AI powered IoT based Real-Time Air Pollution Monitoring and Forecasting

Geetha Mani¹, Joshi Kumar Viswanadhapalli¹, P Sriramalakshmi²,*

¹ School of Electrical Engineering, Vellore Institute of Technology, Vellore, India.
² School of Electrical Engineering, Vellore Institute of Technology, Chennai, India

*sriramalakshipalanidoss@gmail.com

Abstract: Air is one of the most fundamental constituents for the sustenance of life on earth. The consumption of non-renewable energy sources and industrial parameters steadily increases air pollution. These factors affect the welfare and prosperity of life on earth; therefore, the nature of Air Quality in our environment needs to be monitored continuously. This paper presents the execution and plan of Internet-of-Things (IoT) based Air Pollution Monitoring and Forecasting utilising Artificial Intelligent (AI) methods. Also, Online Dashboard was created for real-time monitoring of Air pollutants (both live and forecasted data) through ‘firebase’ from the Google cloud server. The air pollutants like Carbon Mono Oxide (CO), Ammonia (NH₃), and Ozone (O₃) layer information are collected from IoT-based sensor nodes in Vijayawada Region. Time Series modelling techniques like the Naive Bayes Model, Auto Regression Model (AR), Auto Regression Moving Average Model (ARMA), and Auto-Regression Integrating Moving Average Model (ARIMA) used to forecast the individual air pollutants aforementioned. The data collected from the IoT sensor node with a time frame is fed as input features for training the model, and optimised model parameters are obtained. The obtained model parameters are again verified with new unseen data for time. The performances of various Time Series models are validated with the help of performance indices like Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The machine learning algorithm flashed in Raspberry Pi-3. It acts as an edge computing device. The current air pollutants data and forecasted data are monitored for the next 4 hours through an online dashboard created in an open-source firebase from Google cloud service.

1. Introduction

Air is one of the essential elements to sustain life on earth. The nature of air increases the lifespan of every species on our earth. Breathing contaminated air causes harmful diseases, such as coronary heart disease (CHD), Chronic Obstructive Pulmonary Diseases (COPD), and lung cancer. The long-term effect of poor air quality leads to worldwide increases in temperature and changes in climatic patterns [1]. Air pollution is one of the risk factors in densely populated regions in the world. New ailments are being analysed each day. Industrialisation and consumption of non-renewable energy are more accounted for the increase in air pollution. Various surveillance stations have been set up worldwide to check the nature of air quality [2, 3]. World Health Organization (WHO) reported about 2.4 million deaths around the globe due to helpless air quality when contrasted with other causes [4]. The Air Quality Index (AQI) has been proposed to address the proportion of air quality in an area. The essential objective of an AQI is to rapidly publicise real-time information on air quality, especially pollutants having short-term impacts.

In literature, various methodologies in Machine Learning & Artificial Intelligence (ML-AI) have been proposed for air quality anticipating. Exclusively, computational procedures like and Time Series models contribute much precision in the prediction of AQI. Some of the notable works regarding AQI...
monitoring are, Implementation of a mobile AQI monitoring system utilising the Gaussian plume model dependent on the neural network [5]. ImgSensingNet, a vision based airborne ground detecting framework for AQI checking and estimating by the combination of pictures taken from the Unmanned Aerial-Vehicle (UAVs) [6]. An aerial-ground wireless sensor (WSN) network for monitoring of real-time PM2.5 using unmanned-aerial-vehicle (UAVs) in an urban city [7]. A 3-Dimensional (3D) Real-time AQI monitoring using the Adaptive Gaussian Plume (AGPM) model with the help of Unmanned Aerial-Vehicle (UAVs) for air pollutin monitoring [8]. An ongoing, fine-grained, power-productive air quality detecting for the smart city compares the aerial and ground sensing data to improve the quality of data collected [9]. The exhibition investigation and examination of the ARIMA model and the Artificial Neural Network (ANN) model for various data sets like a sunspot, Canadian lynx, and dollar exchange to forecast future values [10]. Air pollution monitoring system with the help of Internet of Thing (IoT) and Gas Detection and Regulating system for automatic detection of gas leaks and valuable for regulation of the source [11]. A unique weather monitoring system was implemented using IoT and raspberry pi, which reduced the power consumption and increased the robustness [12]. A detailed review of the IoT system in the environment was provided. Various sub-domains of IoT and research challenges are listed out [13]. Various ARIMA models were used for the prediction and forecasting of air pollutants [14].

