International Journal of Engineering
Research and Sustainable Technologies
Volume 1, No.2, December 2023, P 19-25
ISSN: 2584-1394 (Online version)

AN EFFICIENT AIR QUALITY INDEX MONITORING AND PREDICTION SYSTEM USING FEMTO SAT TECHNOLOGY AND MACHINE LEARNING

Kanishka Selvam, University of Essex, London.
* **Corresponding author email address**: kanishkaagu@gmail.com

https://doi.org/000000/000000/

Abstract

In recent times, the global population surge has led to a parallel increase in air pollution levels, posing significant threats to the economy, environment, and public well-being. Traditionally, air pollution is measured by placing sensors on buildings spaced a certain distance apart. Nevertheless, there are disadvantages to this strategy, such as higher power consumption for sensor operation in each house and restricted application in remote areas with inadequate infrastructure. Using satellite data from FEMTOSAT for autonomous air pollution monitoring, analysis and mitigation emerges as a ground-breaking solution to these problems. This innovative approach circumvents the limitations associated with traditional methods, offering a more comprehensive and versatile solution. FEMTOSAT leverages the capabilities of LoRa (Long Range) communication, a crucial component in the Internet of Things (IoT), enabling data transmission over extended distances while consuming minimal power. LoRa communication is essential to the FEMTOSAT installation because it makes information sharing between the satellite and ground stations easier. Data is transmitted and received using radio frequency (RF) impulses, which make for a dependable and effective form of communication. The Air Quality Index (AQI) for a particular place is then determined using the environmental data that was gathered by the satellite. An important metric for measuring air quality, the AQI provides information about the degree of pollution or cleanliness in a particular place. The AQI is determined by a number of factors, including the concentrations of pollutants including NO2, CO, O3, PM2.5, SO2, and PM10. Making use of machine learning (ML) techniques such as Support Vector Machines (SVM), direct regression, time series analysis and logistic regression to forecast and analyse AQI trends makes it very efficient. One of the distinctive features of this study is the development of AQI mappings, which are derived from the comprehensive AQI data collected at specific locations. It has been shown via thorough investigation that ML-based AQI prediction models are more consistent and dependable overall. Accuracy and precision have been ensured in the data collection process through the simplified integration of new technology and smart sensors. Only machine learning algorithms can manage the complex analysis needed to produce safe and reliable predictions from large datasets in the field of environmental monitoring. The main objective of this effort is exemplified by the integration of Integrated Sensors as the payload in the FEMTOSAT mission. The benefits of the system—such as its affordability, lightweight construction, resilience, redundancy and low power consumption—highlight FEMTOSAT's applicability and effectiveness in handling the intricate problems related to air pollution monitoring.

Keywords: Femto sat, LoRa Communication, Sensors, IOT, Air Pollution, AQI, ML

1. Introduction

The main cause of adulterants entering the atmosphere and contributing to poor air quality is urbanisation, booming industrialization and linked mortal conditioning. Air is one of the most basic components for life on Earth. The government and other relevant organisations can take the appropriate action to protect the most vulnerable from exposure to air quality hazards thanks to the capacity to forecast air quality. The success of traditional approaches to this endeavour has been severely constrained due to a lack of access to enough longitudinal data for similar styles. A number of distinct factors, including the accumulation of NO2, CO, O3, PM2.5, SO2, and PM10, influence the Air Quality Index (AQI), which indicates the quality of the air (Iyer and Turakhia, no date).

Exposure to air pollution has been linked in numerous studies to negative health effects in the general population. One of the most intriguing methods for AQI analysis and forecasting is data mining (Halsana, 2020; Gupta *et al.*, 2023). In the study done by (Palanichamy *et al.*, 2022) they compare supervised machine learning algorithms to accurately estimate PM 2.5 concentrations in Malaysia's smart cities, thereby reducing the harmful consequences. This article investigated machine learning models for PM 2.5 concentration forecasts using air quality data sets from Malaysia from 2017 to 2018. Data cleaning and normalisation were among the preprocessing procedures for the dataset. It was then reduced to an educational dataset with time and place elements by the feature extraction process. The dataset was put into three supervised machine learning classifiers: long short-term memory (LSTM), artificial neural network (ANN) and random forest (RF). Ultimately, the confusion matrix was used to assess their output, and the results were compared to determine which model produced the most accurate PM 2.5 prediction (Hill *et al.*, 1994; Pasupuleti *et al.*, 2020).

