International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 09 | Sep 2019 www.irjet.net p-ISSN: 2395-0072

A COMPREHENSIVE REVIEW ON AIR POLLUTION DETECTION USING DATA MINGING TECHNIQUES

Pawandeep Kaur

KCET, Dept Computer Science Engineering, Punjab, India

---***---

Abstract: Air pollution is one of the biggest public health issues confronting the world today. Air pollution is increasing at rapid rate in the world. Many countries have declared it as major threat to human life. Currently air pollution is measured by utilizing spatially distributed sensors. However, due to sensor expenses and size limits the operational efficiency, many researchers have designed air pollution detection system using data mining tools without deploying any particular kind of sensors. It reduces the cost of air pollution monitoring system. The overall objective of this paper is to draw competitive analysis of data mining techniques that can be used to predict the benzene (C_6H_6) in the air. The reason behind this review is that benzene is considered most hazardous for humans as it causes blood cancer to our society. Therefore, prediction of benzene will ultimately lead to better citizen health. Finally, the comparisons have been drawn between the competitive data mining based air pollution detection techniques.

Index terms: Air pollution, Benzene, Blood cancer, Data mining.

1. Introduction

The atmosphere is an invisible cover of gases that envelops the earth. The air present in the atmosphere is a mixture of gases such as nitrogen, oxygen, argon, carbon dioxide etc. Air pollution refers to contamination of air by waste products and other impurities which are harmful to human life, other creatures, vegetation and buildings [1]. The primary cause of air pollution is both human activities and natural processes [2]. The sources responsible for air pollution are of two categories which are natural sources and man-made sources. The natural sources include forest fires, volcanic eruption, and wind erosion of soil, natural radio activity and decomposition of organic matter by bacteria [3].

The manmade sources are much diversified. These include automobile, industries, thermal power plants and agricultural activities [4]. The fossils fuels (coal, oil, natural gas) are burnt in industries, thermal power plants and automobiles. Due to this carbon monoxide (C0), carbon dioxide (CO_2), sulphur dioxide (SO_2), sulphur trioxide (SO_3) and nitrogen oxides are emitted. Different hydrocarbons (methane, butane, ethylene, benzene) and suspended particulate matters (dust, lead cadmium, chromium, arsenic salt etc.) are also present in these emissions [5]. These gases and suspended particulate matter (SPM) produced as result of burning fossils fuels are the greatest source of air pollution [6].

The pollutants released from natural sources of air pollution are dispersed in a vast area and do not cause any serious damage. Most of the health-related air pollutants come from man-made sources of air pollution [7]. In large cities, breathing the polluted air proves harmful to human health. Carbon monoxide, a serious air pollutant, reduces the oxygen carrying capacity of blood and causes nausea, headache, muscular weakness and slurring of speed. Oxides of nitrogen can damage the lungs, heart and kidneys of man and other creatures [8]. The presence of hydrocarbon in air causes irritation to eyes, bronchial construction, sneezing and coughing. In densely populated cities, the air pollution may take the form of industrial smog and photo chemical smog [9].

Benzene is considered to be a threat for various kinds of diseases. Therefore, an efficient monitoring of benzene becomes a challenging issue. Air pollutants such as benzene (C6H6) have accelerated the rate of cancer among human beings [10]. Currently, atmospheric contamination is measured using spatially separated networks with limited sensors. However, the expenses involving multiple sensors with varying sizes limit the operational efficiency [11]. Therefore, this study aims to design and develop a novel metaheuristic techniques-based data mining models to predict the concentration of benzene in the air, without deployment of actual sensors for benzene detection. It is possible because there is a relation among various atmospheric gasses and thus regression can be performed to measure C6H6 if the concentration level of other gasses is known [12].

2. Data mining techniques

Data mining (DM) is an important field of today's world of science that carries out a task without explicitly programmed. Data mining creates a function, model or algorithm; to learn and make predictions on the basis of existing datasets known as training datasets. Training dataset acts as an input to an already known output.

Training datasets are studied under supervised learning to predict future events when historical data is available The different data mining algorithms are as follow:

A. J48

It is a decision tree is often a predictive machine-learning model of which chooses the marked cost (dependent variable) of a whole new taste based on a variety of characteristic ideals of possible data. The interior nodes of a determination hardwood imply different features, the actual divisions between the nodes tell us the potential ideals that these features can have from the observed samples, as the terminal nodes tell us one more cost (classification) with the reliant variable.

