REAL TIME MONITORING OF AIR POLLUTION USING AI BASED IOT AND NOVEL SARIMA TECHNIQUE IN SIVAKASI TAMILNADU

Nirmalan R[1, *], Thendral Puyalnithi[2], Karkuzhali S[3], Satish CJ[4]

[1]Assistant Professor, Department of Computer Science and Engineering, Mepco Schlenk Engineering College, Sivakasi, Virudhunagar, Tamilnadu, India.

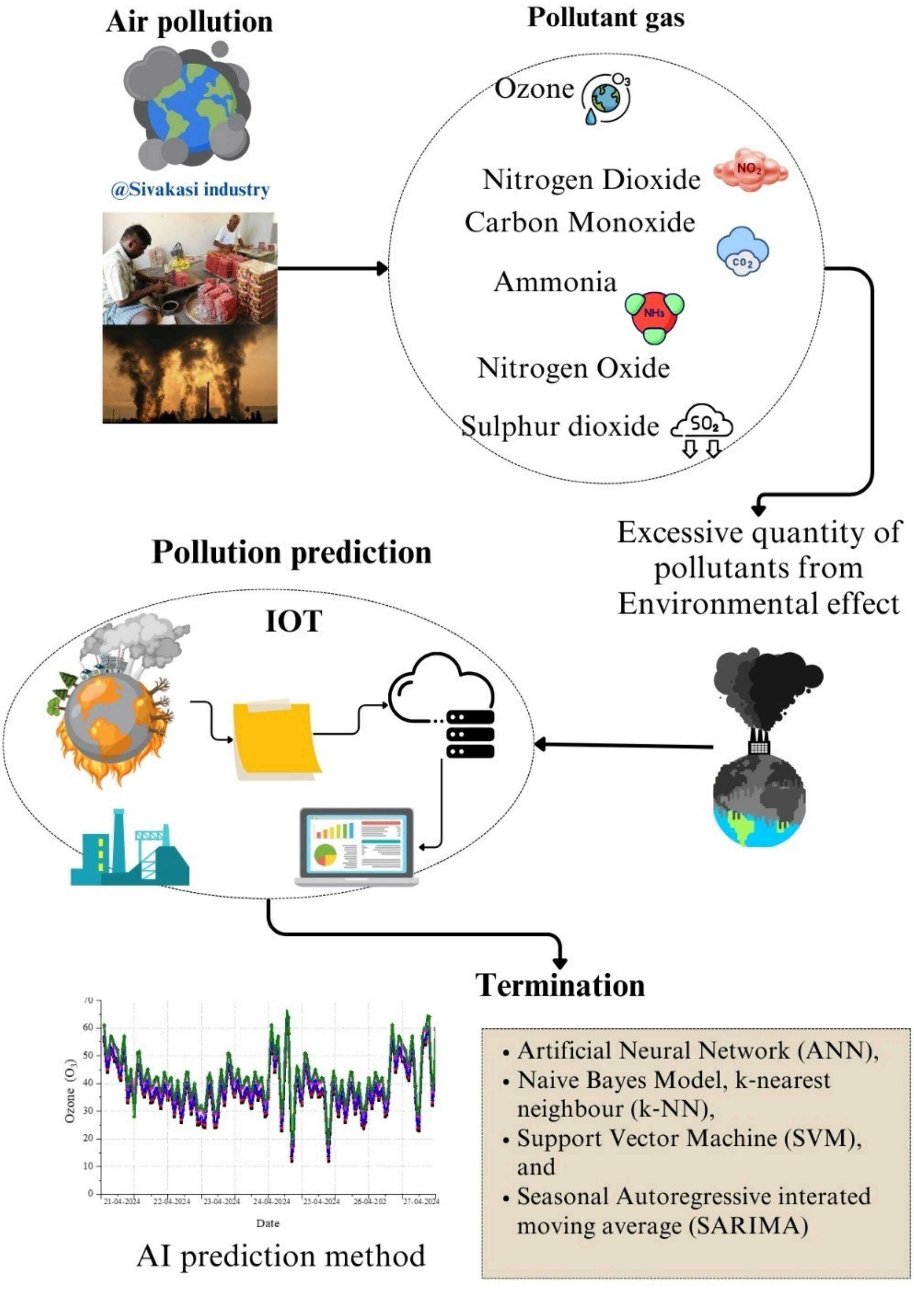
[2]Associate Professor, Department of Artificial Intelligence and Data Science, Mepco Schlenk Engineering College, Sivakasi, Virudhunagar, Tamilnadu, India.

[3]Assistant Professor, Department of Computer Science and Engineering, Mepco Schlenk Engineering College, Sivakasi, Virudhunagar, Tamilnadu, India.

[4]Associate Professor, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamilnadu, India

*Corresponding author E-mail: nirmalan@mepcoeng.ac.in

GRAPHICAL ABSTRACT


Air pollution
@Sivakasi industry
Pollutant gas
Ozone
Nitrogen Dioxide
Carbon Monoxide
Ammonia
Nitrogen Oxide
Sulphur dioxide
Excessive quantity of pollutants from Environmental effect
Pollution prediction
IOT
Termination
• Artificial Neural Network (ANN),
• Naive Bayes Model, k-nearest neighbour (k-NN),
• Support Vector Machine (SVM), and
• Seasonal Autoregressive interated moving average (SARIMA)
Ozone (O3)
Date
AI prediction method

ABSTRACT

Air pollution, a harmful or excessive quantity of pollutants from natural sources and human activities, poses risks to human health, the environment, and ecosystems. AI breakthroughs have allowed for the incorporation of technologies into performance indices, resulting in the development of an AI-based air quality system that evaluates water quality in real time using WHO-defined parameters. This article describes the implementation and planning of AI-based IoT for air pollution tracking and forecasting utilizing AI methodologies, as well as a dashboard on the internet for real-time tracking of air pollutants via Google Cloud servers. Air pollutants such as NO_2, NO_x, NH_3, CO, SO_2, and O_3 are gathered from IoT sensor nodes in Sivakasi, Tamil Nadu, India, utilizing artificial intelligence algorithms. Individual pollutants are forecasted using time series modeling approaches such as Artificial Neural Network (ANN), Naive Bayes Model, k-nearest neighbour (k-NN), Support Vector Machine (SVM), and Seasonal Autoregressive Interated Moving Average (SARIMA). The data from the IoT sensor node is utilized to train the model, resulting in optimal parameters. The derived model parameters are validated using new, previously unknown data for time. The performances of several Time Series models are examined using performance metrics such as Mean Absolute Error (MAE), coefficient of determination (R^2), and Root Mean Square Error (RMSE). An AI-based algorithm has been flashed in the Raspberry Pi 3. The present air pollution data and anticipated data are monitored throughout a 7days from 10 p.m. to 4 a.m. using a digital dashboard built in an open-source using Google cloud services. Finally comparing to all above AI based algorithms, SARIMA performed well and h+ad a 95% accuracy level.

