

# Air Quality Prediction using Artificial Intelligence



P. Shree Nandhini, P. Amudha, S. Sivakumari

**Abstract**—Due to the critical impacts of air pollution, prediction and monitoring of air quality in urban areas are essential tasks. However, because of the dynamic nature and high spatio-temporal variability, the prediction of the air pollutant concentrations is a complex spatio-temporal problem. The data is collected in specific area such as climate condition and vehicular pollutant occurring in the peak hours. the predication process is used to compare the algorithm artificial neural network and support vector machine process. This paper presents a survey on Air quality prediction using artificial intelligence

**Keywords:** Spatio-temporal, Ambient air pollution, Support vector machine,

## I. INTRODUCTION

Due to human activities air pollution is more in the surrounding environment and cause many health issues the weather forecasting changed day by day process due to air pollutant s and damage the atmosphere process. The predication of pollutant brings together a monitor system of air pollutant and It makes use of global meteorological and crop data such as type of agriculture, country, fertilization, etc. Before data modeling, during data preprocessing, several measures were taken into consideration to improve the overall performance of process including the removal of extreme values in the dataset, exclusion of boreal ecosystem ranking of essential features in the dataset; and supply of controlled number input variables. For a fair assessment, the linear and nonlinear regression models are the results.

The result shows that the accuracy performance of the RF-based model was 20-24% higher than that of both regression models. Whereas another RF-based model It is used to predict the AQI by using urban public data based on road information, air quality and meteorological datasets of all the regions of Shenyang, China. Over proficient assessment of the model, it was verified that RF with its highest precision ( $R^2 = 0.81$ ) and lowest error value ( $RAE = 36.9\%$ ) had outperformed the state of the art classifiers such as Naïve Bayes, Logistic Regression, single decision tree and

ANN.A simplified regression technique based on Quito Ecuador.

## II. LITERATURE SURVEY

Y. Jiang, K .et.al [1] Proposed the crowd sensing-based method [1] is a predicate air pollutant such as vehicular area in the pollutant The period of sensing air pollutant is hard .and the predication of air quality take more time of process.

Grange et al. (2018) Proposed RF as a classification technique to build a model by using meteorological conditions, atmospheric pollutant data, and temporal factors to analyze the trend of the particular matter of air pollution as well to make a long term prediction. The methodology of the study involves the creation and training of many out of the bag (OOB) samples to grow Decision Tree, which later was combined to make a prediction. The study makes use of daily input variables named: wind speed and direction, threshold and time, and synoptic-scale, were obtained from 31 monitoring stations of Switzerland for 20 years. The average accuracy calculated in terms of correlation coefficient ( $R^2$ ) at all 31 stations is to be 0.62, with wind speed and boundary layer being the best input predictors and synoptic-scale the worst. The prediction performance obtained under the model at different locations varied from  $R^2 = 0.53$  to  $R^2 = 0.71$ . The lower values were important recorded at stations located near rural mountainous areas.

Diez et al. [3] proposed the method of statistical approach for predicate the air pollutant in the international airport the fast and accurate method is analyzed for observed data analytical methods. for reducing air quality issue in aircraft operation in airport and hard to find the actual operation.

K. D. Kuhn[4] Proposed the machine learning method for predicate international airport. The Air quality index is a substitute for particular pollutant. The international civil aviation configuration and actual performance for meteorological data The developing and predicate the unsupervised learning method is the best model and accuracy.

Li et al. [5] Proposed the weather forecasting datasets. The first is the influence of meteorological conditions such as atmospheric currents, and the second is the change in the average astigmatism coefficient caused by the absorption and scattering of light due to fine particles and significant air pollutants. Xiaoyu.et.al[6][7]Proposed the Comprehensive Pollution Index Method determines the air quality level on based the calculated pollution index value and finally calculated the average level o air pollutant and the damage level of a single contaminant in the calculation, which is simple and easy to conduct. The air pollution is increasing due to weather, and wind that should be cut off value and finally evaluate the result.

Revised Manuscript Received on March 30, 2020.

\* Correspondence Author

**P. Shree Nandhini\***, Research Scholar, Department of Computer Science and Engineering, Avinashilingam Institute for Home Science and Higher Education for Women, School of Engineering, Coimbatore, India.

**Dr. P. Amudha**, Associate Professor, Department of Computer Science and Engineering, Avinashilingam Institute for Home Science and Higher Education for Women, School of Engineering, Coimbatore, India.

**Dr. S. Sivakumari**, Professor and Head, Department of Computer Science and Engineering, Avinashilingam Institute for Home Science and Higher Education for Women, School of Engineering, Coimbatore, India

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Meanwhile, the calculation result depends on the ratio of the observed highest pollutant concentration value to the corresponding standard cost, which would result in a higher Comprehensive Pollution Index if the perceived amount of a particular pollutant is relatively higher.

