IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Air Quality Monitoring and Forecasting Using IoT and ML: A Survey on Methodologies and Challenges

[1]Samartha S, [2]Prateek M Gummaraju, [3]Ruchi Agarwal, [4]Pallavi B

[1]Student, [2]Student, [3]Student, [4]Assistant Professor
Department of Information Science and Engineering,
B.M.S. College of Engineering, Bengaluru, India.

Abstract : Air pollution is the presence of compounds in the atmosphere that are hazardous to human beings and the health of other living organisms, or that can impair climate and materials. Chemical compounds such as carbon monoxide, ozone, nitrogen dioxide etc., are common air pollutants and they lower the quality of air. Air Quality Index (AQI) is a measure of the level of air pollution. Due to the rising levels of air pollution, there is a great need to implement effective air quality monitoring systems that gather data on the concentration of various air pollutants and provide not only the current assessments of the level of pollution, but also predict the level of pollution in the near future. This paper aims to discuss the current state of applicability and advancements in the implementation of various IoT based architectures for monitoring the levels of air pollution. We will examine the various machine learning and deep learning models that have been used to forecast AQI levels. Additionally, this paper will delve into the challenges and limitations faced during the implementation of IoT and ML for monitoring and predicting AQI. Overall, this survey paper aims to provide a comprehensive overview of the current state of IoT and Machine Learning implementation for measuring and forecasting AQI values.

***IndexTerms* - Air pollution, AQI, IoT, Arduino, Machine Learning, AutoRegression, ARMA, ARIMA, Sensors**

I. INTRODUCTION

Air pollution is defined as the presence of compounds in the atmosphere that are hazardous to human beings and the health of other living organisms, or that can impair climate and materials. These compounds can come from various sources, including industrial and agricultural emissions, transportation, and natural sources such as wildfires. The health of both humans and animals can be negatively impacted by air pollution. Respiratory conditions like asthma and bronchitis can be brought on by the minute particles in contaminated air that can enter the lungs. Heart disease and lung cancer risk can grow with prolonged exposure to contaminated air. Additionally, because it damages crops and impairs sight, air pollution can have a negative impact on the ecosystem.
Air pollution is not a new problem. It has been affecting the world for a long time. But its importance has only been recognized in recent years due to the humongous increase in the levels of air pollution and its severe impacts. Over the years, few people and organizations have tried implementing air pollution monitoring systems. But these systems did not receive the necessary traction and they are used in only very few parts of the world. Based on our research, we have come to understand that there are 3 main reasons which are preventing the global usage and implementation of these systems. They are: High costs, bulky equipment and low awareness.
In this paper, we will be surveying the existing methodologies, discussing challenges, and highlighting potential solutions. The insights gathered from this survey can pave the way for developing robust and intelligent systems that contribute to healthier and more sustainable environments.

II. MOTIVATION

It is now believed that air pollution is the largest environmental health risk on Earth, with 7 million deaths globally annually being attributable to it. It has a severe impact on the lives of human beings and animals, it has a negative impact on the environment, ecosystem and biodiversity of the world, it also impacts the economy and social well-being of the world. Air pollution is a very challenging and complex problem that requires a multifaceted solution that includes the reduction in the usage of fossil fuels, promoting the use of renewable sources of energy, alternative sources of energy and clean energy sources, implementing stricter

regulations that govern industrial emissions, and raising awareness amongst the masses. People all around the world must come together and adopt a better lifestyle that will help us tackle the issues of air pollution.
The main causes of the current rapid rise in air pollution, particularly in urban and industrial regions, are vehicle emissions, manufacturing facilities and industries accumulating a large number of dangerous pollutants. In many industrial and urban areas today, maintaining and monitoring air quality has become a top priority. Numerous elements, such as time, location, and uncertain variables have an impact on air quality. Due to the rising levels of air pollution, there is a great need to implement effective air quality monitoring systems that gather data on the concentration of various air pollutants and provide not only the current assessments of the level of pollution, but also predict the level of pollution in the near future.

