Journal of Environmental Management 327 (2023) 116911

Contents lists available at ScienceDirect

Journal of Environmental Management

journal homepage: www.elsevier.com/locate/jenvman

Predictive modeling and analysis of air quality – Visualizing before and during COVID-19 scenarios

Jinil Persis [a,*], Amine Ben Amar [b]

[a] Indian Institute of Management (IIM), Kozhikode, Kerala, India
[b] Excelia Business School, La Rochelle, France

ARTICLE INFO

Keywords:
COVID-19
Air pollution
Air quality index
Machine learning

ABSTRACT

Quality air to breathe is the basic necessity for an individual and in recent times, emission from various sources caused by human activities has resulted in substantial degradation in the air quality. This work focuses to study the inadvertent effect of COVID-19 lockdown on air pollution. Pollutants' concentration before- and during-COVID-19 lockdown is captured to understand the variation in air quality. Firstly, multi-pollutant profiling using hierarchical cluster analysis of pollutants' concentration is performed that highlights the differences in the cluster compositions between before- and during-lockdown time periods. Results show that the particulate matter (PM_{10} and $PM_{2.5}$) in air that formed the primary cluster before lock-down, came down to close similarity with other clusters during lockdown. Secondly, predicting air quality index (AQI) based on the forecasts of pollutants' concentration is performed using neural networks, support vector machine, decision tree, random forest, and boosting algorithms. The best-fitted models representing AQI is identified separately for before- and during-lockdown time periods based on its predictive power. While deterministic method reactively evaluates present AQI when current pollutants' concentration at a particular time and place are known, this study uses the best fitted data-driven model to determine future AQIs based on the forecasts of pollutant's concentration accurately (overall RMSE<0.1 for before lockdown scenario and <0.3 for during lockdown scenario). The study contributes to visualize the variation in pollutants' concentrations between the two scenarios. The results show that the reduced economic activities during lockdown period had led to the drop in concentration of PM_{10} and $PM_{2.5}$ by 27% and 50% on an average. The findings of this study have practical and societal implications and serve as a reference mechanism for policymakers and governing bodies to revise their actions plans for regulating individual air pollutants in the atmospheric air.

1. Introduction

Rapid industrialization and urbanization in low- and middle-income countries have led to increased air pollution causing environmental degradation and health hazards (Balietti et al., 2022). Combustion of fossil fuels, industrial emissions, and natural calamities are potential causes of air pollution (Fromer et al., 2019; Jiang and Yu, 2020). Human activities contribute much to air pollution. $PM_{2.5}$, PM_{10} (grade of particulate matter having diameter less than or equal to 10, 5, 2.5, 1 μm is shown in subscript) in cities are emitted predominantly from vehicles (X. Ma et al., 2020; Mukherjee et al., 2020; Yang et al., 2020). Pollutants like carbon monoxide (CO), hydrocarbon (HC), Nitrogen oxides (NO_X), $PM_{2.5}$, $PM_{1.0}$ and volatile organic compounds (VOCs) are extensively present in business zones of urban area and NO_X, $PM_{2.5}$ and $PM_{1.0}$ are highly present in air in the industrial zones (Song et al., 2019). Further, diffusion and persistence of these pollutants in air is greatly affected by meteorological factors such as wind speed, direction, turbulence, stability, humidity, temperature, radiation and rainfall (Hewson, 1956). Pollutants in air when inhaled causes serious health hazards and give rise to ecological disturbances against which eco-friendly industrial practices, agricultural practices, plant designs and control technologies are suggested by many researchers (Jiang and Yu, 2020; Liu et al., 2020; Wang and Lu, 2020; Xiong et al., 2020; C. Zhang et al., 2020; Zhuang et al., 2020).

As part of preventive and control measures to combat air pollution, regulations for vehicles, industries, and power plants are enforced; alternative fuels and zero-emission vehicles are introduced (Bakır et al., 2022; Guo et al., 2020; Sahu et al., 2021; Saz-Salazar et al., 2020; Yadav

* Corresponding author.
E-mail addresses: jinil@iimk.ac.in, jinil@iimk.ac.in (J. Persis), amine.benamar@gmail.com (A. Ben Amar).

https://doi.org/10.1016/j.jenvman.2022.116911
Received 23 May 2021; Received in revised form 26 September 2022; Accepted 26 November 2022
Available online 1 December 2022

et al., 2019). (Perman and Stern, 2003) argues that the environmental improvement in developed countries is due, at least in part, to the fairly strict environmental regulations. However, such regulations are not in line with the priorities of the low- and middle-income countries that specialize in natural resources and low-skilled labor (Grossman and Krueger, 1995; Tsurumi and Managi, 2010). have highlighted the economic determinants of environmental quality since large-scale economic activities, which exacerbates industrial combustion, result in high emission volume of VOC and PM in the atmospheric air. India rigorously focus on their mission toward “clean air” to tackle the nexus between economic development and environmental degradation (Sahu et al., 2021). The effect of globalization on air pollution are alarmingly felt in Indian cities and India has six out of ten most polluted cities in the world (as per 2021 IQAir records). For instance, World Health Organization (WHO) has already indicated that Delhi has exceeded the maximum PM_{10} limit by ten times in 2011 itself and the residents face lot of health difficulties due to polluted air. Hence, there is a need to control air pollution caused by individual pollutants at its source and regulate air quality to build a healthy nation.

While preventive measures through regulations and policies have been focused on, reactive strategies too are adopted conditionally based on the current value of the air quality index (AQI) which is estimated from the pollutants’ concentration in the air sensed at an instant of time (Tan et al., 2021). When AQI drops alarmingly down, factories adopt mitigation plan to clean the air through the use of filters and purifiers (Jiang and Yu, 2020; Minet et al., 2018; Vijayan et al., 2015). However, appropriate proactive strategies help prevent or control pollution that would be possible if AQI is predicted accurately, and further, the situation can be safely handled with necessary precautions and preparedness which is the focus of this study (Qiu et al., 2020; Ying Wang et al., 2020).

According to NASA (2020), during the COVID-19 lockdown, AQI level was found to be lesser than normal and pollutants’ concentration such as nitrogen dioxide in the atmospheric air was detected to be less (Dang and Trinh, 2021). elucidates the overall reduction in concentration of nitrogen dioxide and $PM_{2.5}$ globally during lockdown and suggests to adopt reduced mobility as much as possible further to sustain these improved air conditions in future. It is noteworthy that forecasts made by the existing climatic models that assumed *business as usual* scenario on pollution during COVID-19 lockdown was very much deviating from the actuals. Hence there is a need to reinvestigate the variation in the air quality levels along with its associated determinants (Streiff, 2020). Air quality levels across India were in the downtrend till 2019, however, in 2020, many cities in India recorded improved air quality (According to Statista report 2021) (Bakır et al., 2022). To control the spread of novel coronavirus disease in early 2020 (COVID-19), social distancing and lockdown were enforced that curbed all non-essential and non-emergency human outdoor activities. During this period, it is found that emission of air pollutants has drastically reduced in atmospheric air and eventually air quality substantially increased in many Indian states due to the suspended industrial activities and reduced vehicular movements (Elsaid et al., 2021; Pal et al., 2021).

Motivated by this, the authors focused on the following research question.

What significant differences COVID-19 lockdown has influenced on the air pollutants’ concentration and in turn on atmospheric air quality?

The authors selected prime locations in India where air pollution is generally reported to be alarmingly high. Pollution control board of both central and state governments (CPCB, SPCB) in India is keen about national air quality index. Existing method determines present AQI at a place by capturing the concentration of individual pollutants in unit volume of atmospheric air using appropriate sensors and uses them to evaluate AQI only if at least three pollutants’ concentration is available that must include either $PM_{2.5}$ or PM_{10} or both. Pollutants’ concentration is sensed every hour and AQI values are so computed. This study attempts to determine future AQI values using time-series analysis of pollutants’ concentration. In order to model air quality depicting both before and during COVID-19 lockdown scenarios, timeseries data of pollutants’ concentrations in selected regions is collected for both the time periods and analyzed through machine learning approaches. Using unsupervised learning approach, multi-pollutant clusters are first formed for these regions for both the scenarios which can reveal variation in pollutants’ concentration. Further, using supervised learning approach, data-driven models are developed to foresee air quality at a particular place. We explore the use of Linear regression modeling (LM), support vector regression (SVR), artificial neural network (ANN), decision tree (DT), random forest (RF) and extreme gradient boosting tree (XGB) to determine future AQI values with the forecasts of pollutants’ concentrations and the best-fitted model with the highest predictive performance is selected for each region. Understanding the variation in pollutants’ concentrations individually before- and during-lockdown scenarios in a particular region not only help policymakers to frame an appropriate pollutant-specific control strategy, but also, in determining future AQI values allowing to adopt appropriate mitigation plan.

2. Related work

Literature evidences show that predicting individual pollutants using learning methods help mitigating them by addressing various emission and non-emission factors associated with them. Forecasts of $PM_{2.5}$ and NO_x concentrations individually under known meteorological and traffic information are obtained accurately using random forest algorithm with high predictive power (Z. Li et al., 2020). Timeseries analysis of O_3, SO_2, CO, NO_2 concentrations using adaptive neuro-fuzzy inference system is performed and the prediction accuracy is found to be higher than the traditional regression methods (Zeinalnezhad et al., 2020). The concentration of O_3 is estimated using extreme gradient boosting algorithm and the performance of this algorithm is found to be high when compared with back propagation neural network, generalized regression neural network, Elman neural network, extreme learning model, linear regression, and random forest algorithm with respect to root mean square error and mean prediction error (R. Li et al., 2020). CH_4, CO and VOCs are the prime antecedents of O_3 and predicting O_3 concentration in atmospheric air is very vital to track and control climate change and global warming effects. This is performed using a long short term memory (LSTM) based hybrid deep learning model and the prediction accuracy is found to be high when compared with the baseline models (H. W. Wang et al., 2020). Shallow multi-output and deep multi-output LSTM network is used to forecast pollutants such as $PM_{2.5}$, PM_{10} and NO_X with high accuracy that yielded low loss function values (Zhou et al., 2019). The concentration of $PM_{2.5}$ under the influence of industry and weather related factors is predicted using gradient boosted decision trees with high precision accuracy compared to conventional methods (Lee et al., 2020). The concentration of $PM_{2.5}$ is forecasted based on meteorological and other factors using deep belief network and spatial lag models that helped urban planning (Yuan et al., 2019). The daily concentrations of $PM_{2.5}$ are determined using partial differential equation model in which the parameters of the equation are determined using Nelder-Mead simplex local optimization method however the computational complexity of large scale problem instances is high (Yufang Wang et al., 2020).

