International Journal of Research (IJR)
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Vol. 9 Issue 10
October 2022

A Review on Prediction of Air Quality Index and Forecasting using Machine Learning Algorithms

MR.VIVEK KUMAR YADAV[1], PROF. VISHWA GUPTA[2]
M.TECH SCHOLAR[1], ASSITANT PROFESSOR[2]
LAKSHMI NARAIN COLLEGE OF TECHNOLOGY & SCIENCE, BHOPAL

Abstract:- Day by day the air pollution becomes serious concern in India as well as in overall world. Proper or accurate prediction or forecast of Air Quality or the concentration level of other Ambient air pollutants such as Sulfur Dioxide, Nitrogen Dioxide, Carbon Monoxide, Particulate Matter having diameter less than 10µ, Particulate Matter having diameter less than 2.5µ, Ozone, etc. is very important because impact of these factors on human health becomes severe. This literature review focuses on the various techniques used forprediction or modelling of Air Quality Index (AQI) and forecasting of future concentration levels of pollutants that may cause the air pollution so that governing bodies can take the actions to reduce the pollution.

Keywords: *AQI, air pollutants, Linear Regression, Prediction, Artificial Neural Network.*

I. INTRODUCTION

Because of new inventions there is a rapid increase in the development, serious population growth and increased number of vehicles will give rise to so many critical problems related to the environment such as acid rain, deforestation, air pollution, water pollution, emission of toxic materials and so on. To fulfil the needs of growing population there is the drastic increase in industrialization that may lead to the emission of harmful gases in the atmosphere from various industries that will cause the serious air pollution problem in urban areas throughout the world. This means that the air we or people breathe is not a clean air but it is polluted as so many harmful gases and particles are present in the air that adversely affect the human health. The quality of air degrades due to the pollution.

In most of the urban areas the air pollution becomes a serious concern. The people should know about the air they breathe. The National Ambient Air Monitoring Network generates the data that includes the concentration of various pollutants present in the air but this data is not easily understood by the common people. So the Central Pollution Control Board (CPCB) develops the national Air Quality Index (AQI) for the cities in India [4]. AQI gives the idea about quality of air or at what extent the air in the particular location is polluted. This means that AQI gives the actual quality of air around us in the qualitative form that is linked with various health impacts.

According to CPCB the AQI is calculated using 12 parameters (Air Pollutants) namely NO2 (Nitrogen Dioxide), SO2 (Sulfur Dioxide), CO (Carbon Monoxide), O3 (Ozone), PM10 (Particulate Matter having diameter 10 micron or less), PM2.5 (Particulate Matter having diameter 2.5 micron or less), NH3 (Ammonia), Pb (Lead), Ni (Nickel), As (Arsenic), Benzo(a)pyrene and Benzene [4]. Most of the time AQI is based on the criteria pollutants (i.e. PM10, PM2.5, SO2, NO2, CO and O3) but while calculatingthe AQI using many pollutants from the list of 12 pollutants is more desirable. However, the selection of pollutants depends on the AQI objectives, averaging µperiod, Data availability, Monitoring frequency and measurement methods [4].

AQI can be defined as it is a numerical value that the governmental agencies used to measure the levels of air pollution in the atmosphere and communicate it with population [2]. If AQI increases then large percentage of population is affected because it adversely affects the human health. As we know that AQI can be calculated by using the concentration of different air pollutants and finally we get the single numerical value as AQI. This can be shown in **figure1.**

Received*: 19 Sept. 2022*
Revised*: 4 Oct. 2022*
Final Accepted for publication*: 5 Oct 2022*

International Journal of Research (IJR)
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Vol. 9 Issue 10
October 2022

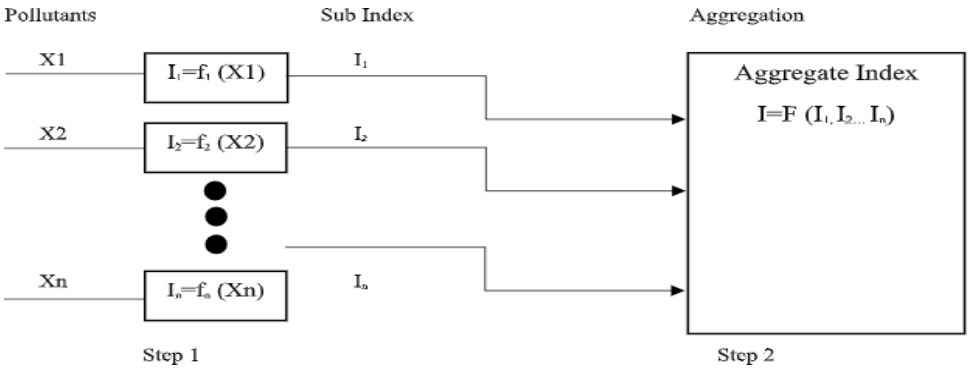

Fig.1:- Formation of an aggregated AQI [4]