III. LITERATURE SURVEY

The authors in [1] discuss building an IOT device to measure the concentration of pollutants such as Sulphur Dioxide (So2), Nitrogen Dioxide (No2), Ozone (O3), Particulate Matter 2.5 and 10 (PM2.5 and PM10). It explains how the Internet of Things is used to connect multiple devices together. IoT refers to all internet connected devices that collect, send, and track information from their surrounding environment using built-in sensors, processors, and communication tools. These "smart" or "connected" devices can occasionally interact with other similar devices and follow up on the information they receive from one another. It offers a Bird’s eye view of the architecture and design of the system and the potential usages and applications of it. The authors make use of Long Short Term Memory (LSTM) cells which is a type of Recurrent Neural Network (RNN) to forecast the future AQI. This method yields mediocre results.

In paper [2], the authors propose an Internet of Things (IoT) system that makes use of a number of sensors, including the MQ135 and MQ131 sensors as well as the PM2.5 sensor, to record the current concentration levels of different pollutants and predicts the future values of the Air Quality Index using various ML and DL models. The authors also go on to discuss the literature survey they have employed, and how they arrived at building the system. The Air Quality Index is predicted for the next 5 hours and they use Machine Learning models, including Linear Regression, time series, SVR, and stacking ensemble models. The Ensemble model has the lowest Root Mean Squared Error value amongst all the models. Their proposed system architecture is to connect the sensors to Aduino Uno, which is connected to the Cloud using Node MCU. Then the prediction is made using the Machine Learning model. This study forecasts the AQI value for the next 5 hours using a variety of machine learning models, including Linear Regression, SVR, time series models, Stacking Ensemble, etc. The Stacking Ensemble model is the chosen model selected for use since it performs the best when the RMSE of all the models is compared.

The authors in papers [3] and [4] introduce the usage of AutoRegression models for forecasting the AQI for the coming days. Paper [3] talks about how severe the issue of air pollution has become and states that one of the biggest challenges that the government faces is managing the air quality levels in urban and industrial areas. The burning of fossil fuels, traffic patterns, and industrial variables all have a big impact on air pollution. These papers explain ARMA and ARIMA models where AR stands for AutoRegression, I stands for Integrated Differencing and MA stands for Moving Average. An AutoRegressive model is a statistical model which forecasts future values based on past values. Just like how linear models make predictions based upon a combination of independent variables, AR models make predictions based on a combination of past values. This is known as autocorrelation. When the data is non-stationary, this method is performed to make the data stationary. That is, prediction is not made upon the data, rather on the differences between every 2 points of data. This may be done multiple times to remove non-stationarity. However, this is just one of the many methods of making the model deal with non-stationarity. Moving average model denotes that the forecast or result of the model depends linearly on the historical data. Additionally, it implies that predicting mistakes are linear functions of prior errors. Keep in mind that statistical moving averages are not the same as moving average models.

In their work [5], the authors talk about AQMS which is Air Quality Multi Sensors Systems. They are IoT devices that make use of inexpensive chemical microsensor arrays that have recently demonstrated their ability to produce reasonably accurate quantitative estimates of air pollutant levels. In order to solve the issue of geographical sparseness impacting the current network of AQ Regulatory Monitoring Systems, the easy accessibility of the proposed components allows the implementation of air quality monitoring networks. Urban authorities and monitoring agencies have recently become interested in the short- to mid-term performances of field data driven calibration models due to seasonal variations in the probability distribution of priors, observables, and hidden context factors. Different combinations of updates periodicity and number of incoming GT labeled samples were investigated, with a focus on network calibration and applying online machine learning components. The results demonstrated the potential for a field deployed AQMS receiving regular or opportunistic input from high accuracy labeled data sources to significantly improve the performance achievable over more than one year. Results also show that the number of labeled samples and update intervals are important factors in influencing final performances. To be more precise, the most uncommon updates need a higher percentage of tagged data to meet the same predetermined performance targets.

In paper [6], the authors propose a system which keeps a track of combinations of gasses such as CO2, methane and dust. The proposed model uses gas sensors which are responsible for detecting the gas level of the environment. Threshold levels for each gasses are decided and the signals detected from the sensors are fed into the microcontroller. ESP8266 can fetch data and load into the Internet of things. The resistance of material inside the sensor was altered, due to which the potential difference based on the gas concentration changed in the sensor which is quantified as the output voltage. An alarm will sound if the concentration of any of the gasses exceeds a predetermined level.