Air pollution at a particular place and time is quantitatively assessed using a commonly used comprehensive measure, air quality index (AQI) that is determined based on the proportion of various contaminants present in the air (Tan et al., 2021). Literature evidence shows that AQI can be forecasted accurately through data-driven models (Kumar and Goyal, 2011; Lee et al., 2020; Liu and Chen, 2020; Song and Fu, 2020; Zhou et al., 2019). The influence of $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO and O_3 concentrations on AQI and their inter-relationships are studied during various meteorological seasons using regression analysis (Q. Zhang et al., 2020) and the concentrations are forecasted for a short-term using machine learning and optimization models (S. Chen et al., 2019). AQI is forecasted using hybrid model of radial basis function neural network, ensemble empirical mode decomposition and autoregressive integrated moving average methods by capturing the influence of $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO and O_3 and other meteorological factors (Song and Fu, 2020). AQI is predicted using a hybrid method that comprised of binary grey wolf optimization-based feature reduction, discrete wavelet packet transform-based decomposition, extreme learning machine and adaptive boosting-based prediction model which is compared with other prediction models such as artificial neural network (ANN) and support vector regression (SVR). The proposed model surpasses ANN and SVR models in terms of robustness, interpretability and adaptability along with accuracy (Liu and Chen, 2020). The relationship of AQI with meteorological, economic, energy, demographic, mobility, and environmental aspects in a region is studied and multi-variate modeling using extreme gradient boosting decision tree is performed. The experimental study shows high accurate results when compared with other machine learning algorithms (J. Ma et al., 2020). The timeseries analysis of concentrations of $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO and O_3 is performed using variational mode decomposition and extreme learning machine to predict AQI in which the parameters of the model are determined using multi-objective grasshopper optimization algorithm and further fuzzy based air quality levels are identified (Ying Wang et al., 2020). However, pattern of AQI and pollutants' concentration are affected by several events that cause changes in the environment and in recent times COVID-19 has brought variations in air quality in many cities. The credibility of the forecasts made by the prediction models is to be examined as pollution data evolves. Moreover, forecasts of AQI values made available in public domain during COVID-19 lockdown showed high deviation (forecast error) from actual due to reduced economic activities (Vasudevan et al., 2021), Therefore, there is a need to develop event-based prediction models for accurate AQI forecasts. Clustering and prediction models using machine learning together could yield multi-pollutant profiling and forecasts of AQI which could help in planning environmental missions, controlling emissions and restoring environmental health. However, such studies are scarcely reported in the literature.

2.1. COVID-19 pandemic and air pollution

A novel coronavirus brought the world to a grinding halt turning up to be a massive global pandemic of this era. Air pollution and its effect on the spread and progression of infection was also one among the major concerns especially in cities (Harvard, 2020) (Becchetti et al., 2022). argued that quality of air has relationship with the spread of novel coronavirus and stringent lockdown measures enabled highly polluted area to mitigate spread and mortality effectively. As a preventive and control measure against this pandemic, lockdown was imposed and in all commercial establishments, only essential and very limited services were permitted to operate. Most countries enforced restrictions on travel and imposed mandatory lockdown which led to negative industrial growth in 2020 all around the world, however, it seems that the negative economic impact of the COVID-19 pandemic implies a positive environmental impact (Sahraei et al., 2021). During pandemic, while Governments' policies, aimed at curbing the virus spread amongst the population, the economic recession deepened and consequently hugely affected the global consumption of fossil fuels and positively influenced the environmental ecosystem and subsequently the air quality (Dang and Trinh, 2021; Tibrewal and Venkataraman, 2022). This highlights the importance of using zero-emission vehicles, alternate fuels and eco-friendly transportation and transshipment practices to sustain this improved air quality further (Rizova et al., 2020). Indeed, NASA reported many evidences indicating the rejuvenation of environment initiated by lockdown measures (Streiff, 2020). (Le Quéré et al., 2020) identified that the global CO_2 emission fell by 17% by early April 2020 relative to the 2019 average level. The effect of COVID-19 pandemic on air quality has been studied by Chinese researchers that has opened up new avenues to combat air pollution in China (Magazzino et al., 2021). Environment-friendly industry practices, logistic practices and green mobility initiatives have to be focused rigorously to sustain air quality (Caspersen and Navrud, 2021; Gonzalez et al., 2022). However, more studies are needed to sustain the improved environmental conditions that prevailed during COVID-19 lockdown without affecting the economic and social aspects.

3. Materials and methods

The present study focuses on evaluating atmospheric air quality based on the air pollutants such as the $PM_{2.5}$, PM_{10}, NO, NO_2, NO_X, NH_3, CO, SO_2, O_3, Benzene, Toluene, Xylene. The study was conducted for selected regions in three Indian states namely Delhi, Telangana and West Bengal. Delhi being highly commercial and highly polluted and Telangana being a newly-found state with huge influx of population having more economical activities and West Bengal being a pivot point for all the economic activities in the eastern region and highly committed towards sustainability are chosen as representative regions to carry out this study. The analysis of air quality before- and during COVID-19 lockdown was carried out for these selected regions to capture the variation in the trends of pollutants' concentration. A framework depicting the research work is shown in Fig. 1. Timeseries data of pollutants' concentration before- and during-lockdown time periods is taken from public repositories. The analysis is carried out in two phases (1) Multi-pollutant profiling for different stations is carried out from which clusters of pollutants based on the similarity in concentration in atmospheric air at a particular region are made using Ward's algorithm. (2) Predicting the air quality index capturing the trend of various pollutants existing at different stations. Timeseries forecasts of pollutants' concentrations are used to compute future AQI values.

3.1. Data

The concentration of $PM_{2.5}$, PM_{10}, NO, NO_2, NO_X, NH_3, CO, SO_2, O_3, Benzene, Toluene and Xylene along with air quality index (AQI) is obtained from online websites published by Central Pollution Control Board (CPCB), Ministry of Environment, Forests and Climate change, Government of India.[1] Here we use nitrogen oxides individually as well as in combination as there is a high chance of reaction of NO with O_3 and conversion of NO to NO_2 (Mazzeo et al., 2005). The timeseries data is available for major stations present in every city in every state. For the purpose of experimental modeling, this study considered nine stations present in Delhi, Hyderabad and West Bengal. Timeseries data of pollutants' concentration and AQI values before and after April 1, 2020 (Hourly records of March and April) is considered to better understand the variation in air quality before and during COVID-19 lockdown respectively.

3.2. Multi-pollutant profiling using cluster analysis

Multi-pollutant profiles are used to study the health effects of distinct pollutant clusters formed based on the concentration of different pollutants present in a particular region (Zanobetti et al., 2014). However, many research studies focused on developing city clusters based on the concentrations of a particular pollutant to frame policies and regulations. Based on the concentration of $PM_{2.5}$, cities in China are clustered taking high dimensional monitoring data evaluating spatial-temporal clustering of $PM_{2.5}$ concentration (Z. Chen et al., 2019; Yufang Wang et al., 2020). *k*-means clustering of dataset collected in Beijing is performed to impart high similarity within clusters that is found to further enhance forecasting accuracy (S. Chen et al., 2019). In this study,

[1] https://app.cpcbccr.com/AQI_India/.

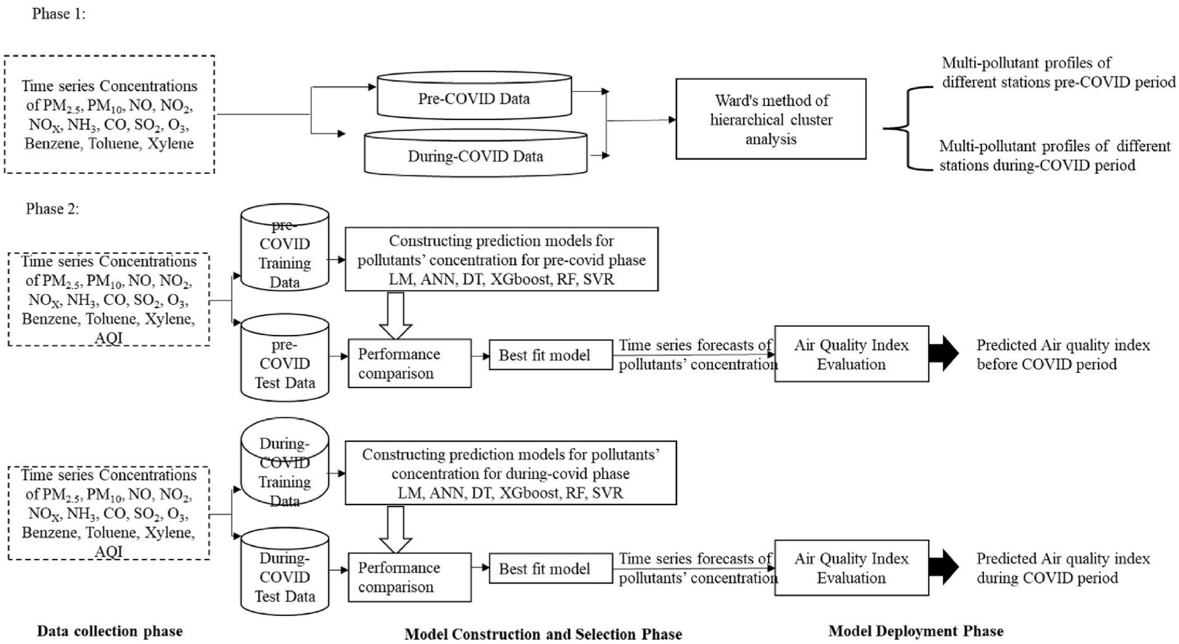

Fig. 1. Research framework.

hierarchical cluster analysis using Ward's method (Amorim, 2015) is used to cluster the pollutants by determining the Euclidean distance between a pair of data points. Initially, the number of clusters is set equal to the number of pollutants and the required number of clusters is also set. Then the pair of clusters (C_i, C_j) that has minimum cost function is iteratively merged based on the following equation till desired number of clusters is reached.

$$ward(C_i, C_j) = \frac{n(C_i) \times n(C_j)}{n(C_i) + n(C_j)} d(c_{C_i}, c_{C_j}) \quad (1)$$

where c_{C_i} denotes the centroid of cluster C_i and $n(C_i)$ denotes the cardinality of cluster C_i. By applying ward's method, multi-pollutant profiles of different regions can be obtained and their effects on human health in those regions can be studied.