In India, air pollutants like CO, NO2, SO2 NH3, and O3 are monitored using ground stations [18]. These air pollutants information is likewise taken from the satellite-based sensor. The error between the ground station and satellite information is around 10% to 26% through the study. Note that the Satellite sends just one picture each day. The information gathered by existing ground stations in India is not continuous. Hence, the average information computation is not adequate for air-quality monitoring. The establishment and support cost of ground stations is additionally extremely high. Considering the above limitations, proposed the Internet-of-Things (IoT) based Air Pollution Monitoring and forecasting system at a low cost. The Air Quality Monitoring System supported IoT utilising Raspberry Pi was proposed [16]. They utilised Node-RED and Thing speak as of their IoT platform for remote monitoring. The research gap that was found is to integrate Machine learning for prediction and forecasting [17]. The fundamental rule of IoT is that objects/things (for example, sensor hubs) recognise, sense, measure, and speak with one another without human interference. IoT is one of the quickest developing advancements over the entirety of the registering innovation. Largely IoT gadgets (sensor hubs) will have long operational lifetimes without manual observation and upkeep; particularly, battery substitution. Because continuous battery substitution is costly, damages the environment, and is frequently not even possible. The paper's primary contribution is planning and actualising an Internet-of-Things (IoT) based Air Pollution Monitoring and forecasting system at a low cost. The Air Quality Monitoring System supported IoT utilising Raspberry Pi was proposed [16]. They utilised Node-RED and Thing speak as of their IoT platform for remote monitoring. The research gap that was found is to integrate Machine learning for prediction and forecasting [17]. The fundamental rule of IoT is that objects/things (for example, sensor hubs) recognise, sense, measure, and speak with one another without human interference. IoT is one of the quickest developing advancements over the entirety of the registering innovation. Largely IoT gadgets (sensor hubs) will have long operational lifetimes without manual observation and upkeep; particularly, battery substitution. Because continuous battery substitution is costly, damages the environment, and is frequently not even possible. The paper's primary contribution is planning and actualising an Internet-of-Things (IoT) based Air Pollution Monitoring and forecasting framework. This article presents the data collection, validation of sensor data, data pre-processing, developing machine learning model, validation of the model using performance indices. Deployment of the optimal model in Edge device like Raspberry pi 3, design of User Interface (UI) using Firebase of Google cloud Server and enable the live and forecasted air pollutants monitoring though remote dashboard. The paper is formulated in such a way that Section 2 presents the proposed methodology, and Section 3 includes the complete hardware requirements, description of each component, and sensor calibration methods. Time Series Model techniques like Naïve Bayes, AR, ARMA and ARIMA, performance indices calculation was discussed in Section 4. Section 5 presents the hardware prototype model developed and an online dashboard creation and website development. Real-time outcomes and definite analysis are introduced in Section 6. At last, Section 7 gathers the conclusion and future scope of the work.

2. Proposed Research Methodology

A hardware prototype model is developed for collecting air pollutants like CO, O3, and NH3. MQ-7, MQ-131, and MQ-135 sensors measure carbon monoxide, Ozone, and ammonia, respectively. The block diagram is shown in Figure 1 of the proposed methodology, and the detailed view of Node in the Arduino method is given in Figure 2.
Figure 1. Block diagram of the given Proposed Model

As shown in Figure 2, first, we have to select the location to place the IoT Node. Three sensors interfaced with Arduino. Sensor calibration was done through Arduino. Calibrated and valid data coming from the IoT node is sent to raspberry pi 3 through Wi-Fi. Raspberry pi 3 acts as a local server cum edge computing device to store data. All kinds of data pre-processing, machine learning algorithms run inside the raspberry pi with the help of python code. Then Online dashboard and website are created using the Firebase IoT database from Google to enable the remote monitoring of Live and Forecasted Air pollutants data.