Keywords: Air pollution, artificial intelligence, Artificial Neural Network (ANN), Naive Bayes Model, k-nearest neighbour (k-NN), Support Vector Machine (SVM), Seasonal Autoregressive Interated Moving Average (SARIMA)

1. Introduction

The rapid expansion of technology, urbanisation, and population increase have made air pollution a global issue. It affects people's life, their employment, the growth of the economy, and it fuels climate change. According to the World Health Organisation (WHO), outdoor air pollution poses a greater threat than previously recognised. Approximately 7 million people die from air pollution each year; Particulate Matter alone was responsible for 5.2 million of these premature deaths in 2020. South and East Asia have the highest death rates. While deaths from air pollution decreased between 2000 and 2023, the mortality rate in developing countries like India increased by 14%. Numerous factors contribute to air pollution, such as vehicle emissions, industrial activities, power plants, burning biomass, construction sites, waste disposal, natural occurrences like wildfires and volcanic eruptions, chemical reactions, and indoor sources like smoking, cooking with solid fuels, and poor ventilation. The extent to which these sources contribute to air pollution varies, therefore finding comprehensive solutions combining collaboration from public and private sectors as well as communities and individuals is necessary. A multifaceted strategy that takes into account the geography, the weather, industrial activity, and population density is needed to address air pollution.

Due to a number of circumstances, Sivakasi, a key centre for the Tamil Nadu matchstick and fireworks industry, is experiencing severe air pollution. Heavy metals, nitrogen oxides, sulphur dioxide, and particulate matter are among the pollutants produced by the industry. Pollution is also caused by other industries, such as printing, packing, and textile manufacture. Pollution is also caused by agricultural practices like burning crop leftovers and using biomass fuels for cooking. Pollution is also a result of the growing number of cars on the road. Sivakasi's geographic location traps airborne contaminants, which are made worse by restricted airflow. Government agencies, business associations, and the neighbourhood must collaborate to enact more stringent laws, support greener production techniques, employ cleaner fuels, and embrace environmentally friendly waste disposal strategies in order to solve the problem. One international organisation in charge of combating air pollution is the World Health Organisation (WHO). It creates regulations and standards for air quality and offers suggestions for acceptable pollution levels to safeguard the general public's health.

The World Health Organisation (WHO) also collects and performs epidemiological studies, health impact assessments, and other research on the impacts of air pollution on human health. Through influencing and public campaigns, the World Health Organisation (WHO) increases public awareness of the health effects of air pollution. The World Health Organisation (WHO) works in partnership with governments, stakeholders, and other international organisations to devise and execute global air pollution mitigation initiatives. In order to combine resources and expertise in the fight against air pollution, the WHO also forms alliances with governments, non-governmental organisations, academic institutions, business, and civil society. Figure 1 shows the frames work of the present study.

Air pollution from firework factories affects plant metabolism in Sivakasi. A study tracking plant metabolism through enzymatic and biochemical measures found significant impacts on Ficus bengalensis, highlighting the alarming rise in air pollution and deterioration of air quality in the region (Thambavani et al., 2009). This underscores the urgent need for pollution control measures to safeguard both the environment and human health. Further emphasizing the importance of pollution control, another study investigated public opinion and awareness of air pollution and control strategies in Sivakasi Taluk, Virudhunagar district (Manikandan et al., 2016). This research sheds light on community perspectives and attitudes toward environmental issues, advocating for informed policy decisions to mitigate air pollution's adverse effects. These studies collectively emphasize the critical need for proactive measures to address air pollution in Sivakasi, aiming to preserve environmental quality and public health in the region.

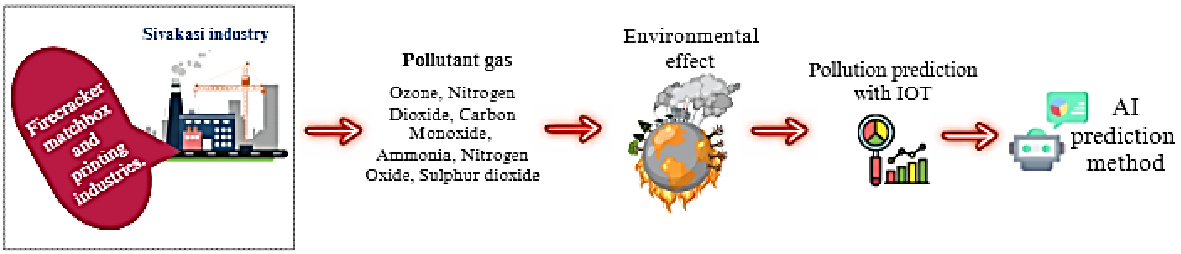


Figure 1: Framework of Air pollution

By continuously monitoring carbon dioxide levels in real-time, this innovative system aims to assess pollution levels in urban and industrial settings. It utilizes data from CO2 sensors, temperature readings, and air quality metrics through IoT technology. An alert mechanism triggers when air quality thresholds are exceeded, promptly notifying traffic management and environmental authorities via GSM communication. This proactive approach enables users to make informed decisions about their travel plans (Raj et al., 2017).

This study introduces an IoT-based system for monitoring and predicting air pollution in specific locations. Employing Long Short Term Memory (LSTM) machine learning, the system forecasts and analyzes air quality trends (Ayele & Mehta, 2018). IoT facilitates a network of smart devices capable of sensing and communicating with their surroundings globally. Proposing a three-phase air pollution monitoring setup utilizing gas sensors, Arduino IDE, and Wi-Fi modules, the solution addresses the global challenge of air pollution. Deployable in urban areas, the system collects real-time air quality data accessible via the IoT-Mobair Android application. It not only measures current pollution levels but also predicts future air quality indices (Dhingra et al., 2019).