The Formation of AQI involves the following two steps

1. Sub Index Calculation
2. Aggregated Index Calculation

1. Sub Index Calculation

The general equation for the sub-index (Ii) for a given pollutant concentration (Cp); as based on 'linear segmented principle' is calculated as

$$Ii = [(I_{HI} - I_{LO} / B_{HI} - B_{LO}) * (Cp - B_{LO})] + I_{LO} \quad [4]$$

Where,

B_{HI} = Breakpoint concentration greater or equal to given Concentration.

B_{LO}= Breakpoint concentration smaller or equal to given Concentration.

I_{HI} = AQI value corresponding to B_{HI}

I_{LO} = AQI value corresponding to BLO

Cp = Pollutant concentration

Breakpoint Concentration for pollutants PM2.5, NO2, SO2, CO and O3 is shown in following figure.

AQI Category (Range)	$PM_{2.5}$ 24-hr	NO_2 24-hr	O_3 8-hr	CO 8-hr	SO_2 24-hr
Good (0–50)	0–30	0–40	0–50	0–1.0	0–40
Satisfactory (51–100)	31–60	41–80	51–100	1.1–2.0	41–80
Moderate (101–200)	61–90	81–180	101–168	2.1–10	81–380
Poor (201–300)	91–120	181–280	169–208	10.1–17	381–800
Very Poor (301–400)	121–250	281–400	209–748	17.1–34	801–1600
Severe (401–500)	250+	400+	748+	34+	1600+

Note: While CO concentrations are expressed in mg m^{-3}; the other pollutants are expressed in µg m^{-3}.

Fig 2:- Breakpoint Concentration for specific pollutant [3]

2. Aggregated Index Calculation

Aggregated Index = AQI/I = Max ($I_1, I_2 \ldots I_n$) [4].

There are several ways or formulae for calculation of AQI but this is one of the most commonly used or popular form.

The AQI range and related health impacts can be shownin following figure .3.

Index Value	Name	Color	Advisory
0 to 50	Good	Green	None
51 to 100	Moderate	Yellow	Unusually sensitive individuals should consider limiting prolonged outdoor exertion
101 to 150	Unhealthy for Sensitive Groups	Orange	Children, active adults, and people with respiratory disease, such as asthma, should limit prolonged outdoor exertion
151 to 200	Unhealthy	Red	Children, active adults, and people with respiratory disease, such as asthma, should avoid prolonged outdoor exertion; everyone else should limit prolonged outdoor exertion
201 to 300	Very Unhealthy	Purple	Children, active adults, and people with respiratory disease, such as asthma, should avoid outdoor exertion; everyone else should limit outdoor exertion
301-500	Hazardous	Maroon	Everyone should avoid all physical activity outdoors.

Fig. 3:- AQI and possible health impacts [7]

Received: *19 Sept. 2022*
Revised: *4 Oct. 2022*
Final Accepted for publication: *5 Oct 2022*

As shown in figure 3 as AQI goes on increasing the risk of adverse or severe health effects will also increase. AQI gives the actual picture of the air around us and it also denotes at what extent the air is polluted. On knowing about AQI people can aware about the quality or pollution in theair in their area or location so that they can take the precautions regarding disease risks related to the airpollution.

India ranks 5th globally among 98 countries in case of air pollution [11]. So it is very important to do study about the air quality of India and make the policies to reduce the air pollution.

II. LITERATURE REVIEW

Anikender Kumar, PramilaGoyal (2011) presented the study that forecasts the daily AQI value for the city Delhi, India using previous record of AQI and meteorological parameters with the help of Principal Component Regression (PCR) and Multiple Linear Regression Techniques. They perform the prediction of daily AQI of the year 2006 using previous records of the year 2000-2005 and different equations. After that this predicted value then compared with observed value of AQI of 2006 for the seasons summer, Monsoon, Post Monsoon and winter using Multiple Linear Regression Technique [1]. Principal Component Analysis is used to find the collinearityamong the independent variables. The Principal components were used in Multiple Linear Regression to eliminate collinearity among the predictor variables and also reduce the number of predictors [1]. The Principal Component Regression gives the better performance for predicting the AQI in winter season than any other seasons. In this study only meteorological parameters were considered or used while forecasting the future AQI but they have not considered the ambient air pollutants that may cause the adverse health effects.

