

A Survey on Air Quality Forecasting Techniques using Machine Learning

Reshma J¹, Ashwini R²

Assistant Professor, Department of Computer Science & Engineering, B N M Institute of Technology,
Bangalore, India ¹

P G Student, Department of Computer Science & Engineering, B N M Institute of Technology,
Bangalore, India ²

ABSTRACT: Air pollution is becoming one of the major threats across the globe. Air quality index is an index used for reporting daily air quality. In India, *eight pollutants are considered to harm the environment. Among them, the fine particulate matter PM2.5 and PM10 are one of the biggest concern to people's health.* The alarming rise in pollution levels calls for predicting air quality. This paper summarizes the various techniques used for predicting air quality.

KEYWORDS: Air Quality Index, PM2.5, Air Pollution.

I. INTRODUCTION

Air is an odorless, colorless, gaseous mixture with varying amount of harmful gases of carbon, sulphur and nitrogen. Every living organism on earth is dependent on air. Air gets contaminated by gases in environment, which pollute atmosphere[1]. Predicting future air quality for a monitoring station is very challenging because air quality has multiple influential factors. As illustrated in Figure 1, air pollutant sources mainly come from vehicle exhaust, industrial emission, coal burning and dust. Each of these sources has a different spatial distribution and temporal pattern. According to surveys, people spend more than 80% of their time in artificial air in offices, factories, public buildings, houses and transport vehicles which is increasingly becoming worse day by day[2]. The main cause of air pollution is by humans. In addition, pollution can be caused naturally due to sudden changes in temperature and seasonal changes. Vehicle emission contains gases like carbon monoxide, nitrogen oxides, nitrogen monoxide (NO) and nitrogen dioxide (NO₂) [3].

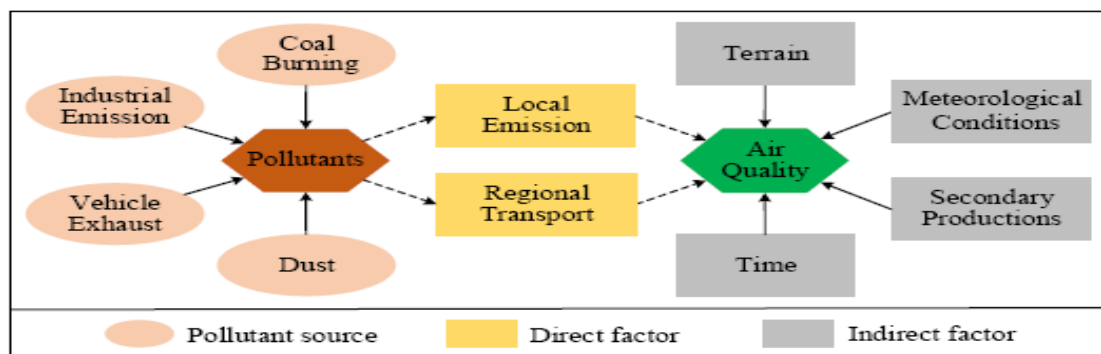


Fig.1 Multiple influential factors

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

As the density of population continues to grow in the urban cities, quality of air has become a serious issue. The fine particulate matter (PM_{2.5}) has raised a series of concerns for public health and has resulted in pollution problem [4]. In residential areas, almost 15% of the pollution is due to CO₂ emission [5]. The growing population of the world and migration of people has become imperative to create an intelligent and sustainable environment which provides citizens with a higher quality of life. The direct effect of this urban drift has resulted in effects on social, economic and ecological systems, causing stresses on the environment and society. The social and economic implications include impacts from human activities such as transport and industrialization and these have an indirect bearing on the environment.

In smart cities, attention is given for large scale development of applications which can be used for monitoring finer-grained air pollution. Air pollution raises a number of concerns ranging from public health to the social economy [6]. The range of pollutants in the atmosphere which have the capacity to cause harm to both humans and the ecosystem includes Carbon monoxide, Nitrogen oxides, Sulphur dioxide, Particular matter, Volatile organic compounds, Ozone and Hydrocarbons.