An Internet of Things (IoT) based real-time air pollution monitoring and forecasting system designed to save hardware costs and enhance environmental protection. Using neural network technology, the device may be installed over wide areas and foresee changes in air pollution. It also improves the effectiveness of traditional air automated monitoring systems by permitting focused emergency disposal actions to reduce losses in real-world applications (Xiaojun et al., 2015). In addition to reviewing IoT-based air quality monitoring systems, this research suggests an intelligent platform for pollution reduction. It makes recommendations for online administration, cloud-based decision making, information tracking, and comprehensive network connectivity while contrasting RFID, M_2M, and sensor networks. The study also looks at how well IoT ambient air quality control platforms function and are available in different scenarios (Zhao et al., 2020).

This study introduces the Environmental Toxicology for Air Pollution Monitoring System powered by artificial intelligence and enabled by the Internet of Things (ETAPM-AIT), aimed at advancing human health. The system employs a sensor array within the Internet of Things to detect eight contaminants, with collected data transmitted to a cloud server for comprehensive analysis.

For efficient categorization and prediction of air quality, the model integrates an Elman Neural Network (ENN) enhanced by Artificial Algae Algorithms (AAA). According to findings from simulation tests, these cutting-edge methodologies demonstrate robust performance under various conditions (Asha et al., 2022) In addition, a novel approach using Internet of Things (IoT) technology for air quality monitoring incorporates edge computing capabilities. This system gathers real-time data through sensors that promptly transmit information to nearby computing nodes for immediate processing and analysis. By employing a strategy that balances local data processing with battery-powered sensing nodes, the model effectively reduces computational burdens. Moreover, algorithms are employed to address cross-sensitivity issues and minimize potential inaccuracies, resulting in a data accuracy rate between seventy-five to eighty percent. Furthermore, through automated sensor calibration, the system achieves enhanced efficiency while reducing power consumption by up to 23%. Experimental evaluations validate the operational effectiveness of this approach (Idrees et al., 2018).

In this case the effects of air pollution on the environment and human health are covered in this overview, with an emphasis on machine learning and artificial intelligence techniques. It focuses on hybrid models for predicting climate change, chronic respiratory conditions, and significant pollutants. Using performance evaluation error measures such as R^2, RMSE, MAE, and MAPE, the study emphasises the higher performance of hybrid models over single AI models (Subramaniam et al., 2022). In Consequence, Several machine learning methods have been proposed by researchers to forecast PM2.5 levels in contaminated urban areas. In Python 3.7.3, the experiment was conducted using Jupyter Notebook. The findings indicated that the models XGBoost, AdaBoost, random forest, and KNN were more accurate in forecasting PM2.5 and air quality levels. With lower error rates, the suggested models performed better than the current ones (Kothandaraman et al., 2022).

The Air Quality Index (AQI), which measures the cleanliness or contamination of the air, is a tool used by the Environmental Protection Agency (EPA) to monitor pollutants such as ozone, sulphur dioxide, particulates matter, carbon monoxide, and nitrogen dioxide (Cordova et al., 2021). The populace becomes more vulnerable as the AQI rises. At more than 4000 locations, the US Environmental Protection Agency keeps an eye on six pollutants: lead, O_3, PM10, PM2.5, NO_2, and SO_2. These pollutants correlate to various air quality criteria (Schürholz et al., 2020).

This study explores the application of SARIMA models specifically in forecasting air pollution levels. It discusses how seasonal ARIMA models can be used to analyze historical air quality data and predict future trends, taking into account the seasonality and other temporal patterns of air pollutants. The paper likely provides insights into methodology, data preprocessing, model selection, and validation techniques relevant to using SARIMA for air quality prediction (Gupta et al 2020).

The primary causes of air pollution, which are combustion processes, traffic, industry, and agriculture, are threats to ecosystems and biodiversity. Using low-emission fuels, green technology, and legal restrictions have the potential to cut air pollutants by at least 40% since 1990 (Abd Rahman et al., 2013). The three air pollutants that have the biggest effects on human health in Germany are ground-level ozone (O_3), nitrogen dioxide (NO_2), and particulate matter (PM). These are the pollutants that the European Environmental Agency concentrates on (Asghari et al., 2016). Reducing air pollution through appropriate measures is the goal of environmental agencies and municipal authorities. It is difficult to predict and simulate air pollution concentrations because of the intricate interactions that exist between contaminants and outside variables like weather, transportation, and land use (Yarragunta et al., 2021). Providing dependable data on air pollution variability to local decision-makers in transportation, urban, and environmental planning is essential (Ren et al., 2022).

Moreover, a mobile pollution sensor platform for enhanced data accuracy, visualisation tools for personalised health, travel, and pollution alerts, a clinical experiment to ascertain the causal relationships between personal pollutants and health perception, and an AI and big data framework for high-resolution, real-time air quality estimation (Geetha Mani et al., 2021). The method is cross-disciplinary and readily adaptable to different fields and nations (Li et al. , 2021). In order to address public health problems, this book examines Internet of Things options for indoor air quality monitoring. It offers case studies and innovative techniques, such as reasonably priced sensors. In this multidisciplinary field, the writers talk about cutting edge technology, applications, algorithms, systems, and future prospects (Saini et al., 2022). With case studies and review questions, it is intended for advanced undergraduate and graduate courses. The book covers issues including emissions, climate change, and ozone profiles and is easily readable by both novice and expert readers (Tiwari, 2018). Finally, With an emphasis on solid waste management, transportation, and healthcare systems, the guide covers the usage of IoT devices in pollution control and health applications. In addition to providing strategies for controlling and reducing pollution sources, it addresses the role of IoT in monitoring industrial pollution, solid waste, and healthcare (Roy et al., 2023).

