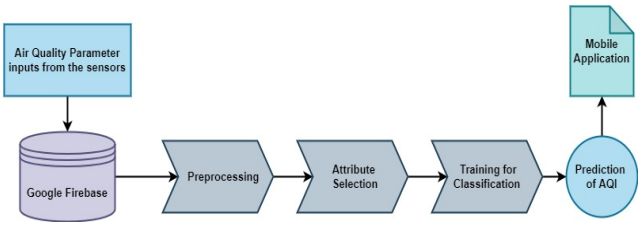
Global NEST Journal, Vol 25, No 3, pp 125-129

Smart air pollution monitoring system

Aswatha S.[1], Deepika R.[1], Dharu Piraba M.[1], Dhaneesh V.P.[1], Madheswari K.[1,*], Saraswathi S.[1], Lokeswari Y.V.[1], K.K.Nagarajan[2]
[1]Department of Computer Science and Engineering, Sri Sivasubramaniya Nadar College of Engineering, Chennai, Tamil Nadu, India
[2]Department of Electrical and Electronics Engineering, Sri Sivasubramaniya Nadar College of Engineering, Chennai, Tamil Nadu, India
Received: 05/07/2022, Accepted: 27/09/2022, Available online: 05/01/2023
*to whom all correspondence should be addressed: e-mail: MadheswariK@ssn.edu.in
https://doi.org/10.30955/gnj.004396

Graphical abstract



Abstract

Air is one of the most basic and important elements for human beings to survive. Clean and Sound air is the key to a good and healthy life. But nowadays in city life it has become the most threatening factor. Pollution of air has become the most concerning and affected issue now for us. As time goes, people are finding it very difficult to cope with the growing levels of air pollution. This proposed model aims to present a smart air pollution monitoring system that calculates and predicts the air quality in real time in the environment. The proposed system measures the concentration of air pollution causing gases in the environment using sensors. The acquired data is processed using Raspberry Pi and the collected information is stored in Firebase. With the help of the training data set, the air quality in forthcoming days can be predicted using the android application and a notification feature is included to indicate very severe pollution levels. The forecasting is done using the machine learning algorithms such as Linear Regression and Random Forest Regression with accuracy of 96.52% and 99.2% respectively.

Keywords: Raspberry Pi, air quality index (AQI), linear regression, random forest regression, forecast

1. Introduction

Air pollution refers to the release of pollutants into the air that are detrimental to human health and the planet as a whole. It can also be defined as an alteration of air quality that can be characterized by measurements of chemical, biological or physical pollutants in the air. Therefore, air pollution means the undesirable presence of impurities or the abnormal rise in the proportion of some constituents of the atmosphere. It can be classified in 2 sections: visible and invisible air pollution. The presence of substances in the atmosphere are harmful to the health of humans and other living beings, or cause damage to the climate or to materials. There are different types of air pollutants, such as gases (such as ammonia NH_3, carbon monoxide CO, Sulphur dioxide SO_2, nitrous oxides NO_x, methane CH_4 and chlorofluorocarbons), particulates (both organic and inorganic), and biological molecules. Due to man-made contributions to the environment and ecosystem such as deforestation and factories that release smog, smoke, and other mentioned chemical components into the atmosphere, the air pollution is increasing day by day. Hence, a need for an automated system to forecast the pollution level of the air is essential.

The objective of our proposed system is to design an automated air quality monitoring and forecasting system with a mobile application, which effectively provides the forthcoming air quality of that particular area and alert the users in case of severely air pollution.

1.1. Related works

Md. Hasib Bin Hossain Khan *et al.* proposed a system which is a combination of android app, servers and gas sensors (CO_2, CO, LPG and CH_4) to sense the air quality of the environment in real time. The system will give the user the indication of air quality and based on given parameters it will let the user know how much the environmental air is polluted or safe.

