therefore, the prediction of the concentration of $PM_{2.5}$ and PM_{10} and the relevance of using transfer learning therein are highlighted in our investigation.

The results obtained from different models for the prediction of the concentration of $PM_{2.5}$ and PM_{10} during normal periods and in case study 1 and case study 2 (during complete lockdown and partial lockdown) are discussed as follows.

Normal periods

In the present study, we have used stacked-bidirectional LSTM (stacked-BDLSTM) for the prediction of concentration of PM_{10} and $PM_{2.5}$ during the normal periods in

Fig. 6 Important pollutants based on ranking using Random Forest algorithm during the complete lockdown, partial lockdown, and normal periods over Kolkata



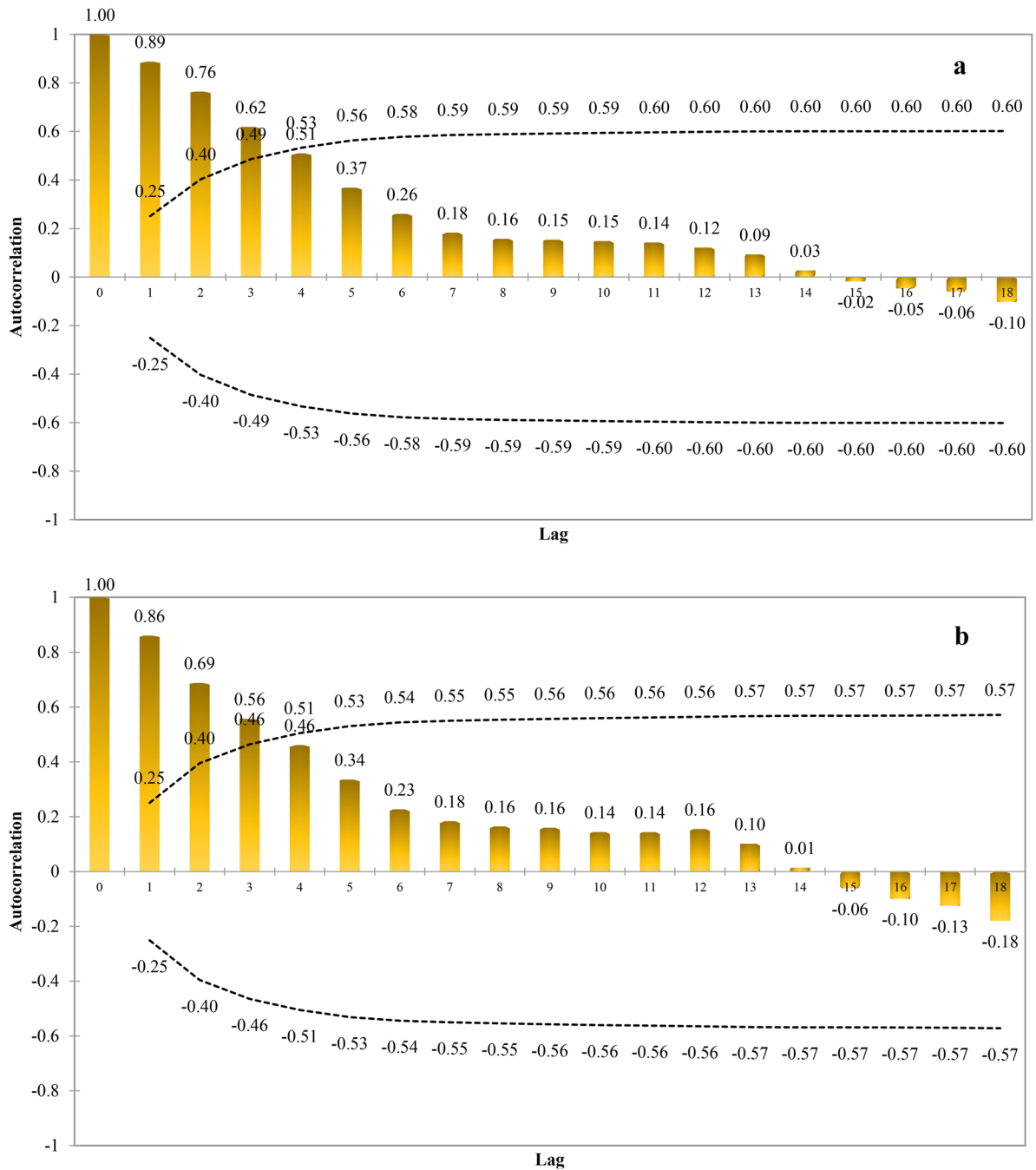


Fig. 7 The autocorrelation coefficients of the concentrations of a $PM_{2.5}$ and b PM_{10} with respects to different time lag values

which data of 2 years have been analyzed. Table 7 depicts a comparative study of stacked-BDLSTM with other models (as mentioned in the sub-section “Three phases of comparison of different models”). The network (stacked-BDLSTM) designed for time series forecast gives better performance as it has the lowest RMSE and MAE for

prediction of concentration of $PM_{2.5}$ (MAE is 10.49 and RMSE is 16.99) and PM_{10} (MAE is 15.72 and RMSE is 24.35) during the normal periods with 24-h lead time. The superiority of stacked-BDLSTM over ordinary LSTM (e.g., MAE and RMSE are 11.00 and 17.67 for $PM_{2.5}$ and 26.19 and 34.29 for PM_{10} , respectively) has been verified.

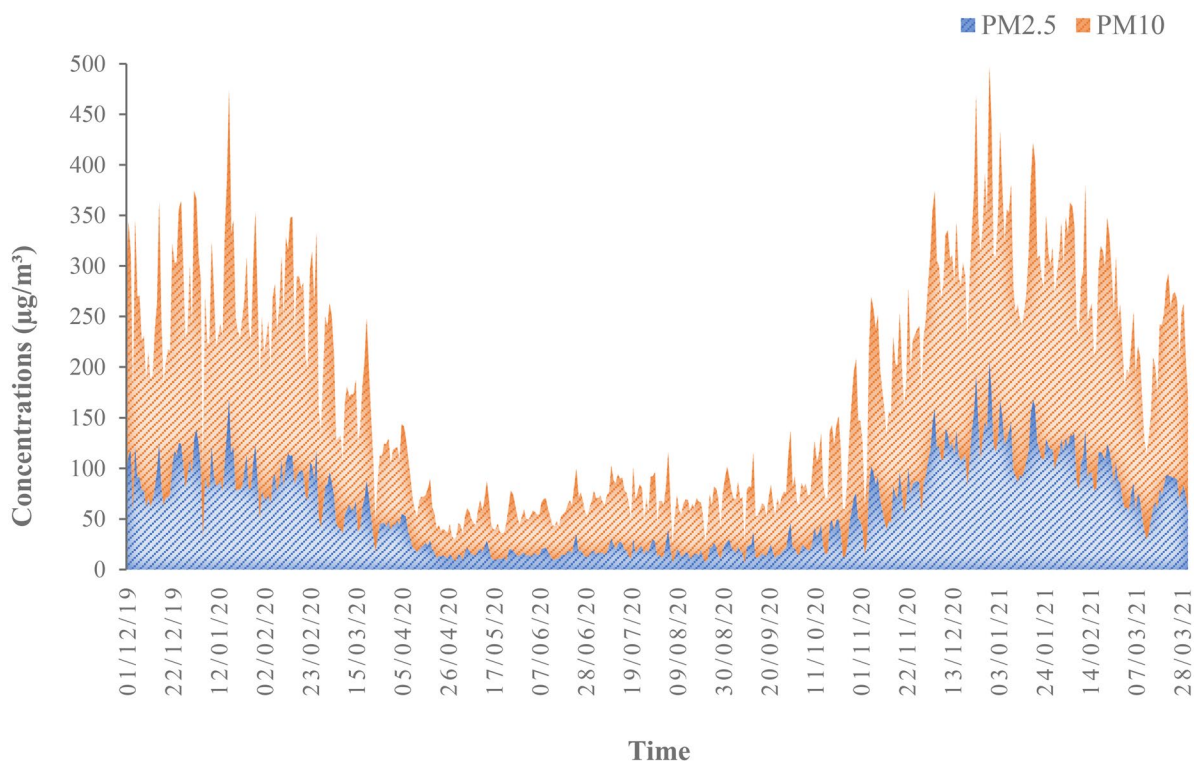


Fig. 8 PM_{2.5} and PM₁₀ concentrations (µg/m³) over Kolkata from 1 December 2019 to 28 March 2021