3.3. *Predictive modeling of air quality*

The effect of pollutants' concentration on air quality is first studied using linear regression method (LM). Linear timeseries models assume linear relationships between the dependent and independent variables (Zeinalnezhad et al., 2020). Learning-based non-linear timeseries analysis is also carried out in this study to explore the possible non-linear relationships among the determinants of air quality (Du et al., 2019; R. Li et al., 2020; J. Ma et al., 2020; Zeinalnezhad et al., 2020). Existing studies show that regression trees constructed using Decision Tree (DT) and its ensembles such as Random Decision Trees or Random Forest (RF) and Extreme gradient boosting trees (XGB) are able to fit a robust regression model (Ayoubloo et al., 2011). Tree based models capture the pattern of data over time and effectively predict the future values. The rules on the decision nodes are made with the selection of independent variables in the order of their significance during the training phase in such a way that the seen instances can be accurately fitted with the established rule. After sufficient training, the decision rules are fine-tuned and the regression tree will eventually become ready to handle unseen instances. Random forest employs simultaneous learning through multiple decision trees and generalizes more quickly especially in hard-to-predict problems. XGB is a boosted tree involving ensemble learning to incorporate dominant relationships among predictors iteratively. Literature also shows evidence that artificial neural network (ANN) and its ensembles are widely used for predictive modeling of air quality (Ayoubloo et al., 2011; Yilmaz and Kaynar, 2011). ANN consists of input, hidden and output neurons that are grouped in layers such that the entire network architecture can be trained to activate appropriate neurons to determine a desired output given an input signal (Song and Fu, 2020). The number of hidden neurons and the number of hidden layers must be experimentally identified as hyper-parameters. There are connections between the layers and the strength of a connection is given as a weight value. The network develops an input-output fitted function iteratively during the training phase with seen instances. The activation function is also chosen experimentally. For each training instance, the error obtained from the deviation between the predicted and actual values is determined and fed back to the input layer through back-propagation and gradient descent method is used to minimize error. The gradient descent method adjusts the weights by an amount that is proportional to the partial derivative of the error function that is back propagated based on a learning rate parameter. This helps to avoid convergence to the saddle (min-max) point and the gradient vector of the approximated error function is obtained through the layers of the network. Literature presents support vector regression (SVR) as another powerful tool that maps the non-linearly related predictors to generalize unseen instances with outstanding performance (Kavousi-fard et al., 2014). Radial basis function incorporated in a Gaussian kernel to model air quality with seen instances of pollutants' concentration during training phase iteratively minimizes the loss function (Smola and Olkapf, 2004). The kernel attempts to determine a statistical fit in the training phase in the predictor space with minimum loss at which the deviation between the actual and obtained values are less than a small tolerance error. A non-linear mapping transforms the input vector onto predictor space performed by kernels and forms mapped support vectors. The variation in pollutants due to COVID-19 and its related events has to be intensely investigated. In this study, every station is investigated using these learning methods to fit regression models for predicting AQI before and during COVID-19 lockdown and the best fitted model is selected which can address COVID-19-induced change in the relationship of pollutants. The predicted pollutants' concentration using the best fitted models further is used to determine national air quality index based on the guidelines of CPCB.

3.3.1. Data-driven modeling of air quality

Best fitted models to determine AQI as a function of its pollutants is to be identified for each region based on the data collected in this study. Linear and non-linear timeseries analysis using support vector regression (SVR), artificial neural network (ANN), decision tree (DT), random forest (RF), and extreme gradient boosting (XGB) tree methods are performed and their prediction error and computational time complexity are obtained for each station before and during COVID-19 lockdown. The forecasts of pollutants' concentration yielded by each model is compared with the actuals in the testing phase and the best fitted models are selected which can be further deployed to forecast AQI accurately in each region (9 stations considered in this study). The timeseries data of pollutants' concentration collected before and after the onset of COVID-19 in these regions are classified into training and test dataset. The training dataset is used to construct the timeseries models to forecast the pollutants' concentration in each region. During the training process, the models exhibit deviation between actual and predicted from which root mean squared error is derived that is most often fed back to the model for finetuning the model parameters iteratively. After sufficient training, the model is tested on its prediction ability with the test data and again root mean squared error is determined. The training error and testing error are evaluated based on the deviation of estimated AQI from actual AQI values to analyze the generalization capability of the machine learning regressors in fitting the models (Refer to Tables 1 and 2).

Most of the models exhibit a high root mean squared error in the testing phase than that in the training phase, especially in the during-lockdown scenario. This is because the models could not capture the complex patterns in the data and are unable to represent the problem with an exact function similar to an actual or target function and hence exhibiting poor generalization capability. The learners have different ability to derive the approximated function and often suffer from underfitting or overfitting issues. The models get fitted to the best approximate function possible with the seen instances and have a deviation between the actual and predicted when confronted with unseen instances. However, the difference in errors during training and testing phases is more in during-lockdown models as it is obvious that the pandemic outbreak created turbulence in air pollutants' concentration that gradually stabilized with time. It can also be observed that the prediction error is high in linear models in most of the cases as there are non-linear relationships in the pollutants which are addressed by the machine learning methods to a great extent. Also, prediction accuracy is found to substantially increase when machine learning models are applied especially in the during-lockdown scenario. To choose an appropriate machine learner, in addition to error, the convergence speed of the regressors is also captured and the execution time in training the models is obtained and presented in Table 3.

It can be observed that the computational time during the training phase in ensemble methods such as random forest and boosting learners is high when compared to other methods. The resultant best fitted models were chosen for each station after modeling are presented in Table 4.

The best fitted prediction model for AQI in each region for before- and during- COVID-19 lockdown is selected based on the prediction accuracy and computational time taken for constructing the models. These models are then further used to analytically determine the significance of the pollutants and the impact of COVID-19 lockdown measures on the environmental pollution in selected nine regions in India.

4. Results and discussions

In this study, the influence of COVID-19 lockdown on various air pollutants and air quality in various regions of India is analyzed using clustering and prediction methods. The best fitted models identified for each region in the previous section are further used to determine AQI based on the pollutants' concentration before and during COVID-19 lockdown.

The multi-pollutant profiles existing during COVID and before the COVID outbreak in each station are obtained using hierarchical cluster analysis using Ward's method and presented in Figs. 2 and 3. The major air pollutants considered in this study viz, $PM_{2.5}$, PM_{10}, NO, NO_2, NO_X, NH_3, CO, SO_2, O_3, Benzene, Toluene, and Xylene are clustered to obtain maximum of five multi-pollutant profiles for every station based on their similarity in variation.

It can be observed that the pollutants are differently clustered before and during COVID-19 lockdown in all the stations. Delhi is the most polluted city in India and one of the top five polluted cities in the world. In the station, DL001, it could be observed that the concentration of PM_{10} in atmospheric air is found to be a persistent pollutant affecting the air quality both pre-and during COVID-19 lockdown. It is a unique pollutant and its concentration is found to decrease as the height of the reference point from ground level increases and has high retention in the air for a longer duration (Q. Zhang et al., 2020). The pollutants CO, benzene, xylene, toluene, and NO form the next important cluster before COVID-19 which during COVID-19 has been significantly shown to reduce effect on air quality. The prime source of Benzene, Toluene, and Xylene (BTX) pollutants is emission from the combustion of fuels in automobiles and industries. Since during COVID-19 outbreak, the lockdown of all non-essential activities has drastically reduced emissions and hence these pollutants became less significant. In the station, DL019, before the outbreak of this pandemic, $PM_{2.5}$ and PM_{10} formed the main cluster which persisted to influence air pollution followed by the cluster composed of NO, NO_2 and cluster composed of SO_2, CO, Benzene, and Xylene, and cluster composed of NH_3, O_3, Toluene. However, after the outbreak of the pandemic, the multi-pollutant profiles have altered their compositions nevertheless PM_{10} remains to be the main threat due to its prolonged endurance. The effect of NO, CO, and BTX turned relatively significant during a pandemic. In both stations in Delhi, the prime contributor of air pollution remained significant though has reduced effect after reducing the unnecessary movements in the city and hence measures should be taken to reduce particulate matter in the air.

PM_{10} is the most significant cluster in stations, TG001 and TG004 located in Telangana and its intensity is lessened during the pandemic outbreak, and [NO, T, CO, Benzene, Xylene] and [Benzene, CO, Xylene] multi-pollutant profiles dominate the air pollution during the pandemic period. In TG002, [NO, CO] is the prominent cluster however during the lockdown period, the [CO, Xylene] cluster became relatively significant in which CO is persistently prominent. In TG003, [Benzene, CO, Xylene] cluster loses its significance to [NH_3, NO_2, NO_X] during the lockdown and in TG006, [SO2, Xylene, CO, Benzene] cluster reduced its influence, and the [Toluene, Benzene, CO and Xylene] cluster became relatively prominent during the lockdown in which CO and BTX pollutants are persistently significant. Generally in Hyderabad, CO, nitrous, and BTX pollutants are prominent which are primarily due to extensive transport and industrial activities (S.K. et al., 2013). The state pollution board projects above 5700 deaths in Hyderabad due to air pollution.[2] The multi-pollutant profiles obtained by this study would be instrumental in developing strategies to combat air pollution.