Lately, public awareness regarding air pollution is created with the aid of air quality index. Agencies are currently tracking the behaviour of pollutants in the air and informing people about environment pollution. This information will aid people who are suffering from illness which are caused through air pollution [7]. To solve some of the difficult problem in the environment, protection of environment and ecological civilization is necessary. Both humans and animals are affected by the ill effects of air pollutions and leads to respiratory, cardiological and pulmonary problems by inhaling certain gases. These changes also is resulting in global warming and climate change. An important issue for government is to improve the quality of life of people by taking into consideration the factors that affect air quality [8]. Given below are some of the air pollutants and its effect.

- Sulphur dioxide and Sulphur oxides- High concentrations of sulphur dioxide can cause inflammation and irritation of the respiratory system.
- Particulate Matter- Particulate Matter can cause lung irritation which leads to increased permeability in lung tissue.
- Lead- Exposure of human to high levels of lead can cause anaemia, weakness, kidney and brain damage.
- Chlorine- Chlorine inhalation may cause irritation to skin and eyes along with a sore throat.
- Fluoride- Fluoride leads to bone disease and also damages sensitive tissues like thyroid glands in the body.
- Nitrogen dioxide- Nitrogen dioxide is an irritant gas and high concentration causes inflammation. Long term exposure can decrease lung function.
- Ozone- Ozone can cause bronchitis, emphysema and asthma leading to increased medical care.
- Carbon Monoxide- Breathing Carbon Monoxide can cause headache, dizziness and vomiting.
- Ammonia - High concentrations of ammonia in air causes immediate burning of the nose, throat and respiratory tract.
- Carbon dioxide- Causes headaches, dizziness and breathing problems.

Air pollutant is used to predict the condition of the atmosphere at different locations. Air quality models perform an important role in all areas of air pollution control [9]. Air quality forecasts provide the public with air quality information which allows people to take precautionary measures to avoid their exposure to unhealthy levels of air pollution[10].

The existence of life on Earth is in danger due to rapid increase of various effects such as global warming and many health disorders are found in human body. There are several substances that leads to respiratory disease by exposure to contaminated air from particulate matter. Children are more likely to be affected by the air environment. Young children inhale more air per unit time with smaller surface area of their lung, indicating more air reaches the lung. There are short-term health effects to children for exposures to ambient air pollutants [11]. Food, fruits and vegetables which are affected by various air pollutants are not safe for consumption. Air pollution is dangerous and damages

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

multiple organs of the human body. Some of gases which are more poisonous are listed as air pollutants and affects human health [12].

Different kinds of monitoring for sites which report data of temperature, wind speed, wind direction and observations are measured for each stations. The major pollutants are NO₂, O₃ and PM₁₀ gathered from different stations. Air quality system also contains meteorological data and descriptive information about each monitoring station which include geographic location and data quality control information [13]. There are two types of standards: primary and secondary.

- Primary standards: These protect against adverse health effects.
- Secondary standards: These protect against welfare effects such as damage to farm crops, vegetation and damage to buildings.

In order to predict air quality status in city, the national air quality index technical regulations stipulates that the air quality index will be used to replace the original air pollution index and it is divided into six grades and the type of pollution is assessed based on the grades [14]. Using air quality index value for the present as well as previous day and weather forecast for the day, the pollution for that day can be predicted.

Air pollution has been a matter of environmental and health concerns, particularly in urban areas. The breakpoints between index values is given separately and the overall index is taken as the maximum value. Different averaging periods are taken for different pollutants. All bands will be advised for at-risk groups and the general population. Table 1 shows the Health Descriptor and reference levels used for health-protection related limit considering the target at national and local level [15].