The superiority is because of the fact that the BDLSTM processes time-series information in two directions, both past and future, for each point in the time series and is capable of learning long-term dependencies. This indicates that this model is effective at capturing long-term dependencies in time series that have an important impact on PM_{2.5} and PM₁₀ concentrations. Furthermore, during normal periods, the two deep learning models, i.e., LSTM and stacked-BDLSTM, outperform the three traditional

(shallow) learning models, viz., MLP, SVM, and KNN, and the statistical model ARIMA. Here the word “shallow” is meant in contrast to “deep” (Pal et al., 2019). One may note that, in the case of MLP, the values of MAE and RMSE are 15.58 and 20.30 for PM_{2.5} and 31.72 and 40.11 for PM₁₀, respectively. On the other hand, these values are 24.44 and 30.89 for PM_{2.5} and 43.46 and 50.55 for PM₁₀ respectively for SVM and 21.68 and 26.69 for PM_{2.5} and 33.91 and 45.05 for PM₁₀ respectively for

Table 7 Stacked-BDLSTM model performance for predicting PM₁₀ and PM_{2.5} concentrations during normal periods

Period	Models	PM _{2.5}		PM ₁₀	
		MAE	RMSE	MAE	RMSE
<i>Normal periods</i>	Stacked-BDLSTM	10.49	16.99	15.72	24.35
	LSTM	11.00	17.67	26.19	34.29
	MLP	15.58	20.30	31.72	40.11
	ARIMA	19.13	27.31	33.01	41.95
	SVM	24.44	30.89	43.46	50.55
	KNN	21.68	26.69	33.91	45.05

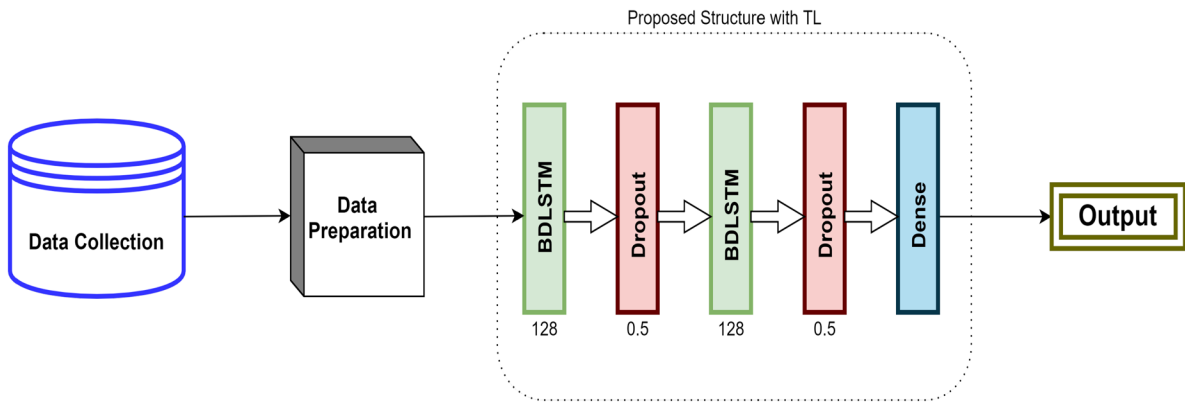


Fig. 9 Proposed structure of TLS-BDLSTM with transfer learning for the study of pandemic situation

KNN. Similarly, MAE and RMSE values for statistical model ARIMA are 19.13 and 27.31 for $PM_{2.5}$ and 33.01 and 41.95 for PM_{10} , respectively.

So, from the above discussion, it is clear that the prediction accuracy of SVM is the lowest for the prediction of the concentrations of $PM_{2.5}$ and PM_{10} during normal periods. SVM is not acceptable for large data, and it will underperform in the cases where the number of features for each data point outstrips the number of training data specimens.

Case study 1 and case study 2

As mentioned earlier, a novel transfer learning approach was used to study the prediction accuracy in the time of

complete lockdown (case study 1) and partial lockdown (case study 2) during pandemic situations when there is an exceptional change in concentrations of $PM_{2.5}$ and PM_{10} than the normal periods (Fig. 8). At the same time, the training data is limited in case study 1 and case study 2 as compared to the normal periods. The architecture of TLS-BDLSTM is presented in Fig. 9 which is obtained from our pre-trained deep learning model stacked-BDLSTM incorporating transfer learning. In the first layer of the new model, BDLSTM is used because it can make use of both forward and backward information about the sequence at every time step, and temporal dependencies of input parameters can be apprehended during the learning process. For further learning, the second BDLSTM network is used.

Table 8 TLS-BDLSTM model performance for predicting PM_{10} and $PM_{2.5}$ concentrations with 24-h lead time during pandemic situation

Period	Models	$PM_{2.5}$		PM_{10}	
		MAE	RMSE	MAE	RMSE
<i>Complete lockdown</i>	TLS-BDLSTM	2.56	6.13	2.47	5.24
	Stacked-BDLSTM	4.89	8.82	5.51	11.22
	LSTM	5.62	9.09	8.99	13.43
	MLP	6.04	10.57	11.02	15.33
	ARIMA	7.10	9.45	7.75	11.33
	SVM	6.81	8.90	9.17	12.69
	KNN	9.66	18.49	12.07	15.21
<i>Partial lockdown</i>	TLS-BDLSTM	3.79	5.02	7.59	9.49
	Stacked-BDLSTM	4.63	6.52	8.00	10.22
	LSTM	5.01	6.86	8.34	11.17
	MLP	5.16	7.05	8.65	11.50
	ARIMA	5.02	6.90	9.31	12.27
	SVM	4.98	6.70	9.16	11.97
	KNN	6.71	8.84	12.85	16.08

Table 9 Computational time for prediction by different machine learning models during normal periods, complete lockdown, and partial lockdown

Period	Models	Time (min) for PM _{2.5}	Time (min) for PM ₁₀
<i>Normal periods</i>	Stacked-BDLSTM	06:03	06.05
	LSTM	03:10	03.10
	MLP	02:36	02:37
	SVM	02:12	02:11
	KNN	02:55	02:55
<i>Complete lockdown</i>	TLS-BDLSTM	01:10	01:10
	Stacked-BDLSTM	01:22	01:21
	LSTM	00:53	00:53
	MLP	00:33	00:35
	SVM	00:31	00:30
<i>Partial lockdown</i>	TLS-BDLSTM	01:39	01:40
	Stacked-BDLSTM	01:57	01:57
	LSTM	01:09	01:09
	MLP	00:49	00:49
	SVM	00:43	00:45
	KNN	00:50	00:50

The dense layer is incorporated at the top for outputting the desired prediction. For optimization of parameters, we have used two dropout layers with probability 0.5 between our BDLSTM layers to avoid over-fitting. Each of the BDLSTM layer is equipped using 128 hidden neurons for learning the temporal features. The learning rate is set to 0.001, and the epochs are set to 300 using Adam Optimizer. The adaptive moment estimation (Adam) algorithm is used to replace stochastic

gradient descent (SGD) to update the weights to get higher accuracy (Guo et al., 2020). The error in forecasting the concentrations of PM_{2.5} and PM₁₀ during the complete lockdown (case study 1) and partial lockdown (case study 2) using this deep TL model, TLS-BDLSTM, is estimated.