In West Bengal, the scenario is different. [Benzene and Toluene] is the prominent cluster that lost its intensity and the [CO, Xylene] cluster became prominent during the pandemic period in the station, WB007. In WB008, PM10 is the persistent pollutant even after reducing unnecessary movements in the lockdown period and the concentration of O_3 has turned up relatively significant. In WB009, the [$PM_{2.5}$, PM_{10}, O_3] cluster turned up significant during the pandemic outbreak whereas the [NO, NO_2, O_3] multi-pollutant profile was formed with a similar effect in both the scenarios. Commonly observed multi-pollutant profiles are

[2] https://www.iqair.com/india/telangana/hyderabad.

Table 1
Training and test root mean square errors for different stations before COVID-19 lockdown.

Regressors	LM	SVR	ANN	DT	RF	XGB	LM	SVR	ANN	DT	RF	XGB
Station ID	Training RMSE						Test RMSE					
WB007	0.089	0.062	0.068	0.290	0.192	0.001	0.097	0.110	0.138	0.115	0.106	0.104
WB009	0.062	0.045	0.051	0.296	0.111	0.001	0.090	0.152	0.087	0.104	0.112	0.093
TG001	0.074	0.054	0.063	0.120	0.005	0.001	0.075	0.111	0.078	0.090	0.081	0.079
TG002	0.042	0.037	0.037	0.107	0.043	0.001	0.032	0.034	0.044	0.039	0.031	0.032
TG003	0.070	0.048	0.060	0.201	0.067	0.001	0.066	0.087	0.115	0.103	0.089	0.074
TG004	0.062	0.041	0.047	0.141	0.045	0.001	0.048	0.069	0.046	0.059	0.049	0.049
TG006	0.069	0.047	0.055	0.149	0.052	0.001	0.075	0.076	0.074	0.073	0.068	0.073
DL001	0.066	0.035	0.044	0.218	0.083	0.001	0.096	0.085	0.071	0.071	0.068	0.073
DL019	0.063	0.034	0.043	0.192	0.061	0.001	0.063	0.101	0.077	0.084	0.074	0.072

Table 2
Training and test root mean square errors for different stations during COVID-19 lockdown.

Regressors	LM	SVR	ANN	DT	RF	XGB	LM	SVR	ANN	DT	RF	XGB
Station ID	Training RMSE						Test RMSE					
WB007	0.0402	0.0416	0.0184	0.26838	0.60922	0.0005	0.2154	0.1356	0.2744	0.1365	0.1791	0.148
WB009	0.0476	0.0272	0.0122	0.23662	0.72568	0.0005	0.1524	0.1165	0.1155	0.1274	0.1355	0.1309
TG001	0.073	0.0594	0.0127	0.72681	0.89294	0.0006	0.7857	0.2508	0.3379	0.1628	0.1584	0.1186
TG002	0.0599	0.0656	0.0137	0.50273	0.97631	0.0006	0.1885	0.155	0.1637	0.1961	0.1903	0.225
TG003	0.1043	0.0725	0.0395	0.22301	0.4974	0.0005	0.1759	0.1869	0.2741	0.3148	0.2156	0.2678
TG004	0.0502	0.0497	0.0131	0.19409	0.90768	0.0006	0.2827	0.2308	0.1748	0.116	0.1605	0.1354
TG006	0.1288	0.0876	0.0159	0.24902	1.07484	0.0006	0.5023	0.304	0.4581	0.2869	0.2346	0.2222
DL001	0.0625	0.0439	0.0163	0.25895	0.7138	0.0005	0.5676	0.1221	0.2654	0.249	0.1562	0.1696
DL019	0.0818	0.122	0.0153	0.24066	0.97345	0.0005	0.1671	0.1046	0.4122	0.0957	0.0909	0.1152

Table 3
Execution time.

Regressors	LM	SVR	ANN	DT	RF	XGB	LM	SVR	ANN	DT	RF	XGB
Station ID	Before COVID Outbreak						During COVID Outbreak					
WB007	0.006	0.017	0.024	0.002	0.128	0.181	0.000	0.008	0.008	0.008	0.008	0.095
WB009	0.000	0.016	0.032	0.008	0.199	0.186	0.000	0.011	0.012	0.006	0.008	0.121
TG001	0.005	0.048	0.069	0.014	7.086	0.730	0.017	0.026	0.018	0.016	0.023	0.127
TG002	0.000	0.053	0.079	0.016	1.037	0.461	0.001	0.713	0.010	0.016	0.019	0.129
TG003	0.006	0.063	0.267	0.016	0.767	0.452	0.000	0.000	0.016	0.000	0.033	0.136
TG004	0.010	0.115	0.105	0.290	1.315	0.592	0.000	0.008	0.016	0.008	0.015	0.135
TG006	0.024	0.104	0.095	0.016	1.774	0.820	0.000	0.008	0.016	0.005	0.018	0.142
DL001	0.000	0.016	0.048	0.008	0.311	0.200	0.000	0.016	0.012	0.008	0.016	0.128
DL019	0.100	0.038	0.072	0.007	0.555	0.333	0.005	0.009	0.008	0.000	0.016	0.122

Table 4
Best fitted models for AQI prediction.

Station ID	During-COVID AQI Prediction model	Pre-COVID AQI Prediction model
WB007	SVR	LM
WB009	ANN	LM
TG001	XGB	LM
TG002	SVR	RF
TG003	LM	LM
TG004	DT	ANN
TG006	XGB	RF
DL001	SVR	RF
DL019	RF	LM

[$PM_{2.5}$, O_3], [NH_3, SO_2], [CO, Benzene, Xylene] which are found to affect air quality. In this study, a representative list of stations is considered to derive the pollutant clusters which could be extended to all other stations also. This study also proves the significance of the combined effect of these profiles on air quality in all the stations.

The variation of the pollutants in air is drawn out from the predictions and presented in Figs. 4 and 5.

It is well known that the outbreak of novel coronavirus disease and its spread has very rapid affected and lockdown regulations are announced and restricted movement is enforced across India. This has drastically reduced emissions from industries and vehicles thereby air pollution has been reduced considerably and a revival of atmospheric air took place. To understand AQI variation in these stations, the mean forecasts of concentrations of these pollutants before and during COVID-19 obtained from the best-fitted prediction models during the testing phase are plotted in Figs. 4 and 5. Generally, pollution due to particulate matter is high in all the regions. However, during COVID lockdown, the concentration of these pollutants including PM in atmospheric air has drastically reduced. The effects of $PM_{2.5}$, $PM_{10,}$ and BTX in atmospheric air continues to be high-priority pollutants even during the lockdown in Delhi however the concentrations of these pollutants were fairly reduced. NH_3, Benzene, and Toluene levels are high during COVID-19 lockdown but comparatively lesser than before lockdown scenario. In Hyderabad, though air pollutants' weakening effect is felt during the lockdown, NO_2 and SO_2 continue to show importance during COVID-19 lockdown which also formed a prime multi-pollutant cluster in the stations present in Hyderabad. Similarly, after the lockdown there is a drastic reduction in air pollutants in atmospheric air in many places of West Bengal. However, NH_3, CO, $SO_{2,}$ and O_3 are relatively significant pollutants to be persistently addressed even after reduced activities. However, the dominant air pollutants in West Bengal have turned out to be NH_3, SO_2, and O_3 in zone WB007 and NH_3, SO_2, CO, NO, and $PM_{2.5}$ in