Ankita P. Dadhich *et al*. (2018) proposed a system that presents the evaluation of air quality in different wards of Jaipur city using geo-spatial and geo-statistical techniques to estimate the seasonal and temporal variations (2004–2015) of gaseous and particulate pollutants. The combination of GIS and statistical analyses contributes to characterize the spatial influences of weather conditions on air pollutants and provide a quick view of the relatively critical areas that need more attention from decision makers to initiate the policies geared towards developing appropriate strategies for reducing air pollution.

Hiren Jethva *et al*. (2018) proposed using NASAs A-train satellite data and back-trajectory calculations, that the Particulate Matter (PM 2.5) measured in New Delhi is strongly affected by the agricultural fires in northwestern Indian states.

Kurubaran *et al*. (2018) proposed a system for collecting values of concentration of Air Pollution from the sensors

and detecting the level of pollution in the environment and displaying the results to the people.

Dr S. Shanthi *et al*. (2019) proposed a water and air quality measuring system that checks the quality of water and air in real time through various sensors using the IoT concept. The Wi-Fi module in the system transfers the data collected by sensors to the microcontroller and transfers the data to the smartphone/PC using the ThinkSpeak API. In addition, an alert system is deployed to alert the PCB officials if the water and air quality is very poor.

V.S. Revathy *et al*. (2016) proposed a pollution monitoring system using wireless sensor networks on a real time basis. A lightweight middleware and a web interface is used to view the live pollution data in the form of numbers and charts. Other parameters like temperature and humidity are also sensed along with the gas concentration to enable data analysis through data fusion techniques.

T. H. Feiroz Khan *et al*. (2018) proposed a air quality measuring system that checks the quality of air in real time through various sensors using the IOT concept.

Col. Rajshekar Hiremath *et al*. (2018) proposed a comparative study of Air Quality Index in Industrial, Residential and Sensitive areas on different locations as study area. They have concluded that the present evaluation of AQI values in Bangalore indicate that, the air quality is “Unhealthy for sensitive group” at high traffic junctions and reaching towards “Unhealthy for sensitive group” in industrial & commercial area, which needs high attention to policy makers to evolve strategy for mitigation. The results have also revealed that even though the AQI values for Residential Area is now “Moderate”, it is showing an increasing trend in the recent past, due to construction activities, which are undertaken in the respective locations.

Pooja Bhalgat *et al*. (2019) proposed a system capable of predicting concentration of Sulphur Dioxide for forthcoming months/years using Machine Learning Algorithms. Artificial Neural Network (ANN), Genetic Algorithm ANN Model, Random Forest, Decision Tree, Deep Belief networks are the algorithms used to predict the Air Quality Index (AQI) of a given area.

Vineeta *et al*. (2019) proposed a system capable of measuring and predicting the level of CO using Random Forest Regression, Decision Tree Regression and Linear Regression from the data collected from government websites and static sensors built using Arduino.

2. Materials and methods

2.1. Machine learning algorithms

Machine learning plays a major role in development of computer algorithms that can take experience from input sources and make data driven expectations on obscure test information. Such algorithms are being separated into two classifications: supervised learning and unsupervised learning. Given labelled input and output sets, managed learning is to discover a mapping rule for foreseeing outputs and inputs sources. Regression is a type of supervised learning, which predicts the continuous quantity output for the given sample. The following are the machine learning algorithms we have used in our model for predicting the AQI.

2.1.1. Multiple linear regression

Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (y) variables, hence called as linear regression. Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable.

Mathematically, we can represent a linear regression as:

$$y = a_0 + a_1 x + \varepsilon \quad (1)$$

Here,

Y = Dependent Variable (Target Variable)

X = Independent Variable (predictor Variable)

a_0 = intercept of the line (Gives an additional degree of freedom)

a_1 = Linear regression coefficient (scale factor to each input value)

ε = random error

The values for x and y variables are training dataset for Linear Regression model representation. If more than one independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Multiple Linear Regression.

2.1.2. Random forest regression

The random forest algorithm is a supervised learning model; it uses labeled data to “learn” how to classify unlabeled data. The Random Forest Algorithm is used to solve both regression and classification problems.