B. Support Vector Machine

Support Vector Machines are supervised learning approaches useful for distinction, in addition to regression. The main benefit of SVM will be that they can utilize particular kernels as a way to change the trouble, so that most of us can apply straight line classification strategies to non-linear data. Applying the kernel equations sets up the info situations to the extent inside the multi-dimensional living space that you have a hyper-plane of which stands between details instances of a single type from that regarding another. These kernel equations can be any kind of perform of which turns the linearly non-separable details available as one website in a different website the location where the situations come to be linearly separable. Kernel equations can be straight line, quadratic, Gaussian, or some different of which attains that purpose.

C. Random Forest

Random forest classifier creates a set of decision trees from randomly selected subset of training set. It then aggregates the votes from different decision trees to decide the final class of the test object. Suppose training set is given as: [X1, X2, X3, X4] with corresponding labels as [L1, L2, L3, and L4], random forest may create three decision trees taking input of subset for example,

- [X1, X2, X3]
- [X1, X2, X4]
- [X2, X3, X4]

So finally, it predicts based on the majority of votes from each of the decision trees made.

This works well because a single decision tree may be prone to a noise, but aggregate of many decision trees reduce the effect of noise giving more accurate results.

3. Review of Literature

This section contains comprehensive review on existing well-known air quality prediction techniques by various researchers.

Siwek and Osowski (2016) [1] have discussed various methods of data mining for prediction of air pollution. The study shows that the pre-selection of the foremost vital attributes, permit increasing the forecasting accuracy of atmospheric pollution in an effective way. The methods of data mining for prediction of air pollution which are discussed are genetic algorithm and the step-wise fit techniques which are further used for feature selection.

Xiaoguang et al. (2015) [2] comprehensively evaluated and improved the daily air pollution prediction for 74 cities in china by various data mining methods. Five different classification algorithms namely, Random forest model, gradient booting model, SVM model, decision tree model, and combined model of the four above models of data mining techniques are adopted with exclusive feature groups which come from WRF-Chem model.

Yeganeh et al. (2017) [3] approximated the concentration of PM 2.5 by developing the satellite based model using ANFIS (Artificial Neuro Fuzzy Inference System). Authors have compared ANFIS with SVM (Support Vector Machine) and Back-Propagation artificial neural network adaptive model identification technique has been used to recognize the optimal predictive model. The different soft computing methods are applied to develop a satellite based model for estimating the spatiotemporal variation of PM2.5.

Sharma et al. (2015) [4] have employed adaptive neuro fuzzy inference system for forecasting air pollutants concentration. Pollutants such as Sulphur Dioxide (SO_2), and Ozone (O_3) in Delhi, India are being taken as environmental agents. A novel application of modified particles swarm optimization for training ANFIS for air pollutants forecasting is successfully investigated. Researchers evaluated an application of modified particles from optimization (MPSO) to train ANFIS(Adaptive Neuro Fuzzy Inference system) for the effective prediction of two major air pollutants. The results obtained are further compared with traditional gradient based method normally used for training ANFIS.

Chen et al. (2016) [5] have discussed various data mining algorithms for forecasting quality of air in urban areas. The discussed data mining models are used to predict the concentration of benzene in air. An effective data mining based model is developed for estimation of concentration of Benzene in the air.

Fu et al. (2015) [6] addresses the prediction of PM2.5 and PM10 by developing and improved Feed Forward Neural Network model. The risk of lung cancer and vision impairment has been one of the major concerns for air quality. The increase in concentration of PM2.5 and PM10 is considered as one of the high priorities issues by health organizations in China.

Yu et al. (2016) [7] described Random Forest Approach for prediction of quality of air in urban sensing system. Random forest approach for predicting air quality (RAQ) is proposed for urban sensing systems. The data generated by urban sensing includes meteorology data, road information, real-time traffic status and point of interest (POI) distribution. The random forest algorithm is exploited for data training and prediction.

De Vito et al. (2008) [8] have evaluated the use of neural networks together with on field data recordings for calibrating a multi-sensor device for benzene estimation. The scenario is characterized by significant correlations among several pollutant species. The proposed sensor fusion subsystem has been selected for exploiting both single sensor specificity and scenario-related correlations.

