IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Revue d'Intelligence Artificielle
Vol. 36, No. 1, February, 2022, pp. 73-78
Journal homepage: http://iieta.org/journals/ria

Machine Learning Approaches Used for Air Quality Forecast: A Review

Saminathan Senthivel*, Malathy Chidambaranathan

School of Computing, SRM Institute of Science and Technology, Kattankulathur 603203, Chengalpattu Dist., TN, India

Corresponding Author Email: saminats@srmist.edu.in

https://doi.org/10.18280/ria.360108

Received: 25 May 2021
Accepted: 25 January 2022

Keywords:
air quality index, forecasting, prediction, classification, machine learning

ABSTRACT

Air Quality Index (AQI) is an indicator of the pollution level of our surroundings and household. Prediction of the AQI values from the historical values can help us analyze and mitigate the pollution levels. The AQI values can be classified into predetermined categories and machine learning algorithms can be made use of to improve the classification accuracy of the Air Quality Index value calculated. The main objective of the paper is to provide the potential researchers, with the importance of various Machine Learning approaches used for the forecast of the Air Quality Index. This paper analyzes various strategies used for the prediction, classification of AQI incorporating machine learning techniques. The air quality index can be calculated using Machine learning-based methods. Some of the methods to be considered are logistic regression, decision tree, support vector regression, support vector classifier, random forest tree, Naive Bayes classifier, and K-nearest neighbor. Application of these methods on the Air Quality Index datasets may yield different Accuracy, Recall, and F1 Score. Different algorithms that can be used for the said purpose with their strengths are summarized in a comparison table.

1. INTRODUCTION

The use of Artificial Intelligence and Machine Learning in recent years is helping in the advancement of technology. Machine learning is an area where an artificial intelligence system learns from data. Various strategies are utilized for the forecast/classification of the Air Quality Index (AQI) utilizing machine learning techniques. Following are some of the machine learning algorithms discussed in this paper for the forecast of Air Quality: logistic regression, decision tree, support vector machine, random forest tree, Naive Bayes classifier, and K-nearest neighbor.

The organization of the paper is like this: Section 2 of this paper discusses the effect of air pollution on the health of the people, and the environment. Air pollution apart from causing various diseases may sometimes lead to the death of people as well. It is hence needed to measure air quality levels. Air Quality Index is a measure of Air Quality. The AQI formula provided by USEPA is considered in this paper. The AQI measure in India is built around six significant pollutants.

Section 3 of this paper outlines existing challenges and related work. It describes the effects of air pollution on the health of human beings and the impact of the COVID-19 pandemic on Air Quality concerning specific locations as a case study.

Section 4 outlines the different formulae used for the performance assessment of air quality forecasting.

Section 5 of the paper gives an overview of the development of machine learning approaches in AQI forecasting. It is broadly categorized into both Prediction and Classification approaches. A summary of machine learning methods for Air Quality forecasting is given in the form of a table in Section 6 that is followed by the conclusions arrived at by studying the various approaches used in air quality forecast.

2. AIR POLLUTION

According to WHO's report on Health [1], millions of people die because of outdoor and household Air pollution every year globally. For reducing the high levels of pollutants, WHO has issued guideline limits. Data provided by WHO shows 90% of people are exposed to air pollution in excess of the prescribed upper limits. In addition to outdoor pollution, exposure to indoor air pollutants also causes health issues for people of all ages. This ranges from respiratory diseases, cancer to eye problems. Air pollution remains a major threat to people's health and the environment. The combined effect of outdoor and household air pollution is the cause of a lot of deaths annually. This is the result of the increased number of deaths from diseases ranging from heart disease to acute respiratory infections caused by air pollution.

India is no exception to the problems caused by air pollution. In many places, pollutant levels are higher than that of the highest possible values assumed fit for the healthy living of human beings. The situation is severe in bigger urban and heavily industrialized areas like Delhi. The central government as well as the state administrations are taking some steps to reduce air pollution levels.

2.1 Air quality index

The definition of the Air Quality Index varies across different parts of the world by different standards-making organizations.

