

# Air Quality Index Detection Using Machine Learning and IOT

Mrs Rajeshwari G L <sup>1</sup>, Jalaja D G <sup>2</sup>, Nandini M <sup>3</sup>, Pooja Malatesha Nadig <sup>4</sup>, R Shyamala <sup>5</sup>

<sup>1</sup>Assistant Professor, Department of Computer Science Engineering, SJB Institute of Technology, Bengaluru, 560060

<sup>2,3,4,5</sup>Student, Department of Computer Science and Engineering, SJB Institute of Technology, Bengaluru, 560060

\*\*\*

**Abstract** - Air quality index (AQI) is a quantitative tool through which air pollution data can be reported, providing information on how clean or polluted the air is. Several countries employ some type of air quality index to communicate the quality of their air. Some of these systems rely on relating measured (monitored) or predicted concentrations of air pollutants (results of modelling system). In this paper, a methodology for developing an air quality index (AQI) is presented. This index is air quality rating system based on the ambient air quality standards. The methodology applied in the calculation of this index can be applied for any country with replacing specific standards

**Key Words:** AQI-air quality index, air quality standards.

## 1.INTRODUCTION

A Polluted Air Detection system is needed to monitor air-quality in metro cities; a precise prediction provides valuable information to protect humans from being damaged by air pollution; a reasonable feature analysis reveals the main relevant factors to the variation of air quality. In general, the solutions to these topics can extract extremely useful information to support air pollution control, and consequently generate great societal and technical impacts. Some of the hazardous indoor pollutants are CO, CO<sub>2</sub>, and VOC & particulate matter. A person spends most of his time indoors, so the air quality control is very much important. National reports have mentioned the increased risks of morbidity and mortality. These studies were conducted in many places around the world and show a co-relation between daily ranges of particulate matter concentration and daily mortality. The poor air quality may cause health issues, less concentration in work, even premature death, etc. Long term health effects from air pollution include heart disease, lung cancer, Chronic asthma, pulmonary insufficiency, and respiratory diseases such as emphysema. Air pollution can also cause long term damage to people's nerves, brain, kidney's, liver, and other organs. Some scientists suspect air pollutants cause birth defects. Though air quality monitoring stations are present, but these are very expensive and not portable. To resolve this issue sensor based device has been developed. Low-cost air quality monitoring mote has been designed by Shaban et al. to measure few pollutants' concentration. Our goal was to design a low cost, high precision, portable device capable of measuring the concentration of most of the indoor environmental pollutants. The sensors are main elements for sensing the onsite environmental data

## 2. AIR QUALITY INDEX(AQI)

The air quality index (AQI) is an index for reporting air quality on a daily basis. It is a measure of how air pollution affects one's health within a short time period. The purpose of the AQI is to help people know how the local air quality impacts their health. The Environmental Protection Agency (EPA) calculates the AQI for five major air pollutants, for which national air quality standards have been established to safeguard public health. The higher the AQI value, the greater the level of air pollution and the greater the health concerns. The concept of AQI has been widely used in many developed countries for over the last three decades. AQI quickly disseminates air quality information in real-time. Different countries use different point scales to report air quality. For instance, the United States uses a 500-point scale, wherein rating between 0 and 50 is considered good. Rating between 301 to 500 range is deemed hazardous. India too follows the point scale. Everyday monitors record concentrations of the major pollutants. These raw measurements are converted into a separate AQI value for each pollutant (ground-level ozone, particle pollution, carbon monoxide, and sulfur dioxide) using standard formulae developed by EPA. The highest of these AQI values are reported as the AQI value for that day. The AQI values range from 0-4. If value is (0-1)air quality is good, (1-2)-moderate, (2- 3)-unhealthy,(3-4) very unhealthy or hazardous.

## 3.CHALLENGES

- Obtaining the correct data sets to train a model
- Selection and implementation of model which gives high precision and accuracy
- Using right concepts for data visualization
- Interfacing and integrating Embedded system, python with Machine Learning and IOT cloud Thing Speak are involved in this system

## 4.MOTIVATION

- According to a report by interdisciplinary journal Lancet Planetary health, 1.7m Indians died due to air pollution.
- Air pollution has cost an estimated 10,000 deaths in Bengaluru in 2021.
- Air pollution caused an estimated 54,000 premature deaths in the Indian capital-New Delhi in 2020.
- Industrial process emits huge amounts of organic compounds like carbon monoxide, carbon dioxide, hydrocarbons and chemicals into the air. A high quantity of carbon dioxide is the reasons for the

greenhouse and many workers working there face a lot of health related issues by inhaling this toxic air.