Random Forest is an ensemble technique capable of performing both regression and classification tasks with the use of multiple decision trees and a technique called Bootstrap and Aggregation, commonly known as bagging. The basic idea behind this is to combine multiple decision trees in determining the final output rather than relying on individual decision trees. Random Forest has multiple decision trees as base learning models. We randomly perform row sampling and feature sampling from the dataset forming sample datasets for every model. This part is called Bootstrap.

The Pros of using Random Forest are it prevents overfitting of data, it is fast to train with test data. The Cons of using Random Forest are it is slow in creating predictions once model is made and we must beware of outliers and holes in the data.

2.2. Experimental Setup

The system consists of several sensors as depicted in Figure 1 such us Mq135, Mq2, Mq9, Dht11 sensors which are directly connected to Raspberry pi device. The analog readings of these sensors are converted into digital readings with help of MCP3008.Then the raw data from the sensor is processed and ppm value of the pollutants is calculated. When one of the calculated values exceed the threshold value, an alert is generated and then the values are uploaded to firebase database simultaneously. The

calculated values are used to predict the AQI level for the next phase and are displayed in an android application along with the calculated value.

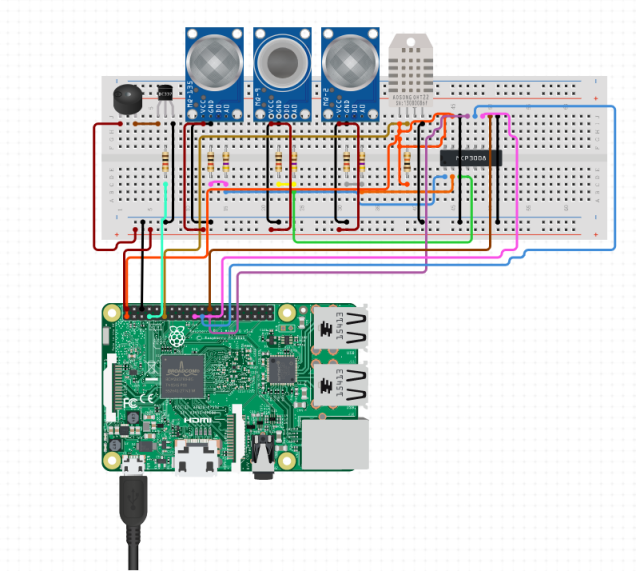


Figure 1. Overall Setup of the Rasperry Pi device.

2.3. Predicting AQI using the proposed system

Air pollution levels are strongly correlated with local weather conditions (temperature and humidity/wind) and nearby pollution emissions. However, long-range transport of pollution - through strong winds - is also a significant influencing factor and must be taken into consideration when forecasting local AQI readings. Predicting air quality, therefore, not only involves the difficulties of weather forecasting, it also requires data on and knowledge of local pollutant concentrations and emissions and pollutant concentrations and emissions from distant locations.

The overall flow of the proposed system is depicted in Figure 2.

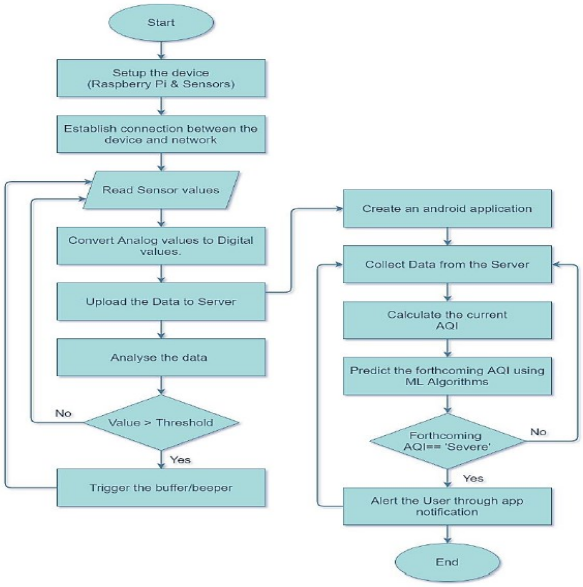


Figure 2: Overall flow of the proposed system.