AQI is calculated by the following formula:

$$I = \frac{I_{high} - I_{low}}{C_{high} - C_{low}} * \left(C - I_{low}\right) + I_{low} \tag{1}$$

where,

I=AQI

C=Concentration of pollutants

C_{low}=the concentration breakpoints < C

C_{high}=the concentration breakpoints >=C

I_{low}=the index breakpoint conforming to C_{low}

I_{high}=the index breakpoint conforming to C_{high}

Source: USEPA

USEPA has categorized the AQI into 6 categories. They range from Good to Hazardous. Central Pollution Control Board, India has come out with a National AQI scheme that was presented as a report in Ref. [2]. Information on air quality can be obtained from AQI in qualitative terms (e.g. good, satisfactory, poor, etc.) as well as its related potential health impacts. There are six AQI categories suggested: ranging from Good to Severe. For AQI calculation, normally eight pollutants are considered. For these pollutants, short-term (up to 24-hour averaging period) standards are prescribed. Nevertheless, AQI can be calculated once data for a minimum of three pollutants are available. These three should include anyone Particulate Matter i.e., PM 2.5 or PM 10. Based on the measured concentrations of the pollutants, corresponding standards, and likely impact on health caused by the pollutant, a sub-index is calculated for each. The worst sub-index among all these is considered as the overall AQI.

In India, the AQI is measured based mainly on 6 pollutants namely Particulate Matters (PM10, PM2.5), and the gases Ammonia (NH_3), Carbon Monoxide (CO), Nitrogen Dioxide (NO_2), and Ozone (O_3). A station monitoring these provide the concentration of specific pollutant over some interval of time. For CO and O_3, the time interval is eight hours, while for the other three pollutants, it is 24 hours. The pollutant level is given in microgram per cubic meter. For CO, it is the milligram.

2.2 Cumulative air quality indices

To indicate the pollutant concentrations, several Air Quality indices have been defined and used. A new fuzzy-based air quality index was suggested by Sowlat et al. [3]. A new Cumulative Index (CI) was proposed by Saxena and Shekhawat [4]. The concentration of four pollutants was utilized to find the CI. They are NO_2, SO_2, PM10, and PM2.5.

3. EXISTING CHALLENGES AND RELATED WORK

The potential impact of big data problems, issues in open research, and different tools related to it were discussed by Acharjya and Ahmed [5]. This paper provides important insights on big data and the principles behind its analysis, which is an important factor for this project because of the large size of the sample data being used to provide conclusions.

3.1 Health effects of air pollution

The effects that air pollution has on hospital admissions were presented by De Leon et al. [6]. This was done for respiratory disease using Poisson regression. Variables of pollution included black smoke (BS), NO_2, SO_2, and ozone (O_3). The hospital admission data for the years 1987-88 and 1991-92 for the City of London was used in the analysis.

Evaluation of the pollution due to fireworks during the festival season of Diwali was done by Ammasi Krishnan et al. [7]. Diwali is celebrated in India from October to November. There was an increase in particle pollutants during the time around Diwali in Chennai in 2017. This increase in pollutant levels results in the increase in health issues of the people during the festival time. Eye irritation was the predominant health issue noticed.

Monitoring the level of pollutants like sulfur dioxide (SO_2) in Kodungaiyur, Chennai, India was done by Krishnan et al. [8]. The study was carried out in 2017 over 6 months, during spring and summer. A survey on the diseases found in the locality Kodungaiyur was done to find out the effects of pollutants particularly PM2.5 on people. The diseases surveyed comprise allergic problems in the eye, lungs, etc. The results established links between pollution acquaintance and an increased number of infections.

3.2 Impact of COVID-19 pandemic on air quality

Few research papers discussed the effect of the COVID-19 pandemic on the values of Air Quality. Agarwal et al. [9] considered 6 cities each in India and China. As expected, Air Quality would have increased as a result of strict lockdown measures and a reduction in major human activities. They have analyzed how Air Quality has improved over 3 months. The results indicated that the drop in AQI of NO_2 was immediate. There was a steady drop in AQI of PM2.5. With an analysis of how these have come done, in the future, measures may be taken to bring pollutant concentrations down.