Using machine literacy (ML) techniques like time series analysis, SVM, logistic retrogression and direct retrogression aids to predict and interpret the AQI. The inconvenience is prevented by the functioning of FEMTO SAT, where satellite data is useful for independent monitoring, analysis, and combating of air pollution. Recent advancements in smartphone electronics over the last decade have greatly facilitated the creation of miniaturised satellite components. The system has the advantage of being cheap, light, robust, redundant, simple and consumes low power. Vibrant models have been used to predict AQI, including statistical, deterministic, physical, and machine learning (ML) models. Traditional methods based on probability and statistics are extremely complex and ineffective.

MLR and supervised machine learning techniques were employed to predict the AQI by many researchers. The ARIMA time series model was employed to project the AQI in the future by (Remus and O'Connor, 2001; Siew, Chin and Wee, 2008; Mani, Viswanadhapalli and S, 2022) . To assess the amount of air pollution at various places in Mumbai and Navi Mumbai, an integrated model utilising artificial neural networks and the Kriging approach was utilised (Kottur and Mantha, 2015; Kleine Deters et al., 2017). In terms of forecast and R value, ANN outperformed simple regression models. 5G and edge computing, which also offer high bandwidth, high-speed gigabit connections, ubiquitous connectivity, and ultra-low latency analytics, enable dense sensor installations at high resolution. Two methods for calibrating low-cost air quality sensors and image processing of photographs captured by hyperspectral cameras to enhance air quality detection are presented in the work of Su et al., 2021. Additionally, it envisions scalable air quality monitoring driven by AI. In these two systems on a 5G edge testbed, we create and implement several AI algorithms.

It has been demonstrated that the ML-based AQI models are more harmonic and dependable. Modern technologies and detectors made data collecting easy and accurate. Only machine learning (ML) algorithms are capable of handling the intensive analysis needed to provide accurate and consistent forecasts based on similar large environmental data **(V et al., 2022).** A Linear Retrogression (LR) model with supervised machine learning is developed in order to predict AQI. The retrogression model receives NO_2, Ozone (O_3), PM2.5, and SO_2 detector affair data obtained from the satellite as input features. The detector affair set is used as a target to train the retrogression model and generate the optimum AQI. New and unseen detector affair is used to validate the obtained model parameters. This design's ultimate goal is to forecast the Air Machine learning-based Quality Index. We validate the derived model parameters using a new and never-before-seen detector affair. This design aims to use machine learning to predict the Air Quality Index. The following system compares the learning techniques of four machines: k nearest neighbours, naive bayes, decision tree, and arbitrary timbers. Similarly, a multitude of cell satellites can be employed to gather environmental data in various areas (data on air adulterants, including PM10, PM2.5, NO_2, SO_2, CO, O_3, NH_3, and Pb), which can then be utilised to determine the vaticination of arterioles. Developing a Femto-CubeSat to gather atmospheric data on an extreme Low Earth Orbit (eLEO) is the task assigned. Not numerous cell satellite operations circumvent this low to the earth, so being suitable to take atmospheric readings from eLEO (Wu, Qu and Zhang, 2019; R.k, 2023; Long, no date) exercising the cargo will give experimenters with precious information from a scientifically rich area not constantly explored.

