Prediction Of Air Quality Index Using Machine Learning Algorithms

Alakananda1, Rajashree Krishna2

1,2 *Department Of Computer Science and Engineering*

*Manipal Institute Of Technology, Manipal Academy Of Higher Education*

Manipal, Karnataka, India 576104

 raji.krish@manipal.edu

 *Abstract*— As a result of industrialization and global urbanization air pollution was increased to a large extent in the past two, three decades. Various air pollutants like carbon monoxide, nitrogen dioxide, sulfur dioxide, suspended particulate matter, benzene that are present in the air affected both the environment and humans. Long and short term exposure to this polluted air caused abrupt climate changes and many health-related problems to humans. Air quality prediction helps both policymakers and people to reduce air pollution. In this paper, we have proposed a model to predict air quality in terms of the Air Quality Index (AQI) based on machine learning techniques. To predict AQI, we have considered major air pollutants like nitrogen dioxide, sulfur dioxide, carbon monoxide, ozone, nitrogen oxides and suspended particulate matter(PM2.5, PM10). The proposed model makes use of Linear regression, Support Vector Machine regression and Gaussian process regression techniques to predict air quality. Meteorological data and historical values of air pollutant concentration were given as input to predict the air quality index. Then we have applied classification algorithms to classify air quality based on Indian Government Standards.

Keywords—air pollution, machine learning, air quality prediction, Air Quality Index(AQI), meteorological data, Linear Regression(LR), Linear Support Vector Regression(LSVR), Exponential Gaussian Process Regression(EGPR).

# Introduction

 Air pollution occurs when harmful or excessive quantities of substances including gases, particulates, and biological molecules are introduced to the earth’s atmosphere. Air pollution may occur as a result of human activity or as an environmental process. Air pollution results in ecological imbalance and causes abrupt climate changes. Carbon monoxide, nitrogen dioxide, sulfur dioxide, suspended particulate matter, and ground-level ozone are some of the major air pollutants that affect the environment, humans and other animals. Long term exposure to this polluted air leads to a variety of health diseases like cancer, asthma, cardiovascular diseases, skin diseases and also damages the human immune system; whereas temporary exposure causes headaches, irritation of eyes and breathing problems.

There are many reasons for air pollution but the most important one is global urbanization and industrialization. Today industries are set up all over the world and they are emitting polluted gases to the atmosphere, as a result, the quality of air is decreasing day by day. Controlling this air pollution is a challenging task for all countries across the world, especially to the developing countries. Thus developing countries like India and China have imposed various acts to control and prevent air pollution, thereby safeguarding millions of people's health. The government had set up many air pollution monitoring stations to assess the quality of air by taking account of air parameters like SO2, NO2, CO, PM2.5, PM10, etc. which are harmful and thereby taking necessary actions to decrease pollution rate. Air quality forecasting allows both government and people to know about the quality of air in advance and helps to take necessary controlling measures. AQI is a numerical value associated with air quality level and can be mapped to different categories ranging from good to severe based on air pollutants concentration. AQI system is used by government agencies to tell common people about how polluted the air is at present and to tell how polluted it will be in the future using numerical ranges associated with it without knowing about the details of pollutants. It also tells about the impact of breathing polluted air on human health. If we predict AQI then people can accordingly plan their indoor and outdoor activities. Predicting air quality is a challenging task as it depends on many factors and research is going on to improve the prediction accuracy.

There are various techniques to predict air quality and they can be generalized to four major categories: statistical forecasting methods, artificial intelligence-based forecasting methods, numerical forecasting methods, and hybrid forecasting methods [1]. Statistical and numerical forecasting methods make use of historical data to predict air quality so they are not efficient. Air quality prediction done using hybrid method have excellent performance than all other methods. Artificial intelligence techniques have better performance than statistical and numerical forecasting methods. In this paper, we are making use of machine learning techniques to predict air quality. Machine learning is a branch of computer science where computers or machines are trained to learn from data without any explicit programming. Since the machine learning method has capacities like fast processing, real-time prediction and ability to handle multi-dimensional, multi-variety data in a dynamic environment we are making use of this technique to predict air quality.