The study in reference [10] aimed to determine the impact of lockdown caused by the COVID-19 pandemic on air quality in the four major cities in India from January to May 2020. The cities were Delhi, Mumbai, Kolkata, and Chennai. This was done by analyzing the pollutant levels of particulate matter and gases like nitrogen dioxide (NO_2). The findings concluded a significant air quality improvement for the studied locations.

4. PERFORMANCE ASSESSMENT OF AIR QUALITY FORECASTING

When the prediction is used, the following parameters are used for the performance assessment of Air Quality Forecasting as to how accurate the predicted values are: Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), etc. [11].

MAE (Mean absolute error) is the average of the absolute errors. Absolute error is given by the difference between the actual and forecasted values.

$$MAE = \sum_{i=1}^{N} | Y_{forecast} - X_{actual} | / N \quad (2)$$

MSE (Mean Squared Error) is the measure of the average of the squares of the errors. The error is given by the difference between the estimated values and the actual value.

RMSE (Root Mean Squared Error) is the error rate given by the square root of MSE.

The equation is shown below:

$$RMSE = \sqrt{\sum_{i=1}^{N} (Y_{forecast} - X_{actual})^2 / N} \quad (3)$$

When classification is used, the following parameters are used for the Performance Assessment of Air Quality Forecasting: Accuracy, Precision, and Recall.

Accuracy: Accuracy is the ratio of the number of observations predicted correctly to the total number of observations.

$$Accuracy = \frac{(TP+TN)}{(TP+FP+FN+TN)} \tag{4}$$

Precision (P): Precision is the ratio of the number of positive observations predicted correctly to the total number of the predicted positive observations.

$$\text{Precision} = \frac{TP}{(TP+FP)} \tag{5}$$

Recall (R): Recall is the attribute of the model that sums up the model's ability to predict all the positive samples.

$$\text{Recall} = \frac{TP}{(TP+FN)} \tag{6}$$

5. DEVELOPMENT OF MACHINE LEARNING APPROACHES IN AQI FORECASTING

Some of the methodologies used to predict/classify the Air Quality Index are analyzed in this section.

5.1 Prediction approaches

Air pollution has a major and pervasive influence in Chengdu [12]. The use of petroleum products increases sulfur dioxide, nitrogen dioxide, and particulate matter. In this study, the spatial circulation fluctuation of the significant air pollutants over the city region of Chengdu in early 2010 was broken down and evaluated. The outcomes showed that the OK strategy was smarter to produce the forecast maps of toxin fixations than the IDW technique.

A semantic ETL (Extract-transform-Load) framework that works on a cloud platform for Air Quality prediction was proposed by Chang et al. [13]. Restful web service was used as the front-end API that retrieves analyzed data. The browser was used to show the visualized result.

Pollution prediction was attempted using the regression techniques [14]. Decision Tree, Multi-Layer Perceptron, Random Forest, and Gradient Boosting regression. Apache Spark was used to conduct the experiments. Different datasets have been used. Random Forest was found to be the best method as per the simulation results.

Another prediction method is based on the large amounts of environment-related data and deep learning methods [15]. A CNN as the base layer was used to automatically extract the features of input data. An LSTM network was used for the output layer. It helped in finding the time dependence of pollutants. Actual features from input data are obtained using the Convolutional Neural Network. Spatial correspondence between the data was estimated after convolution and pooling. The training data was taken from multiple places from one city (Shanghai). To generalize the work data from other cities can be used.

Ozone concentration prediction was done using deep learning [16]. A Deep Neural Network was used. An increase in accuracy was resulted compared to GLM, NN, SVM. They have used GLM, NN, SVM, DNN, for O3 prediction.

Sakarkar et al. [17] have compared different machine learning algorithms. After comparing 7 algorithms, the authors found Boosted Random Forest algorithm gives an accurate prediction. Nagpur was used as a case study. They chose Nagpur as the number of vehicles registered in Nagpur is more. This will increase air pollution.