The MAE and RMSE with 24-h lead time are observed to be 2.56 and 6.13 for PM_{2.5} concentrations during the complete lockdown and 3.79 and 5.02 during partial lockdown with TLS-BDLSTM. The MAE and RMSE using TLS-BDLSTM for the prediction of PM₁₀ concentrations are 2.47 and 5.24 during complete lockdown and 7.59 and 9.49 for partial lockdown with 24-h lead time. Table 8 shows that with transfer learning approaches, TLS-BDLSTM shows better prediction than stacked-BDLSTM without transfer learning during the COVID pandemic for both the case study periods. Using stacked-BDLSTM, the MAE and RMSE for PM_{2.5} and PM₁₀ concentrations during complete lockdown are 4.89 and 8.82, respectively, and 5.51 and 11.22, respectively. Similarly, the MAE and RMSE for PM_{2.5} and PM₁₀ concentrations during partial lockdown are 4.63 and 6.52, respectively, and 8.00 and 10.22, respectively, using stacked-BDLSTM model with a 24-h lead time. Thus, the major finding is that the knowledge gained from the previous study could result in a fair measure of accuracy for the prediction of mostly affected pollutants PM_{2.5} and PM₁₀. Therefore, fine-tuning and changing layers of the pre-trained model provide a significant benefit leading to higher accuracy during the pandemic situations with a shortage of data and change in concentration of pollutants. The results from Table 8 further reveal the superiority of the TLS-BDLSTM model over the other four supervised machine learning models, like SVM, KNN, MLP, and LSTM, as well as statistical time series forecasting model ARIMA. Thus, consistency is also maintained by TLS-BDLSTM during the complete

Table 10 TLS-BDLSTM model performance for the multi-step prediction of PM_{2.5} concentrations during validation

Models	MAE				RMSE			
	24 h	48 h	72 h	96–120 h	24 h	48 h	72 h	96–120 h
TLS-BDLSTM	3.52	4.21	5.39	8.10	4.59	5.40	7.81	12.41
Stacked-BDLSTM	4.22	5.25	7.48	11.44	5.53	6.99	10.94	15.42
LSTM	4.87	5.73	9.05	13.14	6.78	7.53	11.98	14.97
MLP	6.47	7.87	9.22	17.21	8.33	14.86	12.96	21.87
ARIMA	7.29	8.52	12.56	17.25	9.97	11.95	15.88	23.10
SVM	6.51	8.79	11.38	13.36	7.88	9.15	13.15	15.82
KNN	4.45	7.47	8.22	19.88	5.64	9.03	11.12	23.35

Table 11 TLS-BDLSTM model performance for the multi-step prediction of PM₁₀ concentrations during validation

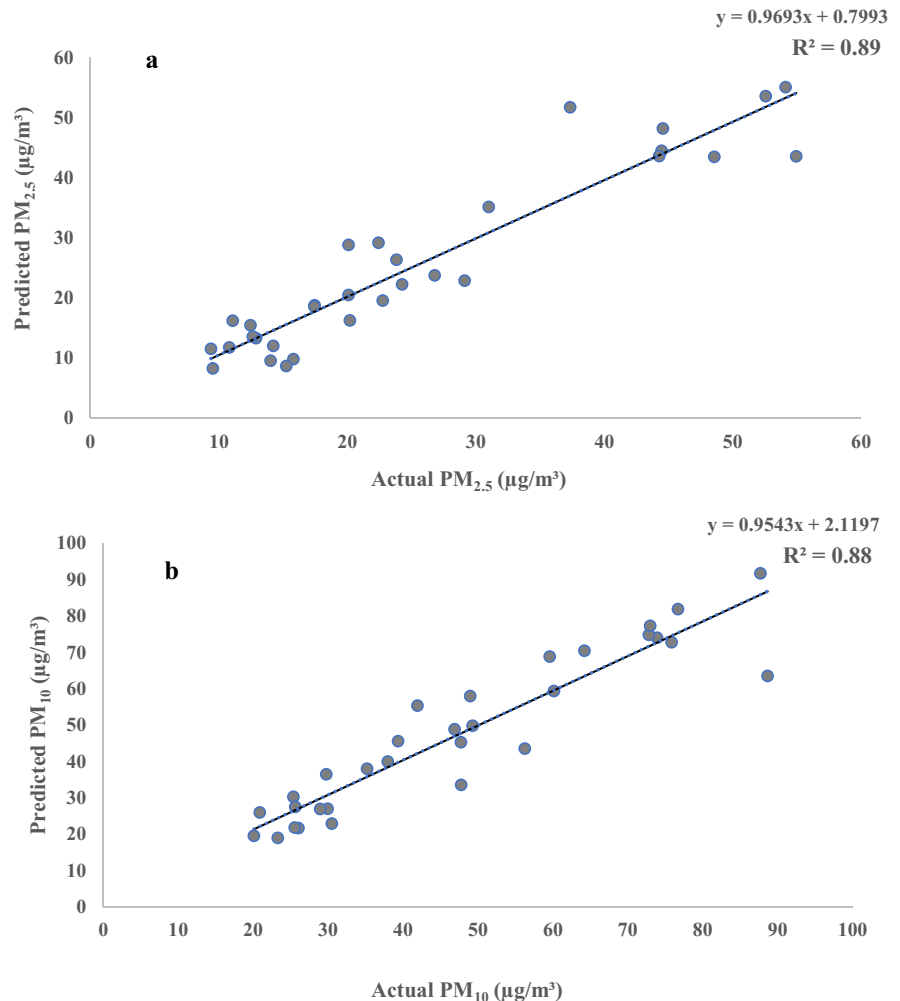
Models	MAE				RMSE			
	24 h	48 h	72 h	96–120 h	24 h	48 h	72 h	96–120 h
TLS-BDLSTM	5.30	7.63	8.79	16.55	7.32	10.37	13.18	25.11
Stacked-BDLSTM	5.80	8.60	9.86	18.13	7.98	10.61	13.99	25.66
LSTM	6.37	9.37	11.34	19.05	9.08	11.95	14.99	26.35
MLP	8.26	10.44	13.79	19.18	11.66	13.92	17.29	28.56
ARIMA	9.77	12.15	19.60	22.71	13.12	22.09	24.96	27.86
SVM	8.72	10.80	12.78	18.38	11.09	12.47	14.90	28.09
KNN	6.98	8.78	13.95	44.71	9.39	12.79	16.11	47.66

lockdown and partial lockdown (Table 8). This signifies the necessity to use and apply a transfer learning approach to improve prediction accuracy.

Table 8 also shows the comparative study of different ML and statistical models where it is found that KNN shows the poorest performance in prediction

of PM_{2.5} and PM₁₀ during pandemic situations, i.e., complete lockdown and partial lockdown periods. As it is mentioned in the sub-section “K-nearest neighbor (KNN) classifier” that KNN needs to be carefully tuned, the choice of K and the metric (distance) to be used are critical. On the other hand, KNN works well

Fig. 10 The diagram shows actual and predicted values of **a** PM_{2.5} and **b** PM₁₀ concentrations during validation period



with the small number of input variables but as the numbers of variables grow, KNN algorithm struggles to predict the output of new data points. The performance of LSTM, MLP, SVM, and ARIMA varies during the above-mentioned two case study periods, so no further inference is drawn from the results. The computational time of these models during complete lockdown, partial lockdown, and normal periods, corresponding to above-mentioned prediction accuracies in Tables 7 and 8, are shown in Table 9.