2. Methodology

The current method of measuring air pollution involves placing sensors on buildings and changing them on a regular basis. A significant disadvantage of this approach is that each building's sensor component needs more power to function. Moreover, this method is unfeasible in rural regions with a low density of buildings. Modern procedures use more energy to transmit signals, which is indicative of a greater dependence on energy-consuming operations. Furthermore, a significant amount of hardware must be set aside for the decoding and tracking of satellite signals under the existing methods. These difficulties highlight the necessity of investigating and putting into practice more sustainable and effective solutions for monitoring air pollution. Satellites in orbit and ground-based instruments offer information on the make-up of our atmosphere. The GOES-R satellites of NOAA For instance, the (Geostationary Operational Environmental Satellites-R) Series keeps an eye on atmospheric particle pollution. Data on airborne particles are also collected by the Joint Polar Satellite System (JPSS). These particles include smoke from wildfires, airborne dust from sandstorms and dust storms, pollution from cities and enterprises, and volcanic ash. Ozone at ground level can also be measured by the JPSS satellite series. During the day, observations of particle pollution may be obtained by GOES-R Series satellites every five minutes. Once per day, JPSS satellites may measure aerosols at a better resolution over the whole planet. Throughout the day, measurements of particle pollution may be provided by GOES-R Series satellites every five minutes. Once a day, the JPSS satellites may provide a higher resolution survey of aerosols across the whole planet. JPSS is able to further monitor the movement of aerosols from one place to another. JPSS is also capable of detecting carbon monoxide, which has been connected to poor air quality produced by wildfires. Traditional methods to this problem have had little success due to a lack of access to extensive longitudinal data (Kalaivani and Mayilvahanan, 2021).

The primary goal of this concept is to create an autonomous air pollution monitoring, analysis, and combat system using a FEMTOSAT (Environmental pollution discovery with integrated detectors as cargo). Particulate matter, ozone, carbon monoxide, moisture, temperature, pressure, wind speed, and rush are all measured in a given region at different heights using inexpensive, colourful monitors. Additionally, a dataset for air pollution forecasting and training will be created using the satellite data. i.e. AQI position in colourful regions (Jha, 2020; Holloway *et al.*, 2021; Brown, no date). A Supervised Machine Learning Model is erected, with this model AQI position is prognosticated. To reduce the ground station workload an integrated is used to store and cover the data entered from the satellite. To cut down on the quantum of time (Processing time) it takes to crack the signal from the satellite. Compute the raw signal to digital data using ML algorithm to broadcast the digital data to any digital platforms like Website, apps etc. The purpose of this design is to use machine learning to predict the Air Quality Index. Using satellite data, the machine learning algorithm is intended to predict the Air Quality Index. In order to do a computation that relies on machine learning to predict the information of impurities in the future, an ML model is developed in this work that can function with the current air contaminant apparatus and with the help of previous poisons. The data that has been found is stored inside the Excel distance for further analysis. Data on air pollution and adulterants are gathered by a variety of detector types on board the Femto satellite. Comparable to the way many femto satellites can be utilised to gather environmental data from various locations (much as air adulterants like as NO_2, SO_2, CO, O_3, NH_3, PM10, PM2.5, and Pb) and utilise that data to predict AQI conditions. Additionally, it seeks to investigate the relationship between using a cheap femtosatellite in the field of remote viewing and providing students with an incredible opportunity to work on an actual engineering design with actual engineering issues.

3. Implementation

The OBC, LoRa Tx transmitter, and microcontroller are coupled to the integrated sensors inside the Femto satellite. (IOT embedded system) The NodeMCU ESP8266 is a wifi module. It is connected to the LoRa receiver Rx. Data processing is done using programming within the NodeMCU, resulting in the digital representation of the data. The information kept in the cloud by Amazon Web Services will be processed and converted into the appropriate file format. The data must be in the form of CSV files, which will then be used as a dataset for training and forecasting the degree of air pollution, or AQI, in various places. This dataset will later be treated as Training dataset for modelling. Following the launch of the Femto Satellite in e-LEO, the data can be viewed on-site using a portable ground station. It can also be stored, processed, and monitored in Thing Speak Server. The architecture of the entire system is represented in Fig 1 below.