Figure 2. IoT Node Architecture

3. Hardware Requirements

Raspberry pi 3-Model B, Arduino Uno board, Carbon monoxide sensor module(MQ-7), Ozone sensor Module (MQ-131), Air Quality sensor Module(MQ-135) are used to implement the hardware set up. The hardware prototype model uses both the Arduino and Raspberry Pi. The connection diagram with the sensor interface is shown in Figure 3. The sensors are calibrated before putting into operation to get better accuracy.
4. Air Quality Forecasting using Machine Learning (ML) Techniques

The various machine learning algorithms such as Time series analysis, Naïve Bayes, Autoregression model, Auto Regression Moving Average Mode, Auto Regression Integrating Moving Average Model are used for forecasting and performance indices are discussed in this section.

4.1 Auto-Regression Integrating Moving Average Model (ARIMA)

In an ARIMA model, data transformation has been performed to make the data stationary. A Stationary is an essential condition for an ARIMA Model. A Stationary is a necessary condition for an ARIMA Model. The stationary of the data in the time series is characterised by mean, standard deviation, and auto-correlation structure. If the data present any trend, then applying the differencing and power transformation trend will be removed. When the ARIMA model is recognized, model parameters are assessed, and the last chosen model is utilized for forecast purposes. The model parameters are recorded in Table 1. The block diagram for the time-series Analysis-ARIMA model is in Figure 4.

| Model coefficients       | Values |
|--------------------------|--------|
| Auto regressive model(p) | 8      |
| Differencing model(d)    | 2      |
| Moving average (q)       | 1      |
4.2 Performance Indices
The statistical criteria RMSE and MAE [15] are used to evaluate the performance measure of each developed model.

4.2.1 Root Mean Squared Error (RMSE)
RMSE is the square root of the mean of the squared errors. RMSE indicates how close the predicted values are to the actual values. Hence, the lower RMSE value signifies that the model performance is good. RMSE is calculated using equation 1.

\[
RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (X_i - \bar{X}_2)^2}
\]  

4.2.2 Mean Absolute Error (MAE)
MAE is the mean or average of the absolute value of the errors, the Predicted – Actual. It is calculated using equation 2.

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |x_i - \hat{x}_i|
\]

5. Online Dashboard Creation and Website Development

5.1 Case Study area: Vijayawada
Vijayawada is a piece of the Capital city of Andhra Pradesh. It is the 42 biggest urban areas. Vijayawada city is an extent of 181 km2 falls under Krishna regions in Andhra Pradesh state. It is one of the most significant social, monetary, and instructive focuses in Andhra Pradesh. Because of quick industrialization and gigantic developments, there is the Real-time outstanding development of air contamination.

5.2 Server Configuration
The RPie computer was configured to establish a remote connection and provide remote access to the administrator. Using Google Firebase, an online database was created and interfaced with the existing sensor code to store values on the database. Finally, a web application was developed to extract values from the online database, plot it graphically, and display the current and forecasted air quality information in the public domain.
5.3 Realtime Data Logging
With the help of the Arduino program, the sensors collect analogue data. For every 15 minutes, one reading was recorded. The data is stored in the Firebase database with the help of raspberry pi. The live data collection is shown in Figure 5.

![Figure 5. Data Collected from IoT Node](image)

With the assistance of python code, the information is utilised for Forecasting future values. With the help of a different Time Series model, the future values are forecasted for the next five hours with a span of 15 minutes for each reading. The obtained forecasted values are stored in the firebase database.

![Figure 6. Firebase database](image)

5.4 Online Database configuration
The data collected from the sensors is sent to the IoT database as shown in Figure 6 (firebase) through
wireless mode to monitor live and forecasting data.

5.5 Online Monitoring of Air Pollutants
The real-time live and forecasting values of air pollutants for the next 4 hours are displayed on the website as shown in Figures 7a and 7b (https://rtaqms.netlify.app/), which is used for Online Monitoring of Air Pollutants from anywhere in the world.