II. RELATED WORK

This paper discusses about the air quality forecast monitoring and its impact on brain health. Brain health quality pre-monitoring has become an urgent need and this is a system of complex engineering. Here, the perspective of intelligent decision-making is based on big data. With the help of intelligent air index prediction and the popular classification algorithms, the hidden information of historical data is mined and the brain health quality prediction is realized. The brain health quality monitoring system is based on Internet of Things and the classification algorithm is used to realize real-time acquisition. This aids in improving the data processing speed and enhancing the real-time performance of brain health quality predicate. Table 1 gives the details of the impact the air pollutants can have on health depending on the numeric value. If the value is less than 50, then the health effects are unlikely to be detected by individuals. If the range is from 51-75, the effects are mild whereas for range of 126-150, significant effects may be noticed. If the range is from 151-175, the effects will worsen.

TABLE 1 Air Quality Index and their health impact

Quality Indicator	Numeric Value	Numeric Index	Health Descriptor
Low	0-50	1-3	Effects are unlikely to be noticed even by individuals who know they are sensitive to air pollutants
Moderate	51-75	4-6	Mild effects unlikely to require action may be noticed amongst sensitive individuals.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

High	126-150	7-9	Significant effects may be noticed by sensitive individuals and action to avoid or reduce these effects may be needed. Asthmatics will find that their 'reliever' inhaler is likely to reverse the effects on the Lung.
Very High	151-175	10	The effects on sensitive individuals described for 'High' levels of pollution may worsen.

In order to enable users to understand the air index anytime and anywhere, it is designed based on the problem of large historical data of air index and real-time data collection. The intelligent brain health quality index forecasting is based on popular information technology and proposes an empty space on big data analysis. The gas quality prediction model has certain feasibility from the experimental verification. The predictive monitoring model also combines big data, Internet of Things and smart devices so that users can easily detect and predict brain health quality conditions, improve traditional single brain health quality testing methods and make detection receipts more accurate [2].

A case study has been presented which characterizes the air quality for urban environments. Recently, the social impact of air quality prediction is increasing in cities and towns and it is helping to creating awareness for accessing the real time air quality levels and its impact on human health. The monitoring stations maintained by local authorities are geographically sparsely and the resultant datasets are featured with missing labels. Data driven mechanisms are used to address the data sparsity challenge. Machine learning techniques are used to accurately predict the air quality index and the same is measured using environmental monitoring data combined with meteorological measurements. Nonlinear auto regressive neural network is used for prediction using exogenous input model. A case study featuring is applied with framework for different monitoring sites in London with a comparison between other standard machine-learning based predictive algorithms so it results in a feasible and robust performance of the proposed method for different locations in the urban region.

Two approaches proposed for air quality index estimation and prediction are based on meteorological and historical pollutant data. The first method learns based on previous air quality index and meteorological data for predicting air quality index whereas the other method learns based on previous pollution data and meteorological data. Both approaches can get good band accuracy based on the evaluations conducted across various datasets. The best approach is the latter approach which uses neural network and obtains better accuracy. But on analyzing the individual pollutant value, it is found that the method using neural network is not optimum in predicting PM10 data. Hence, linear regression can be used for prediction of air quality index [3].

Ling-Jyh Chen and Yao Hua Ho devised a method for the Large-Scale PM2.5 Sensing Systems. Studies indicate that the fine particulate matter PM2.5 has raised concerns in public health. Hence large scale low cost PM2.5 monitoring systems have been deployed in several international smart city projects. One of the major challenges for such environmental sensing systems is ensuring the data quality. An abnormal detection framework is used for large-scale, real-world environmental sensing system. The framework is composed of four modules 1) time-sliced anomaly detection (TSAD) which detects spatial, temporal and spatio-temporal anomalies in the real-time sensor measurement data stream 2) real-time emission detection which detects potential regional emission sources 3) device ranking which provides a ranking for each sensing device and 4) malfunction detection which identifies malfunctioning devices. Real world measurement of data from the AirBox project has demonstrated that the framework can effectively identify outliers in the raw measurement data as well as infer anomalous events that are perceivable by the general public and government authorities. Because of its simple design, anomaly detection framework is highly extensible to other advanced applications and it can be exploited to support various large-scale environmental sensing systems.