### 5. ARCHITECTURE

In the proposed systems we are designing the system for air quality index detection using machine learning and iot. These Machine Learning approaches help in data visualization as well as developing prediction models.

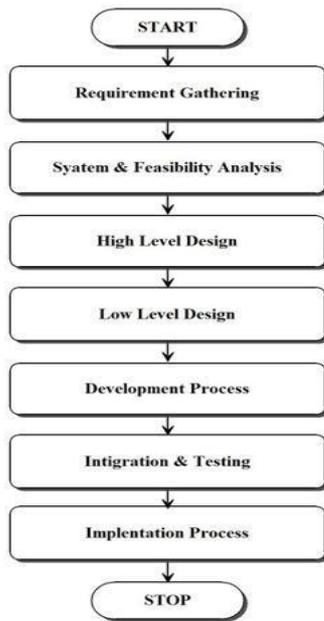


Fig 1 Architecture for the Proposed Model

### 6. METHODOLOGY

Microcontroller unit is interfaced with various sensors like Gas Sensors, temperature, humidity, etc and telemetry operation is performed. Wi-Fi Transceiver module is used to transmit all the sensed parameter to the IOT cloud THINKSPEAK and data is relayed from cloud to Python Machine Learning application. Using air pollution dataset and with Machine learning technology using python this system must build a trained model which has to receive input parameters from IOT cloud thing speak and predict the Air Quality Index

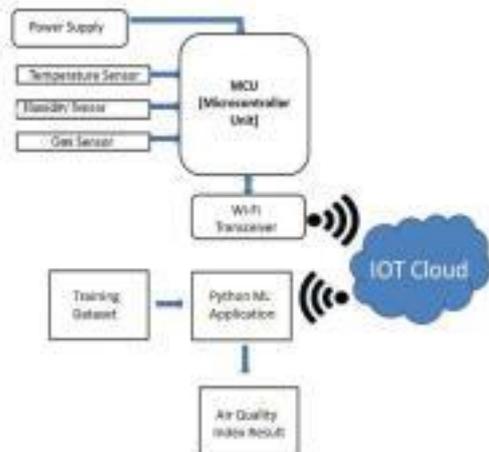


Fig 2 Methodology used for prediction

### Data Flow Diagram

This model predicts the AQI using embedded unit by taking input from all sensors attached. Data-visualization techniques use data seaborn and matplotlib.