2. Proposed Research Methodology

2.1 Data collection

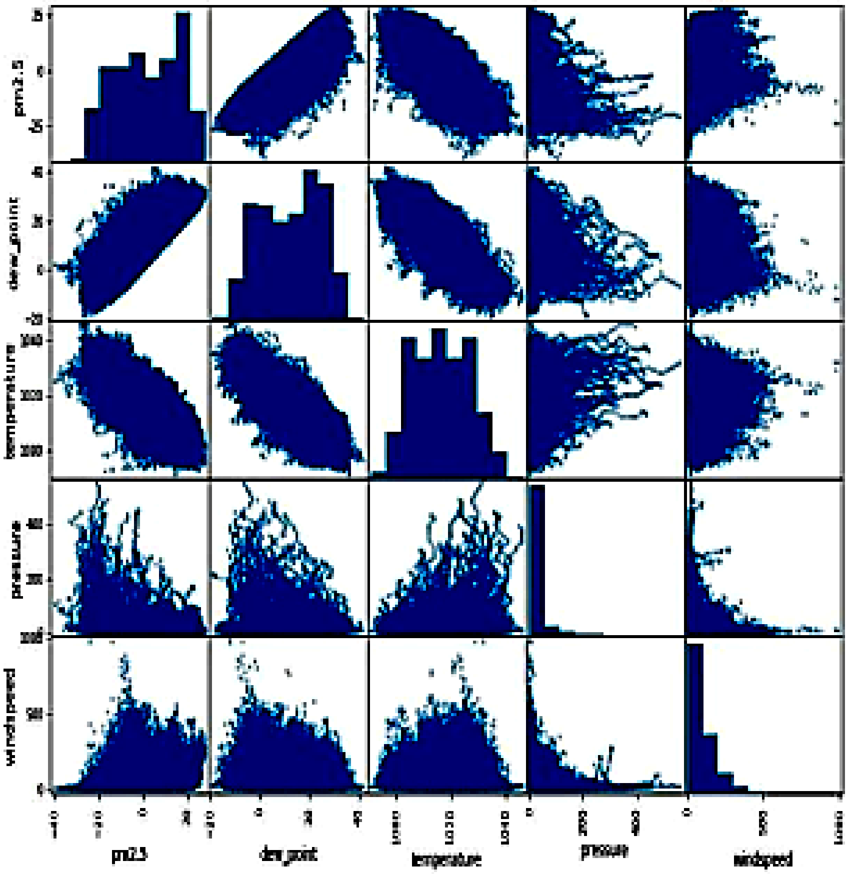


Figure 2 Scatter plot for the relation among attributes

Data is acquired from the UCI AI repository using PM2.5 concentrations. The data set was collected between April 21, 2024 and April 27, 2024 from 10PM to 4AM in Sivakasi, TamilNadu, India as well as from https://www.windy.com/-Menu/menu?21.997,79.001,5. The data set's major characteristics are time, date, hour, NO_2, NO_x, NH_3, CO, SO_2, O_3, pressure, temperature, combined wind direction, and cumulative wind speed. Each recorded sample has a length of one hour (Figure 2). By identifying and removing redundant values throughout preparation stages, the performance of proposed models is assessed using the UCI AI repository data set.

2.2Data measurement

An Internet of Things-based Air Quality Index (AQI) system monitors and measures air quality in real time. The following steps describe the data measurement of IoT-based AQI, which is examined in this study.

1. At first, Sensors are deployed across different locations to measure pollutants and environmental parameters.
2. Data is collected and transmitted to a central server or cloud platform via wireless communication protocols such as wifi, Bluetooth and so on.
3. Real-time data processing and analysis using algorithms, often with artificial intelligence and machine learning techniques, detect patterns, trends, and anomalies.
4. An AQI value is calculated using standardized formulas or algorithms, providing a numerical or color-coded scale indicating the overall air quality level. The calculated

AQI values are presented to users through user-friendly interfaces, such as mobile apps, websites, or dashboards. In case of poor air quality conditions or when AQI thresholds are exceeded, alerts are sent to relevant stakeholders.

5. Data collected from IoT devices can be shared with other systems for decision-making and policy formulation. This system enables proactive measures to mitigate air pollution and protect public health, empowering individuals, communities, and authorities to make informed decisions regarding outdoor activities, transportation, and environmental policies.

These data measures, obtained from monitoring stations strategically located across Sivakasi, can give useful insights into air quality trends, pollutant sources, and possible health hazards connected with air pollution in the area. Continuous surveillance and evaluation of these indicators are required for successful air quality management and the deployment of mitigation strategies. Figure 3 demosntrates the present archtectures of the research.