Huixiang Liu (et al.2019) have taken two different cities Beijing and Italian city for the study purpose. They have forecasted the Air Quality Index (AQI) for the city Beijing and predicting the concentration of NOxin an Italian City depending on two different publicly available datasets. The first Dataset for the period of December 2013 to August 2018 having 1738 instances is made available from the Beijing Municipal Environmental Centre [5] which contains the fields like hourly averaged AQI and the concentrations of PM2.5, O3, SO2, PM10, and NO2 in Beijing. The second Dataset with 9358 instances is collected from Italian city for the period of March 2004 to February 2005. This dataset contains the attributes as Hourly averaged concentration of CO, Non methane Hydrocarbons, Benzene, NOx, NO2 [5]. But they focused majorly on NOxprediction as it is one of the important predictor for Air Quality evaluation. They used Support Vector Regression (SVR) and Random Forest Regression (RFR) techniques for AQI and NOx concentration prediction. SVR shows better performance in prediction of AQI while RFR gives the better performance in predicting the NOx concentration.

Ziyue Guan and Richard O. Sinnot (2018) used the various machine learning algorithms to predict the PM2.5 concentration. Data were collected from the official website of Environment Protection Agency (EPA) for the city Melbourne that contains PM2.5 air parameter and they have also collected the unofficial data from Airbeam which is the mobile device used to measure PM2.5 value [8]. The machine Learning Algorithms Artificial Neural Network (ANN), Linear Regression (LR) and Long Short Term Memory (LSTM) recurrent neural network were used for the PM2.5 prediction but out of these algorithms LSTM givesthe best performance ad predict the high PM2.5 value with reasonable Accuracy.

HeidarMaleki (et al.2019) predicted the hourly concentration values for the ambient air pollutants NO2,SO2, PM10, PM2.5, CO and O3 for the stations Naderi, Havashenasi, MohiteZist and Behdasht in Ahvaz, Iran which is the most polluted city in the world. They have also calculated and predicted Air Quality Index (AQI) and Air Quality Health Index (AQHI) for the four air quality monitoring stations in Ahvaz mentioned above. They used Artificial Neural Network (ANN) machine learning algorithm for the prediction of air pollutants concentration (hourly) and two air quality indices AQI and AQHI over the August 2009 to August 2010. Input to ANN algorithms involves the factors such as meteorological parameters, Air pollutants concentration, time and date.

Aditya C R (et al.2018) employed the machine algorithms to detect and forecast the PM2.5 concentration level on the basis of dataset containing atmospheric conditions in a specific city. They also predicted the PM2.5 concentration level for a particular date [10]. First of all they classify the air as polluted or not polluted by using Logistic Regression algorithm and then Auto Regression algorithm was used to predict the future value of PM2.5 depending upon previous records.

Nidhi Sharma (et al.2018) had gone through the detailed data analysis of air pollutants from 2009-2017 and also proposed the critical observation of 2016-1017 air pollutants trend in Delhi, India [14]. They have predicted the future trends of various pollutants as Sulfur Dioxide (SO2), Nitrogen Dioxide (NO2), Suspended Particulate Matter (PM), Ozone (O3), Carbon Monoxide (CO) and Benzene. By using data analytics Time series Regression forecasting they have predicted the future values of the pollutants mentioned earlier on the of previous records. According to this study results the AnandVihar and Shadipur monitoring stations of Delhi are under the study. The result shows that there is a drastic increase in PM10 concentration level, NO2 and PM2.5 are evidently increased showing the increased pollution in Delhi [14]. CO is predicted to reduce by 0.169mg/m^3, there is increase in NO2 concentration level forcoming years by 16.77 µg/m^3, Ozone is predicted to increaseby 6.11mg/m^3, Benzene reduce by 1.33mg/m^3 and SO2 is forecasted to increase by 1.24µg/m^3

Received: *19 Sept. 2022*
Revised: *4 Oct. 2022*
Final Accepted for publication: *5 Oct 2022*

Mohamed Shakir and N.Rakesh (2018) have analysed the proportion of various air pollutants (NO, NO2, CO, PM10 and SO2) with respect to the time of the day and the day of the week and estimated the effect of environmental parameters as temperature, wind speed and humidity on the air pollutants mentioned above with the helpof WEKA tool [15]. The data was collected from pollution control board of Karnataka. By using ZeroR algorithm in WEKA tool the study come up with the results that shows that the concentration levels of air pollutants increase during the working days and especially during the peak hours of theday and decrease during week-ends or holidays [15]. Using Simple K-means Clustering algorithms the study shows the relationship or dependencies between the environmental factorslike Temperature, wind speed and humidity and the air pollutants like NO, NO2, PM10, CO and SO2.