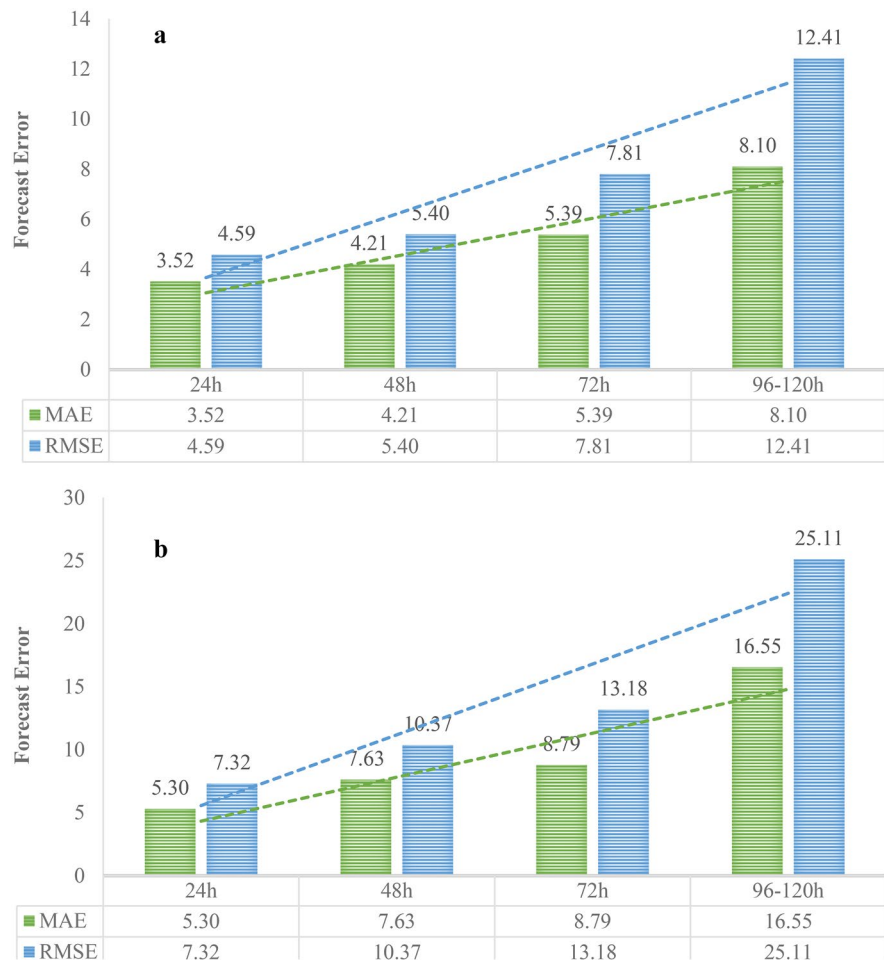
Validation

Validation of TLS-BDLSTM is executed with real-time data of complete lockdown, 2021, when the devastating second wave of coronavirus hit the country. This is a short time data with 1 month from 16 May–15 June 2021. During validation, we analyze the performance of our proposed model, i.e., TLS-BDLSTM, with the

multi-step prediction of $PM_{2.5}$ and PM_{10} concentrations using real-time data. The errors in predicting the concentrations with our proposed approach and other algorithms with 24 h, 48 h, 72 h, and 96–120-h lead time are computed. The prediction results from TLS-BDLSTM and six other models in terms of MAE and RMSE are compared in Tables 10 and 11, where several interesting observations are found. TLS-BDLSTM outperforms the other models when it comes to both $PM_{2.5}$ and PM_{10} multi-step prediction performance.

The result shows that the forecast errors (MAE and RMSE) with the TLS-BDLSTM model in estimating the concentration of $PM_{2.5}$ during validation are 3.52 and 4.59 (for 24-h lead time), 4.21 and 5.40 (for 48-h lead time), 5.39 and 7.81 (for 72-h lead time), and 8.10 and 12.41 (96–120-h lead time) (Table 10). While in forecasting PM_{10} concentrations during validation, the forecast errors (MAE, RMSE) with the TLS-BDLSTM are 5.30 and 7.32 (for 24-h lead time), 7.63 and 10.37 (for 48-h lead time), 8.79 and 13.18 (for 72-h lead time), and 16.55 and 25.11 (for 96–120-h lead time) (Table 11).

Fig. 11 The mean absolute error (MAE) and root mean square error (RMSE) for prediction of **a** $PM_{2.5}$ and **b** PM_{10} concentrations using TLS-BDLSTM during validation time



(for 48-h lead time), 8.79 and 13.18 (for 72-h lead time), and 16.55 and 25.11 (96–120-h lead time) (Table 11). For PM_{10} , 24-h prediction results of TLS-BDLSTM and stacked-BDLSTM are comparable (Table 11); however, with the increase in time, TLS-BDLSTM shows better performance than stacked-BDLSTM. The performance accuracy of traditional models, like MLP, SVM, KNN, and ARIMA, is comparatively lower than the deep learning models, like TLS-BDLSTM, stacked-BDLSTM, and LSTM. However, TLS-BDLSTM with transfer learning consistently outperforms the other methods in modeling long-term dependency for effective prediction of the concentration of $PM_{2.5}$ and PM_{10} . The scatter plot of observed and predicted values of $PM_{2.5}$ and PM_{10} through TLS-BDLSTM model with 24-h lead time is shown in Fig. 10, and the results are quite satisfactory, as the R^2 values are 0.89 and 0.88 for $PM_{2.5}$ and PM_{10} , respectively.

When comparing the other models, it is found that maximum error in terms of MAE and RMSE is observed in the case of KNN with 96–120-h prediction. So, it may be inferred that KNN fails to predict the long-term dependency in multivariate time series data. From the rest of the models, no unique inferences are reached. However, the performance of all models deteriorates as the time to predict becomes longer. Figure 11 depicts the MAE and RMSE for the prediction of the concentrations of $PM_{2.5}$ and PM_{10} during validation time using TLS-BDLSTM where it is observed that the forecast error of the model increases as the lead time increases. Furthermore, it is found that our model TLS-BDLSTM shows the minimum prediction error (RMSE and MAE) and maintains consistency in terms of best performance as the prediction time step increases when compared with other existing models (Tables 10 and 11). Therefore, TLS-BDLSTM with transfer learning can effectively learn long-term temporal dependency of multivariate time series data and has the best accuracy.

Conclusions

Kolkata metropolitan city, although not as polluted as Delhi, is still emerging as the second most polluted metro city of India (WHO, 2018). Like other metro cities, Kolkata has recognized several issues that contribute to air pollution. The coronavirus pandemic (COVID-19) unfolds unexpectedly around the world in early 2020, altering anthropogenic activities permanently.

Lockdown because of COVID-19 substantiates a “silver lining in the dark clouds” when the concentrations of air pollution have come down beneath their permissible limits which had been unusual in current previous and an uncommon expectation in the future too. During the lockdown period, the improved air quality standard ensures a healthy environment in Kolkata.

In the present study, we identify the most important air pollutants influencing the air quality of Kolkata during three aforesaid different periods where it is observed that $PM_{2.5}$ and PM_{10} are the main pollutants of Kolkata. The effect of the lockdown on $PM_{2.5}$ is the most significant which come up within the air of Kolkata due to diesel-driven vehicles, domestic and commercial combustion activities, road dust, and open burning. During the investigation, transfer learning is incorporated in the deep learning (DL) network (viz., stacked-bidirectional LSTM) to learn during the normal periods and forecast for concentrations of $PM_{2.5}$ and PM_{10} during the pandemic situation of two cases such as complete lockdown (case study 1) and partial lockdown (case study 2). This deep transfer learning-based model (TLS-BDLSTM) shows the least amount of forecast error in terms of RMSE and MAE for both the concentrations of $PM_{2.5}$ and PM_{10} than other machine learning, statistical models, and even stacked-BDLSTM without transfer learning.

The results obtained from TLS-BDLSTM reveal the following: the MAE and RMSE with 24-h lead time are respectively 2.56 and 6.13 for $PM_{2.5}$ concentrations and 2.47 and 5.24 for PM_{10} concentrations during the complete lockdown. The same during partial lockdown are 3.79 and 5.02 for $PM_{2.5}$ concentrations and 7.59 and 9.49 for concentrations of PM_{10} . The forecast errors all through real-time validation are minimum. During the validation with 24-h lead time, the MAE and RMSE are respectively 3.52 and 4.59 for $PM_{2.5}$ and 5.30 and 7.32 for PM_{10} . The observed and predicted concentrations during validation are quite acceptable, as the correlations are fairly high with R^2 values of 0.89 and 0.88 for $PM_{2.5}$ and PM_{10} , respectively. The model has superior forecasting ability not only in single step but also in multi-step for the prediction of the concentrations of both $PM_{2.5}$ and PM_{10} as compared to other approaches. The accuracy of the model, as expected, decreases as time to forecast becomes longer.

Forecast of short-term air quality can give accurate information about impending air pollution episodes and

to make the general people more aware of the potential changes in their city's air quality. Thus, it will be possible to implement the necessary mitigation strategies and take systematic action to improve the air quality and handle any associated health problems. Our proposed approach contributes to the above. The findings of our study help to increase the citizen awareness regarding how air pollution changes during different phases of lockdown and normal periods, resulting in significant consequences.