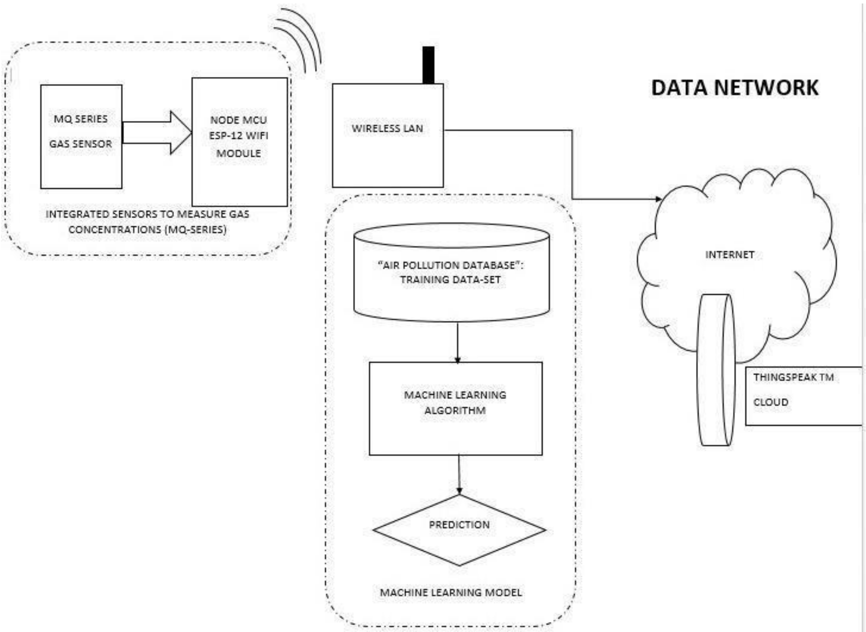

Fig1: Architecture diagram

Thing Talk is an IoT analytics system that makes it possible to collect, visualise, and analyse real-time data streams in the cloud. Thing Speak rapidly visualises data that is sent in by devices to Thing Talk. utilising to store and retrieve data from objects use the open-source Thing Talk Internet of Things (IoT) application and API to access the HTTP protocol over the

Internet or a Local Area Network. This technique allows data to be seen remotely from any location in the world. To calculate the air quality index (AQI), data for a minimum of three pollutants must be provided, with one of those pollutants having to be either PM2.5 or PM10. Every pollutant has a unique concentration on the index, which spans from 0 to 500, and each has associated health effects. This work aims to build a machine learning (ML) model that can predict the Air Quality Index for Pollution Monitoring using a variety of ML techniques, such as SVM, Linear Regression, Random Forest, Logistic, KNN, Decision tree, and Nave Bayes. The levels of air pollution in that specific area are forecasted for the present and the future using this machine learning model. A set of input variables (x) is forecasted into an output variable (y) by machine learning. There is a connection between the input and output variables. ML aims to put this relationship into numbers.

The user must first submit the dataset for the prediction model. The machine learning model is trained using the dataset that was provided to it. Each new piece of information entered into the application form serves as a test data set. The training data set, which consists of 5 feature columns and 1 target column, is used to train the model (AQI). The testing data is then predicted using feature analysis. Out of the five feature columns in the testing data, the AQI Index is the target column that our model has to forecast. Data visualisation is the next stage and is an essential field of knowledge in applied statistics. Software analyses are helpful tools for generating qualitative understanding, but machine learning statistics mostly rely on quantitative reasons. This facilitates learning about a dataset, identifying fraudulent data transfer patterns, and other tasks.

The following stage involves pre-processing and validation. The data changes carried out before the algorithm is processed are referred to as pre-processing. Data from the Femto Sat cannot be assessed once it has been collected since it is raw. Pre-processing and validation are therefore essential. The dataset is trained and the AQI is predicted using a range of machine learning techniques, such as decision trees, Random forests, Naive Bayes, Logistic Regression, SVM, and others, following the data's preprocessing. Finally, the Model Accuracy is evaluated by comparing several machine learning techniques. The most accurate machine learning technique is used to build the prediction model which is shown in Fig 2 below.