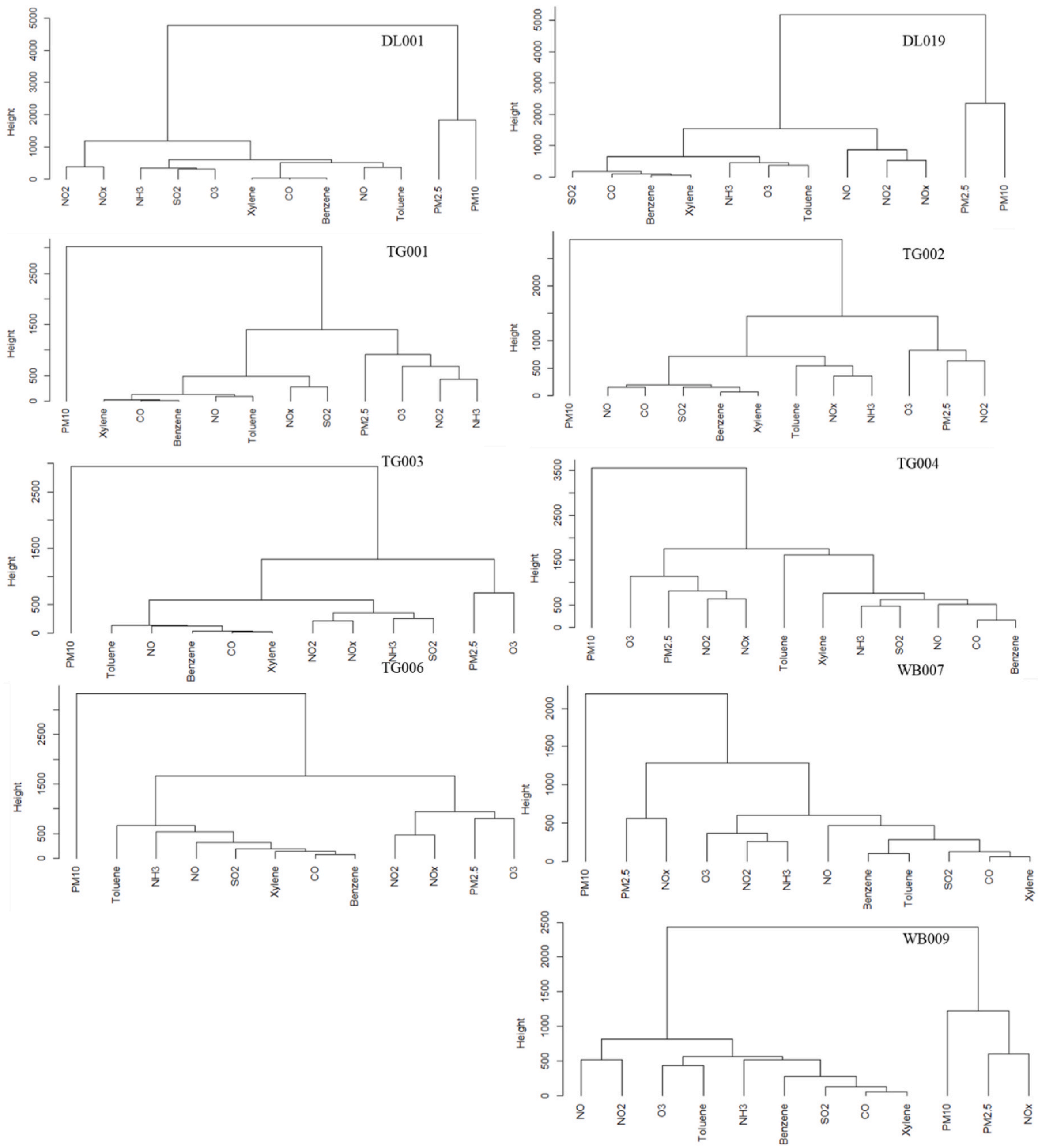

Fig. 2. Pre-COVID Clusters of pollutants.

WB009. The effect of O_3 is more in West Bengal due to climatic conditions. Similar foresights about future pollutants' concentration enable handling these zones with appropriate pollutant specific proactive and preventive measures to mitigate pollution in a timely manner.

Linear timeseries analysis yielded high prediction error compared to ensemble learning-based non-linear timeseries analysis of pollutants with better accuracy of forecasts made on AQI. The findings from the analysis (Refer to Tables 1–3) show that the performance of prediction through non-linear timeseries modeling is high. Hyper-parameter tuning and selection of kernel and activation functions are carefully performed experimentally. The prediction error for different stations using the selected models is minimized by 9% for the before COVID-19 scenario and 50% in during-COVID-19 lockdown compared with that exhibited by linear regression. Especially in COVID-19 scenario, the uncertain changes in the release of air pollutants in the environment have been well addressed by the selected models. For all the regions under consideration, the concentration of particulate matter has reduced due to the COVID-19 outbreak and subsequent lockdown. Emissions from vehicles have greatly reduced and shut-down enforcement of industries during this period had influenced this drastic change.

5. Deployment and limitations

This study demonstrated predictive modeling of air quality index using the past pollutants' concentration data. The timeseries data of pollutants' concentration is obtained from the official website of CPCB, India. This data is used in this study to train the prediction models for making forecasts of pollutants' concentration and the best fit model obtained during the test phase is further used to evaluate National air quality index as per the guidelines of CPCB. National air quality index is updated on hourly basis in official application developed by CPCB with a latency. However, this study helps evaluate future values based on the

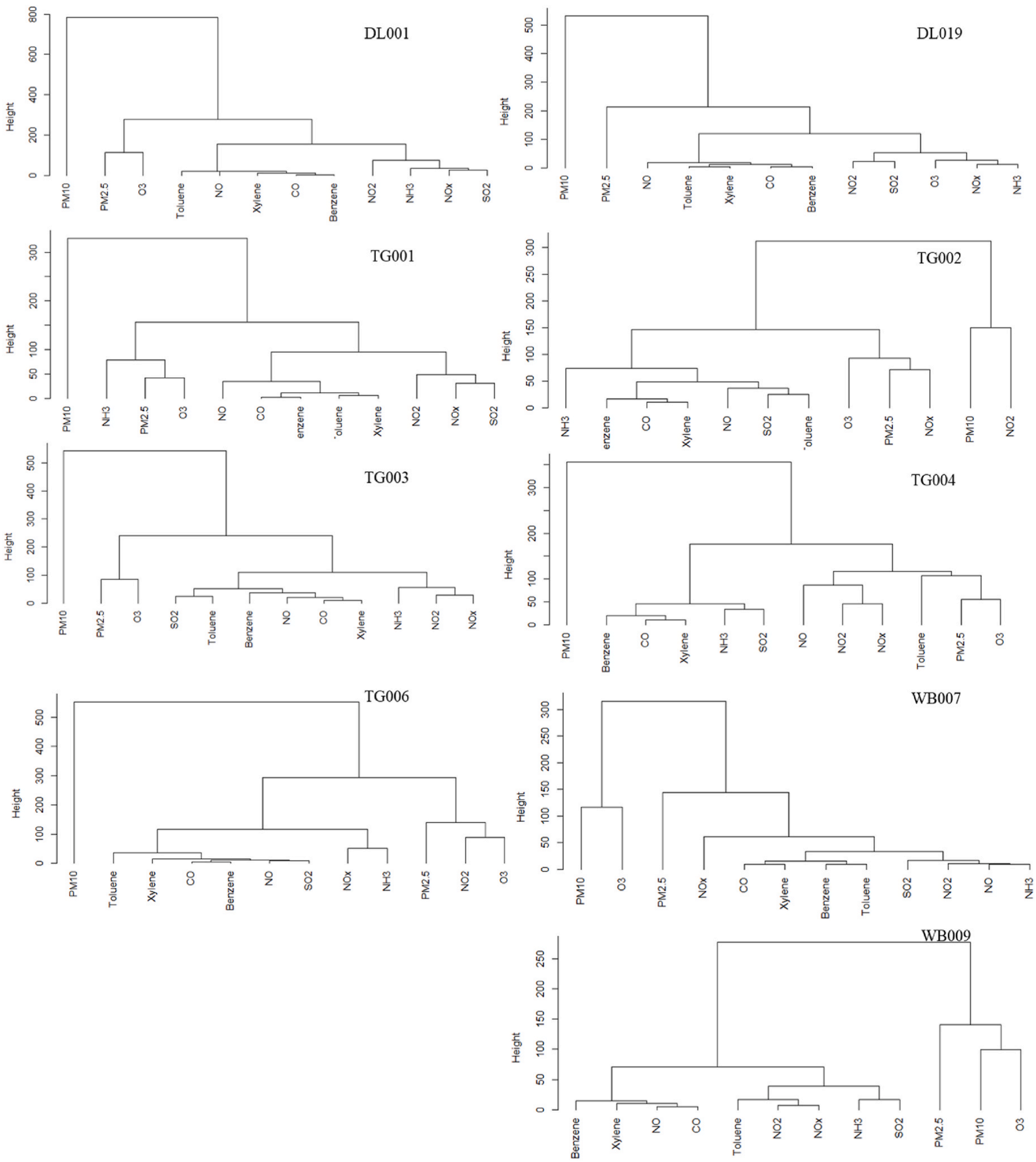

Fig. 3. During-COVID clusters of pollutants.

forecasts of pollutants' concentration obtained from the prediction models. The prediction modeling and selection process has to be often performed at regular intervals or whenever there is an event or whichever is the earliest. In this study, we have presented two scenarios, viz., models before and during- COVID lockdown since the lockdown period in India has witnessed substantial air quality improvement due to reduced economic activities. The impact of events induced by meteorological factors, nature-induced factors, perceptual and situational factors along with pollutants also affect air quality and hence has to be incorporated in the model time-to-time. The deployment model is presented in Fig. 6.

The findings of this study are limited to the stations chosen in Delhi, Hyderabad, and Kolkata. The data for this study is obtained through public repositories and hence we had to tackle incompleteness and distortion to a larger extent. The deployment model is scalable to other stations with sufficient dataset. AQI prediction made in this study is based on pollutants' concentration assuming consistent meteorological conditions (during March–April 2020).

6. Implications

The present study used a learning approach to forecast air quality using timeseries analysis of pollutants' concentration in a particular region. It has significant implications.

6.1. Theoretical contributions

This study offers a significant contribution to theory. The study uses event-driven timeseries data of pollutants' concentration for machine learning analysis to predict future values that is used further to

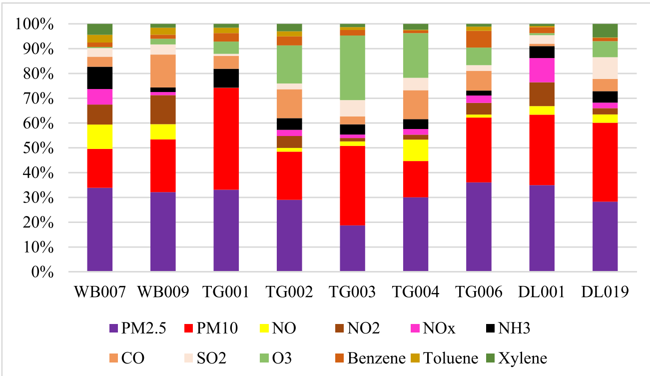

Fig. 4. Impact of air pollutants on AQI before the COVID outbreak.

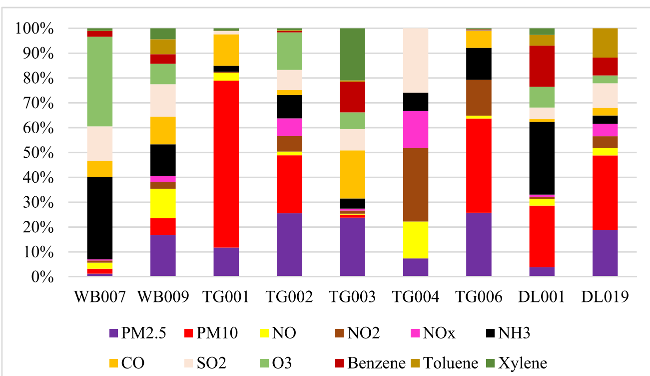

Fig. 5. Impact of air pollutants on AQI during COVID.

determine future AQI values. In reality, pollution data is unbounded and evolving with numerous events in the environment. The trend in timeseries of pollutants' concentration captured analytically often gets distorted due to these events and increase the prediction error. Therefore, this study performs prediction modeling of pollutants' concentration for the scenarios, viz., before and during COVID lockdown separately which forecasts AQI accurately during these periods. The best fitted models selected for these scenarios for each geographical region also elucidate the variation in the air quality in both the time periods.