Vlachokostas et al. (2011) [9] have discussed that state of the art epidemiological research has found consistent associations between traffic-related air pollution and various outcomes, such as respiratory symptoms and premature mortality. However, many urban areas are characterised by the absence of the necessary monitoring infrastructure, especially for benzene (C6H6), which is a known human carcinogen. The use of environmental statistics combined with air quality modelling can be of vital importance in order to assess air quality levels of traffic-related pollutants in an urban area in the case where there are no available measurements.

Yildirim and Bayramoglu (2006) [10], have proposed a new methodology based on neural fuzzy method to estimate the concentrations of daily SO2 and TSP pollution over an urban area. Effective input variables in the model are ranked as temperature, pollutant (SO2 or TSP) concentration of the previous day, wind speed, relative humidity, pressure, solar radiation and precipitation. It is demonstrated that the temperature and previous day's pollutant (SO2 or TSP) concentrations are indispensable parameters for an acceptable performance of the model.

Kocadaglı (2015) [11] have discussed that Bayesian neural networks are useful tools to estimate the functional structure in the nonlinear systems. However, they suffer from some complicated problems such as controlling the model complexity, the training time, the efficient parameter estimation, the random walk, and the stuck in the local optima in the high-dimensional parameter cases.

Liu et al. (2016) [12] have selected Beijing, Tianjin and Shijiazhuang as three cities from the Jingjinji Region for the study to come up with a new model of collaborative forecasting using Support Vector Regression (SVR) for Urban Air Quality Index (AQI) prediction in China. The present study is aimed to improve the forecasting results by minimizing the prediction error of present data mining algorithms by taking into account multiple city multi-dimensional air quality information and weather conditions as input.

Fernández-Camacho et al. (2015) [13] focuses on correlations between total number concentrations, road traffic emissions and noise levels in an urban area in the southwest of Spain during the winter and summer of 2009. The high temporal correlation between sound pressure levels, traffic intensity, particle number concentrations related to traffic, black carbon and NOx concentrations suggests that noise is linked to traffic emissions as a main source of pollution in urban areas.

Gass et al. (2015) [14] illustrated the approach by investigating the joint effects of CO, NO2, O3, and PM2.5 on emergency department visits for paediatric asthma in Atlanta, Georgia. Pollutant concentrations were categorized as quartiles. Researchers concluded that regression trees can be used to hypothesize about joint effects of exposure mixtures and may be particularly useful in the field of air pollution epidemiology for gaining a better understanding of complex multi-pollutant exposures.

Pandey et al. (2015) [15]have employed a range of data mining techniques to predict UFP and PM1.0 levels based on a dataset consisting of observations of weather and traffic variables recorded at a busy roadside in Hangzhou, China. Based upon the thorough examination of over twenty five classifiers used for this task, researchers found that it is possible to predict PM1.0 and UFP levels reasonably accurately and tree-based classification models (Alternating Decision Tree and Random Forests) perform the best for both these particles.

Hu et al. (2014) [16] evaluated the association between exposure to criteria air pollutants and the risks of hypertensive disorders of pregnancy. Based on the findings from various studies, researcher's evaluation suggested that ambient air pollution exposure during pregnancy may be associated with increased risk of hypertensive disorders of pregnancy and preeclampsia.

Lana et al. (2016) [17] presented a data-based method to inspect the interplay among traffic, meteorological conditions and pollution in Madrid. Researchers have examined the coupling between traffic, meteorological features and different pollutants over districts of this city. Background pollution is found to be scarcely influenced by local traffic emissions. The aim of this study is to further explore this relationship by using high-resolution real traffic data.

Xu et al. (2014) [18] discussed PM2.5 has a significant influence on human health. Researchers proposed an ensemble learning method for PM 2.5 prediction. The assumption is that the information inside the historical data of PM2.5 in the selected station and other stations orderly from the one can be beneficial for the prediction of PM2.5. The results show that the more information, the more accurate the predictions are. Moreover, there are a balance between the good performance and the costs of modelling.

Singh et al. (2013) [19] developed tree ensemble models for seasonal discrimination and air quality prediction. PCA (Principal Component Analysis) used to identify air pollution sources; air quality indices used for health risk. Bagging and boosting algorithms enhanced predictive ability of ensemble models. Ensemble classification and regression models performed better than SVMs. Proposed models can be used as tools for air quality prediction and management.