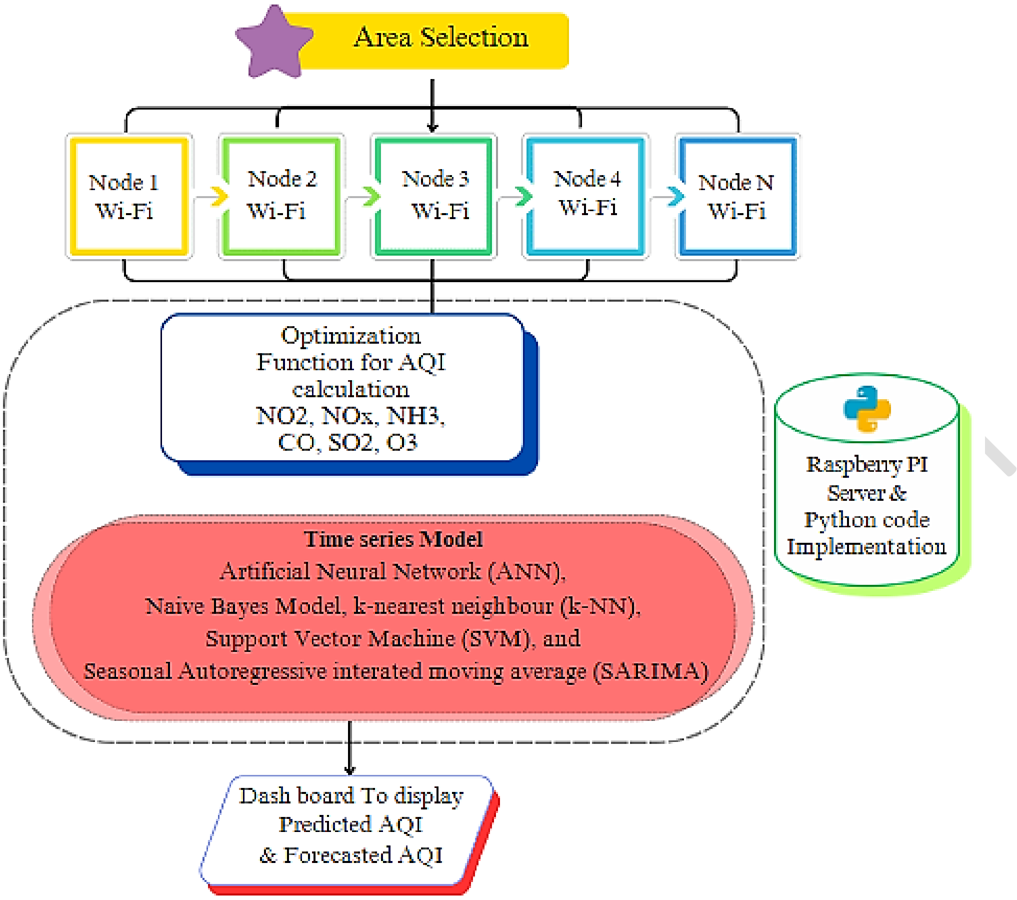


Figure 3. Proposed Model

2.3 IoT node architecture

As shown in figure 4, the IoT Node was set up using three sensors that were linked to an Arduino and calibrated using the Arduino. The data was wirelessly transferred to a Raspberry Pi 3, which works as both a local server and an edge computing device for data storage. On the Raspberry Pi, techniques for artificial intelligence and data preprocessing were implemented using Python code. To create an air pollution monitoring system using a Raspberry Pi 3 Model B, Arduino Uno board, and gas sensor modules, it needs the following hardware components: a Raspberry Pi 3 Model B board, microSD card, power supply,

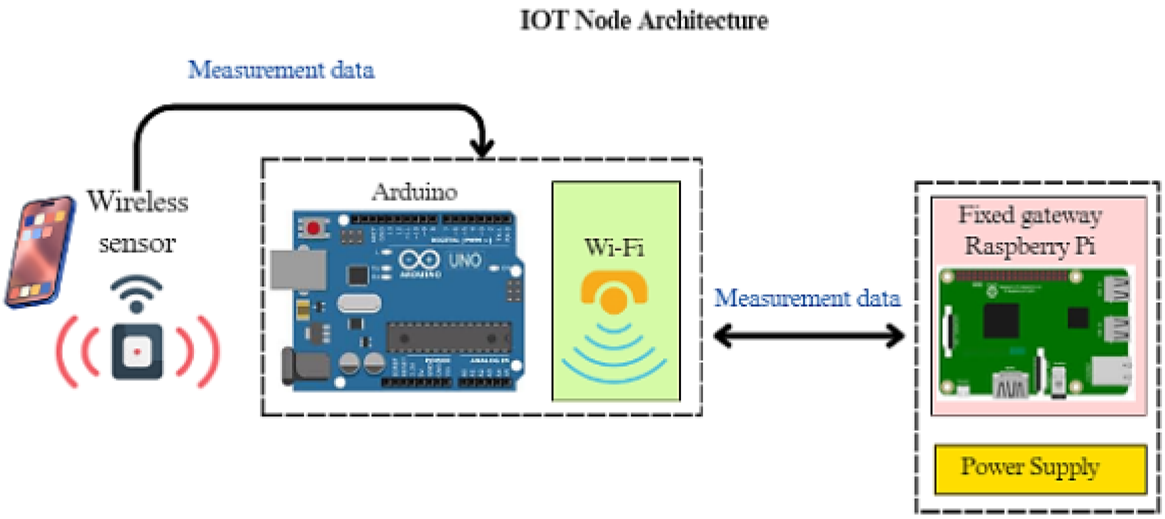


Figure 4. Iot node architecture

HDMI cable, monitor, Arduino Uno board, USB cable, and gas sensor modules. Wiring and connectors include jumper wires, board or perfboard, stable power supply, waterproof enclosure, communication interface, computer with development environment, optional components like LCD display, LEDs, resistors, and capacitors. Mounting hardware ensures secure placement of sensors and boards. Follow safety precautions and calibration routines for accurate data collection (Figure 5).

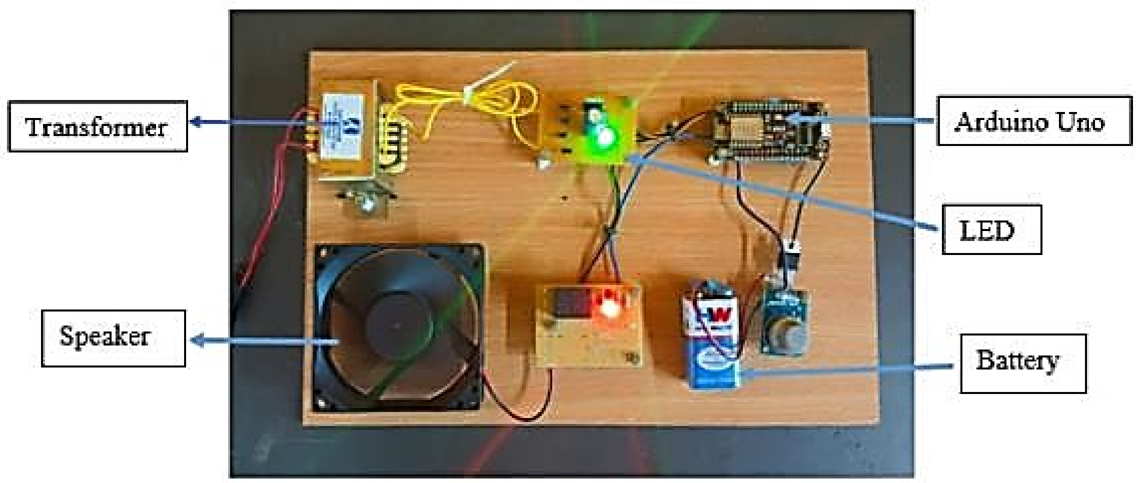


Figure 5. Hardware set up

2.4 Seasonal Autoregressive Integrated Moving Average (SARIMA)

Using time-series data, the Seasonal Autoregressive Integrated Moving-Average (SARIMA) model is a technique for identifying air pollution. It entails gathering, preparing, and examining data to look for patterns and seasonal variances. Plots of the autocorrelation function and partial autocorrelation function are used to identify the model, and methods such as maximum likelihood estimation are used to estimate it. Metrics including accuracy, coefficient, mean absolute error, and root mean square error are used to assess its predictions.