6.2. Practical implications

Collaborative governance helps effective control of air pollution (Wang and Zhao, 2021). Several attempts are made by the government to provide real-time awareness of pollution status in a particular area that paves way for control and mitigation. The present study attempts to provide prior information about air quality that enables prevention and control. This study developed an evaluation mechanism for AQI accurately through learning-based modeling that can be adopted in industrial and traffic zones to understand the trend in air quality. The variation of pollutants' concentration observed in the COVID-19 lockdown phase allows researchers to analyze the pollutants' concentration before and after the COVID-19 outbreak which can potentially lead to practical solutions and sustainable business practices to retain the improved air conditions further. However, there is a need to detect any event that would bring significant variation in pollutants' concentration in real-time and the prediction model should be evolving in nature to maintain its predictive performance during these changes which is more challenging and requires more data. It is even more challenging to address these variations in the predictive models if such variations are temporary. This study clearly shows the changes in the concentration of air pollutants during the COVID-19 lockdown and their impact on improved air quality levels. It also highlights those pollutants that can be mitigated by addressing fuel combustion and industrial emission properly. Further, the amount of data generated by monitoring air pollution is vast and highly unstructured. Therefore, the present study serves as the walkthrough to the pollution management team of every city in developing countries to adopt the proposed model (Fig. 1) for evaluating the air quality and introduce reforms in the regulations.

6.3. Societal implications

The results of this study have important societal implications. The prevalence of the pollutants in the air in selected regions in India is ranked before and during the COVID-19 lockdown to capture variation in the air quality. The presence of $PM_{2.5}$ and NO_X in the air at Alipur in Delhi has got reduced by 89% and 91% respectively during the COVID-19 lockdown which helped people inhale the air of better quality. $PM_{2.5}$, CO, NH_3, NO, and O_3 reduced by 33.5%, 38%, 28%, 13.5%, and 51% respectively in the Mandir Marg region of Delhi during the COVID-19 lockdown. The reduction in the concentration of $PM_{2.5}$ and nitrogen oxides during the COVID-19 lockdown in Delhi helped the environment rejuvenate which in turn offered better air for the society. Further, a lot of pharmaceutical companies are located in these industrial zones that emit harmful pollutants like PM, ozone, BTX, Sulfur oxides, and carbon monoxide, and proper source management and air filtration methods should be adopted to avoid potential health hazards. Sulfur dioxide and carbon monoxide in the air lead to acid rain which is a major threat to farmers and residents living in this area. In the state of Telangana, in the Bollaram industrial area, $PM_{2.5}$, O_3, NO_X, NH_3, benzene, toluene, and

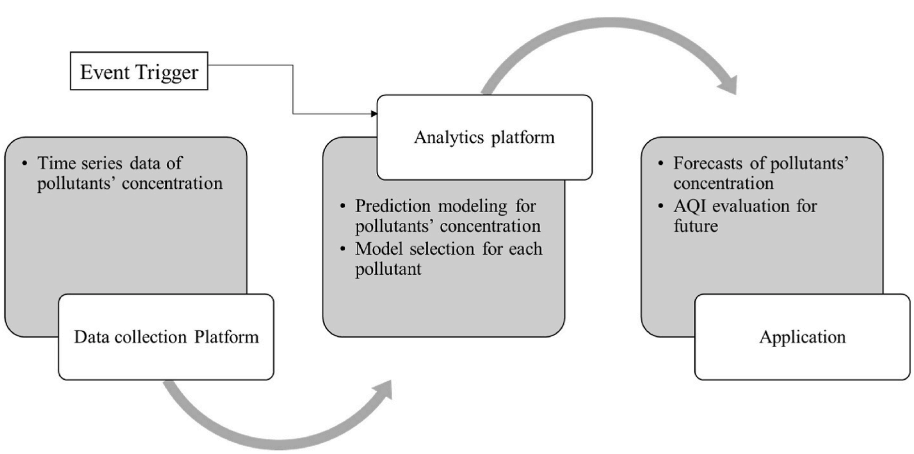

Fig. 6. Deployment model.

xylene are observed to be reduced by 64.5%, 99.9%, 27%, 68%, 99.7%, 99.6%, 39% respectively, in the Patancheru industrial zone, PM_{10}, NO, O_3, Toluene got reduced by 96%, 78%, 74%, 52% respectively and in Pashamylaram industrial zone, it was observed that the pollutants, PM_{10}, CO, Ozone, BTX reduced completely and $PM_{2.5}$ reduced by 75%. In the university zone, CO and BTX reduced by 82%, 86%, 81%, and 76% respectively due to reduced vehicle emissions. In West Bengal, the Ballygunge locality of South Kolkata recorded much-improved air quality and shows a reduction of $PM_{2.5}$, PM_{10}, NO, NO_2, NO_X, Toluene, and Xylene by 96%, 87%, 75%, 90%, 92%, 94%, 78% respectively. Pollutants such as $PM_{2.5}$, PM_{10}, NO_2, CO reduce by 48%, 69%, 77% respectively in Fort William area in Kolkata city. The major contributor to air pollution in Kolkata city is vehicular emission and during the COVID-19 lockdown, the pollutants' concentration dropped due to travel restrictions. The results of this study enable policymakers to regulate air pollution effectively by detecting the variation in air quality occurring from time to time that brings a positive impact on social life. Improved air quality level is possible during COVID-19 lockdown due to public commitment and people across the three states followed restrictions to the possible extent which in turn improved the air quality level and decreased the pollutants' concentration in the different stations.

6.4. Economic implications

In the field of environmental economics, the Environmental Kuznets Curve (EKC) suggests that, in the short-run, as the economy shifts towards development, the environment worsens, but, in the long run, as the society becomes aware of the social cost of this negative externality, it introduces, through government regulations, changes in environmental standards and policies and reinvests part of its income to improve the environment and restore the ecosystem. In line with this, many studies support the Environmental Kuznets Curve hypothesis for India, for example (Kanjilal and Ghosh, 2013; Usman and Jahanger, 2021), and our results suggest that pollution mitigation in India is not attributed to a slowdown in economic growth, but regulatory efforts to limit pollution as well as the country will to ensure the energy transition necessary for sustainable development. Indeed, aware of the social cost of this negative externality, the Indian society has introduced, through regulations, changes in environmental standards and policies and has recently started to reinvest part of its revenues to improve the environment and restore the ecosystem. Specifically, the main environmental laws in India include (a) the Wild Life Act (1972), (b) the Water Act (1974), (c) the Forest Act (1980), (d) the Air Act (1981), (e) the EP Act (1986) and (f) the National Green Tribunal (NGT) Act. Moreover, from its inception, the NGT ordered the CPCB and the SPCB, which are the Indian environmental regulatory authorities, to strictly enforce the Comprehensive Environmental Pollution Index (CEPI) criteria which were initiated in 2009 (and updated in 2016) to categorize polluting industries as well as polluted industrial areas. In 2019, the NGT monitored the enforcement of the CEPI criteria by the CPCB and the SPCB. To foster the environment rejuvenation instigated by COVID -19 lockdown and to internalize the negative externality, the Indian environmental regulatory authorities must now start thinking not only about seeking compensation from polluting industries (e.g. via a Pigovian Tax, a lump-sum tax, and/or restricting the amount of pollution …), but above all about doing so in an optimal way to improve social welfare.

6.5. Implications for policies

Understanding current trend in air quality and its associated pollutants' is instrumental for developing various mitigation plans and control strategies. Ministry of Housing & Urban Affairs of the Indian government launched the National clean air program in 2019 with a mission to achieve a reduction in PM concentration by 20–30% by 2024 from the concentration observed in 2017 in cities which include Delhi, Kolkata, and Hyderabad. Under this, city-specific action plans need to be devised to reduce emissions, strengthen the monitoring networks, bring mobility and industry policies to regulate air pollution, and create public awareness. Since the major source of air pollution in Delhi and Kolkata is vehicular emissions, strict compliance to exhaust emission standards is enforced and all vehicles must possess valid Pollution under control (PUC) certificates. The mass rapid transport system is introduced for convenient conveyance to reduce traffic. Under a clean fuel program, governments also strictly monitor the quality of petrol and diesel supplied to these places to control harmful emissions during combustion. Particulate filters in diesel vehicles are mandated and the supply of adulterated fuels is curbed. Initiatives toward electric-only vehicles are also being developed by the government to establish zero-emission zones in the cities. Small-, medium- and large-scale pharma industries are under strict monitoring by the pollution control board and safe discharge of emissions is enforced. During the lockdown, the decrease in harmful pollutants in the air has unraveled the potential activities causing large-scale pollution in these cities. It was understood that more than 50% of air pollution is from vehicles and in the case of Hyderabad, emissions from pharma industries also play a major role. The pollution control action plans of these cities need to be revised based on the lessons learned from the pandemic. Further, the COVID-19 pandemic has brought potential initiatives for public well-being such as, no-car road and proving cycling path along streets that are sustainable to the environment and reduce the carbon footprint. Public awareness about everyday air quality and lifestyle modifications will take these initiatives forward for better wellbeing.