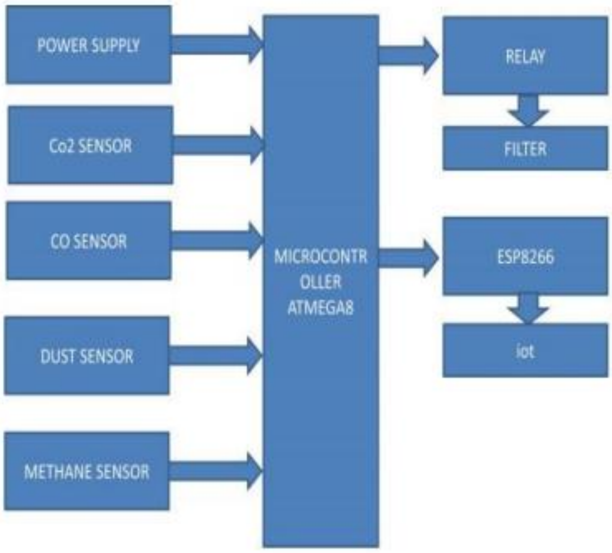

Fig 1. Block diagram of the model proposed in [6]

The authors in [7] propose 3 statistical models to forecast the daily AQI in Delhi. The three statistical models—principal component regression (PCR), time series auto regressive integrated moving average (ARIMA), and a combination of both (model 3)—are used to forecast the daily AQI for each season. Pollutant concentrations are used to evaluate the performance of all three models, and it is clear that model 3 performs the best in terms of agreement with observed values when compared to model 1 and model 2. The statistical parameters also support the same. The performance of Model 3, which combines ARIMA and PCR to tackle the problem of utilizing the autocorrelation and collinearity in the variables, is satisfactory. As a result, this model 3 can be used to anticipate air quality in additional Indian urban centers. Although the model is functioning satisfactorily, it still has a lot of uncertainty.

In paper [8], the authors discuss that the current models used for forecasting or predicting pollution in the air such as AutoRegressive Integrated Moving Average, Support Vector Regression and Multiple Linear Regression are not adequate. They are unable to understand the underlying trends in the time series data. As a result, it is necessary to provide new, efficient methods for predicting air pollution indexes. To overcome the aforementioned issues and improve forecasting accuracy, the major goal of the current research is to create efficient forecasting models for regional Air Quality Indexes. In order to forecast AQI data, two hybrid models are proposed. These are EMD - SVR - Hybrid and EMD - IMFs - Hybrid. The following are the main steps of the EMD - SVR - Hybrid model: The original AQI data is sorted using the data pre-processing technique EMD (empirical mode decomposition), yielding one group of smoother IMFs and a noise series, where the IMFs hold the crucial data (level, fluctuations, and others) from the original AQI series. The sum of the IMFs is predicted using LS - SVR, and the residual sequence of LS - SVR is predicted using S-ARIMA (seasonal ARIMA). Additionally, EMD - IMFs - Hybrid forecasts each IMF individually using statistical models before combining the findings as EMD - IMFs.

In their work [9], the authors proposed using a combination of fixed and mobile IOT sensors for monitoring Air Quality and consequently predicting it. In this experiment, the authors deployed 3 fixed IoT sensors and 3 mobile IoT sensors fitted onto cars which were roaming around in a given locality. A communication System was built to enable the different sensors to send back data to a central server through Voice over Long-Term Evolution (VoLTE). The collected data was Geo-tagged and stored in a Database. This data was used to predict the Air Quality metrics for the next day and compared with the ground truth to see how the models were performing. This paper also mentions the importance of being able to model the difference between weekends and weekdays. By merging stationary and mobile sensors, a novel method for forecasting the immediate air quality around people was investigated in this research. The outcomes of the experiments demonstrate how well our suggested hybrid distributed fixed and IoT sensor system predicts the quality of the air near humans. Additionally, by using public transportation systems like buses and taxis—which are already fitted with IoT sensor devices to assess various areas—the proposed system can be made practically realizable. The system's forecasted air quality data can be used in a variety of situations, such as when making outdoor activity plans.