Liu et al. [18] used support vector regression (SVR) and random forest regression (RFR) for forecasting the AQI values in Beijing and to find the NO_X level in an Italian city. SVR-based model acted better for the prediction of the AQI. RFR-based model functioned better for the prediction of the NO_X level.

Chang et al. [19] proposed a model named ALSTM (Aggregated LSTM model). This model is based on the deep learning LSTM method. Three of the LSTM models were aggregated into a combined model for prediction. The authors used the data provided by Taiwan Environmental Protection Agency. They have compared the new ALSTM model with other models in the prediction of PM2.5. The model proposed is improving the accuracy of prediction, they have found.

5.2 Classification approaches

Machine Learning methods can be used for classifying air quality into different categories ranging from Good to Severe. Decision tree (J48) and Naive Bayes algorithm were used in reference [20]. 91% Accuracy was observed for the decision tree. A short data amount was used. The data set used was for US cities. Decision trees cannot act as good classifiers for time series.

The challenges of forecasting the Air Quality Index (AQI) were addressed by Mahalingam et al. [21]. The paper aimed to minimize pollution. Neural Networks and Support Vector Machines were used for the prediction of AQ. The air pollution data set obtained from the CPCB, India was used. Data for Delhi city was considered. The results showed increased prediction accuracy of the proposed model over other models.

A Support Vector Machine (SVM) based classifier was proposed in reference [4]. This classifier is used to classify the air quality into good or harmful. The authors stated that the classifier performed well to classify the values of the air quality. The calculated values of Cumulative Indices were used as potential input to the classifier. Real data from three cities namely Kolkata, Delhi, and Bhopal were used to test the classifier.

A predictive air quality map for the next 24 hours in Tehran was done efficiently [22]. They have used Apache Hadoop, Naive Bayes, and Logistic regression. They find Logistic Regression to be the best estimator. Logistic Regression can perform well for predicting classes. Here they use it for classification to produce the Predictive Air Quality Risk Map. So, both Naive Bayes, Logistic regression can be used to classify AQI.

Air pollution trends in various cities in India were done by Sharma et al. [23]. Pollution data provided by the CPCB in India was considered. These data demonstrated the annual growth of pollutants like SO_2, NO_x, and PM2.5 over three years from 2015 to 2018. Data for three cities were used. The cities were Delhi, Bengaluru, and Chennai. Sources of pollution were analyzed. It was observed that NOx, SO_2, and

PM 2.5 resulted from outdoor sources like various transport modes, power generation plants, industries, construction activities, and indoor sources like domestic cooking. The areas considered were categorized into the following four classes: critically, highly, moderately, and low polluted. The results help to estimate the pollution for the cities.

6. FINDINGS

A summary of machine learning methods for Air Quality forecasting discussed in the previous section is given in the following table (Table 1).

Discussion of results obtained from some significant works in this area is summarized here.

Pollution prediction with the correlation of pollutants with other metrological variables for 5 cities of China was done in reference [14]. PM2.5 was the pollutant that has been used. RF was found to be the best method among the four methods that have been used.

PM 2.5 values have been predicted by Qin et al. [15] using CNN and LSTM. In reference [19], also it is done with Aggregated LSTM model. NO_2, SO_2, CO, and O_3 levels have been classified with the Decision tree and Naive Bayes algorithm [20]. Neural Networks, Support Vector Machines have been used in reference [21] for classifying air quality for 8 pollutants.