KazemNaddaf (et al.2012) used the AirQ software proposed by WHO that provides the quantitative data on the impact of PM10, SO2, NO2 and O3 on the health of the people in Tehran City of Iran which is the most populated city in Iran [16]. Health impacts under the consideration were all cause mortality, cardiovascular diseases and therespiratory diseases. "The results of the study shows that the air pollutant PM10 had the highest heath impact on the 8,700,000 inhabitants of Tehran City and also caused an excess of total mortality of 2194 deaths out of 47284 in a year" [16]. The total number of excess cases of mortality due to SO2, NO2 and Ozone are 1458, 1050 and 819 respectively. These results shows that Tehran suffered from critical problem of air pollution and for Tehran there is a need to reduce the health burden of air pollution.

YusefOmidiKhaniabadi (et al.2016) the main aim of this study is to discover the relation or association between health impacts such as mortality rate of cardiovascular diseases and the air pollutants as PM10, NO2 and O3 over the years 2014 and 2015 for Kermanshah city in Iran. They used the AirQ software proposed by WHO for this purpose. The number of prematuredeaths for cardiovascular diseases is of 188 related to PM10, 33 related to NO2 and 83 related to O3[17].The results of this study indicates that if there is $10\mu/m^3$ increase in PM10, NO2 and O3 concentration level the mortality risk will increase by 1.066, 1.012 and 1.020 respectively.

R. Gunasekaran (et al.2012) the main objective of this study is to monitor the air quality of Salem Swadeswari College, Tamil Nadu area for the period of April 2011 to March 2011 and it has been shown that this area has no serious pollution issues related to the pollutants as Sulfur Dioxide, Oxides of Nitrogen and Suspended Particulate Matter because their annual average concentration are within the range of national standards. But the annual average concentration of the pollutant PM10 is slightly higher than the levels of national standard. Also the monthly 24-hour average concentration of PM10 in the same year were crossed the national standard level except during July to October [18].

S.TikheShruti (et al.2013) the research employed two soft computing algorithms Artificial Neural Network (ANN) and Genetic Programming (GP) for the prediction of future concentration levels of air pollutants such as Oxides of Sulfur (SOx), Oxides of Nitrogen (NOx) and RespirableSuspended Particulate Matter (RSPM) over the year 2005 to 2011 for Pune city in Maharashtra which is at the second position I list of polluted cities in India. They have developed total six models (three of each algorithm ANN and GP) based on hourly average data values of pollutants concentration spanning greater than 7 years. Out of these two algorithms GP algorithms gives the better performance than ANN.

ArchontoulaChaloulakou (et al.2003) the research have implemented Artificial Neural Network (ANN) and Multiple Linear Regression (MLR) algorithms to forecast the PM10 concentration over two year time period for the city Athens, Greece. "Before applying input to ANN the dataset is divided into three unequal subsets as the training dataset contains two third of the available records or cases and the remaining cases were equally divided into validationand test set" [20]. Comparison between ANN and MLR was also done in this study that indicates ANN is better in performance than MLR. According to this study ANN will give the adequate prediction solutions or results as per the requirement if it is properly trained.

III. CONCLUSION

The purpose of this literature review paper is to know in detail about the Air Quality Index (AQI) as AQI tells whether the air around us is polluted or not. It is important to know about AQI because unless and until the people know the worst impacts or hazards of air pollution they will not become that much aware about the air pollution and try to reduce it. As per this review most of the researchers worked on AQI and pollutants concentration level forecasting that will give the actual idea about AQI. Artificial Neural Network (ANN), Linear and Logistic Regression are the choices of many researchers for the prediction of AQI and air pollutants concentration. The future scope may include consideration of all parameters that is meteorologicalparameters, air pollutants while predicting AQI or forecasting the future concentration level of different pollutants.