#  Related Work

 Air pollution severely affected human health in the last few decades. A variety of air pollution monitoring and air quality prediction techniques are developed so that necessary controlling measures can be taken to reduce this air pollution. Neural network, K-nearest neighbor, support vector machines (SVM) and decision tree algorithms are used to predict AQI and are compared with other air prediction models to measure the performance. To the developed model concentration of SO2, NO2, O3, PM2.5, PM10, and CO was given as input and three classifiers low, medium, high was used to predict the air pollution levels [2]. Out of these four models, neural network model with one input, one hidden and one output layer showed accurate results. To the input layer, 6 air parameters were given as input, the hidden layer consisted of 10 neurons and the output layer had 3 classifiers. Here it should be noticed that in order to predict air quality index they used previous day air pollutants concentration but air pollutants concentration is highly affected by various meteorological data, so results were not accurate. A simple neural network and improved neural network-based model was developed to predict AQI[3][4]. The improved neural network model used a BP neural network and genetic algorithm for prediction. To the developed model meteorological data and AQI of the previous day were given as input. Using the algorithms the model generated 6 AQI classifiers as output. Later developed model was compared with a simple neural network-based air quality prediction model for its accuracy. The result showed that an improved BP neural network-based air quality prediction model had higher accuracy than the neural network. But to predict AQI they are not considering the concentration of air pollutants so we can say that the predicted AQI is not accurate.

It is important to consider both meteorological parameters and air parameters to predict AQI, so later many models are developed to predict air pollutants concentration considering both meteorological data and air pollutants. Yves Rybarczyk and Rasa Zalakeviciute [5] developed a model where initially meteorological data and concentration of PM2.5 were collected and later the collected data were carefully analyzed to determine the dependency between each meteorological parameter and particulate matter. An ANFIS (adaptive neuro-fuzzy inference system) based model was developed by Sanda Florentina Mihalache, Marian Popescu and Mihaela Oprea to predict particulate matter concentration for next one hour [6]. ANFIS combines features of FIS and Artificial Neural Network (ANN) to predict particulate matter (PM) concentration. The developed model took the last 4 hours' concentration of PM as input to predict next hour PM concentration. The FIS was formed by a fuzzification unit, defuzzification unit, database unit, rule base unit, and decision unit. The result showed that the prediction was poor when the output type changed from constant to linear. Jean-Michel Poggia and Bruno Portierb[7] developed a cluster wise regression-based model to predict PM10 concentration by considering meteorological data and the average concentration of PM10 in the previous day. The model made use of random forest technique for variable selection. Later the cluster wise regression model was compared with persistence, generalized additive non-linear model for measuring accuracy of PM10 prediction. It is observed that the cluster wise regression-based model had higher accuracy than the other two models. An ensemble method for forecasting nitrogen dioxide, ozone and PM10 was developed on the Prev'Air operational platform. Prev' Air is an operational platform used for predicting air pollutants concentration in France and aims to inform people about pollutants concentration in advance [8]. This platform consists of a variety of operational models where different weights are assigned before each simulation to predict pollutant concentration and the performance of each model varies with inputs and time. Initially, air pollutant concentration was predicted and later ensembled to get high accuracy. But this method requires aggregation method to calculate the uncertainty involved in the aggregated forecast and also requires to improve weight assigning methods. Support Vector Machine (SVM) and its variant, Least Square Support Vector Machines (LS-SVM) are the most common and simple machine learning techniques based on statistical learning theory used for regression, time series prediction. An LS-SVM based model was developed to predict air quality[9]. SVM based models made use of quadratic programs(QP) to solve nonlinear regression problems but LS-SVM made use of a set of linear equations for solving regression problems so the LS-SVM algorithm-based model was developed to predict air quality. For the developed model meteorological data and air pollutants are given as input but the prediction was good only for SO2, NO2, SPM and it showed a large discrepancy for other pollutants, especially for ozone. Zhongshan Yang and Jian Wang developed a hybrid air pollution monitoring and early warning system [10]. This model consists of an air pollutant evaluation and prediction unit. The evaluation unit made use of a comprehensive fuzzy technique to determine the main pollutants. The developed hybrid model made use of the Cuckoo search algorithm and differential evaluation algorithms to initialize weight and threshold. This hybrid model was later used to predict pollutants. MLP and RBF algorithm-based ANN model was developed by W. Kaminski et al. to predict air quality classes [11]. To the developed model meteorological data and air pollutant data were given as input. But this model showed poor results as it considered only one air pollutant that is particulate matter(PM) to predict air quality classes. A recurrent air quality prediction model was developed to predict air pollutants by Ke GU et al. [12]. In order to predict pollutants concentration for the next one-hour historical air pollutant and meteorological data were given as input. This model showed good results as it successfully addressed high correlation between air pollutants and meteorological data but prediction can be improved if we use ensemble learning. An extreme learning machine based model was proposed by Jiangshe Zhang and Weifu Ding to predict air pollutants considering previous days pollutant concentration and it showed better results compared to other machine learning algorithms [13].