Table 1. Summary of machine learning methods for air quality forecasting

Sl. No.	Ref. No.	Topic discussed	Location / Datasets	Pollutants considered
1	De Leon et al. [6]	Effect of AQI on hospital admissions for respiratory diseases	The hospital admission data for 1987-88 and 1991-92 for London	Black smoke (BS), nitrogen dioxide (NO_2), Sulphur dioxide (SO_2), and ozone (O_3)
2	Agarwal et al. [9]	Effect of COVID-19 pandemic on AQI	6 cities each in India and China	PM2.5 and NO_2
3	Li et al. [12]	Generation of the prediction maps of pollutant	Chengdu	SO_2, NO_2, PM10
4	Ameer et al. [14]	Pollution prediction, Correlation of pollutants with other metrological variables	5 cities of China	PM2.5
5	Qin et al. [15]	Pollution prediction	Shanghai	PM2.5
6	Ghoneim et al. [16]	Ozone concentration prediction	Aarhus	Ozone
7	Sakarkar et al. [17]	Pollution prediction	Nagpur	SO_2, NO_2, RSPM, SPM, PM2.5
8	Liu et al. [18]	Forecasting AQI, Finding NO_X level	Beijing and an Italian city	NO_X
9	Chang et al. [19]	Prediction of pollutants	Data are given by Taiwan Environmental Protection Agency	PM2.5
10	Gore and Deshpande [20]	Classifying air quality	Data set US cities	NO_2, SO_2, CO, and O_3
11	Mahalingam et al. [21]	Classifying air quality	Delhi	8 pollutants including PM10 and Lead
12	Asgari et al. [22]	Predictive air quality map (classification)	Tehran	AQI, Wind direction, etc.
Sl. No.	**Ref. No.**	**Methods used**	**Inference**	**Remarks**
1	De Leon et al. [6]	Poisson regression	Time series analysis of hospital admissions for different age groups pollutant levels and the temperature was found	Effect on other diseases can be found.
2	Agarwal et al. [9]	A simple comparison of pollutant levels	The drop in AQI of NO_2 was immediate. There was a steady fall in AQI of PM2.5. The seaside areas recorded a more significant decline.	In the future, measures may be taken to bring pollutant levels down.
3	Li et al. [12]	Two spatial interpolation methods based on GIS, Inverse Distance Weighted Interpolation (IDW) and Ordinary Kriging	OK method was better	The work can be extended to a larger geographical area.
4	Ameer et al. [14]	DT, RF, MLP, Gradient Boosting	RF was found to be the best method	Data from other cities can be used.
5	Qin et al. [15]	CNN, LSTM	Superior prediction accuracy of the proposed model	Geomorphic conditions have to be considered
6	Ghoneim et al. [16]	Deep Neural Network	An increase in accuracy was resulted compared to GLM, NN, SVM.	Other pollutants can be considered

Sl. No.	Ref. No.	Topic discussed	Location / Datasets	Pollutants considered
7	Sakarkar et al. [17]	7 algorithms	Boosted Random Forest algorithm gives accurate prediction	Only one city's data used
8	Liu et al. [18]	Support vector regression (SVR) and random forest regression (RFR)	For AQI, SVR was better. For predicting NO_X, RFR- was better	Only two cities were considered for different aspects i.e., AQI, NO_X level
9	Chang et al. [19]	Aggregated LSTM model	The model proposed is improving the accuracy of prediction	Better preprocessing of data to deal with missing data has to be done.
10	Gore and Deshpande [20]	Decision tree and Naive Bayes algorithm	Decision tree The algorithm gives more accuracy than that of Naïve Bayes algorithm	Training data does not have values > 300. This has to be addressed.
11	Mahalingam et al. [21]	Neural Networks, Support Vector Machines	Increased classification accuracy for the proposed model	NN and SVM can be used for the classification
12	Asgari et al. [22]	Apache Hadoop, Naive Bayes, Logistic regression.	Logistic Regression is the best estimator	Naive Bayes, Logistic regression can be used for classification

7. CONCLUSION

To prevent air pollution, an AQI forecast is needed. A handful of machine learning techniques can be deployed to do this job with improved accuracy. We discussed the use of machine learning algorithms for pollution level prediction or classification. Some methods are more suitable for the prediction, classification of AQI or they give more accuracy.

By comparing the methods discussed in various research papers discussed here, the following inferences were made; Naive Bayes, Logistic regression, SVM, Decision trees can be used to classify AQI each with its advantages and disadvantages. As an example, decision trees can be used but they cannot act as good classifiers for time series data. Logistic Regression can perform well for predicting classes. DT, RF, MLP, GB regression, Deep Neural Network can be used to forecast the Air Quality values. Regarding the data set, the different authors have used data sets for a select city or area. The work can be extended to a larger geographical area. Similarly, the number and type of pollutants considered were different for different studies.