Anomaly detection framework to ensure the data quality for large-scale environmental sensing systems. The framework comprises four modules, namely, time-sliced anomaly detection, real-time emission detection, device ranking and

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

malfunction detection. Using the AirBox PM_{2.5} monitoring system as an example it is analyzed the data to identify the intrinsic properties of the measurement datasets. First time-sliced anomaly detection divide the continuous real time sensor measurement data into a sequence of fixed-length discrete representations using the time-slicing technique. Once the parameters to be used in the time-sliced anomaly detection module are determined, the real-time emission detection module reports potential regional emission sources every hour using the real-time AirBox data streams, the device ranking module ranks the consistency of each sensing device over time and the cross-device consistency with its neighboring devices. Finally the malfunction detection module identifies malfunctioning devices based on the daily results of the time-sliced anomaly detection module. The analyzed results of the proposed approach are available to the public in open data format. They have been used by various parties for data visualization, on-demand responses and future policy-making. Because of its simple design, anormal detection framework is highly extensible to other advanced applications and it can be exploited to support other large-scale environmental sensing systems [6].

Weisong Hu et al have analyzed and depicted that interpolation, prediction and feature analysis as the three important parameters to be considered for fine-grained air quality. The solution to the above three parameters provides extremely useful information to support air pollution control and has societal and technical benefits. It is proposed to have a general and effective approach to solve the these problems in one model called the deep air learning. The main idea of deep air learning lies in embedding feature selection and semi-supervised learning in different layers of the deep learning network. The approach utilizes the information pertaining to the unlabeled spatio-temporal data to improve the performance of the interpolation and prediction, perform feature selection and association analysis to reveal the main relevant features for the variation of air quality. In order to improve the performance of interpolation and prediction, the intrinsic characteristics of the spatio-temporal data and the information contained in the unlabeled data are embedded to spatio-temporal semi supervised learning on the output layer of neural network. To improve the performance, redundant or irrelevant features are removed while selecting features in the input layer. Association analysis when combined with feature selection reveals some of the inner mechanism of the blackbox deep network models [7].

This paper deals with how PM_{2.5} affects the health and leads to acute respiratory infection in children and adults. It performs a comparative analysis on the health effect of PM_{2.5} of patients who are admitted in hospitals with hospital admissions with acute respiratory infections. A total of 39,604 hospital admission records were collected from 98 hospitals and then matched with air pollution data from 19 monitoring stations from Jan 1 to Dec 31 in 2014. Spatial correlation test was used to estimate the spatial dependency and a time series analysis was designed with distributed lag non-linear model which was used to evaluate the associations between PM_{2.5} pollution and acute lower respiratory infections.

Significant effect distinctions were detected in children and adults, mainly revealing acute lower respiratory infections (ALRI). The spatial correlation coefficient between population exposure and hospital admissions was 0.69 for children and 0.34 for adults, significant associations between PM_{2.5} and acute lower respiratory infections admissions was found at quite low concentrations and lasted for six days with each 10- $\mu\text{g}/\text{m}^3$ increase in PM_{2.5} corresponding to a 4.3% increase in the number of admissions. While for adults, no significant association emerged until the concentration of PM_{2.5} exceeded a threshold value and lasted no more than 3 days, results suggested that short-term exposures to PM_{2.5} were associated with increased risk of acute lower respiratory infections admissions for children and adults in various ways and emphasized the needs for specific preventive measures for different age groups. Adverse health effect of PM_{2.5} on acute lower respiratory infections hospital admissions for children was estimated to be more significant than for adult considering both minimum toxic concentration and lag period. The estimated numerical result may provide valuable information for conducting health risk assessment [11].

III. CONCLUSION

The present study majorly indicates that particulate matter in air pollutant is high in urban areas. Industrial and transportation activities are mainly responsible for high pollution load in the ambient air of the area. The present study

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

further suggests that public awareness has a major role in planning and developing innovative ways to solve health, transport and related air pollution problems and the strategic plan for their implementation.