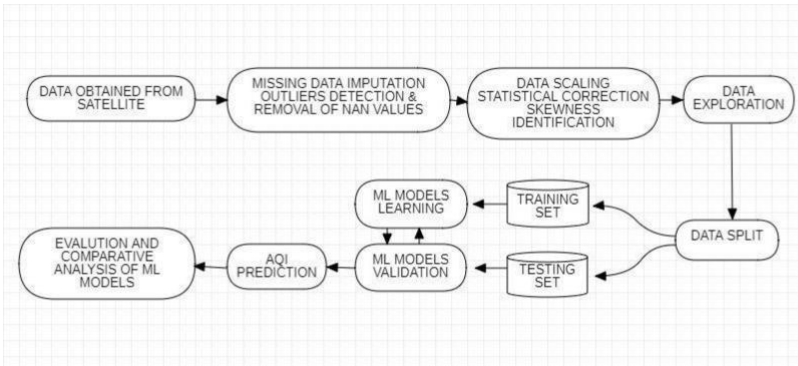

Fig 2 Proposed ML model

4. Results and Discussions

The schematic representation in Figure 3 illustrates the utilization of a BMP280 sensor to measure temperature, pressure, and altitude. The acquired data is consistently updated and made accessible through a website. This process provides real-time information about environmental parameters. Figure 4 showcases the integration of sensors such as MQ135 and MQ7 in the Arduino Integrated Development Environment (IDE). This configuration is employed for testing the functionality of these sensors, with the output meticulously monitored and timestamped in the serial monitor of the Arduino IDE. A programme that uses an ESP8266 and a Gas MQ135 gas sensor to monitor gas levels online is presented. The gathered information is sent to the cloud and then observed on the Thingspeak server, improving the centralization and accessibility of the gathered information. The utilization of the Thingspeak server as a conduit for delivering the measured percentage of gas via the internet is shown. This approach enables remote access to the data, facilitating real-time monitoring from any location worldwide. Figure 5 introduces a sophisticated system designed to anticipate pollutant and particulate levels while air quality index prediction (AQI). This model improves the precision and dependability of AQI predictions by utilising a variety of machine learning methods, such as Decision Tree (DT), Support Vector Machine (SVM), k-Nearest Neighbour (k-NN), Random Forest (RF), and Logistic Regression.The inclusion of these machine learning algorithms underscores the advanced analytical capabilities employed to provide insightful forecasts in the realm of air quality monitoring.