Wang et al. (2017)[20] presented the early-warning system. This study consists of prediction and assessment modules, and its effectiveness is verified by performing a case study. Accordingly, researcher's early-warning system can produce more reasonable and comprehensive analyses of air pollution, thus providing trustworthy reference data for use by environmental supervisors in air pollution monitoring and management and providing the public with more information about the negative effect of air pollution.

Sun and Sun (2016) [21] presented a novel hybrid model based on principal component analysis (PCA) and least squares support vector machine (LSSVM) optimized by cuckoo search (CS). First PCA is adopted to extract original features and reduce dimension for input selection. Then LSSVM is applied to predict the daily PM2.5 concentration.

Shamsoddini et al. (2017) [22] examined the performance of the Random Forest feature selection in combination with multiple-linear regression and Multilayer Perceptron Artificial Neural Networks methods, in order to achieve an efficient model to estimate carbon monoxide and nitrogen dioxide, sulphur dioxide and PM2.5 contents in the air. The results indicated that Artificial Neural Networks fed by the attributes selected by Random Forest feature selection method performed more accurate than other models for the modeling of all pollutants. The estimation accuracy of sulphur dioxide emissions is lower than the other air contaminants whereas the nitrogen dioxide is predicted more accurate than the other pollutants.

Qi et al. (2017) [23] proposed a general and effective approach to solve the three problems in one model called the Deep Air Learning (DAL). The main idea of DAL lies in embedding feature selection and semi-supervised learning in different layers of the deep learning network.

Nieto et al. (2017) [24] designed a multilayer perceptron network (DMP) and the M5 model tree were also fitted to the experimental dataset for comparison purposes. Finally, the predicted results show that the hybrid proposed model is more robust than the DMP and M5 model tree prediction methods in terms of statistical estimators and testing performances.

Zhang et al. (2018)[25] found that the particulate matters such as PM10, PM2.5 may contain heavy metal oxides and harmful substances that threaten human health and environmental quality. An integrated neural network algorithm which based on Elman, echo state network (ESN) and cascaded BP neural network (CBP) has been designed to predict PM10 and PM2.5. To improve the performance of the prediction result, the simulated annealing algorithm (SA) has also been used to optimize the parameters in the combination method to form the optimal combination model.

Gu et al. (2018)[26] designed a new picture-based predictor of PM2.5 concentration (PPPC), which employs the pictures acquired using mobile phones or cameras to make a real-time estimation of PM2.5 concentration. First, using a large body of pictures which were captured under the good weather conditions, i.e., very low PM2.5 concentration, naturalness statistics (NS) models are built upon entropy features in spatial and transform domains. Second, for a novel picture, we measure its deviation degree from the above-mentioned NS models, considering the fact that the naturalness of a picture tends to reduce with the PM2.5 concentration increased. Third, a simple nonlinear function is introduced to map the deviation degree to the PM2.5 concentration.

Wang et al. (2018)[27] shown that the air pollution is still a big problem in China, and research on PM2.5 related issues is still a hot topic. Based on the data of Changsha City from 2016.01-2018.04, this paper uses Genetic algorithm optimization BP neural network to construct the evolutionary model of PM2.5 in Changsha. The validity, versatility and reliability of the model are verified by comparison with actual data.

Huang et al. (2018)[28] studied that the a high-performance machine-learning model was developed directly at monthly level to estimate PM2.5 levels in North China Plain. A random forest model has been implemented using the latest Multi-angle implementation of atmospheric correction (MAIAC) aerosol optical depth (AOD), meteorological parameters, land cover and ground PM2.5 measurements from 2013 to 2015. A multiple imputation method was applied to fill the missing values of AOD.