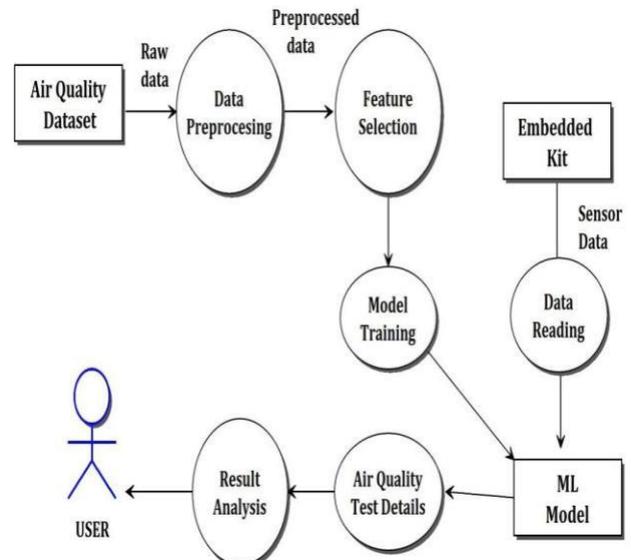


Fig 3 Data flow diagram

### Use case Diagram

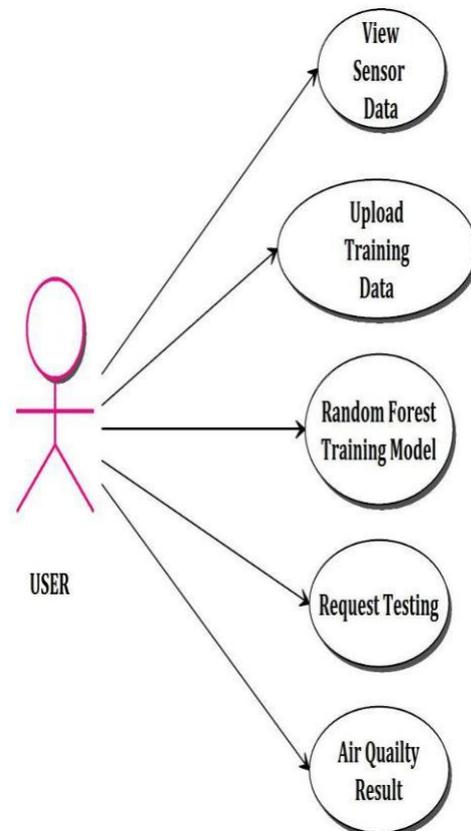


Fig 4 Use case diagram

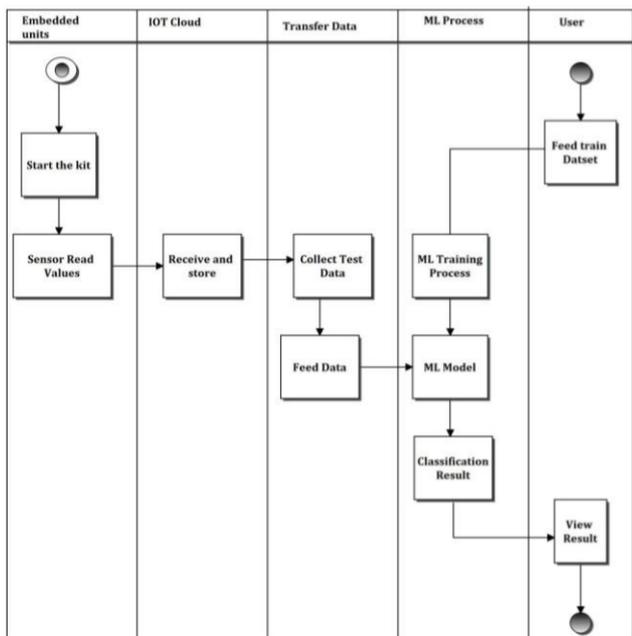


Fig 5 Activity diagram

REFERENCES

1. A. Priyadarshini, N. Dehury, and A. K. Samantaray, "A real time portable embedded system design for particulate matter monitoring," in 2015 IEEE Bombay Section Symposium (IBSS). IEEE, 2015, pp. 1–5.
2. K. B. Shaban, A. Kadri, and E. Rezk, "Urban air pollution monitoring system with forecasting models," IEEE Sensors Journal, vol. 16, no. 8, pp. 2598–2606, 2016.
3. P. K. Sharma, B. Poddar, S. Dey, S. Nandi, T. De, M. Saha, S. Mondal, and S. Saha, "On detecting acceptable air contamination in classrooms using low-cost sensors," 9th International Conference on COMMunication Systems & NETworkS, COMSNETS 2017.
4. R. Shaughnessy, U. Haverinen Shaughnessy, A. Nevalainen, and D. Moschandreas, "A preliminary study on the association between ventilation rates in classrooms and student performance," Indoor air, vol. 16, no. 6, pp. 465–468, 2006

The Machine Learning Algorithm used in our Models are:

**Support Vector Machine:** Support Vector Machine” (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. However, it is mostly used in classification problems In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then we perform classification by finding the hyper-plane that differentiate the two classes very well.

**Random Forest:** Random Forest is a flexible, easy to use machine learning algorithm that produces, even without hyper-parameter tuning, a great result most of the time. It is also one of the most used algorithms, because its simplicity and the fact that it can be used for both classification and regression tasks. It is considered as a very handy and easy to use algorithm, because its default hyper-parameters often produce a good prediction result. The number of hyper-parameters is also not that high and they are straightforward to understand

7. CONCLUSION

While there have been similar projects in the past, this project aims to bring a more refined and accurate outcome, and its portable. The goals have been satisfied with the code developed.

The proposed system in data visualization and data predictions using various prediction models such as the Null Data Handling, Random Forests, Shuffling and splitting data model helps in analysing the large datasets and predicting with high accuracy. These systems give a better way of approaching analytics as well as predictions. These systems save a large amount of time and have better accuracy.