REFERENCES

[1]. Anikender Kumar, PramilaGoyal, "Forecasting of air quality in Delhi using principal component regression technique", Atmospheric Pollution Research, 2 (2011) 436-444.
[2]. https://www.aqi.in/blog/aqi/
[3]. https://www.researchgate.net/profile/Shovan_Sahu/publication/315725810/figure/tbl1/AS:668795018440728@1536464566616/Breakpoints-of-different-pollutants-in-IND-AQI-CPCB-2014.png
[4]. https://app.cpcbccr.com/ccr_docs/FINAL-REPORT_AQI_.pdf

Received: *19 Sept. 2022*
Revised: *4 Oct. 2022*
Final Accepted for publication: *5 Oct 2022*

[5]. Huixiang Liu, Qing Li, Dongbing Yu, Yu Gu, "Air Quality Index and Air Pollutant Concentration Prediction Based on Machine Learning Algorithms", Applied Sciences, ISSN 2076-3417; CODEN: ASPCC7, 2019, 9, 4069; doi:10.3390/app9194069.

[6]. PoojaBhalgat, SejalPitale, Sachin Bhoite, "Air Quality Prediction using Machine Learning Algorithms", International Journal of Computer Applications Technology and Research Volume 8–Issue 09, 367-370, 2019, ISSN:-2319–8656.

[7]. https://www.lung.org/clean-air/outdoors/air-quality-index

[8]. Ziyue Guan and Richard O. Sinnot, "Prediction of Air Pollution through Machine Learning on the cloud", IEEE/ACM5th International Conference on Big Data Computing Applications and Technologies (BDCAT), 978-1-5386-5502-3/18/$31.00 ©2018 IEEE DOI 10.1109/BDCAT.2018.00015.

[9]. Heidar Malek, Armin Sorooshian, Gholamreza Goudarzi, Zeynab Baboli, Yaser Tahmasebi Birgani, Mojtaba Rahmati, "Air pollution prediction by using an artifcial neural network model", Clean Technologies and Environmental Policy, (2019) 21:1341–1352.

[10]. Aditya C R, Chandana R Deshmukh, Nayana D K, Praveen Gandhi Vidyavastu, "Detection and Prediction of Air Pollution using Machine Learning Models", International Journal of Engineering Trends and Technology (IJETT) – volume 59 Issue 4 – May 2018

[11]. https://www.iqair.com/us/india

[12]. Pallavi Pant, Raj M. Lal, Sarath K. Guttikunda, Armistead G. Russell, Ajay S. Nagpure, AnuRamaswami, Richard E. Peltie, "Monitoring particulate matter in India: recent trends and future outlook", Air Quality, Atmosphere & Health, 2018.

[13]. M. Bahattin CELIK, Ibrahim KADI, "The Relation Between Meteorological Factors and Pollutants Concentrations in Karabuk City", G.U. Journal of Science, 20(4): 87-95 (2007).

[14]. Nidhi Sharma , ShwetaTaneja , VaishaliSagar , Arshita Bhatt, "Forecasting air pollution load in Delhi using data analysis tools", ScienceDirect, 132 (2018) 1077–1085.

[15]. Mohamed Shakir, N. Rakesh, "Investigation on Air Pollutant Data Sets using Data Mining Tool", IEEE Xplore Part Number:CFP18OZV-ART; ISBN:978-1-5386-1442-6.

[16]. KazemNaddafi, Mohammad SadeghHassanvand, MasudYunesian, , FatemehMomeniha, RaminNabizadeh, SasanFaridi, Akbar Gholampour, "Health impact assessment of air pollution in megacity of Tehran, Iran", IRANIAN JOURNAL OF ENVIRONMENTAL HEALTH SCIENCE & ENGINEERING, 2012, 9:28

[17]. YusefOmidiKhaniabadi, GholamrezaGoudarzi, Seyed Mohammad Daryanoosh, Alessandro Borgini, Andrea Tittarelli, Alessandra De Marco, "Exposure to PM10, NO2, and O3 and impacts on human health", Environ SciPollut Res, 2016.

[18]. R. Gunasekaran, K. Kumaraswamy, P.P. Chandrasekaran, R. Elanchezhian, "MONITORING OF AMBIENT AIR QUALITY IN SALEM CITY, TAMIL NADU", International Journal of Current Research, ISSN: 0975-833X, Vol. 4, Issue, 03, pp.275-280, March, 2012

[19]. S.TikheShruti, K.C.Khare, S.N.Londhe, "Forecasting Criteria Air Pollutants Using Data Driven Approaches: An Indian Case Study", International Journal of Soft Computing 8 (4), 305-312, 2013, ISSN: 1816-9503.

[20]. ArchontoulaChaloulakou, GeorgiosGrivas, Nikolas Spyrellis, "Neural Network and Multiple Regression Models for PM10 Prediction in Athens: A Comparative Assessment", Journal of the Air & Waste Management Association, 2012.

Received: *19 Sept. 2022*
Revised: *4 Oct. 2022*
Final Accepted for publication: *5 Oct 2022*