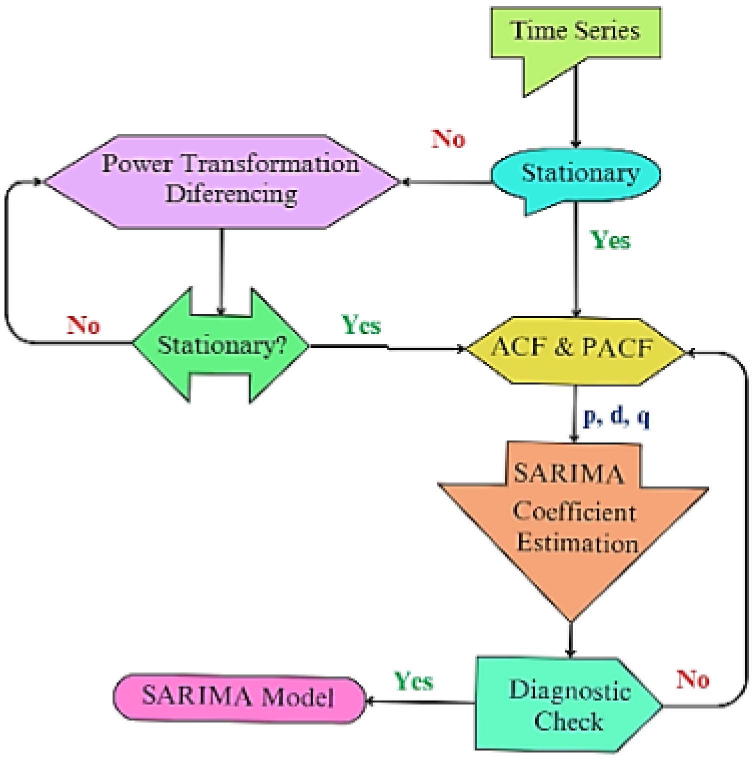


Figure 6 SARIMA Model

The trained model is used to forecast future air pollution levels, considering both short-term and long-term trends. An alerting system is set up to notify stakeholders when pollution levels exceed certain thresholds or when the forecast indicates deteriorating air quality. The model is regularly updated to improve accuracy and adapt to changing patterns. It is integrated with IoT-based air quality monitoring systems to automate data collection and analysis. This approach allows for the development of robust tools for detecting air pollution trends, making informed decisions, and implementing effective pollution control measures. Table 1 contains a record of the model parameters. Figure 6 shows the block diagram for the time-series Analysis-ARIMA model.

Table 1 SARIMA model coefficient

Model coefficients	Auto regressive model(p)	Differencing model(d)	Moving average (q)
Values	6	4	2

2.5 PM2.5 level prediction of pollutants (NO_2, NO_x, NH_3, CO, SO_2, O_3) using AI based algorithms

Predicting PM2.5 levels based on other pollutant concentrations (NO_2, NO_x, NH_3, CO, SO_2, O_3) using various AI models such as Naive Bayes, Artificial Neural Network, Support Vector Machine, k-Nearest Neighbors can be applied to this task:

3. Model Training and Evaluation:

3.1 Naive Bayes:

The process of predicting PM2.5 levels using Naive Bayes involves several steps. First, historical data on PM2.5 levels and pollutants is collected from various sources. Data is preprocessed to clean, handle missing values, and transform into a suitable format for analysis. Relevant features are selected, such as pollutant concentrations (NO_2, NO_x, NH_3, CO, SO_2, O_3), to predict PM2.5 levels. The Naive Bayes classifier is trained on the training data once the dataset is split into training and testing sets. The model is used to forecast new pollutant concentrations and its performance is assessed using testing data and metrics such as MAE and MSE.

3.2 Artificial Neural Network (ANN):

There is a step-by-step method for utilising an Artificial Neural Network (ANN) to estimate PM2.5 levels depending on pollutant concentrations. The first step in the process is gathering data, which includes past information on PM2.5 concentrations and other pollutants. The next step is data preparation, which cleans and formats the data so that it can be analysed. After that, features are chosen with the concentrations of contaminants serving as the primary features. The architecture of the ANN, comprising of the number of neurons, layers, activation functions, and other hyperparameters, is designed. The model is trained using the training data once the dataset has been split into two sets: testing and training. The testing data is used to evaluate the model's performance and can be applied to forecast new data.

3.3 Support Vector Machine

Support Vector Machine (SVM) is a popular supervised learning technique for forecasting PM2.5 levels using on pollutant concentration. To use SVM, gather historical data on PM2.5 levels and pollutants, prepare it by cleaning, handling missing values, and transforming it into a suitable format for analysis. Select relevant features from the dataset, such as pollutant concentrations, to predict PM2.5 levels. A dataset that has been split into training and testing sets is used to train the SVM model, and metrics like MSE, R-squared, and MAE are used to assess the model's performance. With fresh data, the model can forecast contaminants; its kernel and hyperparameters are tuned according to the features and performance demands of the dataset.

3.4 k-Nearest Neighbors

The k-Nearest Neighbors (k-NN) approach is a classification algorithm that detects air pollution. It entails gathering historical information on pollutant concentrations and accompanying pollution levels. The data is preprocessed to remove errors, manage missing numbers, and convert it to an appropriate format for analysis.

The training data is used to train the k-NN classifier, which is then assessed using the testing data. Accuracy, precision, recall, and F1-score are some of the most used assessment measures for classification tasks. Once trained and assessed, the model may be used to forecast new data using pollutant concentrations as input. This approach provides a step-by-step guidance to applying k-NN in air pollution detection. By following these steps and considerations, you can develop predictive models to estimate PM2.5 levels based on other pollutant concentrations using Naive Bayes, Artificial Neural Network, Support Vector Machine, and k-Nearest Neighbors algorithms.

3.5 Performance Indices

Each created model's performance measure is evaluated using the statistical criteria of RMSE, coefficient of determination (R^2), and MAE.

3.5.1 MAE

MAE is the average of expected and actual errors, determined via an equation.

$$MAE = \frac{1}{n}\sum_{l=1}^{n} |x_i - \hat{x}_l|$$

3.5.2 Root Mean Squared Error (RMSE).

The Root Mean Square Error (RMSE) is a statistical measure that compares predicted values to actual values, with a smaller number indicating better performance, and is calculated using Equation .

$$RMSE = \sqrt{\frac{1}{N}\sum_{l=1}^{N} \left(X_1 - \widehat{X_2}\right)^2}$$

3.5.3 Coefficient of determination (R^2)

The coefficient of determination (R) is a statistical metric used to evaluate a regression model's fit, reflecting the proportion of variation in the dependent variable that can be predicted from the independent variables. It is especially important in air pollution prediction, as R^2 reflects the model's ability to forecast PM2.5 levels based on pollutant concentrations.