Studies showed that in order to predict air quality or air pollutants concentration accurately we need to consider the Spatio-temporal correlation between meteorological data and air pollutants. So later several models were developed considering their Spatio-temporal correlation. A Spatio-temporal prediction of PM2.5 was done by Yu Zhan et al. using a Geographically weighted Gradient Boosting Machine(GW-GBM) [14]. This GW-GBM model successfully handled missing data and Spatio-temporal dependencies of PM2.5 with other factors. Lei Song et al. developed a spatial data aided incremental support vector regression model was to predict Spatio-temporal PM2.5 concentration. In order to predict PM2.5, the model was supplied with historical values of daily averaged PM10 and PM2.5.the model first performed temporal prediction to gain knowledge about spatial dimensions and later spatial data aided prediction learning was done to gain Spatio-temporal cross-dimensional knowledge [15]. Even though several algorithms are developed to predict Spatio-temporal air quality their accuracy can be further improved if we consider all depending meteorological and air parameters. Thus we can tell that predicting air quality by considering all dependent parameters is a difficult task.

#  Methodology

 There are various methods to predict AQI. Most of the methods use historical values of various pollutants concentration as input variables to predict AQI[16][17][18] and some use meteorological parameters as inputs to predict AQI[19][20].

 In this paper, we are using machine learning algorithms to predict AQI. We have considered three different sets of input and these inputs are given to machine learning algorithms to predict AQI. Later we compare the results obtained by different sets of input and machine learning algorithms.

##  Data descripton

 The data used in this study consists of meteorological data and air parameter. Meteorological data used in this study are atmospheric temperature(degree c), pressure(mmHg) ,relative humidity(%) ,wind degree(degree) and wind speed(m/s). The air parameters considered are carbon monoxide(CO), nitrogen oxides(NOx), ozone(O3), suspended particulate matter (PM2.5),particulate matter(PM10), nitrogen dioxide(NO2) and sulphur dioxide(SO2). Carbon monoxide is measured in mg/m3 and all other parameters are measured in ug/m3.

 As Delhi is suffering from more air pollution compared to other states we considered it as the target area of the study. We collected meteorological data and air pollutants concentrations from RK Puram monitoring stations of Delhi as an average of 24 hours for three years. The total number of sample data collected in RK Puram station is 900 and out of these 750 samples are used as training set data and 150 samples are used as testing data ie. 85% as training data and 15% as test data. These meteorological data and air parameters concentration of Delhi are collected from the central pollution control board(CPCB), Govt of India website( <https://app.cpcbccr.com/ccr/#/caaqm-dashboard-all/caaqm-landing>).