The whistleblowers proposed the method air pollution is a direct victim in the environment protection. For long term process air pollutant is affected by whistleblowers due to issue in air pollution exposure and finally report the process of whistleblowers.

Website of Urban Air [11], is an icon on the map stands for a checking and monitoring all area in the station. The number consider with image air quality index and the number lesser than and better than air quality The color of an image is denoted by air quality, e.g., "green" means a "good" and "yellow" means "moderate" by Chinese AQI standards. The time line is used to analyze air quality prediction particular time interval.

K. V. Hamilton Ron .et.al due to risk in the environment problem cause many health diseases.in the urban air quality process. Such as respiratory diseases and lung cancer and asthma and many issue in the building tools [12].

R.C.Abernethy .et.al. Proposed the method of environment pollution prediction. Day by day pollution level are changed because pollutant is caused by vehicular areas in the meteorological data [13–14]. The environment data is many datasets are used for statistical tool. The multivariate analysis methods are used many methods for high amount of data.

L. V. Perez-Arribas .et.al Proposed the Concept of clustering data process in the weather forecasting data such as wind direction in the hours and monthly data. The environment data is dependent upon another process data. Different parameter is used to multidimensional analysis in the same time end with many Pollutant. [15].

### III. COMPARATIVE ANALYSIS OF METHODOLOGY

Ong et al. [16] proposed the method of time series air quality prediction tasks for automatic encoder pre-training method. The VENUS system accuracy goes beyond the training process. The sites of pollutant avoid the previous machine learning process. There is no input variable detecting and cause impact of air quality prediction

Cho K, Merrienboer [17] proposed the method of recurrent neural network (RNN) deals with spatial –temporal data are used to analyses the air quality process. The recurrent neural network model of GRU in the encoder decoder process to overcome the problem of overfitting data in the GRU and used less parameter it provides more predicated result and accuracy is good [17].

Kim PS [18] Tang Y, HuangY[19] Wang Y[20] proposed the method of prediction air quality process. Hours data predication is calculated in the first stage process the prediction data in the meteorological condition Due to forecasting sequence learning method are known in the deep learning technique. The real time point value. Denoted in the point process.

Several health problem is caused by air pollutant such as asthma dying people in the world.it performs European has performs thirty less than Percentage pollution. Due to respiratory infection cause 13 %percentage death because of increasing vehicular pollutant.

The spatial-temporal feature proposed the method of air quality prediction is the variable of internal tasks. It focused on method of time-series data are used to design for by air pollutant interactions and transmissions as well.

Due to climate process and forecast the process is a framework for improve accuracy prediction.

Due to forecast air quality data increases is used to improve to predict data in air quality pollution Air planning is a method is used to develop the forecasting. [8][9][10].

The real-time emission data, is an input, for decision-makers to analyze the performance of air planning decisions.

The tree model is a decision tree is used to represent branch node and represent decision tree.fig1.the main node describe the process and end node are leaves.

In S. Deleawe et al. proposed the method decision tree model for classification process and select the predict the values in the air, and air quality class are used to the attribute process. The confidence factor is 0.25 are used in decision tree process. Implementation of process is Weka algorithm.

Besides a fair performance during regular days, it was interesting to see that improved results during the weather conditions. The concept of lazy learning technique captures the difference PM10 emissions and AQI Hourly dataset of SO<sub>2</sub>, NO<sub>x</sub>, CO, PM10, and Ammonia (NH<sub>3</sub>) for one year period from an area called Lombard in Italy

(Carnevale et al., 2016). Other specifications of the study include the application of the Dijkstra algorithm for large scale data processing. The ration between data 80% and 20% is used for training and validation purposes, respectively. The validation phase results and compared with deterministic models instead of comparing it with the targeted values. The results obtainable through this approach in the form of R<sup>2</sup>were nearly the same as achieved by the Transport Chemical Aerosol Model (TCAM), i.e., R<sup>2</sup>=0.99. By the way, TCAM is a costly computational method commonly used in decision making.