The basic flow to predict the forthcoming air quality index using the machine learning algorithms such as Linear Regression and Random Forest Regression is depicted in Figure 3.

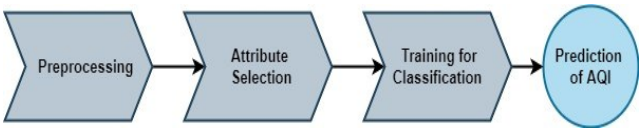


Figure 3: Flow of Prediction using ML algorithm

2.3.1. Dataset from the model

During the real time monitoring using the model, the data gets collected in a firebase. Study of data is the significant step where the data in the data set has noisy data (missing values) and redundant (duplicate values). By performing investigation of data, we can do data cleaning, data integration for filling up the missing values, for removing the redundant.

2.3.2. Preprocessing

The data considered for the training of the model is preprocessed by first estimating unknown data points between two known data points using interpolation and then removing the missing values.

2.3.3. Attribute selection

It is the initial step that helps to reduce the dimensionality by increasing the performance and accuracy. Our main features used for training the data were SO_2, NO_2, CO, NH_3.

2.3.4. Training and classification

The data which is being given as Input is taught and some categorization methods are being functioned such that we can take out the data which is hidden so that we can get exact outcomes. The raw data for training the system for predicting the AQI is obtained from a reliable resource. The table show that raw data sample obtained in the form of CSV for training the system. The machine learning algorithms are used to train the system using the data set obtained and predict the forthcoming AQI with the help of the current sensor values.

PM2.5	PM10	NO	NO2	NOX	NH3	CO	SO2	AQI
43.23	50.83	7.14	1.07	10.48	28.95	0.57	4.53	62
33.82	41.03	7.09	0.36	9.73	28.41	0.48	4.63	70
27.14	35.04	5.63	2.32	8.09	23.98	0.5	4.71	54
27.32	35.75	3.07	2.14	3.41	24.57	0.48	4.84	40
31.76	41.51	3	1.48	5.24	23.42	0.47	5.04	51
43.8	53.59	2.97	1.31	4.97	23.41	0.48	5.3	63
35.48	44.02	3.01	0.83	4.64	24.85	0.49	5.32	69
51.27	60.05	3.01	0.88	4.62	24.44	0.49	5.63	70
33.24	40.33	2.93	0.13	3.65	18.55	0.37	5.41	71
35.34	41.55	2.98	0.16	3.71	20.24	0.32	5.63	53
47.34	56.65	2.96	0.14	3.66	20.21	0.37	6.17	66
53.64	62.86	2.96	0.17	3.77	25.43	0.44	6.75	92
54.9	62.7	2.96	0.13	3.69	25.48	0.42	6.8	89
44.64	53.86	2.94	0.11	3.59	28.26	0.4	7.34	82
51.33	60.03	2.92	0.35	3.89	29.89	0.44	7.69	71
34.54	41.95	2.93	0.13	3.64	28.47	0.36	7.8	81
39.47	48.07	2.93	0.24	3.79	29.05	0.42	7.67	60
27.27	33.9	2.95	0.25	3.8	30.99	0.4	7.9	51

Figure 4: Training Dataset

The sensor values as in Figure 4, were obtained from the server in JSON format, was later converted into array for

applying machine learning algorithm and the forthcoming AQI was predicted.

2.3.5. Calculating AQI

The AQI is one of the important tools available for analyzing and representing air quality status uniformly. The cumulative effect of concentration of individual pollutants in ambient air is often expressed through a single value in the form of AQI. The AQI index for the training dataset has been computed by using the following equation.

$$AQI = \frac{1}{4}\left[\frac{SPM_{actual}}{SPM_{standard}} + \frac{PM10_{actual}}{PM10_{standard}} + \frac{SO_{2actual}}{SO_{2standard}} + \frac{NO_{xactual}}{NO_{xstandard}}\right]$$

Where actual pre-script indicate observed value of parameter for a duration and standard pre-script indicates standard value as per Central Pollution Control Board (CPCB) norms. The Table 1 represents the AQI category with the corresponding pollutants and its health break.