##  Air Quality Index Calculation and Classification

 An air quality index is deﬁned as an overall scheme that transforms the weighed values of individual air pollution-related parameters (for example, pollutant concentrations) into a single number. In India Air Quality Index value is calculated according to National Ambient Air Quality Standard (CPCB 2009) and it specified that 12 air parameters: PM2.5, PM10, NO2, SO2, CO, O3, Pb, NH3, As, Ni, Benzo(a) pyrene and Benzene should be considered for calculating AQI based on availability of data. The selection of parameters is based on AQI objective, data availability, averaging period, measurement method and frequency of monitoring. PM2.5, PM10, SO2, NO2, Pb and NH3 have 24 hourly as well as annual averaging standards.CO and O3 have short-term standards (1 and 8 hours averaging standard) and As, Ni, Benzo(a) pyrene and Benzene have an annual averaging standard. Usually, 6 parameters(PM2.5, PM10, SO2, NO2, CO, and O3) are considered for calculating AQI. The number of parameters used in calculating AQI value can be increased or decreased based on the availability of data but to calculate AQI value there must be a minimum of 3 parameters and one of them should be either PM10 or PM2.5.

 Air quality index calculation involves two steps[24]. First, we need to calculate the sub-index value for each pollutant and later AQI value is computed as the aggregate function of these sub-indices as shown in Fig. 1.

 Each air pollutant is associated with a sub-index value (Ii) based on its concentration value (Xi) value in air. The general equation used for calculating sub-index for a given pollutant concentration is shown in (1) below:

 Ii = [{(IHI - ILO)/(BHI -BLO)} \* (Cp-BLO)]+ ILO (1)

where,

BHI = Breakpoint concentration greater or equal to given concentration.

BLO = Breakpoint concentration smaller or equal to given concentration.

IHI = AQI value corresponding to BHI.

ILO = AQI value corresponding to BLO.

Ip = Pollutant concentration.

 Once sub-index associated with each pollutant is calculated then AQI value is calculated using (2):

 I = MAX (I1,I2,I3……In) (2)

where,

I = AQI value.

I1,I2,…In = sub-index value associated with n pollutants.

 Fig. 2 shows the AQI value classification and breakpoint concentration for each pollutant according to Indian government standards.



Fig. 1. Formation of aggregated Air Quality Index.



 Fig. 2. AQI classes, pollutants, and breakpoint concentration.



 Fig. 3. Steps involved in prediction.

## Machine Learning Algorithms

 Machine learning is a branch of artificial intelligence, where the system automatically learns to analyze the input with minimal human intervention. Machine learning can be broadly classified as supervised learning, unsupervised learning and re-enforcement learning. To predict AQI, we are making use of supervised learning. Supervised learning can be further classified as regression and classification learning. In regression, we predict the output value for a given input value based on previous learning but in classification, we categorize input data to one or more groups based on previous learning or training. To predict pollutant concentration and AQI value we are using regression algorithms; for classifying predicted AQI we are using classification algorithms.

 The regression methods used here are linear regression, linear support vector regression, and exponential Gaussian process regression. Linear regression is the most commonly used simple supervised learning technique of machine learning where we predict the dependent variable based on the values of an independent variable. Support vector machines can be used for both regression and classification problems. When SVM is used for classification it will classify the data to two or more categories but during regression, it will predict the response variable values based on input data and training. Gaussian processes are powerful non-parametric tools widely used in supervised learning. These processes have the ability to learn the noise and to estimate the uncertainty involved based on the training data.

 The classification methods used in this paper to classify the air quality based on pollutant concentration are decision trees, SVM and ensemble classifiers. For decision trees we are considering only fine trees, in SVM we are using linear SVM and the ensemble classifier used here is ensemble bagged trees.

## Methodology

The general flow of the methodology is shown in figure 3. Initially, we perform preprocessing on collected data of the monitoring station. Later we calculate the sub-index for each pollutant. Then we calculate the AQI value. Then data is divided into training set data and test data. Machine learning algorithms like LR, SVR, and EGPR are applied to processed input data to calculate AQI value. Since AQI value and associated air quality classes are difficult to remember for common people we then apply classification algorithms that classify the air into different classes based on each pollutant concentration. In this paper, we are giving three different sets of input to ML algorithms. The details about these three types of inputs are given below.

#### Giving predicted pollutants concentration as input

In all previous studies in order to predict AQI researchers used historical values of pollutant concentration. In this study, we predict pollutant concentration by considering current meteorological data parameters[7][22]. Initially, meteorological parameters like temperature, pressure, humidity, wind speed, wind direction are given as input to regression algorithms to predict each pollutant concentration. Once all different pollutants like PM2.5, PM10, SO2, and NO2 concentration are predicted then these predicted air pollutants concentration values are given as input to three regression algorithms to predict current AQI value.