The authors in [10] discuss how the AQI can be calculated using the concentrations of various pollutants. AQI is expressed as a number between 0 and 500. The lower the AQI value, the better the quality of air. The concentration levels of various pollutants is averaged and is used to calculate the air quality index (AQI). For instance, the 24-hour average concentration is used to calculate the AQI for PM2.5. By utilizing Lora nodes and Lora Gateway, the acquired data may be updated on a cloud platform. The Google map API used in the development of the online application allows for frequent updates to the pollution status.

In [11], the author proposes an independent real-time air quality monitoring system. A cutting-edge method for better managing data from various sensors may be found in the Internet of Things and cloud computing. This data is collected and transferred by the low-cost, ARM-based Raspberry Pi minicomputer. The system is tested in Delhi, where measurements are tabulated and compared with information from the regional environment control authority. The sensor-based air quality monitoring system is very accurate, reasonably priced, and simple to operate. A PM sensor, DSM501A, is interfaced to Arduino's digital pin 5, while DHT22 and BMP180 are attached to its digital pins 3 and 4, and MQ135 and MQ9 are connected to its analogue pins 2 and 3. A USB cable connects Arduino to Raspberry Pi. A Wi-Fi adapter is used to connect Raspberry Pi to the internet, and Raspberry Pi is connected to the adapter through a USB port.

$$I_{p}= [\{(I_{HI} - I_{LO})/ (B_{HI} - B_{LO})\} * (C_{p} - B_{LO})] + I_{LO}$$

B_{HI}= Break point concentration greater or equal to given conc.
B_{LO}= Break point concentration smaller or equal to given conc.
I_{HI} = AQI value corresponding to B_{HI}
I_{LO} = AQI value corresponding to B_{LO}

Fig 2. Formula for calculating AQI

In [12], the authors propose a deep learning framework which uses a neural network with a temporal sliding Long Short Term Memory extended model (TS-LSTME) to predict the air quality over every hour of the next day. Through a bidirectional long short-term memory (LSTM), which also happens to be multi-layered, that included the hourly historical PM2.5 concentration, temporal data, and meteorological data, the model incorporated the ideal time lag to realize sliding prediction. The suggested approach was used to forecast the upcoming 24-hour average PM2.5 concentration in China's most severely polluted region, Jing-Jin-Ji. In comparison to multiple linear regression, support vector, the conventional LSTM, and the LSTME, the suggested model performed better and had a higher correlation coefficient R Squared. The suggested approach has great applicability and is suitable for tasks requiring long-term forecasting.

IV. PROPOSED SYSTEM

We would like to propose a implementation with which, we hope to demonstrate the viability of a low-cost, Internet of Things based system constructed using Arduino and a variety of sensors that can not only monitor the current air quality but also predict the future and offer recommendations on how to reduce pollution and protect people from dangerous pollutants. We hope this system will help regulate and bring down the levels of air pollution and improve the quality of air, which in turn will improve the quality of lives of humans. We are suggesting an IoT device because IoT has tremendous potential in this field. IoT devices make it possible to gather and analyse data in real-time, which will help to optimise operations and enhance decision-making, which results in improved efficiency and accuracy.

Our system will have 3 parts:

1. Monitoring

An IoT device will be built using Arduino as the microcontroller. We will use MQ135 Air Quality sensor, MQ131 Ozone sensor, MQ7 Carbon Monoxide sensor, DHT 11 sensor and PM2.5 dust sensor. These sensors will measure the current concentrations of the various pollutants and will output the current AQI. The values of the concentrations of the gases is to be stored in a database, which will be used to forecast the future levels of AQI.

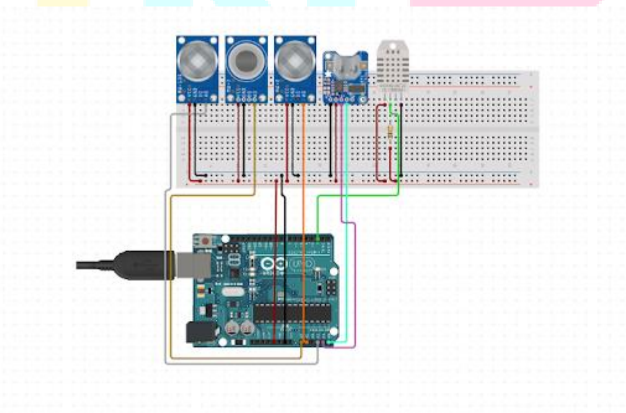

Fig 3. IoT implementation of the proposed system

2. Forecasting

Using the data collected in the previous module, an ML model is to be trained to predict the future levels of AQI. Based on the literature survey, the AutoRegression model provided the best results.