Bougoudis et al. (2016) built a hybrid system by using ANN, RF, fuzzy logic along with unsupervised clustering for the prediction of various pollutants. The experiment conducted makes use of 12 years' hourly dataset of air pollutants the performance of R<sup>2</sup> model is evaluated using Unsupervised clustering Mamdani proposed the method of fuzzy logic system was applied to enquire about factors affecting the quality of air. The model performed well to estimate the concentration of CO and NO with R<sup>2</sup> =0.95 when FIS ensemble with RF and ANN respectively while using RF alone it could predict NO and O<sub>3</sub> with accuracy equal to 0.91 MartínezEspaña et al. (2018) conducted work in Murcia, Spain. The study aimed at the prediction accuracy of ground-level O<sub>3</sub> by using five different classification models, namely Bagging, RF, Decision Tree, k-Nearest Neighbor, and Random Committee. The work applied two years' air pollution dataset of NO, NO<sub>x</sub>, PM10, O<sub>3</sub>, SO<sub>2</sub>, C<sub>6</sub>H<sub>6</sub>, C<sub>7</sub>H<sub>8</sub>, XIL, and environmental parameters such as pressure, solar radiations, temperature, relative humidity, wind speed, and direction. Two phases in the experimental design. At first, the prediction accuracy of all the models was tested out and compared against each other, whereas to study the number of models required for O<sub>3</sub> modeling in the Murcia region, during the second phase,

The work adopted a hierarchical clustering approach. Results obtained in terms of the correlation coefficient are as follow: the performance of RF has been superior, having an R2 value equal to 0.85 as compared to Random Committee (0.83), Bagging (0.82), and Decision Tree (0.82). Among the five models, CNN performed the worst (0.78), whereas NOx, threshold, climate condition is the top predictors. In the end, using a clustering approach suggested that the study region only requires two models for accurate modeling of O3.

Sayegh et al., (2014) aimed at capturing the variability of PM10 by employing several statistical and machine learning models such as Boosted Regression Tree (BRT), Generalized Additive Model, Linear, and Quantile Regression models (QRMs). The (observed/ predicted values) is the performance of the coefficient of determination. Considering the role of different quantiles instead of the central tendency of PM10, the QRM had performed better as compared to other data mining tools to predict the hourly concentrations PM10.

Nieto et al. (2015) conducted research using different air pollutant dataset. is collected. The experimental setup employs MLP and Multivariate Adaptive Regression Splines as modeling tools. The proficient assessment of modeling results inferred that the estimation of NO2 (R2 = 0.85) has been pretty good, followed by a slightly lower accuracy obtained for predicate parameter. Meanwhile, to predict the concentration of a greenhouse gas N2O, another study adopts RF algorithm

Kleine Deters et al. (2017) Proposed method predictions used to understand the effects of meteorological factors on the precise prediction of PM2.5. During collection of metrological data training and tested datasets it performance of the technique cross-validation and coefficient of determination R2. The meteorological parameter is a best choice of predicating air quality.

Eldakhly et al. (2018) to make 1 hour ahead forecast of PM10 concentrations. The study uses fuzzy logic and chance weight value to handle fuzziness of the data; the target minimizes an outlier point in the training process. The different results obtained are one of the rare examples, as the study establishes that the proposed approach can outperform ensemble learning algorithms.

Xu et al. (2017). It proposed to estimating ozone profile shapes. In this work, the author developed a KNN based algorithm, namely Physics Inverse Learning Machine. The working principle of the model involves five significant steps: 1) application of k-means clustering to group different ozone profiles based on its concentration values; 2) generation of simulated UV spectra from each cluster of respective ozone profiles; 3) input predictive feature selection by using PCA to enhance classification effectiveness; 4) spectrum of uv rays are classification models (5) the shape of ozone process are fit for ozone column. The predicted and observed values in the tested. Results obtained indicate that, altogether, a total of 11 clusters and estimation error lower than 10%.different technique based on less reliable dense mobile sensors data collected from sparse monitoring stations a fine granularity (Hu et al., 2017). For this Sydney, Australia based study, seven regression models are method to predict the concentration of CO. Before modeling and validation, and the study involves three main steps. During first regression models is used ten years of data, including previous years off data is collected. Comparison and validation in the

Proficient model during the second and third steps, respectively. Out of seven modeling techniques, Support Vector Machine for regression (SVR), RF, and Decision Tree regression achieved the best results, with SVR having the highest spatial resolution and precise demarcation of pollution boundaries as compared to other models.

#### IV. RESULT

In this Section The efficiency of Existing algorithm are shown in this Table1

**Table 1.Database Characteristics**

Dataset Characteristics	Multivariate, Time Series
Attribute Characteristics	Categorical, Integer
Area	Computer
No. of Instances	9358
No. of Attributes	15

#### V. CONCLUSION

Here the paper review the air quality prediction using artificial intelligence method, and compares present and existing research work on air quality evaluation based on machine learning algorithm and techniques.