#### b)Giving air pollutants concentration

In this type of input, the present day’s air pollutant concentration values are directly given as input to LR, SVR and EGPR algorithms to predict AQI value[23]. In this study, we have considered pollutants like CO, NO, O3, SO2, NO2, PM2.5, PM10. These pollutant concentrations are then given as input to regression algorithms to calculate AQI value and to classification algorithms like fine tree, SVM and ensemble bagged tree to classify air quality to different classes.

#### c)Giving air pollutants concentration along with meteorological data values

Here we are considering both pollutant concentration value as well as meteorological parameters as input to the

Table I. Aqi Prediction Result For Predicted Pollutants concentration

|  |  |  |
| --- | --- | --- |
|  | **RMSE** | **R-squared** |
| **LR** | 112 | 0.27 |
| **SVR** | 112 | 0.27 |
| **EGPR** | 146 | -0.21 |

Table II. Aqi Prediction Result For Pollutants concentration

|  |  |  |
| --- | --- | --- |
|  | **RMSE** | **R-squared** |
| **LR** | 34.36 | 00.93 |
| **SVR** | 38.61 | 0.91 |
| **EGPR** | 25.79 | 0.96 |

Table II. Aqi Prediction Result For Pollutants concentration and Meteorological data

|  |  |  |
| --- | --- | --- |
|  | **RMSE** | **R-squared** |
| **LR** | 34.96 | 00.93 |
| **SVR** | 38.64 | 0.91 |
| **EGPR** | 28.55 | 0.95 |

regression and classification algorithms to classify air quality. The pollutants considered are the same as the above method. Meteorological parameters considered are temperature, pressure, relative humidity, wind speed, wind direction, and sun fall ratio.

# Result

The results obtained using three types of input with three types of regression and classification algorithms are expressed in terms of Root Mean Square error values and R-squared values in this section. RMSE value is an indicator that specifies the difference between an actual observed value and the predicted value. R-squared value specifies the relationship between the independent variable and the dependent variable.

 Table I shows the RMSE and R-squared values of LR, SVR, and EGPR when predicted pollutant concentration is given as input to the regression algorithm. The RMSE and R-squared values are very high, in fact, it is above 100% which indicates that predicted pollutant concentration is completely incorrect when compared with actual concentrations. So we can say that meteorological parameters completely fail to predict pollutant concentration. In other words, we can say that pollutant concentrations have no dependency on meteorological parameters.

 RMSE and R-squared values in Table II indicate the results of regression algorithms when current air pollutant concentration is given as input. EGPR algorithm has an RMSE value of 25.79 which is very less compared to the other two regression algorithms and the R- squared value of 0.96 indicated that there is a 96% dependency of the dependent variable with the independent variable. The RMSE value in

Fig .4. AQI prediction results when pollutant concentration is given as input.

Fig.5. AQI prediction results when pollutant concentration and meteorological parameters are given as input.

Table II is very much less and R-squared values are very high compared to that of Table I which indicates that if pollutant concentration is predicted properly then predicted AQI has a very less error rate. Fig.4 shows the graph of predicted AQI values when pollutant concentration is given as input to regression algorithms.

 Table III shows the RMSE and R-squared values of the regression algorithm when both meteorological parameters and current pollutants concentration are given as input to ML algorithms. The result shows that the SVR algorithm has an RMSE value of 38.64 which is high when compared to other algorithms and EGPR has an RMSE value of 28.55 which is very less compared to other regression algorithms. Fig. 5 shows the graph of actual and predicted AQI values when input data to regression algorithms consists of both meteorological parameters as well as pollutant parameters.

 When we compare the results of 3 types of input with 3 types of regression algorithm we can say that meteorological parameters do not affect predicting pollutant concentration as well as predicting AQI value. This fact is noticeable if we compare the results of Table II and III. There is no much difference between RMSE and R-squared values of table II and III. So we say that AQI values have a high dependency on pollutant concentrations. If we predict pollutants concentration correctly then predicted AQI value has a very less error rate. We can also notice that the EGPR algorithm has better results compared to LR and SVM which has poor results among considered regression algorithms.