Table 1: Comparative analysis of the existing techniques

S.NO	Year	Technique	Target class	Ensemble	Optimization	K-cross validation	Feature selection
[1]	2017	PSO-ANFIS	C_6H_6	✗	✓	✓	✗
[2]	2016	Genetic algorithm	$PM_{2.5}$	✗	✓	✓	✗
[3]	2011	Multilayer perception	$PM_{10.0}$	✗	✗	✗	✓
[4]	2008	ANFIS	$PM_{2.5}$	✓	✗	✗	✓
[5]	2006	Ensemble	$PM_{2.5}$	✓	✗	✓	✓
[6]	2013	Regression	C_6H_6	✓	✗	✗	✗
[7]	2016	Random forest	C_6H_6	✓	✓	✓	✓
[8]	2017	Decision tree	$PM_{2.5}$	✗	✗	✓	✗
[9]	2018	Simulated annealing	$PM_{2.5}$ and $PM_{10.0}$	✓	✓	✓	✗
[10]	2018	Neural network	$PM_{2.5}$, $PM_{10.0}$, and C_6H_6	✓	✗	✗	✓
[11]	2018	Genetic algorithm	$PM_{10.0}$	✗	✓	✓	✓
[12]	2018	Random forest	$PM_{2.5}$ and $PM_{10.0}$	✓	✓	✓	✗

4. Conclusion and future work

Many countries have declared air pollution as a major threat to human life as it causes several diseases. Currently air pollution is measured by utilizing spatially distributed sensors, but due to sensor expenses and range limits the operational efficiency. Therefore, many researchers have proposed air pollution detection system using data mining tools without deploying any particular kind of sensors. Thus, it reduces the cost of air pollution monitoring system. The reason behind the study is to predict the concentration of benzene, which is considered as most hazardous for human being as it causes blood cancer to individuals.

Therefore, in a near future, a novel particle swarm optimization-based data mining models will be designed to predict the concentration of benzene in the air. Additionally, to modify particle swarm optimization-based data mining models genetic operators will also be used.

References

[1] K. Siwek and S. Osowski, "Data mining methods for prediction of air pollution", International Journal of Applied Mathematics and Computer Science, vol. 26, no. 2, 2016.

[2] X. Xi, Z. Wei, R. Xiaoguang, W. Yijie, B. Xinxin, Y. Wenjun and D. Jin, "A comprehensive evaluation of air pollution prediction improvement by a data mining method", 2015 IEEE International Conference on Service Operations And Logistics, And Informatics (SOLI), 2015.

[3] B. Yeganeh, M. Hewson, S. Clifford, L. Knibbs and L. Morawska, "A satellite-based model for estimating PM 2.5 concentration in a sparsely populated environment using soft computing techniques", Environmental Modelling & Software, vol. 88, pp. 84-92, 2017.

[4] S. Sharma, U. Kalra, S. Srivathsan, K. Rana and V. Kumar, "Efficient air pollutants prediction using ANFIS trained by Modified PSO algorithm", 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), 2015.

[5] M. Fu, W. Wang, Z. Le and M. Khorram, "Prediction of particular matter concentrations by developed feed-forward neural network with rolling mechanism and gray model", Neural Computing and Applications, vol. 26, no. 8, pp.

[6] R. Yu, Y. Yang, L. Yang, G. Han and O. Move, "RAQ–A Random Forest Approach for Predicting

Air Quality in Urban Sensing Systems", Sensors, vol. 16, no. 12, p. 86, 2016.

[7] A. Nebot and F. Mugica, "Small-Particle Pollution Modeling Using Fuzzy Approaches", Advances in Intelligent Systems and Computing, pp. 239-252, 2014.

[8] S. De Vito, E. Massera, M. Piga, L. Martinotto and G. Di Francia, "On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario", Sensors and Actuators B: Chemical, vol. 129, no. 2, pp. 750-757, 2008.

[9] C. Vlachokostas, C. Achillas, E. Chourdakis and N. Moussiopoulos, "Combining regression analysis and air quality modelling to predict benzene concentration levels", Atmospheric Environment, vol. 45, no. 15, pp. 2585-2592, 2011.

[10] Y. Yildirim and M. Bayramoglu, "Adaptive neuro-fuzzy based modelling for prediction of air pollution daily levels in city of Zonguldak", Chemosphere, vol. 63, no. 9, pp. 1575-1582, 2006.

[11] O. Kocadağlı, "A novel hybrid learning algorithm for full Bayesian approach of artificial neural networks", Applied Soft Computing, vol. 35, pp. 52-65, 2015.

[12] L. Chen et al. "Spatially fine-grained urban air quality estimation using ensemble semi-supervised learning and pruning", ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 1076–1087, 2016.