Table 1. AQI Category, Pollutants and Health Breaks

AQI Category (Range)	PM_{10} (24 hours)	$PM_{2.5}$ (24 hours)	NO_2 (24 hours)	O_3 (8 hours)	CO (8 hours)	SO_2 (24 hours)
Good (0 – 50)	0 – 50	0 – 30	0 – 40	0 – 50	0 – 1.0	0 – 40
Satisfactory (51 – 100)	51 – 100	31 – 60	41 – 80	51 – 100	1.1 – 2.0	41 – 80
Moderately polluted (101-200)	101 – 250	61 – 90	81 – 180	101 – 168	2.1 – 10	81 – 380
Poor (201 – 300)	251 - 350	91 – 120	181 – 280	169 – 208	10 – 17	381 – 800
Very Poor (301 – 400)	351 – 430	121 – 250	281 – 400	209 – 748	17 – 34	801 – 1600
Severe (401 – 500)	430+	250+	400+	748+	34+	1600+

2.3.6. Prediction

Model predicts the accuracy of the air quality index level in future with the help of python machine learning library functions and the regression models such as Linear Regression and Random Forest Regression. The accuracy obtained from both the models are compared.

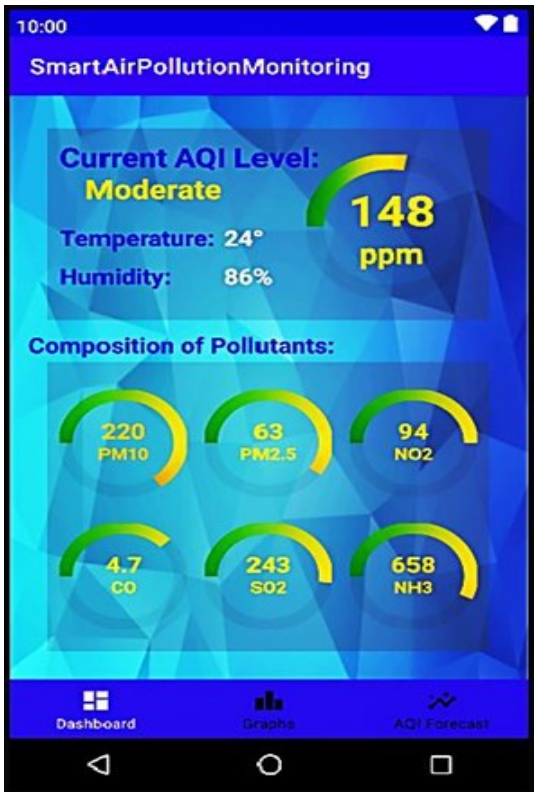


Figure 5. Dashboard Screen of the Android Application

2.3.7. App development

A mobile application is developed to display the current AQI value, the current temperature and humidity values and the composition of all the pollution causing gases. The app also displays the forthcoming AQI that has been predicted by the ML model and alerts the user in case of severe air pollution.