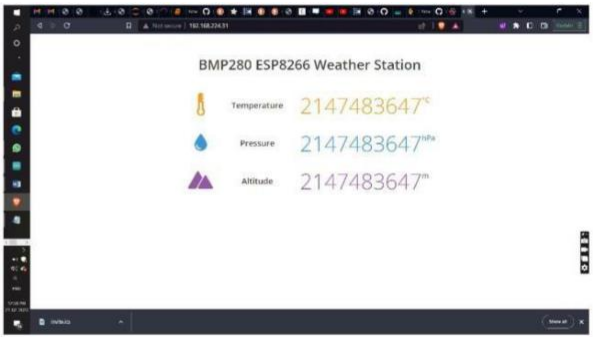

Fig. 3 BMP280 Sensor Output

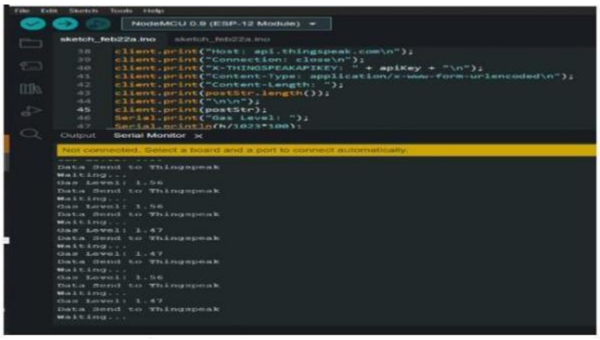

Fig 4 Online Monitoring of gas levels

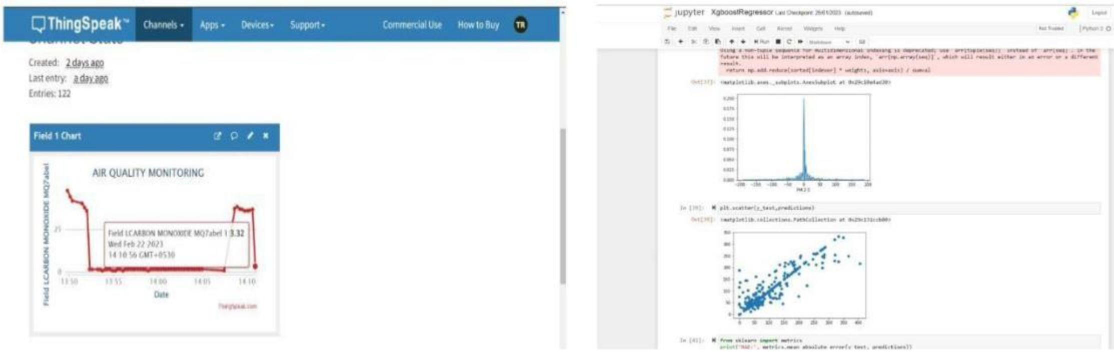

Fig. 5 Machine Learning prediction of Pollutant Levels

5. Conclusion

Estimating pollution levels is inherently challenging since the data is unstable, dynamic, and changing in space and time. However, given the effects of pollution on the environment and population, it is more crucial than ever to forecast pollutant levels. In this study, machine learning (ML) models were used to estimate levels of pollutants such NO2, SO2, PM2.5, and PM10 as well as the Air Quality Index (AQI) using publically available data. Future research utilising the obtained satellite data will also make use of these models. With the aid of satellite data, FEMTOSAT provides an additional method for dealing with large-scale air pollutant detection that will assist in monitoring pollutants and taking additional measurements for preventing air pollution in the atmosphere. This technique uses a 1U femtosat (which has sensors built in) to detect the presence of these poisons and independently monitor the air quality in a given area. The femto satellite has been used to measure the convergences of PM 2.5, O3, CO, and NH3 pollution. The study's findings indicate that femtosats can be utilised for automated flying pollutant identification, however further investigation is required to improve the system. The location's particular Air Quality Index is determined using the environmental data that was gathered from the satellite. The payload of Integrated Sensors serves as an example of the Femto Sat mission.

The solution is to use a FEMTO SAT (with integrated sensors as payload) to measure pollution levels in various locations. The device is portable, making it easy to track data at different heights from different locations. Low-cost chipset-based modules such the ATMEGA328P, MQ-7 Gas Sensor, MQ-135 Gas Sensor, and others are given a lot of consideration in its development. The study employs machine learning techniques, such as Decision Tree (DT), Support Vector Machine (SVM), k-Nearest Neighbour (k-NN), Random Forest (RF), and Logistic Regression, to forecast the air quality index and assess pollutant and particle levels. Based on real observations, the ML model forecasts the AQI level and most pollutants using training data and satellite data. Building a Femto Sat will use an extremely low Earth orbit (eLEO) to collect atmospheric data.was the aim of the Femto Sat analysis project. Since there aren't many cube satellite missions that orbit this close to the earth, collecting air data from eLEO with the payload will provide researchers with vital information from a scientifically rich area that isn't frequently explored. After that, this project will advance and include cutting-edge concepts in space science and engineering.