Traditionally, machine learning (ML) performance depends on a large amount of training data. ML-based prediction relies on one crucial assumption: the training and testing data are chosen from the same distribution. In many real-world problems, however, this assumption may not always hold true. As a result, most traditional machine learning algorithms face three major challenges: insufficient data, incompatible computation power, and distribution mismatch. The same is true for deep learning (DL) which is an advanced form of ML. Transfer learning (TL) has been able to address these issues in many critical application areas in case of insufficient well-labeled training data. Here, the time and the computational resources required to train a model are also considerably reduced, since pre-learned knowledge is reused. In our study concerning COVID-19 pandemic data, the size of labeled pollution data available during COVID lockdown periods is considerably smaller than that of normal period. Further, one may note that the prediction accuracy of pollutants $PM_{2.5}$ and PM_{10} gets significantly affected due to the complexity in their characteristics, such as non-linear properties in time and space. TL approach is found here to be a reliable and suitable technique for air quality forecasting in the course of the remarkable change in concentrations of pollutants and data shortage. Therefore, the operational organizations can use it as an additional model along with the conventional ones for prediction of air quality.

The present investigation has certain concerns because of some assumptions made, and these may lead to several scopes for future research in the areas we have highlighted. (1) TL considers that the target and source domains are different but related; some common instances or attributes can be transferred between domains. However, it prevents TL from being used in situations when the source and the target are only loosely connected. (2) Negative transfer occurs when there is a larger domain discrepancy. There are no specific standards on which tasks are related or how algorithms can determine which

tasks are related, thereby making the problem of negative transfer difficult to solve. (3) Detailed source apportionment study of air pollutants is not considered in the present study. Our findings encourage further research into combining these approaches.

Further, in the present study, we only considered the applications of some AI and ML techniques, although there exist several numerical models and fused (integrated) AI techniques. Studying them in the light of TL may be explored in the future. Similarly, some judicious integration between these ML techniques and other modern uncertainty handling models, say, soft computing paradigm, may be attempted. Moreover, in the present study, we have selected a deep architecture to be used for transfer learning; one may go for any other conventional ML algorithms, depending on the requirements and problems, to result in a suitable ML-based transfer learning model. As a whole, transfer learning broadens the scope of deep learning applications in the area of air quality prediction. Although our study focuses on the COVID-19 pandemic, the reliability of our predictions supports the concept that transfer learning-based modeling can quantify the impact of any event on air pollution. In addition, the proposed approach developed for Kolkata can be used for studying the same in any other urban city.

Acknowledgements The work was done when Prof. S.K. Pal held a National Science Chair, SERB-DST, Govt. of India.

Author contribution DD formulated the research problem, wrote the programs, and made the first draft. SKP is the mentor and principal investigator who gave the guidance in machine learning aspects of the paper and provided corrections of the overall manuscript for better organization and understanding.

Funding SERB-National Science Chair, Govt. of India, awarded to Prof. Sankar K. Pal.

Data availability The corresponding author will provide data supporting the findings of this study upon reasonable request.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Yes.

Competing interests The authors declare no competing interests.

References

Arief-Ang, I. B., Hamilton, M., & Salim, F. D. (2018). A scalable room occupancy prediction with transferable time series decomposition of CO 2 sensor data. *ACM Transactions on Sensor Networks*, 14(3–4), 1–28. <https://doi.org/10.1145/3217214>

Atyabi, A., Luerssen, M. H., & Powers, D. M. W. (2013). PSO-based dimension reduction of EEG recordings: Implications for subject transfer in BCI. *Neurocomputing*, 119, 319–331. <https://doi.org/10.1016/j.neucom.2013.03.027>

Bauwens, M., Compernelle, S., Stavrakou, T., Müller, J.-F., van Gent, J., Eskes, H., Levelt, P. F., vanderA, R., Veeffkind, J. P., Vlietinck, J., Yu, H., & Zehner, C. (2020). Impact of coronavirus outbreak on NO2 pollution assessed using TROPOMI and OMI observations. *Geophysical Research Letters*, 47(11), e2020GL087978. <https://doi.org/10.1029/2020GL087978>

Bayraktar, E. P., Isinkaralar, O., & Isinkaralar, K. (2022). Usability of several species for monitoring and reducing the heavy metal pollution threatening the public health in urban environment of Ankara. *World Journal of Advanced Research and Reviews*, 14(3), 276–283. <https://doi.org/10.30574/wjarr.2022.14.3.0553>

Bel, N., Koster, C. H. A., & Villegas, M. (2003). Cross-lingual text categorization. In *Research and Advanced Technology for Digital Libraries* (pp. 126–139). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-45175-4_13

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2), 157–166. <https://doi.org/10.1109/72.279181>

Bera, B., Bhattacharjee, S., Shit, P. K., Sengupta, N., & Saha, S. (2020). Significant impacts of COVID-19 lockdown on urban air pollution in Kolkata (India) and amelioration of environmental health. *Environment Development and Sustainability*, 23(5), 6913–6940. <https://doi.org/10.1007/s10668-020-00898-5>

Berman, J. D., & Ebisu, K. (2020). Changes in U.S. air pollution during the COVID-19 pandemic. *The Science of the Total Environment*, 739(139864), 139864. <https://doi.org/10.1016/j.scitotenv.2020.139864>

Blitzer, J., McDonald, R., & Pereira, F. (2006). Domain adaptation with structural correspondence learning. *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing - EMNLP '06*. <https://doi.org/10.3115/1610075.1610094>

Boulesteix, A.-L., Janitza, S., Kruppa, J., & König, I. R. (2012). Overview of Random Forest methodology and practical guidance with emphasis on computational biology and bioinformatics: Random Forests in bioinformatics. *Wiley Interdisciplinary Reviews. Data Mining and Knowledge Discovery*, 2(6), 493–507. <https://doi.org/10.1002/widm.1072>

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/a:1010933404324>

Brook, R. D., Rajagopalan, S., Pope, C. A., III., Brook, J. R., Bhatnagar, A., Diez-Roux, A. V., Holguin, F., Hong, Y., Luepker, R. V., Mittleman, M. A., Peters, A., Siscovick, D., Smith, S. C., Jr., Whitsel, L., & Kaufman, J. D. (2010). Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association. *Circulation*, 121(21), 2331–2378. <https://doi.org/10.1161/cir.0b013e3181d8bec1>

Castelli, M., Clemente, F. M., Popovič, A., Silva, S., & Vanneschi, L. (2020). A machine learning approach to predict air quality in California. *Complexity*, 2020, 1–23. <https://doi.org/10.1155/2020/8049504>

Cesur, A., Zeren Cetin, I., Cetin, M., Sevik, H., & Ozel, H. B. (2022). The use of Cupressus arizonica as a biomonitor of Li, Fe, and Cr pollution in Kastamonu. *Water, Air, and Soil Pollution*, 233(6). <https://doi.org/10.1007/s11270-022-05667-w>

Cetin, M. (2016). A change in the amount of CO2 at the center of the examination halls: Case study of Turkey. *Studies on Ethno-Medicine*, 10(2), 146–155. <https://doi.org/10.1080/09735070.2016.11905483>

Cetin, M., & Sevik, H. (2016). Measuring the impact of selected plants on indoor CO2 concentrations. *Polish Journal of Environmental Studies*, 25(3), 973–979. <https://doi.org/10.15244/pjoes/61744>

Cetin, M., Onac, A. K., Sevik, H., & Sen, B. (2019). Temporal and regional change of some air pollution parameters in Bursa. *Air Quality, Atmosphere, & Health*, 12(3), 311–316. <https://doi.org/10.1007/s11869-018-00657-6>

Cetin, M., Aljama, A. M. O., Alrabiti, O. B. M., Adiguzel, F., Sevik, H., & Zeren Cetin, I. (2022). Determination and mapping of regional change of Pb and Cr pollution in Ankara city center. *Water, Air, and Soil Pollution*, 233(5). <https://doi.org/10.1007/s11270-022-05638-1>