 Air quality level represented in terms of AQI value is useful for researchers and other people who are working on that domain but these AQI values are not easily understandable to common people if they do not know AQI value and its associated classes so we are again applying classification algorithm that classifies air quality into different classes. In our study, we are classifying air quality into 6 classes according to the Indian government standard.

 When we gave pollutants concentration and meteorological parameters as input to 3 classification algorithms like fine tree, SVM and ensemble bagged tree, Fine tree and ensemble bagged tree produced results with an accuracy of 98% and SVM has accuracy of 91.4%. When input containing only pollutant concentration is given as input to classification algorithm, the fine tree showed an accuracy of 98.5%, ensemble bagged tree showed accuracy of 97.9% and SVM showed accuracy of 91.7%.

 The results of classification algorithms have better accuracy compared to results obtained by regression algorithms. This is because in regression algorithm we are predicting the AQI value and small deviation in predicted value with actual value results in error but in classification algorithm, we are having range of AQI values that belong to each class so small deviation in AQI value does not result in error only if predicted class is wrong then only it causes decrease inaccuracy

# Conclusion

Prediction of AQI is very important so that people can take necessary preventive measures to avoid health-related problems that are associated with poor quality air. In this paper, we had considered 3 regression models like LR, LSVR, and EGPR to predict AQI value and later we had classified the air quality into six classes using 3 classification algorithms like fine tree, linear support vector regression and ensemble bagged tree. The performance of the EGPR model was good in predicting AQI value out of the other two regression algorithms. Fine tree and ensemble bagged tree's performance was good out of considered classification algorithms.

In this paper, we have considered meteorological parameters and pollutants concentration as input but in future all possible independent variables that can influence AQI value like vehicle emission and poisonous air emissions from power plants, industries, etc. can be considered for predicting AQI value.