REFERENCES

- [1] Ibrahim Kok, Mehmet Ulvi Simsek, Suat Ozdemir, "A deep learning model for air quality prediction in smart cities", IEEE International Conference on Big Data, Boston, USA, 15 January 2018.
- [2] Yuan Huang, Qiang Zhao, Qianyu Zhou, "Air Quality Forecast Monitoring and its Impact on Brain Health Based on Big Data and the Internet of Things", New Trends in Brain Signal Processing and Analysis, Vol.6, pp. 78678 - 78688, 05 December 2018.
- [3] Yuchao Zhou, Suparna De, Gideon Ewa, Charith Perera, Klaus Moessner, "Data-Driven Air Quality Characterization for Urban Environments: A Case Study", Advanced Software and Data Engineering for Secure Societies, Vol.6, pp. 77996 - 78006, 03 December 2018.
- [4] Liliana Romo-Melo, Beatriz Aristizabal, Mauricio Orozco-Alzate, "Air-Quality Monitoring in an Urban Area in the Tropical Andes", IEEE Potentials, Vol. 37, Issue. 1, pp. 34 - 39, 2018.
- [5] Abdelaziz El Fazziki, Djamel Benslimane, Abderrahmane Sadiq, Jamal Ouarzazi, Mohamed Sadgal, "Agent Based Traffic Regulation System for the Roadside Air Quality Control", Heterogeneous Crowdsourced Data Analytics, Vol.5, pp. 13192 - 13201, Electronic ISSN. 2169-3536, 11 July 2017.
- [6] Ling-Jyh Chen, Yao Hua Ho, "An Anomaly Detection Framework for Large-Scale PM2.5 Sensing System", IEEE Internet of Things Journal, Vol. 5, Issue. 2, pp. 559 - 570, April 2018.
- [7] Weisong Hu, Xi Li, Zhongfei Zhang, "Deep Air Learning: Interpolation, Prediction, and Feature Analysis of Fine-grained Air Quality", IEEE Transactions on Knowledge and Data Engineering, Vol.30, Issue.12, 1 December 2018.
- [8] Akin Yasin Ayturan, Zeynep Cansu Ayturan, "Air Pollution Modelling with Deep Learning: A Review", International Journal of Environmental Pollution & Environmental Modelling, Albania, Vol.1, pp. 58- 62, 20 September 2018.
- [9] Min Chen, Jun Yang, Long Hu, M. Shamim Hossain, Ghulam Muhammad, "Urban Healthcare Big Data System Based on Crowdsourced and Cloud-Based Air Quality Indicators", IEEE Communications Magazine, Vol. 56, Issue. 11, pp. 14 - 20, 2018.
- [10] Yuxiang Zhao, Li Fu, Lingling Wang, "Design of PM2.5 Monitoring System Under the Human Micro-Environment", IEEE 27th International Symposium on Industrial Electronics, pp. 218 - 223, 2018.
- [11] Ling Yao, Xiaolin Xia, "Spatio-temporal differences in health effect of ambient PM2.5 pollution on acute respiratory infection between children and adults", Engineering Technology Center of Remote Sensing Big Data Application of Guangdong Province, Guangzhou, China, vol. 7, pp. 25718 - 25726, February 2019.
- [12] Ping-Wei Soh, Jia-Wei Chang, Jen-Wei Huang, "Adaptive Deep Learning-Based Air Quality Prediction Model Using the Most Relevant Spatial-Temporal Relations", IEEE Journals & Magazines, Vol. 6, pp. 38186 - 38199, 2018.
- [13] Xiangyu Zhao, Tong Xu, Yanjie Fu, Enhong Chen, Hao Guo, "Incorporating Spatio-Temporal Smoothness for Air Quality Inference", IEEE International Conference on Data Mining, pp. 1177 - 1182, 2017.
- [14] Wang Zhenghua, Tian Zhihui, "Prediction of Air Quality Index Based on Improved Neural Network", International Conference on Computer Systems, Electronics and Control, pp. 200 - 204, 2017.
- [15] Vladimir Ivanov, Ivelina Georgieva, "Air quality index evaluations for Sofia city", IEEE 17th International Conference on Smart Technologies, pp. 920 - 925, 2017.