Cetin, M., & Jawed, A. A. (2022). Variation of Ba concentrations in some plants grown in Pakistan depending on traffic density. *Biomass Conversion and Biorefinery*. <https://doi.org/10.1007/s13399-022-02334-2>

Chaudhuri, S., & Dutta, D. (2014). Mann-Kendall trend of pollutants, temperature and humidity over an urban station of India with forecast verification using different ARIMA models. *Environmental Monitoring and Assessment*, 186(8), 4719–4742. <https://doi.org/10.1007/s10661-014-3733-6>

Chaudhuri, S., Dutta, D., Goswami, S., & Middey, A. (2015). Track and intensity forecast of tropical cyclones over the North Indian Ocean with multilayer feed forward neural nets: Track and intensity forecast of tropical cyclones over NIO. *Meteorological Applications*, 22(3), 563–575. <https://doi.org/10.1002/met.1488>

Cifuentes, J., Marulanda, G., Bello, A., & Reneses, J. (2020). Air temperature forecasting using Machine Learning techniques: A review. *Energies*, 13(16), 4215. <https://doi.org/10.3390/en13164215>

Chakraborty, D. B., & Pal, S. K. (2021). Granular video computing: With rough sets, deep learning and in IoT. *World Scientific Publishing*. <https://doi.org/10.1142/12013>

CPCB. (2020). A report on impact of lockdown on ambient air quality. <https://static.pib.gov.in/WriteReadData/userfiles/AAQ%20-Final%2022-09-2%20A.pdf>. Accessed 17 Jun 2021

CSIR-NEERI. (2019). PM₁₀ and PM_{2.5} source apportionment study and development of emission inventory of twin cities Kolkata

- and Howrah of West Bengal. Sponsored by West Bengal Pollution Control Board, Kolkata. https://www.wbpcb.gov.in/writereaddata/files/SA_Kol-How_Final%20Report.pdf. Accessed 17 Jun 2021
- Dai, W., Xue, G.-R., Yang, Q., & Yu, Y. (2007). Co-clustering based classification for out-of-domain documents. *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '07*. San Jose, California, 210–219. <https://doi.org/10.1145/1281192.1281218>
- Dhole, A., Ambekar, I., Gunjan, G., & Sonawani, S. (2021). An ensemble approach to multi-source transfer learning for air quality prediction. *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*.
- Ding, J., Eskes, H. J., Mijling, B., Stavrakou, T., Geffen, J. H. G. M., & Veeffkind, J. P. (2020). NOx emissions reduction and rebound in China due to the COVID-19 crisis. *Geophysical Research Letters*, 47(19). <https://doi.org/10.1029/2020gl089912>
- Dutta, D., & Chaudhuri, S. (2015). Nowcasting visibility during wintertime fog over the airport of a metropolis of India: Decision tree algorithm and artificial neural network approach. *Natural Hazards (dordrecht, Netherlands)*, 75(2), 1349–1368. <https://doi.org/10.1007/s11069-014-1388-9>
- Dutta, D., & Pal, S. K. (2022a). Interpretation of black box for short-term predictions of pre-monsoon cumulonimbus cloud events over Kolkata. *Journal of Data, Information and Management*, 4(2), 167–183. <https://doi.org/10.1007/s42488-022-00071-9>
- Dutta, D., & Pal, S. K. (2022b). Z-number-based AQI in rough set theoretic framework for interpretation of air quality for different thresholds of PM2.5 and PM10. *Environmental Monitoring and Assessment*, 194(9), 653. <https://doi.org/10.1007/s10661-022-10325-z>
- Dragomir, E. G. (2010). Air quality index prediction using K-nearest neighbor technique. *Bulletin of PG University of Ploiesti, Series Mathematics, Informatics, Physics, LXII 1(2010)*, 103–108.
- Elsunousi, A. A. M., Sevik, H., Cetin, M., Ozel, H. B., & Ozel, H. U. (2021). Periodical and regional change of particulate matter and CO2 concentration in Misurata. *Environmental Monitoring and Assessment*, 193(11), 707. <https://doi.org/10.1007/s10661-021-09478-0>
- Esteves, J. T., de Souza Rolim, G., & Ferraudo, A. S. (2019). Rainfall prediction methodology with binary multilayer perceptron neural networks. *Climate Dynamics*, 52(3–4), 2319–2331. <https://doi.org/10.1007/s00382-018-4252-x>
- Fong, I. H., Li, T., Fong, S., Wong, R. K., & Tallón-Ballesteros, A. J. (2020). Predicting concentration levels of air pollutants by transfer learning and recurrent neural network. *Knowledge-Based Systems*, 105622, 105622. <https://doi.org/10.1016/j.knsys.2020.105622>
- Fraivan, L., & Alkhdari, M. (2020). Investigating the use of uni-directional and bi-directional long short-term memory models for automatic sleep stage scoring. *Informatics in Medicine Unlocked*, 20(100370), 100370. <https://doi.org/10.1016/j.imu.2020.100370>
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>
- García Nieto, P. J., Combarro, E. F., del Coz Díaz, J. J., & Montañés, E. (2013). A SVM-based regression model to study the air quality at local scale in Oviedo urban area (Northern Spain): A case study. *Applied Mathematics and Computation*, 219(17), 8923–8937. <https://doi.org/10.1016/j.amc.2013.03.018>
- Gautam, S. (2020). The influence of COVID-19 on air quality in India: A boon or Inutile. *Bulletin of Environmental Contamination and Toxicology*, 104(6), 724–726. <https://doi.org/10.1007/s00128-020-02877-y>
- Gers, F. A., Schmidhuber, J., & Cummins, F. (1999). Continual prediction using LSTM with forget gates. In *Neural Nets WIRN Vietri-99* (pp. 133–138). Springer London. https://doi.org/10.1007/978-1-4471-0877-1_10
- Ghose, M. K., Paul, R., & Banerjee, S. K. (2004). Assessment of the impacts of vehicular emissions on urban air quality and its management in Indian context: The case of Kolkata (Calcutta). *Environmental Science & Policy*, 7(4), 345–351. <https://doi.org/10.1016/j.envsci.2004.05.004>
- Gilik, A., Ogrenci, A. S., & Ozmen, A. (2022). Air quality prediction using CNN+LSTM-based hybrid deep learning architecture. *Environmental Science and Pollution Research International*, 29(8), 11920–11938. <https://doi.org/10.1007/s11356-021-16227-w>
- Goldberg, D. L., Anenberg, S. C., Griffin, D., McLinden, C. A., Lu, Z., & Streets, D. G. (2020). Disentangling the impact of the COVID-19 lockdowns on urban NO2 from natural variability. *Geophysical Research Letters*, 47(17), e2020GL089269. <https://doi.org/10.1029/2020GL089269>
- Guo, C., Liu, G., & Chen, C.-H. (2020). Air pollution concentration forecast method based on the deep ensemble neural network. *Wireless Communications and Mobile Computing*, 2020, 1–13. <https://doi.org/10.1155/2020/8854649>
- Heydari, A., MajidiNezhad, M., Astiaso Garcia, D., Keynia, F., & De Santoli, L. (2021). Air pollution forecasting application based on deep learning model and optimization algorithm. *Clean Technologies and Environmental Policy*. <https://doi.org/10.1007/s10098-021-02080-5>
- Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A.-R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T., & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6), 82–97. <https://doi.org/10.1109/msp.2012.2205597>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahuvaroglu, T., Morrison, J., & Giovis, C. (2005). A review and evaluation of intraurban air pollution exposure models. *Journal of Exposure Analysis and Environmental Epidemiology*, 15(2), 185–204. <https://doi.org/10.1038/sj.jea.7500388>
- Ji, X., Shang, X., Dahlgren, R. A., & Zhang, M. (2017). Prediction of dissolved oxygen concentration in hypoxic river systems using support vector machine: A case study of Wen-Rui Tang River, China. *Environmental Science and Pollution Research International*, 24(19), 16062–16076. <https://doi.org/10.1007/s11356-017-9243-7>
- Jo, J., Jo, B., Kim, J., Kim, S., & Han, W. (2020). Development of an IoT-based indoor air quality monitoring platform.