References

1. Brown, A. (no date) 'Satellite measurements of air quality and greenhouse gases: application to regulatory activities'.
2. Gupta, N.S. *et al.* (2023) 'Prediction of Air Quality Index Using Machine Learning Techniques: A Comparative Analysis', *Journal of Environmental and Public Health*, 2023, p. e4916267. Available at: https://doi.org/10.1155/2023/4916267.
3. Halsana, S. (2020) 'Air Quality Prediction Model using Supervised Machine Learning Algorithms', *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, pp. 190–201. Available at: https://doi.org/10.32628/CSEIT206435.
4. Hill, T. *et al.* (1994) 'Artificial neural network models for forecasting and decision making', *International Journal of Forecasting*, 10(1), pp. 5–15. Available at: https://doi.org/10.1016/0169-2070(94)90045-0.
5. Holloway, T. *et al.* (2021) 'Satellite Monitoring for Air Quality and Health', *Annual Review of Biomedical Data Science*, 4(1), pp. 417–447. Available at: https://doi.org/10.1146/annurev-biodatasci-110920-093120.
6. Iyer, D.R. and Turakhia, M.T. (no date) 'ASSESSMENT OF AIR POLLUTION OF COLLEGE CAMPUS'.
7. Jha, R. (2020) 'Air Quality Sensing and Reporting System Using IoT', in, pp. 790–793. Available at: https://doi.org/10.1109/ICIRCA48905.2020.9182796.
8. Kalaivani, G. and Mayilvahanan, P. (2021) 'Air Quality Prediction and Monitoring using Machine Learning Algorithm based IoT sensor- A researcher's perspective', in *2021 6th International Conference on Communication and Electronics Systems (ICCES)*. *2021 6th International Conference on Communication and Electronics Systems (ICCES)*, pp. 1–9. Available at: https://doi.org/10.1109/ICCES51350.2021.9489153.
9. Kleine Deters, J. *et al.* (2017) 'Modeling $PM_{2.5}$ Urban Pollution Using Machine Learning and Selected Meteorological Parameters', *Journal of Electrical and Computer Engineering*, 2017, p. e5106045. Available at: https://doi.org/10.1155/2017/5106045.
10. Kottur, S. and Mantha, S. (2015) 'An Integrated Model using Artificial Neural Network (ANN) and Kriging for Forecasting Air Pollutants using Meteorological Data', *IJARCCE*, pp. 146–152. Available at: https://doi.org/10.17148/IJARCCE.2015.4130.
11. Long, J.R. (no date) 'LEVERAGING THE NPS FEMTO SATELLITE FOR ALTERNATIVE SATELLITE COMMUNICATION NETWORKS'.
12. Mani, G., Viswanadhapalli, J.K. and S, A.A. (2022) 'Prediction and Forecasting of Air Quality Index in Chennai using Regression and ARIMA time series models', *Journal of Engineering Research*, 10(2A), pp. 179–194. Available at: https://doi.org/10.36909/jer.10253.
13. Palanichamy, N. *et al.* (2022) 'Machine learning methods to predict particulate matter PM 2.5', *F1000Research*, 11, p. 406. Available at: https://doi.org/10.12688/f1000research.73166.1.
14. Pasupuleti, V.R. *et al.* (2020) 'Air Quality Prediction Of Data Log By Machine Learning', *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pp. 1395–1399. Available at: https://doi.org/10.1109/ICACCS48705.2020.9074431.
15. Remus, W. and O'Connor, M. (2001) 'Neural Networks for Time-Series Forecasting', in J.S. Armstrong (ed.) *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Boston, MA: Springer US (International Series in Operations Research & Management Science), pp. 245–256. Available at: https://doi.org/10.1007/978-0-306-47630-3_12.
16. R.k, T. (2023) 'FEMTOSAT-Based Air Quality Monitoring: Leveraging Satellite Data and LoRa Communication for Improved AQI Predictions', *Acceleron Aerospace Journal*, 1(3), pp. 47–53. Available at: https://doi.org/10.61359/11.2106-2311.
17. Siew, L.Y., Chin, L.Y. and Wee, P.M.J. (2008) 'ARIMA AND INTEGRATED ARFIMA MODELS FOR FORECASTING AIR POLLUTION INDEX IN SHAH ALAM, SELANGOR', 12(1).
18. Su, X. *et al.* (2021) 'Intelligent and Scalable Air Quality Monitoring With 5G Edge', *IEEE Internet Computing*, PP. Available at: https://doi.org/10.1109/MIC.2021.3059189.
19. V, S. *et al.* (2022) 'Energy-Efficient Markov-Based Lifetime Enhancement Approach for Underwater Acoustic Sensor Network', *Journal of Sensors*, 2022. Available at: https://doi.org/10.1155/2022/3578002.

20. Wu, T., Qu, D. and Zhang, G. (2019) 'Research on LoRa Adaptability in the LEO Satellites Internet of Things', *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, pp. 131–135. Available at: https://doi.org/10.1109/IWCMC.2019.8766462.