7. Conclusions

The work focuses on evaluating air pollution and pollutants' impact before and during COVID-19 lockdown. The present study contributes to research, by evaluating Air Quality Index (AQI) for the pre-and during the COVID-19 lockdown time periods. The concentrations of pollutants such as $PM_{2.5}$, PM_{10}, NO, NO_2, NO_X, NH_3, CO, SO_2, O_3, Benzene, Toluene, and Xylene along with air quality index (AQI) are obtained as time-series data that is made publicly available for major stations present in every city and states in India. This study considered nine representative stations covering Delhi, Hyderabad, and Kolkata. The pollutant clusters derived from cluster analysis based on their similarity in concentration levels highlight their combined impact on air quality across different stations. This study intends to make a comparison of multi-pollutant clusters formed before and during COVID-19 lockdown to show the effect of reduced human outdoor activities during COVID-19 due to lockdown. Although literature evidences regarding the prediction models for forecasting AQI are existing, the deviation of forecasts made by these models from actuals during the COVID-19 lockdown was alarmingly high, and hence revised models have to be developed. In this study, we performed statistical modeling of time series data collected before and during COVID-19 lockdown at selected regions to capture the change in trends of pollutants' concentration and subsequently AQI. The selected models representing AQI in these regions that resulted from machine learning analysis showed minimal prediction errors and high accuracy. The data used is comprised of hourly data points collected by the pollution monitoring systems installed in these stations. Smart pollution monitoring systems incorporating the Internet of Things, machine learning, and cloud computing technologies integrated into pollution monitoring platforms can be developed in the future to sense these variations in pollutants' concentration. Real-time Affordable Multi-Pollutant (RAMP) sensor packages can be used to capture the data effectively and seamlessly. The results of clustering and prediction modeling play a vital role in the visualization of insights in such platforms. Further, the google mobility data can also be integrated with these systems to assign events such as the closure of restaurants, shopping centers, theme parks, cafés, pandemic, rainfall and cinemas with variations in air quality. Furthermore, it would be interesting to have future studies using time-frequency econometric tools to examine the environmental impact of the very recent Indian government policies during the COVID-19 aiming at flattering the epidemiological curve, as well as the relationship between economic activity and pollution in India.

Credit author statement

Jinil Persis: Conceptualization, Methodology, Analysis, Results, Writing, Amine Ben Amar: Writing, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Data availability

Data is available in online repositories

References

Amorim, R.C. de, 2015. Feature relevance in ward's hierarchical clustering using the Lp norm. J. Classif. 32, 46–62. https://doi.org/10.1007/s00357.

S, K., G, A.W., G R V K P D, 2013. Receptor model-based source apportionment of particulate pollution in Hyderabad, India. Environ. Monit. Assess. 185, 5585–5593.

Ayoubloo, M.K., Azamathulla, H.M., Jabbari, E., Zanganeh, M., 2011. Predictive model-based for the critical submergence of horizontal intakes in open channel flows with different clearance bottoms using CART, ANN and linear regression approaches. Expert Syst. Appl. 38, 10114–10123. https://doi.org/10.1016/j.eswa.2011.02.073.

Bakır, H., Ağbulut, Ü., Gürel, A.E., Yıldız, G., Güvenç, U., Soudagar, M.E.M., Hoang, A.T., Deepanraj, B., Saini, G., Afzal, A., 2022. Forecasting of future greenhouse gas emission trajectory for India using energy and economic indexes with various metaheuristic algorithms. J. Clean. Prod. 360, 131946 https://doi.org/10.1016/j.jclepro.2022.131946.

Balietti, A., Datta, S., Veljanoska, S., 2022. Air pollution and child development in India. J. Environ. Econ. Manag. 113, 102624 https://doi.org/10.1016/j.jeem.2022.102624.

Becchetti, L., Conzo, G., Conzo, P., Salustri, F., 2022. Understanding the heterogeneity of COVID-19 deaths and contagions: the role of air pollution and lockdown decisions. J. Environ. Manag. 305, 114316 https://doi.org/10.1016/j.jenvman.2021.114316.

Caspersen, E., Navrud, S., 2021. The sharing economy and consumer preferences for environmentally sustainable last mile deliveries. Transport. Res. Transport Environ. 95, 102863 https://doi.org/10.1016/j.trd.2021.102863.

Chen, S., Wang, J. qiang, Zhang, H. yu, 2019. A hybrid PSO-SVM model based on clustering algorithm for short-term atmospheric pollutant concentration forecasting. Technol. Forecast. Soc. Change 146, 41–54. https://doi.org/10.1016/j.techfore.2019.05.015.

Chen, Z., Chen, D., Xie, X., Cai, J., Zhuang, Y., Cheng, N., He, B., Gao, B., 2019. Spatial self-aggregation effects and national division of city-level PM2.5 concentrations in China based on spatio-temporal clustering. J. Clean. Prod. 207, 875–881. https://doi.org/10.1016/j.jclepro.2018.10.080.

Dang, H.A.H., Trinh, T.A., 2021. Does the COVID-19 lockdown improve global air quality? New cross-national evidence on its unintended consequences. J. Environ. Econ. Manag. 105, 102401 https://doi.org/10.1016/j.jeem.2020.102401.

Du, Y., Wan, Q., Liu, Haimeng, Liu, Hao, Kapsar, K., Peng, J., 2019. How does urbanization influence PM 2.5 concentrations? Perspective of spillover effect of multi-dimensional urbanization impact. J. Clean. Prod. 220, 974–983. https://doi.org/10.1016/j.jclepro.2019.02.222.

Elsaid, K., Olabi, V., Sayed, E.T., Wilberforce, T., Abdelkareem, M.A., 2021. Effects of COVID-19 on the environment: an overview on air, water, wastewater, and solid waste. J. Environ. Manag. 292, 112694 https://doi.org/10.1016/j.jenvman.2021.112694.

Fromer, N., Mi, Z., Guan, D., Liu, Z., Liu, J., Vigui, V., Wang, Y., 2019. Cities : the core of climate change mitigation. J. Clean. Prod. 207, 582–589. https://doi.org/10.1016/j.jclepro.2018.10.034.

Gonzalez, J.N., Gomez, J., Vassallo, J.M., 2022. Do urban parking restrictions and Low Emission Zones encourage a greener mobility? Transport. Res. Transport Environ. 107, 103319 https://doi.org/10.1016/j.trd.2022.103319.

Grossman, G.M., Krueger, A.B., 1995. Economic growth and the environment. Q. J. Econ. 110, 353–377. https://doi.org/10.2307/2118443.

Guo, J., Zhang, X., Gu, F., Zhang, H., Fan, Y., 2020. Does air pollution stimulate electric vehicle sales? Empirical evidence from twenty major cities in China. J. Clean. Prod. 249, 1–12. https://doi.org/10.1016/j.jclepro.2019.119372.

Harvard, T.H., 2020. Chan School of public health. Coronavirus and Air Pollution 1, 1000.

Hewson, E.W., 1956. Meteorological factors affecting causes and controls of air pollution. J. Air Pollut. Control Assoc. 5, 235–241. https://doi.org/10.1080/00966665.1956.10467718.

Jiang, Z., Yu, X.B., 2020. Impact of Visible-Solar-Light-Driven photocatalytic pavement on air quality improvement. Transp. Res. Part D 84, 102341. https://doi.org/10.1016/j.trd.2020.102341.

Kanjilal, K., Ghosh, S., 2013. Environmental Kuznet's curve for India: evidence from tests for cointegration with unknown structuralbreaks. Energy Pol. 56, 509–515. https://doi.org/10.1016/j.enpol.2013.01.015.

Kavousi-fard, A., Samet, H., Marzbani, F., 2014. A new hybrid modified firefly algorithm and support vector regression model for accurate short term load forecasting. Expert Syst. Appl. 41, 6047–6056. https://doi.org/10.1016/j.eswa.2014.03.053.

Kumar, A., Goyal, P., 2011. Forecasting of air quality in Delhi using principal component regression technique. Atmos. Pollut. Res. 2, 436–444. https://doi.org/10.5094/APR.2011.050.

Le Quéré, C., Jackson, R.B., Jones, M.W., Smith, A.J.P., Abernethy, S., Andrew, R.M., De-Gol, A.J., Willis, D.R., Shan, Y., Canadell, J.G., Friedlingstein, P., Creutzig, F., Peters, G.P., 2020. Temporary reduction in daily global CO2 emissions during the COVID-19 forced confinement. Nat. Clim. Change 10, 647–653. https://doi.org/10.1038/s41558-020-0797-x.

Lee, M., Lin, L., Chen, C., Tsao, Y., Yao, T., Fei, M., 2020. Forecasting air quality in taiwan by using machine learning. Sci. Rep. 1–13. https://doi.org/10.1038/s41598-020-61151-7.

Li, R., Cui, L., Hongbo, F., Li, J., Zhao, Y., Chen, J., 2020. Satellite-based estimation of full-coverage ozone (O3) concentration and health effect assessment across Hainan Island. J. Clean. Prod. 244, 118773 https://doi.org/10.1016/j.jclepro.2019.118773.

Li, Z., Yim, S.H.L., Ho, K.F., 2020. High temporal resolution prediction of street-level PM2.5 and NOx concentrations using machine learning approach. J. Clean. Prod. 268, 121975 https://doi.org/10.1016/j.jclepro.2020.121975.

Liu, H., Chen, C., 2020. Spatial air quality index prediction model based on decomposition, adaptive boosting, and three-stage feature selection: a case study in China. J. Clean. Prod. 265, 121777 https://doi.org/10.1016/j.jclepro.2020.121777.

Liu, J., Wang, J., Cheng, J., Zhang, Y., Wang, T., Pan, W.P., 2020. Distribution and emission of speciated volatile organic compounds from a coal-fired power plant with ultra-low emission technologies. J. Clean. Prod. 264, 121686 https://doi.org/10.1016/j.jclepro.2020.121686.

Ma, J., Ding, Y., Cheng, J.C.P., Jiang, F., Tan, Y., Gan, V.J.L., Wan, Z., 2020. Identification of high impact factors of air quality on a national scale using big data and machine learning techniques. J. Clean. Prod. 244, 118955 https://doi.org/10.1016/j.jclepro.2019.118955.

Ma, X., Li, C., Dong, X., Liao, H., 2020. Empirical analysis on the effectiveness of air quality control measures during mega events : evidence from Beijing , China. J. Clean. Prod. 271, 122536 https://doi.org/10.1016/j.jclepro.2020.122536.