Figure 7(a). Overview of Air Quality Monitoring website

Figure 7(b). Location Selection View form website

6. Results and discussions
The information which is gathered and put away in the data set with the assistance of Raspberry pi is utilized for forecasting with the assistance of the python program and different time series models. The gathered three-sensor information is partitioned into training data and test data. Naïve Bayes, AR, ARMA, and ARIMA calculations are trained with the help of training data, and results are forecasted for the following 4 hours with 15 minutes. Then, at that point, the forecasted information was contrasted with test data to check the model exactness of the Naïve Bayes, AR, ARMA, and ARIMA models. The performance indices are also calculated to validate the same. Figures 8 and 9 show the comparative performance analysis of Forecasting of CO during the training and testing phases. The obtained results have been inferred that the ARIMA algorithm has more accuracy for forecasting CO. The various model performance indices for test data of CO are given in Table 2. The confidence interval is 95%.
Figure 8. Forecasting of CO using Training Data

Figure 9. Forecasting of CO using Test Data

Table 2. Comparative Performance indices of CO data

| Performance Indices | MAE   | RMSE  |
|---------------------|-------|-------|
| Naive Bayes         | 0.3933| 0.4317|
| AR Model            | 0.3319| 0.3654|
| ARMA Model          | 0.3359| 0.3699|
| ARIMA Model         | 0.3269| 0.3603|

Figures 10 and 11 show the comparative performance analysis of Forecasting of NH3 during the training and testing phases. The obtained results have been inferred that the ARIMA algorithm has more accuracy for forecasting NH3 and least MAE and RMSE. The various model performance indices for test data of NH3 are given in Table 3. The confidence interval is 95%.
Figure 10. Forecasting of NH₃ using Training data

Figure 11. Forecasting of NH₃ using Test Data

Table 3. Performance indices of NH₃ data

| Performance Indices | MAE   | RMSE  |
|---------------------|-------|-------|
| Naive Bayes         | 3.1679| 3.6464|
| AR Model            | 1.3155| 1.8201|
| ARMA Model          | 1.5031| 1.9851|
| ARIMA Model         | 1.1946| 1.6751|

Figures 12 and 13 show the comparative performance analysis of Forecasting of O₃ during the training and testing phases. The obtained results have been inferred that the ARIMA algorithm has more accuracy for forecasting O₃. The various model performance indices for test data of O₃ are given in Table 4. The confidence interval is 95%.
Figure 12. Forecasting of O₃ using Training Data

![Figure 12. Forecasting of O₃ using Training Data](image1)

Figure 13. Forecasting of O₃ using Test Data

![Figure 13. Forecasting of O₃ using Test Data](image2)

Table 4 Comparative Performance indices of O₃ data

| Performance Indices | MAE  | RMSE   |
|---------------------|------|--------|
| Naive Bayes         | 1.5843 | 1.9109 |
| AR Model            | 1.5007 | 1.8765 |
| ARMA Model          | 1.4452 | 1.8032 |
| ARIMA Model         | 1.4025 | 1.7991 |

7. Conclusion

To empower the practical Real-time Air Quality Monitoring and Forecasting framework from any place, an IoT with AI strategy was proposed. IoT based hardware prototype was developed to validate the functionality of the proposed technique. Realtime air pollution data of the Vijayawada region are collected with the help of a designed IoT node from March 2020 to April 2020. The collected data is
assessed and validated using the developed machine learning models. The various Time-series Model is formulated, and they are applied to forecast individual air pollutants like CO, NH3, and O3. Firstly, the models are trained with the help of training data and assessed with the help of unseen test data. All the chosen ML techniques forecasted the values for the next four hours within a 95% confidence interval. The performance indices for the chosen models with test data were found to be acceptable. The obtained results show that ARIMA has the least MAE and RMSE values and is more accurate among other three techniques in all case studies. Hence, the ARIMA model was found as more suitable forecasting technique to predict future air pollutant values. Also, the online dashboard has been established through Firebase from Google cloud Server to enable the remote monitoring of both live and forecasted data for a selected region in India Map. For this experimental study, we tested only one IoT node located in Vijayawada, India. In the future, the proposed idea can be feasible to extend and scale up to required locations. It is a promising solution to enable lower cost and accurate real-time air pollution tracking and forecasting framework across the country.