[13] B. Liu, A. Binaykia, P. Chang, M. Tiwari and C. Tsao, "Urban air quality forecasting based on multi-dimensional collaborative Support Vector Regression (SVR): A case study of Beijing-Tianjin-Shijiazhuang", PLOS ONE, vol. 12, no. 7, p. e0179763, 2017.

[14] R. Fernández-Camacho, I. Brito Cabeza, J. Aroba, F. Gómez-Bravo, S. Rodríguez and J. de la Rosa, "Assessment of ultrafine particles and noise measurements using fuzzy logic and data mining techniques", Science of The Total Environment, vol. 512-513, pp. 103-113, 2015.

[15] K. Gass, M. Klein, H. Chang, W. Flanders and M. Strickland, "Classification and regression trees for epidemiologic research: an air pollution example", Environmental Health, vol. 13, no. 1, 2014.

[16] G. Pandey, B. Zhang and L. Jian, "Predicting submicron air pollution indicators: a data mining approach", Environmental Science: Processes & Impacts, vol. 15, no. 5, p. 996, 2013.

[17] H. Hu, S. Ha, J. Roth, G. Kearney, E. Talbott and X. Xu, "Ambient air pollution and hypertensive disorders of pregnancy: A systematic review and meta-analysis", Atmospheric Environment, vol. 97, pp. 336-345, 2014.

[18] I. Laña, J. Del Ser, A. Padró, M. Vélez and C. Casanova-Mateo, "The role of local urban traffic and meteorological conditions in air pollution: A data-based case study in Madrid, Spain", Atmospheric Environment, vol. 145, pp. 424-438, 2016.

[19] W. Xu, C. Cheng, D. Guo, X. Chen, H. Yuan, R. Yang and Y. Liu, "PM2.5 Air Quality Index Prediction Using an Ensemble Learning Model", Web-Age Information Management, pp. 119-129, 2014.

[20] K. Singh, S. Gupta and P. Rai, "Identifying pollution sources and predicting urban air quality using ensemble learning methods", Atmospheric Environment, vol. 80, pp. 426-437, 2013.

[21] J. Wang, X. Zhang, Z. Guo, and H. Lu, "Developing an early-warning system for air quality prediction and assessment of cities in China," Expert Syst. Appl., vol. 84, pp. 102–116, 2017.

[22] W. Sun and J. Sun, "Daily PM2.5 concentration prediction based on principal component analysis and LSSVM optimized by cuckoo search algorithm," J. Environ. Manage., vol. 188, pp. 144–152, 2017.

[23] A. Shamsoddini, M. R. Aboodi, and J. Karami, "Tehran air pollutants prediction based on Random Forest feature selection method," Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch., vol. 42, no. 4W4, pp. 7–10, 2017.

[24] Z. Qi, T. Wang, G. Song, W. Hu, X. Li, Zhongfei, and Zhang, "Deep Air Learning: Interpolation, Prediction, and Feature Analysis of Fine-grained Air Quality," no. 38, 2017.

[25] P. J. García Nieto, E. García-Gonzalo, A. Bernardo Sánchez, and A. A. Rodríguez Miranda, "Air

Quality Modeling Using the PSO-SVM-Based Approach, DMP Neural Network, and M5 Model Tree in the Metropolitan Area of Oviedo (Northern Spain),” Environ. Model. Assess., pp. 1–19, 2017.

[26] Zhang, Hairui, Zhihao Shang, Yanru Song, Zhaoshuang He, and Lian Li. "A novel combined model based on Echo State Network-A case study of PM10 and PM2. 5 prediction in China." Environmental technology just-accepted (2018): 1-23.

[27] Gu, Ke, Junfei Qiao, and Xiaoli Li. "Highly Efficient Picture-Based Prediction of PM2. 5 Concentration." IEEE Transactions on Industrial Electronics (2018).

[28] Wang, Zhencheng, and Zou Long. "Pm2. 5 Prediction Based on Neural Network." In 2018 11th International Conference on Intelligent Computation Technology and Automation (ICICTA), pp. 44-47. IEEE, 2018.

[29] Huang, Keyong, Qingyang Xiao, Xia Meng, Guannan Geng, Yujie Wang, Alexei Lyapustin, Dongfeng Gu, and Yang Liu. "Predicting monthly high-resolution PM2. 5 concentrations with random forest model in the North China Plain." Environmental pollution 242 (2018): 675-683.