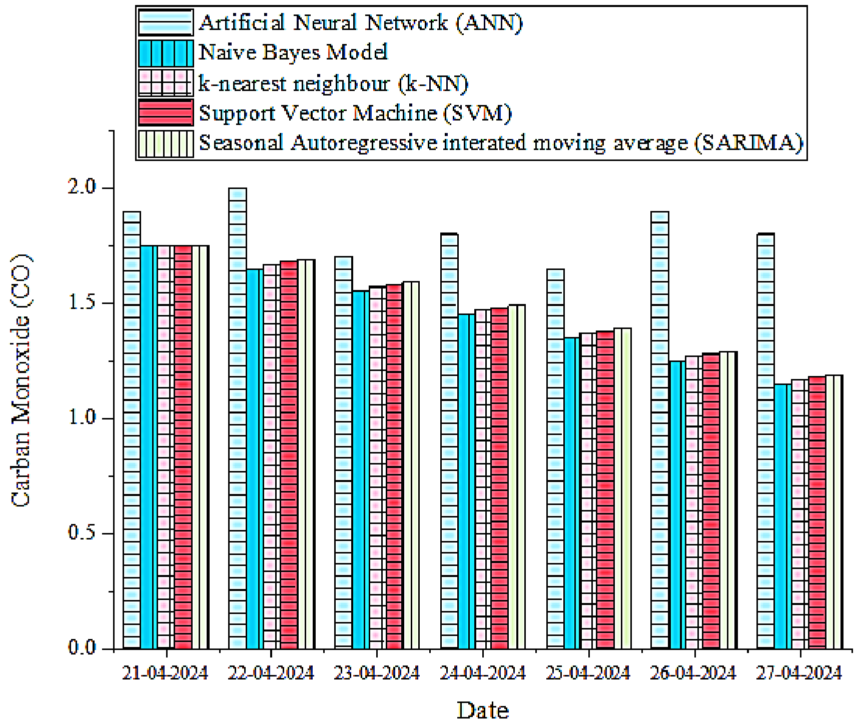
$$\mathrm{R}^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

4. Results and Discussion

The study uses Raspberry Pi for data collection and forecasting using Python programs and time series models. Artificial Neural Network (ANN), Naive Bayes Model, Support Vector Machine (SVM), k-nearest neighbor (k-NN), and Seasonal Autoregressive Interacted Moving Average (SARIMA) calculations are trained using training data. Results are anticipated for one hour every seven days. To ensure model correctness, anticipated information is compared to test data. Performance indices are used to validate the outcomes. The results reveal that the

SARIMA algorithm is more accurate at estimating CO levels. Table 2 and Figures 7 (a) and (b) show the model performance metrics for test data, together with the 95% confidence intervals.

a



b

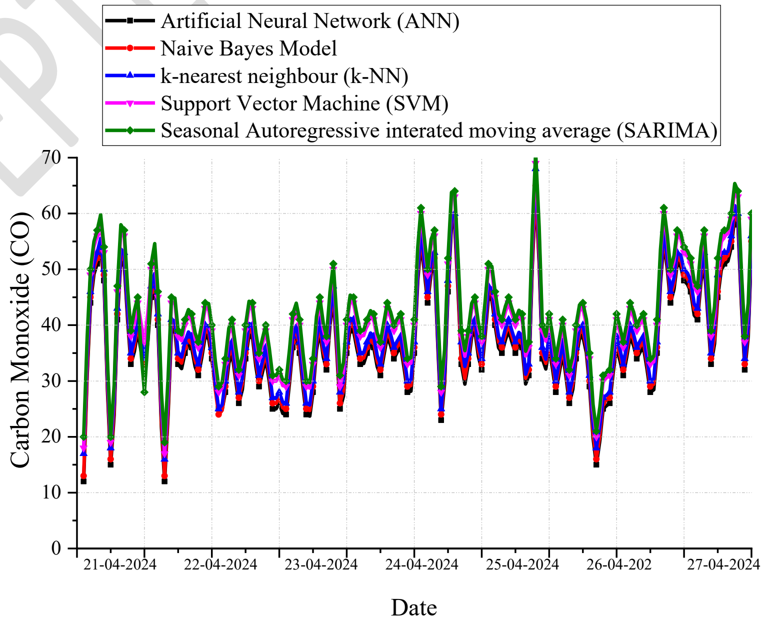


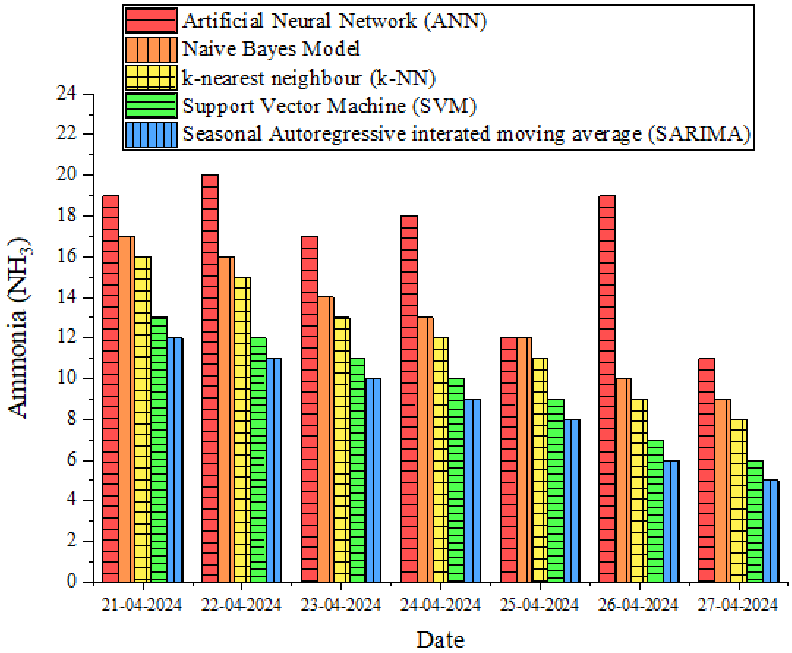
Figure 7 Prediction of CO (a) Training Data (b) Test Data