- Journal of Sensors*, 2020, 1–14. <https://doi.org/10.1155/2020/8749764>
- Kharroubi, S., & Saleh, F. (2020). Are lockdown measures effective against COVID-19? *Frontiers in Public Health*, 8, 549692. <https://doi.org/10.3389/fpubh.2020.549692>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90. <https://doi.org/10.1145/3065386>
- Kumar, P., Hama, S., Omidvarborna, H., Sharma, A., Sahani, J., Abhijith, K. V., Debele, S. E., Zavala-Reyes, J. C., Barwise, Y., & Tiwari, A. (2020). Temporary reduction in fine particulate matter due to “anthropogenic emissions switch-off” during COVID-19 lockdown in Indian cities. *Sustainable Cities and Society*, 62(102382), 102382. <https://doi.org/10.1016/j.scs.2020.102382>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444. <https://doi.org/10.1038/nature14539>
- Lepioufle, J.-M., Marsteen, L., & Johnsrud, M. (2021). Error prediction of air quality at monitoring stations using Random Forest in a total error framework. *Sensors (basel, Switzerland)*, 21(6), 2160. <https://doi.org/10.3390/s21062160>
- Lei, T. M. T., Siu, S. W. I., Monjardino, J., Mendes, L., & Ferreira, F. (2022). Using machine learning methods to forecast air quality: A case study in Macao. *Atmosphere*, 13(9), 1412. <https://doi.org/10.3390/atmos13091412>
- Li, Z., Guo, J., Ding, A., Liao, H., Liu, J., Sun, Y., Wang, T., Xue, H., Zhang, H., & Zhu, B. (2017). Aerosol and boundary-layer interactions and impact on air quality. *National Science Review*, 4(6), 810–833. <https://doi.org/10.1093/nsr/nwx117>
- Ling, X., Dai, W., Xue, G.-R., Yang, Q., & Yu, Y. (2008). Spectral domain-transfer learning. *Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD 08*. <https://doi.org/10.1145/1401890.1401951>
- Liu, B.-C., Binaykia, A., Chang, P.-C., Tiwari, M. K., & Tsao, C.-C. (2017). Urban air quality forecasting based on multi-dimensional collaborative support vector regression (SVR): A case study of Beijing-Tianjin-Shijiazhuang. *PLoS ONE*, 12(7), e0179763. <https://doi.org/10.1371/journal.pone.0179763>
- Liu, K., Hu, X., Zhou, H., Tong, L., Widanage, W. D., & Marco, J. (2021). Feature analyses and modeling of lithium-ion battery manufacturing based on Random Forest classification. *IEEE/ASME Transactions on Mechatronics*, 26(6), 2944–2955. <https://doi.org/10.1109/tmech.2020.3049046>
- Lu, J., Behbood, V., Hao, P., Zuo, H., Xue, S., & Zhang, G. (2015). Transfer learning using computational intelligence: A survey. *Knowledge-Based Systems*, 80, 14–23. <https://doi.org/10.1016/j.knsys.2015.01.010>
- Ma, J., Ding, Y., Gan, V. J. L., Lin, C., & Wan, Z. (2019a). Spatiotemporal prediction of PM_{2.5} concentrations at different time granularities using IDW-BLSTM. *IEEE Access: Practical Innovations, Open Solutions*, 7, 107897–107907. <https://doi.org/10.1109/access.2019.2932445>
- Ma, J., Cheng, J. C. P., Lin, C., Tan, Y., & Zhang, J. (2019). Improving air quality prediction accuracy at larger temporal resolutions using deep learning and transfer learning techniques. *Atmospheric Environment (Oxford, England: 1994)*, 214(116885), 116885. <https://doi.org/10.1016/j.atmosenv.2019.116885>
- Ma, J., Li, Z., Cheng, J. C. P., Ding, Y., Lin, C., & Xu, Z. (2020). Air quality prediction at New Stations using spatially transferred bi-directional long short-term memory network. *Science of the Total Environment*, 705, 135771. <https://doi.org/10.1016/j.scitotenv.2019.135771>
- Masood, A., & Ahmad, K. (2021). A review on emerging artificial intelligence (AI) techniques for air pollution forecasting: Fundamentals, application and performance. *Journal of Cleaner Production*, 322(129072), 129072. <https://doi.org/10.1016/j.jclepro.2021.129072>
- Mahato, S., Pal, S., & Ghosh, K. G. (2020). Effect of lockdown amid COVID-19 pandemic on air quality of the megacity Delhi, India. *The Science of the Total Environment*, 730(139086), 139086. <https://doi.org/10.1016/j.scitotenv.2020.139086>
- Mhawish, A., Banerjee, T., Broday, D. M., Misra, A., & Tripathi, S. N. (2017). Evaluation of MODIS Collection 6 aerosol retrieval algorithms over Indo-Gangetic Plain: Implications of aerosols types and mass loading. *Remote Sensing of Environment*, 201, 297–313. <https://doi.org/10.1016/j.rse.2017.09.016>
- Mitra, S., Sarbadhikari, S. N., & Pal, S. K. (1996). An MLP-based model for identifying qEEG in depression. *International Journal of Bio-Medical Computing*, 43(3), 179–187. [https://doi.org/10.1016/s0020-7101\(96\)01203-2](https://doi.org/10.1016/s0020-7101(96)01203-2)
- Montgomery, D. C., & Johnson, L. A. (1976). *Forecasting and time series analysis*. McGraw Hill Higher Education.
- Motesaddi, S., Hashempour, Y., & Nowrouz, P. (2017). Characterizing of air pollution in Tehran: Comparison of two air quality indices. *Civil Engineering Journal*, 3(9), 749–758. <https://doi.org/10.21859/cej-030911>
- Navinya, C., Patidar, G., & Phuleria, H. C. (2020). Examining effects of the COVID-19 national lockdown on ambient air quality across Urban India. *Aerosol and Air Quality Research*, 20(8), 1759–1771. <https://doi.org/10.4209/aaqr.2020.05.0256>
- Ong, B. T., Sugiura, K., & Zetsu, K. (2015). Dynamically pre-trained deep recurrent neural networks using environmental monitoring data for predicting PM_{2.5}. *Neural Computing and Applications*, 27, 1553–1566. <https://doi.org/10.1007/s00521-015-1955-3>
- Pal, S. K., Bhoumik, D., & Chakraborty, B. (2019). Granulated deep learning and Z-numbers in motion detection and object recognition. *Neural Computing and Applications*, 32, 16533–16548. <https://doi.org/10.1007/s00521-019-04200-1>
- Pal, S. K., & Mitra, S. (1999). *Neuro-fuzzy pattern recognition: Methods in soft computing*. Wiley.
- Pal, S. K., Meher, S. K., & Dutta, S. (2012). Class-dependent rough-fuzzy granular space, dispersion index and classification. *Pattern Recognition*, 45(7), 2690–2707. <https://doi.org/10.1016/j.patcog.2011.12.027>
- Pal, S. K., Pramanik, A., Maiti, J., & Mitra, P. (2021). Deep learning in multi-object detection and tracking: State of the art. *Applied Intelligence*, 51(9), 6400–6429. <https://doi.org/10.1007/s10489-021-02293-7>
- Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345–1359. <https://doi.org/10.1109/tkde.2009.191>
- Plaia, A., & Bondi, A. (2006). Single imputation method of missing values in environmental pollution data sets. *Atmospheric Environment*, 40(38), 7316–7330. <https://doi.org/10.1016/j.atmosenv.2006.06.040>