Magazzino, C., Mele, M., Sarkodie, S.A., 2021. The nexus between COVID-19 deaths, air pollution and economic growth in New York state: evidence from Deep Machine Learning. J. Environ. Manag. 286, 112241 https://doi.org/10.1016/j.jenvman.2021.112241.

Mazzeo, N.A., Venegas, L.E., Choren, H., 2005. Analysis of NO, NO2, O3 and NOx concentrations measured at a green area of Buenos Aires City during wintertime. Atmos. Environ. 39, 3055–3068. https://doi.org/10.1016/j.atmosenv.2005.01.029.

Minet, L., Stokes, J., Scott, J., Xu, J., Weichenthal, S., Hatzopoulou, M., 2018. Should tra ffi c-related air pollution and noise be considered when designing urban bicycle networks. Transp. Res. Part D 65, 736–749. https://doi.org/10.1016/j.trd.2018.10.012.

Mukherjee, A., McCarthy, M.C., Brown, S.G., Huang, S.M., Landsberg, K., Eisinger, D.S., 2020. Influence of roadway emissions on near-road PM2.5: monitoring data analysis and implications. Transport. Res. Transport Environ. 86, 102442 https://doi.org/10.1016/j.trd.2020.102442.

Pal, S., Das, P., Mandal, I., Sarda, R., Mahato, S., Nguyen, K.A., Liou, Y.A., Talukdar, S., Debanshi, S., Saha, T.K., 2021. Effects of lockdown due to COVID-19 outbreak on air quality and anthropogenic heat in an industrial belt of India. J. Clean. Prod. 297, 126674 https://doi.org/10.1016/j.jclepro.2021.126674.

Perman, R., Stern, D.I., 2003. Evidence from panel unit root and cointegration tests that the Environmental Kuznets Curve does not exist. Aust. J. Agric. Resour. Econ. 47, 325–347. https://doi.org/10.1111/1467-8489.00216.

Qiu, Q., Wang, Y., Qiao, S., Liu, R., Bian, Z., Yao, T., Son, T., 2020. Does air pollution affect consumer online purchasing behavior ? The effect of environmental psychology and evidence from China. J. Clean. Prod. 260, 120795 https://doi.org/10.1016/j.jclepro.2020.120795.

Rizova, M.I., Wong, T.C., Ijomah, W., 2020. A systematic review of decision-making in remanufacturing. Comput. Ind. Eng. 147, 106681 https://doi.org/10.1016/j.cie.2020.106681.

Sahraei, M.A., Kuşkapan, E., Çodur, M.Y., 2021. Public transit usage and air quality index during the COVID-19 lockdown. J. Environ. Manag. 286 https://doi.org/10.1016/j.jenvman.2021.112166.

Sahu, S.K., Zhu, S., Guo, H., Chen, K., Liu, S., Xing, J., Kota, S.H., Zhang, H., 2021. Contributions of power generation to air pollution and associated health risks in India: current status and control scenarios. J. Clean. Prod. 288, 125587 https://doi.org/10.1016/j.jclepro.2020.125587.

Saz-Salazar, S. del, Feo-Valero, M., Vázquez-Paja, B., 2020. Valuing public acceptance of alternative-fuel buses using a Latent Class Tobit model: a case study in Valencia. J. Clean. Prod. 261 https://doi.org/10.1016/j.jclepro.2020.121199.

Smola, A.J., Olkapf, B., 2004. A tutorial on support vector regression. Stat. Comput. 14, 199–222.

Song, C., Fu, X., 2020. Research on different weight combination in air quality forecasting models. J. Clean. Prod. 261, 121169 https://doi.org/10.1016/j.jclepro.2020.121169.

Song, J., Zhao, C., Lin, T., Li, X., Prishchepov, A.V., 2019. Spatio-temporal patterns of traffic-related air pollutant emissions in different urban functional zones estimated by real-time video and deep learning technique. J. Clean. Prod. 238, 117881 https://doi.org/10.1016/j.jclepro.2019.117881.

Streiff, L., 2020. NASA Model Reveals How Much COVID-Related Pollution Levels Deviated from the Norm (Nasa Tv).

Tan, X., Han, L., Zhang, X., Zhou, W., Li, W., Qian, Y., 2021. A review of current air quality indexes and improvements under the multi-contaminant air pollution exposure. J. Environ. Manag. 279, 111681 https://doi.org/10.1016/j.jenvman.2020.111681.

Tibrewal, K., Venkataraman, C., 2022. COVID-19 lockdown closures of emissions sources in India: lessons for air quality and climate policy. J. Environ. Manag. 302, 114079 https://doi.org/10.1016/j.jenvman.2021.114079.

Tsurumi, T., Managi, S., 2010. Decomposition of the environmental Kuznets curve: scale, technique, and composition effects. Environ. Econ. Pol. Stud. 11, 19–36. https://doi.org/10.1007/s10018-009-0159-4.

Usman, M., Jahanger, A., 2021. Correction to: heterogeneous effects of remittances and institutional quality in reducing environmental deficit in the presence of EKC hypothesis: a global study with the application of panel quantile regression (Environmental Science and Pollution Resear. Environ. Sci. Pollut. Res. 28, 37311 https://doi.org/10.1007/s11356-021-13791-z.

Vasudevan, M., Natarajan, N., Selvi, S.M., Ravikumar, K., Rajendran, A.D., Bagavathi, A. B., 2021. Correlating the trends of COVID-19 spread and air quality during lockdowns in Tier-I and Tier-II cities of India—lessons learnt and futuristic strategies. Environ. Sci. Pollut. Res. https://doi.org/10.1007/s11356-021-16028-1.

Vijayan, V.K., Paramesh, H., Salvi, S.S., Dalal, A.A.K., 2015. Enhancing indoor air quality -The air filter advantage. Lung India 32, 473–479. https://doi.org/10.4103/0970-2113.164174.

Wang, Y., Lu, Y., 2020. Evaluating the potential health and economic effects of nitrogen fertilizer application in grain production systems of China. J. Clean. Prod. 264, 121635 https://doi.org/10.1016/j.jclepro.2020.121635.

Wang, Y., Zhao, Y., 2021. Is collaborative governance effective for air pollution prevention? A case study on the Yangtze river delta region of China. J. Environ. Manag. 292.

Wang, H.W., Li, X.B., Wang, D., Zhao, J., He, H. di, Peng, Z.R., 2020. Regional prediction of ground-level ozone using a hybrid sequence-to-sequence deep learning approach. J. Clean. Prod. 253, 119841 https://doi.org/10.1016/j.jclepro.2019.119841.

Wang, Yufang, Wang, H., Zhang, S., 2020. Quantifying prediction and intervention measures for PM2.5 by a PDE model. J. Clean. Prod. 268, 122131 https://doi.org/10.1016/j.jclepro.2020.122131.

Wang, Ying, Wang, J., Li, Z., 2020. A novel hybrid air quality early-warning system based on phase-space reconstruction and multi-objective optimization : a case study in China. J. Clean. Prod. 260, 121027 https://doi.org/10.1016/j.jclepro.2020.121027.

Xiong, F., Pan, J., Lu, B., Ding, N., Yang, J., 2020. Integrated technology assessment based on LCA: a case of fine particulate matter control technology in China. J. Clean. Prod. 268, 122014 https://doi.org/10.1016/j.jclepro.2020.122014.

Yadav, M., Sahu, S.P., Singh, N.K., 2019. Multivariate statistical assessment of ambient air pollution in two coalfields having different coal transportation strategy: a comparative study in Eastern India. J. Clean. Prod. 207, 97–110. https://doi.org/10.1016/j.jclepro.2018.09.254.

Yang, G., Zhang, Y., Li, X., 2020. Impact of gasoline upgrade policy on particulate matter pollution in China. J. Clean. Prod. 262, 121336 https://doi.org/10.1016/j.jclepro.2020.121336.

Yilmaz, I., Kaynar, O., 2011. Multiple regression, ANN (RBF, MLP) and ANFIS models for prediction of swell potential of clayey soils. Expert Syst. Appl. 38, 5958–5966. https://doi.org/10.1016/j.eswa.2010.11.027.

Yuan, M., Song, Y., Huang, Y., Shen, H., Li, T., 2019. Exploring the association between the built environment and remotely sensed PM2.5 concentrations in urban areas. J. Clean. Prod. 220, 1014–1023. https://doi.org/10.1016/j.jclepro.2019.02.236.

Zanobetti, A., Austin, E., Coull, B.A., Schwartz, J., Koutrakis, P., 2014. Health effects of multi-pollutant profiles. Environ. Int. 71, 13–19. https://doi.org/10.1016/j.envint.2014.05.023.

Zeinalnezhad, M., Gholamzadeh, A., Kleme, J., 2020. Air pollution prediction using semi-experimental regression model and Adaptive Neuro-Fuzzy Inference System. J. Clean. Prod. 261, 1–16. https://doi.org/10.1016/j.jclepro.2020.121218.

Zhang, C., Wang, H., Bai, L., Wu, C., Shen, L., Sippula, O., Yang, J., Zhou, L., He, C., Liu, J., Ristovski, Z., Morawska, L., Wang, B., 2020. Should industrial bagasse-fired boilers be phased out in China? J. Clean. Prod. 265, 121716 https://doi.org/10.1016/j.jclepro.2020.121716.

Zhang, Q., Gao, T., Liu, X., Zheng, Y., 2020. Exploring the influencing factors of public environmental satisfaction based on socially aware computing. J. Clean. Prod. 266, 121774 https://doi.org/10.1016/j.jclepro.2020.121774.

Zhou, Y., Chang, F., Chang, L., Kao, I., Wang, Y., 2019. Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts. J. Clean. Prod. 209, 134–145. https://doi.org/10.1016/j.jclepro.2018.10.243.

Zhuang, M., Zhang, J., Kong, Z., Fleming, R.M., Zhang, C., Zhang, Z., 2020. Potential environmental benefits of substituting nitrogen and phosphorus fertilizer with useable crop straw in China during 2000–2017. J. Clean. Prod. 267, 122125 https://doi.org/10.1016/j.jclepro.2020.122125.