##### References

1. Lu Bai, Jianzhou Wang, Xuejiao Ma, and Haiyan Lu, "Air Pollution Forecasts: An Overview" Int. J. Environ. Res. Public Health,vol 15, 4780, 17 April 2018, doi:10.3390/ijerph15040780.
2. Kostandina Veljanovska, Angel Dimoski, "Air Quality Index Prediction Using Simple Machine Learning Algorithms" International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 7, Issue 1, January – February 2018,pp.25-30.
3. Mukesh Sharma, Sachin Aggarwal, Purnendu Bose, and Ashok Deshpande, "Meteorology-based Forecasting of Air Quality Index Using Neural Network", unpublished
4. Wang Zhenghua, Tian Zhihui, "Prediction of air quality index based on improved neural network", 2017 International Conference on Computer Systems, Electronics, and Control (ICCSEC), Dalian, China•Dec25-27,2017.
5. Yves Rybarczyk, Rasa Zalakeviciute(2016). Machine Learning Approach to Forecasting Urban Pollution A case study of Quito, Ecuador. In 2016 IEEE Ecuador Technical Chapters Meeting, ETCM 2016 [7750810] IEEE. Available: https://doi.org/10.1109/ETCM.2016.7750810
6. Sanda Florentina Mihalache, Marian Popescu, Mihaela Oprea (2015, October 14-16). Particulate Matter Prediction using ANFIS Modelling Techniques,2015 19th International Conference on System Theory, Control and Computing (ICSTCC).
7. Jean-Michel Poggia, Bruno Portierb, "PM10 forecasting using clusterwise regression", ELSeVIER, Atmospheric Environment.Vol 45, Issue 38, pp. 7005-7014, December 2011.
8. E. Debry a, V. Mallet, "Ensemble forecasting with machine learning algorithms for ozone, nitrogen dioxide and PM10 on the Prev’Air platform", ELSEVIER, Atmospheric Environment. Vol 91, pp. 71-84, July 2014.
9. W.F.Ip, C.M. Vong, J.Y. Yang and P.K.Wong(June 20-23, Harbib, China). Forecasting Daily Ambient Air Pollution Based on Least Squares Support Vector Machines, Proceedings of the 2010 IEEE, International Conference on Information and Automation.
10. Zhongshan Yang, Jian Wang, "A new air quality monitoring and early warning system: Air quality assessment and air pollutant concentration prediction", ELSEVIER, Environmental Research.Vol 158, pp. 105–117, October 2017.
11. W. Kaminski J. Skrzypski, E. Jach-Szakie (August 2008). Application of Artificial Neural Networks (ANNs) to Predict Air Quality Classes in Big Cities. (ICSENG)19th International Conference on Systems Engineering. Available: https://doi.org/ 10.1109/ICSEng.2008.14
12. Ke Gu, Junfei Qiao, and Weisi Lin, "Recurrent Air Quality Predictor Based on Meteorology- and Pollution-Related Factors”. IEEE Trans. Industrial Informatics. Vol 14, pp 3946-3955, 2018.
13. Jiangshe Zhang and Weifu Ding, "Prediction of Air Pollutants Concentration Based on an Extreme Learning Machine: The Case of Hong Kong" Int. J. Environ. Res. Public Health. 2017. 14(2), 114; DOI:10.3390/ijerph14020114.
14. Yu Zhan, Yuzhou Luo, Xunfei Deng, Huajin Chen, Michael L. Grieneisen, Xueyou Shen, Lizhong Zhu, Minghua Zhang, "Spatiotemporal prediction of continuous daily PM2.5 concentrations across China using a spatially explicit machine learning algorithm", ELSEVIER.Atmospheric Environment. Vol 155, pp. 129-139, 2017.
15. Lei Song, Shaoning Pang, Ian Longley, Gustavo Olivares, and Abdolhossein Sarrafzadeh (2014, July6-11, Beijing, China). Spatio-temporal PM2.5 Prediction by Spatial Data Aided Incremental Support Vector Regression. (IJCNN) International Joint Conference on Neural Networks.
16. Suleiman A, Tight M.R., Quinn A.D, ” Applying machine learning methods in managing urban concentrations of traffic-related particulate matter (PM10 and PM2.5)”. ScienceDirect: Atmospheric Pollution Research. Vol 10, Issue 1, pp. 134-144, January 2019. Available:<https://doi.org/10.1016/j.apr.2018.07.001>
17. Khaled Bashir Shaban, Abdullah Kadri and Eman Rezk, "Urban Air Pollution Monitoring System With Forecasting Models ", IEEE SENSORS JOURNAL, Vol. 16, No. 8, APRIL 15, 2016
18. Hui Xie, Fei Ma, Qingyuan Bai( 2009). Prediction of Indoor Air Quality Using Artificial Neural Networks. 2009 Fifth International Conference on Natural Computation. DOI 10.1109/ICNC.2009.502
19. Tapiwa M. Chiwewe and Jeofrey Ditsela (July 2016). Machine Learning Based Estimation of Ozone Using Spatio-Temporal Data from Air Quality Monitoring Stations. INDIN 2016 IEEE International Conference on Industrial Informatics. DOI: 10.1109/INDIN.2016.7819134
20. Xia Xi, Zhao Wei, Rui Xiaoguang, Wang Yijie, Bai Xinxin, Yin Wenjun, Don Jin (2015). A Comprehensive Evaluation of AirPollution Prediction Improvement by a Machine Learning Method.2015 IEEE International Conference on Service Operations And Logistics, And Informatics.
21. Bo Liu, Shuo Yan, Jianqiang Li, Yong Li (2016). Forecasting PM2.5 Concentration using Spatio-Temporal Extreme Learning Machine. 2016 15th IEEE International Conference on Machine Learning and Applications. DOI 10.1109/ICMLA.2016.100
22. Mahanijah Md Kamal, Rozita Jailani and Ruhizan Liza Ahmad Shauri (June 2006). Prediction of Ambient Air Quality Based on Neural Network Technique. (SCOReD 2006) 4th Student Conference on Research and Development.
23. www.indiaenvironmentportal.org.in/files/file/Air%20Quality%20Index.pdf