#### REFERENCES

1. Y. Jiang, K. Li, L. Tian, R. Piedrahita, X. Yun, O. Mansata, Q. Lv, R. P. Dick, M. Hannigan, and L. Shang. Maqs: A personalized mobile sensing system for indoor air quality. In Proc. of UbiComp 2011.
2. Grace .Et.Al Machine Learning Algorithm Using Air Quality Modeling 2018
3. D.M.Diez,F.Dominici,D.Zarubiak,andJ.I.Levy,“Staticallyapproaches for identifying air pollutant mixtures associated with aircraft departures at Los Angeles international airport,” Environmental Science& Technology, vol.46, no.15, pp.8229– 8235,2012
4. K. D. Kuhn, “A methodology for identifying similar days in air trafficflowmanagementinitiativeplanning,”TransportationResearchPartC:EmergingTechnologies,vol.69,pp.1–15,2016.
5. Li, H.; Tan, X.; Guo, J.; Zhu, K.; Huang, C. Study on an Implementation Scheme of Synergistic Emission Reduction of CO2 and Air Pollutants in China’s Steel Industry. Sustainability 2019, 11, 352.
6. Xiaoyu, L.; Jun, Y.; Pengfei, S. Structure and Application of a New Comprehensive Environmental Pollution Index. Energy Procedia 2011, 5, 1049–1054.
7. Ga, siorek, M.; Kowalska, J.; Mazurek, R.; Paja, k, M. Comprehensive assessment of heavy metal pollution into soil of historical urban parkonan example of the Planty ParkinKrakow(Poland). Chemosphere2017,179, 148–158
8. [Ministry of Environmental Protection of the People’s Republic of China. The Ministry of Environmental Protection Will Open the “010-12369” Environmental Whistleblowing Hotline on the World EnvironmentDay(5June).Availableonline:http://www.mee.gov.cn/gkml/sthjbgw/qt/200910/t20091023\_179602.htm (accessed on 17 November 2018).
9. [Ministry of Environmental Protection of the People’s Republic of China. The “12369 Environment Whistleblowing Platform” on WeChat Has Been Working Well Since Its Opening in 2015. Availableonline:http://www.mee.gov.cn/gkml/sthjbgw/qt/201603/t20160322\_334063.htm (accessed on 17 November 2018).



10. Environmental Emergency and Accident Investigation Center of the Ministry of Environmental Protection. 12369: Promoting Environmental Information Disclosure and Public Participation. World Environ. 2017, 4, 76–77.
11. Urban Air Website: <http://urbanair.msra.cn/>
12. K. V. Hamilton Ron, T. Johan, and J. Watt, The Effects of Air Pollution on Cultural Heritage, Springer-Verlag, Boston, MA, USA, 2009.
13. R.C.Abernethy, R.W.Allen, I.G.McKendry, and M.Brauer, "A land use regression model for ultrafine particles in Vancouver, Canada," Environmental Science and Technology, vol. 47, no. 10, pp. 5217–5225, 2013.
14. M.Rivera, X.Basagaña, I.Aguilera et al., "Spatial distribution of ultrafine particles in urban settings: a land use regression model," Atmospheric Environment, vol. 54, pp. 657–666, 2012.
15. L. V. Pérez-Arribas, M. E. León-González, and N. Rosales Conrado, "Learning principal component analysis by using data from air quality networks," Journal of Chemical Education, vol. 94, no. 3, pp. 458–464, 2017.
16. Ong B T, Sugiura K, and Zettsu K, "Dynamic pretraining of Deep Recurrent Neural Networks for predicting environmental monitoring data," IEEE International Conference on Big Data. IEEE, 2015, 760765.
17. Cho K, Merriënboer B V, Gulcehre C, Bahdanau D, Bougares F, and Schwenk H, "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," Computer Science, 2014.
18. Kim P S, Lee D G, and Lee S W, "Discriminative Context Learning with Gated Recurrent Unit for Group Activity Recognition," Pattern Recognition, 2017, 76.
19. Tang Y, Huang Y, Wu Z, Meng H, Xu M, and Cai L, "Question detection from acoustic features using recurrent neural network with gated recurrent unit," IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2016:6125-6129
20. Wang Y, Liu M, Bao Z, and Zhang S, "Short-Term Load Forecasting with Multi-Source Data Using Gated Recurrent Unit Neural Networks", Energies, 2018.

## AUTHORS PROFILE



**P. Shree Nandhini**, Research scholar, Department of Computer Science and Engineering, Avinashilingam Institute for Home Science and Higher Education for Women, School of Engineering, Coimbatore, India.



**Dr. P. Amudha**, Associate Professor, Department of Computer Science and Engineering, Avinashilingam Institute for Home Science and Higher Education for Women, School of Engineering, Coimbatore, India.



**Dr. S. Sivakumari**, Professor and Head, Department of Computer Science and Engineering, Avinashilingam Institute for Home Science and Higher Education for Women, School of Engineering, Coimbatore, India