A study of the different techniques on a larger data set for all major pollutants using all possible ML methods can be carried out as future work and the results can be compared to obtain the best possible method that gives more accuracy.

REFERENCES

[1] World Health Organization Health Topic on Air Pollution. https://www.who.int/health-topics/air-pollution#tab=tab_1, accessed on 25 April 2021.

[2] Report on National Air Quality Index. (2015). Central Pollution Control Board, Ministry of Environment, Forests and Climate Change, Government of India. https://app.cpcbccr.com/ccr_docs/FINAL-REPORT_AQI_.pdf, accessed on 25 April 2021.

[3] Sowlat, M.H., Gharibi, H., Yunesian, M., Mahmoudi, M. T., Lotfi, S. (2011). A novel, fuzzy-based air quality index (FAQI) for air quality assessment. Atmospheric Environment, 45(12): 2050-2059. https://doi.org/10.1016/j.atmosenv.2011.01.060

[4] Saxena, A., Shekhawat, S. (2017). Ambient air quality classification by grey wolf optimizer based support vector machine. Journal of Environmental and Public Health, 2017: 3131083. https://doi.org/10.1155/2017/3131083

[5] Acharjya, D.P., Ahmed, K. (2016). A survey on big data analytics: Challenges, open research issues and tools. International Journal of Advanced Computer Science and Applications, 7(2): 511-518. https://dx.doi.org/10.14569/IJACSA.2016.070267

[6] De Leon, A.P., Anderson, H.R., Bland, J.M., Strachan, D.P., Bower, J. (1996). Effects of air pollution on daily hospital admissions for respiratory disease in London between 1987-88 and 1991-92. Journal of Epidemiology & Community Health, 50(S1): s63-s70. https://doi.org/10.1136/jech.50.suppl_1.s63

[7] Ammasi Krishnan, M., Devaraj, T., Velayutham, K., Perumal, V., Subramanian, S. (2020). Statistical evaluation of PM2.5 and dissemination of PM2.5, SO_2 and NO_2 during Diwali at Chennai, India. Natural Hazards, 103(3): 3847-3861. https://doi.org/10.1007/s11069-020-04149-8

[8] Krishnan, M.A., Jawahar, K., Perumal, V., Devaraj, T., Thanarasu, A., Kubendran, D., Sivanesan, S. (2019). Effects of ambient air pollution on respiratory and eye illness in population living in Kodungaiyur, Chennai. Atmospheric Environment, 203: 166-171. https://doi.org/10.1016/j.atmosenv.2019.02.013

[9] Agarwal, A., Kaushik, A., Kumar, S., Mishra, R.K. (2020). Comparative study on air quality status in Indian and Chinese cities before and during the COVID-19 lockdown period. Air Quality, Atmosphere & Health, 13(10): 1167-1178. https://doi.org/10.1007/s11869-020-00881-z

[10] Pant, G., Garlapati, D., Gaur, A., Hossain, K., Singh, S. V., Gupta, A.K. (2020). Air quality assessment among populous sites of major metropolitan cities in India during COVID-19 pandemic confinement. Environmental Science and Pollution Research, 27(35): 44629-44636. https://doi.org/10.1007/s11356-020-11061-y

[11] Senthil, K.P. (2019). Improved prediction of wind speed using machine learning. EAI Endorsed Transactions on Energy Web, 6(23). https://doi.org/10.4108/eai.13-7-2018.157033

[12] Li, S., Song, S., Fei, X. (2011). Spatial characteristics of air pollution in the main city area of Chengdu, China. In 2011 19th International Conference on Geoinformatics, Shanghai, China, pp. 1-4. https://doi.org/10.1109/GeoInformatics.2011.5981082