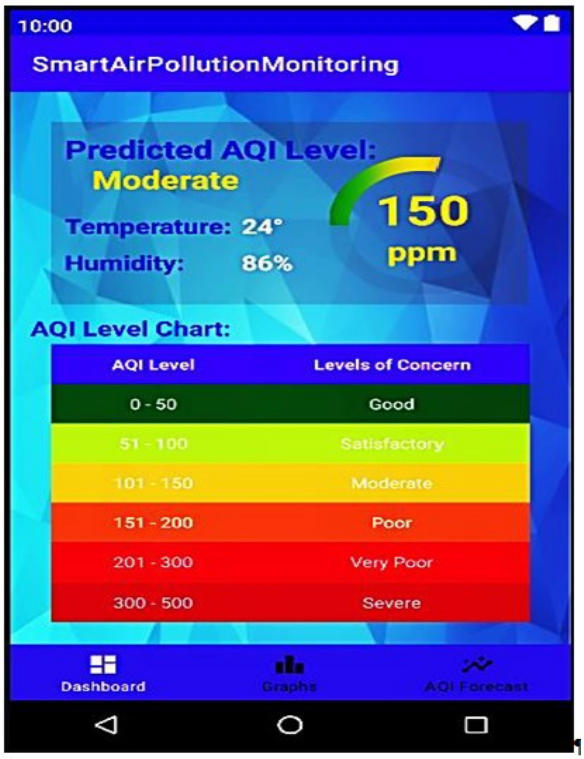


Figure 6. Forecast Screen of the Android Application

3. Results and discussion

The machine learning models - linear regression and random forest regression are used to predict the forthcoming AQI where the accuracies obtained by the linear regression model and the random forest model is 96.52% and 99.2% respectively, concluding that the performance of random forest regression is better than linear regression model.

The Android Application prompts for authentication at the beginning. The user can provide his/her username and password to login to the app. The UI contains three navigation panels: Dashboard, Graphs and AQI Forecast. The Dashboard of the android application as in Figure 5 displays the current AQI level in ppm along with composition of each pollutant like PM, NO_x, SO_2, CO, NH. The temperature and humidity are also displayed on the dashboard. The Graphs navigation of the android application contains the pictorial and numerical representation of the training data sets used for analyzing the real time data values.

The AQI Forecast navigation of the android application, as shown in Figure 6, displays the predicted AQI of the forthcoming days and the AQI level chart, for the convenience of the user to get awareness on the AQI levels.

4. Conclusion and Future Work

In this paper, we have given detailed information on how we have designed and implemented a smart air pollution monitoring system that constantly keeps track of air-quality in an area and analyses the same with the help of a training data set and predicts the level of pollution in the forthcoming days. This system creates awareness of the quality of air, as the data is displayed graphically in the android application. The implementation of the forecasting model in larger areas with more complex machine learning models can be carried out as future work.

References

Bhalgat P., Pitale S., and Bhoite S. (2019). Air Quality Prediction using Machine Learning Algorithms, International Journal of Computer Applications Technology and Research. **8** issue 9. 368-369.

Dadhich A.P., Goyal R., and Dadhich P.N. (2018). Assessment of spatio-temporal variations in air quality of Jaipur city, Rajasthan, India, The Egyptian Journal of Remote Sensing and Space Sciences. 174-180.

Hiremath R.C., Ravi D.R., Indira B., and Prof Maya Naik C. (2018). Inventorization of Air Quality Index (AQI) – A comparative study of Different Locations, International Conference on Impact of Global Atmospheric Changes on Natural Resources. 7-10

Jethva H., Chand D., Torres O., Gupta P., Lyapustin A., and Patadia F. (2018). Agricultural Burning and Air Quality over Northern India: A Synergistic Analysis using NASAs A-train Satellite Data and Ground Measurements, Aerosal and Air Quality Research. 1758-1767.

Khan F.T.H., Bhardwaj D., Ranjan Kumarn R.R., and Jadav S.K. (2018). Air Quality Monitoring System, Journal of Network Communications and Emerging Technologies. **8** issue 10. 86-89.

Khan H.B.H.MD., Ur Rahman R., Arafat Y.Md., and Nayem Jony M.A. Air Pollution Monitoring System for Smart City.

Kurubaran A., Krishna V.S., Vignesh K., and Bhaskaran A. (2018). Air Pollution Monitoring System using Android Application, International Journal of Current Research and Review. 112-115.

Revathy V.S., Ganesan K., Rohini K., Chindhu T.S, and Boobalan T. (2016). Air Pollution Monitoring System, IOSR Journal of Electronics and Communication Engineering. **11** issue 2. 27-40.

Shanthi S.Dr., Gopi T., and Vidhyesh C. (2019). IoT Based Water and Air Quality Monitoring System and Analysis, International Journal of Applied Engineering Research. **14** number 6. 1405-1409.

Vineeta, Bhat A, Manek A.S., and Mishra P. (2019). Machine Learning based Prediction System for Detecting Air Pollution, International Journal of Engineering Research and technology **8** issue 9. 155-159.