- Pope, C. A., 3rd., Burnett, R. T., Thurston, G. D., Thun, M. J., Calle, E. E., Krewski, D., & Godleski, J. J. (2004). Cardiovascular mortality and long-term exposure to particulate air pollution: Epidemiological evidence of general pathophysiological pathways of disease: Epidemiological evidence of General pathophysiological pathways of disease. *Circulation*, *109*(1), 71–77. <https://doi.org/10.1161/01.CIR.0000108927.80044.7F>
- Prasad, K., Gorai, A. K., & Goyal, P. (2016). Development of ANFIS models for air quality forecasting and input optimization for reducing the computational cost and time. *Atmospheric Environment (Oxford, England: 1994)*, *128*, 246–262. <https://doi.org/10.1016/j.atmosenv.2016.01.007>
- Qi, Z., Wang, T., Song, G., Hu, W., Li, X., & Zhang, Z. (2018). Deep air learning: Interpolation, prediction, and feature analysis of fine-grained air quality. *IEEE Transactions on Knowledge and Data Engineering*, *30*(12), 2285–2297. <https://doi.org/10.1109/tkde.2018.2823740>
- Qin, Z., Cen, C., & Guo, X. (2019). Prediction of air quality based on KNN-LSTM. *Journal of Physics. Conference Series*, *1237*(4), 042030. <https://doi.org/10.1088/1742-6596/1237/4/042030>
- Sánchez-Sánchez, P. A., García-González, J. R., & Coronell, L. H. P. (2020). Encountered problems of time series with neural networks: Models and architectures. In *Recent Trends in Artificial Neural Networks - from Training to Prediction*. IntechOpen. <https://doi.org/10.5772/intechopen.88901>
- Sarkar, M., Das, A., & Mukhopadhyay, S. (2021). Assessing the immediate impact of COVID-19 lockdown on the air quality of Kolkata and Howrah, West Bengal, India. *Environment Development and Sustainability*, *23*(6), 1–30. <https://doi.org/10.1007/s10668-020-00985-7>
- Sevik, H., Cetin, M., Uzunozel, H., Ozel, H. B., Mossi, M. M., & Zeren Cetin, I. (2020). Determination of Pb and Mg accumulation in some of the landscape plants in shrub forms. *Environmental Science and Pollution Research International*, *27*(2), 2423–2431. <https://doi.org/10.1007/s11356-019-06895-0>
- Sharma, S., Zhang, M., Anshika, Gao, J., Zhang, H., & Kota, S. H. (2020). Effect of restricted emissions during COVID-19 on air quality in India. *The Science of the Total Environment*, *728*(138878), 138878. <https://doi.org/10.1016/j.scitotenv.2020.138878>
- Shi, X., & Brasseur, G. P. (2020). The response in air quality to the reduction of Chinese economic activities during the COVID-19 outbreak. *Geophysical Research Letters*, *47*(11), e2020GL088070. <https://doi.org/10.1029/2020GL088070>
- Stabile, L., Dell'Isola, M., Russi, A., Massimo, A., & Buonanno, G. (2017). The effect of natural ventilation strategy on indoor air quality in schools. *The Science of the Total Environment*, *595*, 894–902. <https://doi.org/10.1016/j.scitotenv.2017.03.048>
- Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, *14*(4), 323–348. <https://doi.org/10.1037/a0016973>
- Sun, Q., & Ulintz, L. (2016). Ambient particulate matter pollution on lipid peroxidation in cardiovascular diseases. *Environmental Disease*, *1*(4), 109. <https://doi.org/10.4103/2468-5690.198616>
- Tao, Q., Liu, F., Li, Y., & Sidorov, D. (2019). Air pollution forecasting using a deep learning model based on 1D convnets and bidirectional GRU. *IEEE Access: Practical Innovations, Open Solutions*, *7*, 76690–76698. <https://doi.org/10.1109/access.2019.2921578>
- Tariq, S., Loy-Benitez, J., Nam, K., Lee, G., Kim, M., Park, D., & Yoo, C. (2021). Transfer learning driven sequential forecasting and ventilation control of PM2.5 associated health risk levels in underground public facilities. *Journal of Hazardous Materials*, *406*(124753), 124753. <https://doi.org/10.1016/j.jhazmat.2020.124753>
- Tatavarti, V. S. N. R., Ganesh, S. S., & Arulmozhivarman, P. (2018). Air quality index forecasting using artificial neural networks - A case study on Delhi. *International Journal of Environment and Waste Management*, *22*(1/2/3/4), 4. <https://doi.org/10.1504/ijewm.2018.10014709>
- Tella, A., & Balogun, A.-L. (2021). GIS-based air quality modeling: Spatial prediction of PM10 for Selangor State, Malaysia using machine learning algorithms. *Environmental Science and Pollution Research International*. <https://doi.org/10.1007/s11356-021-16150-0>
- Torrey, L., & Shavlik, J. (2010). Transfer learning. In: *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques* (pp 242–264). IGI Global. <https://doi.org/10.4018/978-1-60566-766-9.ch011>
- Tu, W., & Sun, S. (2012). A subject transfer framework for EEG classification. *Neurocomputing*, *82*, 109–116. <https://doi.org/10.1016/j.neucom.2011.10.024>
- Turias, I. J., González, F. J., Martín, M. L., & Galindo, P. L. (2008). Prediction models of CO, SPM and SO(2) concentrations in the Campo de Gibraltar Region, Spain: A multiple comparison strategy. *Environmental Monitoring and Assessment*, *143*(1–3), 131–146. <https://doi.org/10.1007/s10661-007-9963-0>
- Wang, P., Chen, K., Zhu, S., Wang, P., & Zhang, H. (2020). Severe air pollution events not avoided by reduced anthropogenic activities during COVID-19 outbreak. *Resources, Conservation, and Recycling*, *158*(104814), 104814. <https://doi.org/10.1016/j.resconrec.2020.104814>
- Wang, Q., & Su, M. (2020). A preliminary assessment of the impact of COVID-19 on environment - A case study of China. *The Science of the Total Environment*, *728*(138915), 138915. <https://doi.org/10.1016/j.scitotenv.2020.138915>
- Wang, Y., Xue, S., & Ding, J. (2021). Research on water pollution prediction of township enterprises based on support vector regression machine. *E3S Web of Conferences*, *228*, 02014. <https://doi.org/10.1051/e3sconf/202122802014>
- World Health Organization (WHO). (2018). WHO global ambient air quality database (update 2018). Ambient Air Quality Database (update 2018).
- Wu, P., Ding, Y., & Liu, Y. (2017). Atmospheric circulation and dynamic mechanism for persistent haze events in the Beijing-Tianjin-Hebei region. *Advances in Atmospheric Sciences*, *34*(4), 429–440. <https://doi.org/10.1007/s00376-016-6158-z>
- Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural language processing [Review Article]. *IEEE Computational Intelligence Magazine*, *13*, 55–75. <https://doi.org/10.1109/mci.2018.2840738>

- Yu, R., Yang, Y., Yang, L., Han, G., & Move, O. A. (2016). RAQ-A Random Forest approach for predicting air quality in urban sensing systems. *Sensors*, *16*(1), 86. <https://doi.org/10.3390/s16010086>
- Yu, Z., Ramanarayanan, V., Suendermann-Oeft, D., Wang, X., Zechner, K., Chen, L., Tao, J., Ivanou, A., & Qian, Y. (2015). Using bidirectional lstm recurrent neural networks to learn high-level abstractions of sequential features for automated scoring of non-native spontaneous speech. *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*. <https://doi.org/10.1109/asru.2015.7404814>
- Zambrano-Monserrate, M. A., Ruano, M. A., & Sanchez-Alcalde, L. (2020). Indirect effects of COVID-19 on the environment. *The Science of the Total Environment*, *728*(138813), 138813. <https://doi.org/10.1016/j.scitotenv.2020.138813>
- Zhang, X., Yang, Y. H., Han, Z., Wang, H., & Gao, C. (2013). Object class detection. *ACM Computing Surveys*, *46*, 1–53. <https://doi.org/10.1145/2522968.2522978>
- Zhao, J., Deng, F., Cai, Y., & Chen, J. (2019). Long short-term memory - Fully connected (LSTM-FC) neural network for PM2.5 concentration prediction. *Chemosphere*, *220*, 486–492. <https://doi.org/10.1016/j.chemosphere.2018.12.128>
- Zhao, N., Fan, Z., & Zhao, M. (2021). A new approach for estimating dissolved oxygen based on a high-accuracy surface modeling method. *Sensors (basel, Switzerland)*, *21*(12), 3954. <https://doi.org/10.3390/s21123954>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.