[13] Chang, Y.S., Lin, K.M., Tsai, Y.T., Zeng, Y.R., Hung, C.X. (2018). Big data platform for air quality analysis

and prediction. In 2018 27th Wireless and Optical Communication Conference (WOCC), Hualien, Taiwan, pp. 1-3. https://doi.org/10.1109/WOCC.2018.8372743
[14] Ameer, S., Shah, M.A., Khan, A., Song, H., Maple, C., Islam, S.U., Asghar, M.N. (2019). Comparative analysis of machine learning techniques for predicting air quality in smart cities. IEEE Access, 7: 128325-128338. https://doi.org/10.1109/ACCESS.2019.2925082
[15] Qin, D., Yu, J., Zou, G., Yong, R., Zhao, Q., Zhang, B. (2019). A novel combined prediction scheme based on CNN and LSTM for urban PM 2.5 concentration. IEEE Access, 7: 20050-20059. https://doi.org/10.1109/ACCESS.2019.2897028
[16] Ghoneim, O.A., Manjunatha, B.R. (2017). Forecasting of ozone concentration in the smart city using deep learning. In 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, India, pp. 1320-1326. https://doi.org/10.1109/ICACCI.2017.8126024
[17] Sakarkar, G., Pillai, S., Rao, C.V., Peshkar, A., Malewar, S. (2020). Comparative study of ambient air quality prediction system using machine learning to predict air quality in smart city. Proceedings of International Conference on IoT Inclusive Life (ICIIL 2019), NITTTR Chandigarh, India, pp. 175-182. https://doi.org/10.1007/978-981-15-3020-3_16
[18] Liu, H., Li, Q., Yu, D., Gu, Y. (2019). Air quality index and air pollutant concentration prediction based on machine learning algorithms. Applied Sciences, 9(19): 4069. https://doi.org/10.3390/app9194069
[19] Chang, Y.S., Chiao, H.T., Abimannan, S., Huang, Y.P., Tsai, Y.T., Lin, K.M. (2020). An LSTM-based aggregated model for air pollution forecasting. Atmos Pollut Res, 11(8): 1451-1463. https://doi.org/10.1016/j.apr.2020.05.015
[20] Gore, R.W., Deshpande, D.S. (2017). An approach for classification of health risks based on air quality levels. In 2017 1st International Conference on Intelligent Systems and Information Management (ICISIM), Aurangabad, India, pp. 58-61. https://doi.org/10.1109/ICISIM.2017.8122148
[21] Mahalingam, U., Elangovan, K., Dobhal, H., Valliappa, C., Shrestha, S., Kedam, G. (2019). A machine learning model for air quality prediction for smart cities. In 2019 International conference on wireless communications signal processing and networking (WiSPNET), Chennai, India, pp. 452-457. https://doi.org/10.1109/WiSPNET45539.2019.9032734
[22] Asgari, M., Farnaghi, M., Ghaemi, Z. (2017). Predictive mapping of urban air pollution using Apache Spark on a Hadoop cluster. In ICCBDC 2017: Proceedings of the 2017 International Conference on Cloud and Big Data, London United Kingdom, pp. 89-93. https://doi.org/10.1145/3141128.3141131
[23] Sharma, R., Kumar, R., Sharma, D.K., et al. (2019). Inferring air pollution from air quality index by different geographical areas: Case study in India. Air Quality, Atmosphere & Health, 12(11): 1347-1357. https://doi.org/10.1007/s11869-019-00749-x

NOMENCLATURE

ALSTM	Aggregated LSTM model
ANN	Artificial Neural Networks
AQI	Air Quality Index
BP	Back Propagation
BS	Black smoke
CNN	Convolutional Neural Network
CO	Carbon Monoxide
CPCB	Central Pollution Control Board, Ministry of Environment, Forests and Climate Change, Government of India
DT	Decision Tree
FN	False Negative
FP	False Positive
GB	Gradient Boosting
GLM	Generalized Linear Mode
LSTM	Long Short Term Memory
MAE	Mean Absolute Error
MSE	Mean Square Error
NH_3	Ammonia
NN	Neural Network
NO_2	Nitrogen Dioxide
NO_X	Nitrogen Oxides
O_3	Ozone
PM2.5	Particulate Matter 2.5
PM10	Particulate Matter 10
RMSE	Root Mean Square Error
MBE	Mean Bias Error
RNN	Recurrent Neural Network
SO_2	Sulfur Dioxide
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
USEPA	United States Environmental Protection Agency
WHO	World Health Organization