3. Controlling

AQI is categorized into 6 categories based on the value. The categories are:

AQI Value	Category
0 to 50	Good
51 to 100	Moderate
101 to 150	Unhealthy for sensitive groups
151 to 200	Unhealthy
201 to 300	Very unhealthy
301 and higher	Hazardous

Table 1. Categorization of AQI

Depending on the category of the current AQI level, suggestions are to be provided to improve the category to 'Good'.

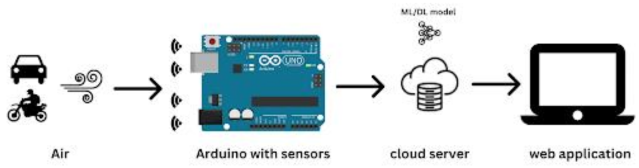

Fig 4. High level diagram of the proposed system

V. CONCLUSION

Air pollution is currently considered the biggest environmental health threat on the planet. It causes millions of deaths worldwide annually. Apart from causing severe issues to human health, it also poses a significant threat to animal life, the environment, ecology and biodiversity of the planet. It also impacts the economy and social well-being of the world. Air pollution is a very challenging and complex problem that requires a multifaceted solution. To tackle this complex issue, we are proposing the implementation of a holistic approach to tackle air pollution. Through the proposed system, we not only aim to reduce the levels of air pollution, but also positively impact the lives of people by helping them change their lifestyle in such a way that it will encourage them to adopt some habits that will help reduce air pollution, which will help them, their pets and their environment to be healthier. We are leveraging the capabilities of powerful technologies like IoT and ML to build a system that will fight against the harmful effects of air pollution and make the world a better place to live in.

REFERENCES

[1] Ayele, Temesegan Walelign, and Rutvik Mehta. "Air pollution monitoring and prediction using IoT." 2018 second international conference on inventive communication and computational technologies (ICICCT). IEEE, 2018.

[2] Balasubramanian, Ssneha, Sneha, Talapala, B, Vinushiya and S, Saraswathi. "Air Pollution Monitoring and Prediction using IOT and Machine Learning." (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 12 (2), 2021.

[3] Bhalgat, Pooja, Sachin Bhoite, and Sejal Pitare. "Air quality prediction using machine learning algorithms." International Journal of Computer Applications Technology and Research 8.9 (2019): 367-370.

[4] Tomar, Nimisha, Durga Patel, and Akshat Jain. "Air Quality Index Forecasting using Auto-regression Models." 2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS). IEEE, 2020.
[5] De Vito, Saverio, et al. "Adaptive machine learning strategies for network calibration of IoT smart air quality monitoring devices." Pattern Recognition Letters 136 (2020): 264-271.
[6] Meivel S, Mahesh M, Mohnish S, Mukesh Pandian S, "Air Pollution Monitoring System Using IOT And Artificial intelligence", IJMA, vol. 10, no. 2, pp. 3085 - 3089, May 2021.
[7] Kumar, Anikender, and P. Goyal. "Forecasting of daily air quality index in Delhi." Science of the Total Environment 409.24 (2011): 5517-5523.
[8] Zhu, Suling, et al. "Daily air quality index forecasting with hybrid models: A case in China." Environmental pollution 231 (2017): 1232-1244.
[9] Zhang, Dan, and Simon S. Woo. "Real time localized air quality monitoring and prediction through mobile and fixed IoT sensing network." IEEE Access 8 (2020): 89584 89594.
[10] Moses, Leeban. "IoT enabled Environmental Air Pollution Monitoring and Rerouting system using Machine learning algorithms." IOP Conference Series: Materials Science and Engineering. Vol. 955. No. 1. IOP Publishing, 2020.
[11] Kumar, Somansh, and Ashish Jasuja. "Air quality monitoring system based on IoT using Raspberry Pi." 2017 International conference on computing, communication and automation (ICCCA). IEEE, 2017.
[12] Mao, Wenjing, et al. "Modeling air quality prediction using a deep learning approach: Method optimization and evaluation." Sustainable Cities and Society 65 (2021): 102567.