References

[1] Ganesh, S.S., Arulmozhivarman, P. and Tatavarti, R., 2019. Forecasting air quality index using an ensemble of artificial neural networks and regression models. Journal of Intelligent Systems, 28(5), pp.893-903.

[2] Ganesh, S.S., Arulmozhivarman, P. and Tatavarti, V.R., 2018. Air quality index forecasting using artificial neural networks-a case study on Delhi. International Journal of Environment and Waste Management, 22(1-4), pp.4-23. [1]

[3] Ganesh, S.S., Arulmozhivarman, P. and Tatavarti, V.R., 2018. Prediction of PM 2.5 using an ensemble of artificial neural networks and regression models. Journal of Ambient Intelligence and Humanized Computing, pp.1-11.

[4] World Health Organization, 2016. Ambient air pollution: A global assessment of exposure and burden of disease.

[5] Yang, Y., Zheng, Z., Bian, K., Song, L. and Han, Z., 2017. Real-time profiling of fine-grained air quality index distribution using UAV sensing. IEEE Internet of Things Journal, 5(1), pp.186-198.

[6] Yang, Y., Hu, Z., Bian, K. and Song, L., 2019, April. Imgsensingnet: Uav vision guided aerial-ground air quality sensing system. In IEEE INFOCOM 2019-IEEE Conference on Computer Communications (pp. 1207-1215). IEEE.

[7] Yang, Y., Bai, Z., Hu, Z., Zheng, Z., Bian, K. and Song, L., 2018, April. AQNet: Fine-grained 3D spatio-temporal air quality monitoring by aerial-ground WSN. In IEEE INFOCOM 2018-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS) (pp. 1-2). IEEE.

[8] Yang, Y., Zheng, Z., Bian, K., Jiang, Y., Song, L. and Han, Z., 2017, December. Arms: A fine-grained 3D AQI real-time monitoring system by UAV. In GLOBECOM 2017-2017 IEEE Global Communications Conference (pp. 1-6). IEEE.

[9] Hu, Z., Bai, Z., Yang, Y., Zheng, Z., Bian, K. and Song, L., 2019. UAV aided aerial-ground IoT for air quality sensing in smart city: Architecture, technologies, and implementation. IEEE Network, 33(2), pp.14-22.

[10] Khashei, M. and Bijari, M., 2011. A novel hybridisation of artificial neural networks and ARIMA models for time series forecasting. Applied soft computing, 11(2), pp.2664-2675.

[11] Valluri, B., 2021. Implementing IoT for Society with GDR System and Air Pollution Monitoring System. Journal of Advanced Research in Microelectronics and VLSI, 2(2), pp.6-9.

[12] Joseph, F.J.J., 2019. IoT based weather monitoring system for effective analytics. International Journal of Engineering and Advanced Technology, 8(4), pp.311-315.
[13] Joseph, F.J.J., 2019. A Review of IoT Implementations in Environment and Agriculture. JARES, 6(1&2), pp.1-5.

[14] Gocheva-Ilieva, S. and Ivanov, A., 2019. Assaying SARIMA and generalised regularised regression for particulate matter PM10 modelling and forecasting. International Journal of Environment and Pollution, 66(1-3), pp.41-62.

[15] Swamynathan, M., 2019. Mastering machine learning with python in six steps: A practical implementation guide to predictive data analytics using python. Apress.

[16] Kumar, S. and Jasuja, A., 2017, May. Air quality monitoring system based on IoT using Raspberry Pi. In 2017 International Conference on Computing, Communication and Automation (ICCCA) (pp. 1341-1346). IEEE.

[17] Balasubramaniyan, C. and Manivannan, D., 2016. Iot enabled air quality monitoring system (AQMS) using raspberry Pi. Indian Journal of Science and Technology, 9(39), pp.32-39.

[18] CPCB. "National Air Quality Index, " Control of urban pollution series, 2014. https://tinyurl.com/y78slqnq