Table 2. Performance metrics of CO

Performance Indices	MAE	RMSE	Coefficient of determination (R^2)
Artificial Neural Network (ANN)	0.3343	0.4677	0.6565
Naive Bayes Model	0.3569	0.3344	0.6232
k-nearest neighbour (kNN)	0.3679	0.3659	0.6785
Support Vector Machine (SVM)	0.3869	0.3783	0.5673
Seasonal Autoregressive interated moving average (SARIMA)	0.3876	0.3862	0.5452

The study compares the performance of various AI based algorithms and time series forecasting models for predicting ammonia (NH_3) levels in the air which is displayed in figure 8 (a) and (b). Data collection involves gathering historical data, preprocessing it, and feature engineering. A larger portion of the dataset is used for training, and the dataset is split into testing and training phases. Model selection and training include k-NN, SVM, SARIMA, Naive Bayes, and neural network model design and training. For NH_3 levels, the optimal model is chosen through performance evaluation.

a



b

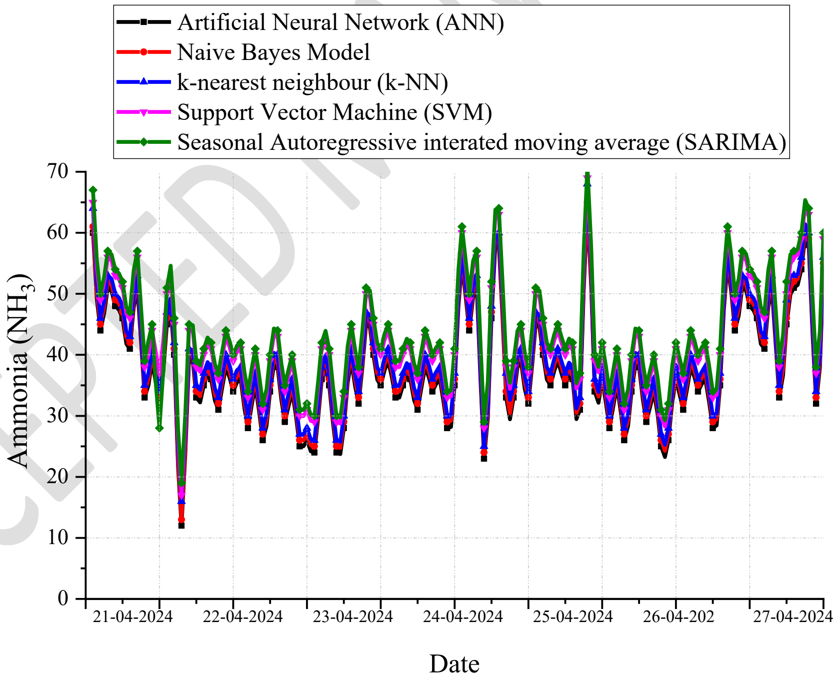


Figure 8. Prediction of NH_3 (a) Training Data (b) Test Data

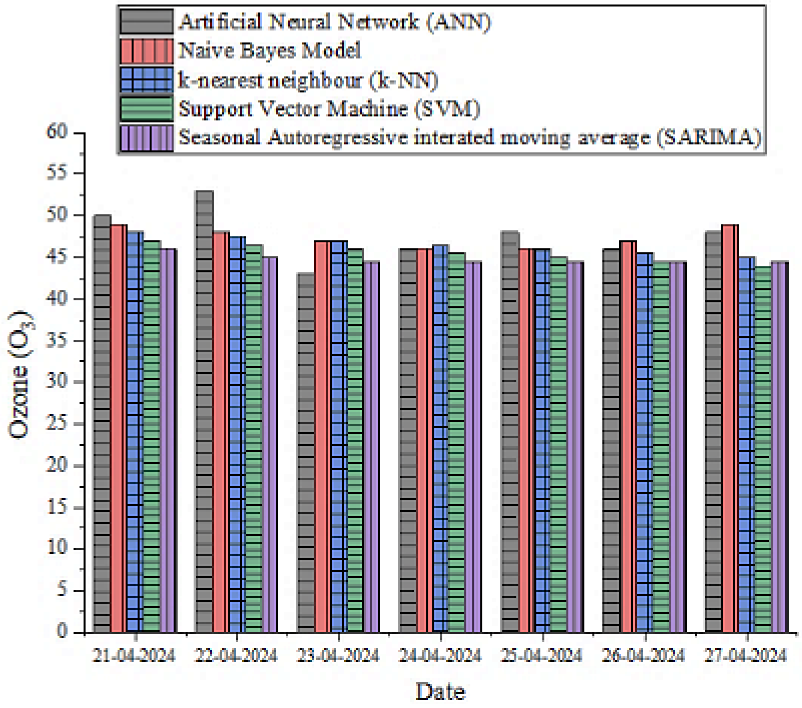
Table 3 shows that the SARIMA method has greater accuracy in forecasting NH_3, with minimum MAE, R^2, and RMSE, as well as a 95% confidence range for test data.

Table 3. Performance metrics of NH_3

Algorithm	MAE	RMSE	Coefficient of determination (R^2)
Artificial Neural Network (ANN)	3.4539	3.6465	3.4567
Naive Bayes Model	1.4545	1.8342	1.8642
k-nearest neighbour (kNN)	1.4631	1.9753	1.7867
Support Vector Machine (SVM)	1.3426	1.3446	1.8986
Seasonal Autoregressive interated moving average (SARIMA)	1.3467	1.7865	1.4758

Figures 9 (a) and (b) illustrates a comparison of the performance of O_3 forecasting throughout the training and testing periods. The study collects data on Raspberry Pi and forecasts it using Python scripts and time series models. Three different detector data is separated into training and test data. AI-based algorithms and Seasonal Autoregressive Interacted Moving Average (SARIMA) computations are trained using training data. The results are anticipated for one hour every seven days from 21-04-2024 to 27-04-2024. The predicted data is compared to test data to ensure model correctness. Performance indicators are computed to validate the findings. The results reveal that the SARIMA algorithm is more accurate for O_3 forecasting.

a



b

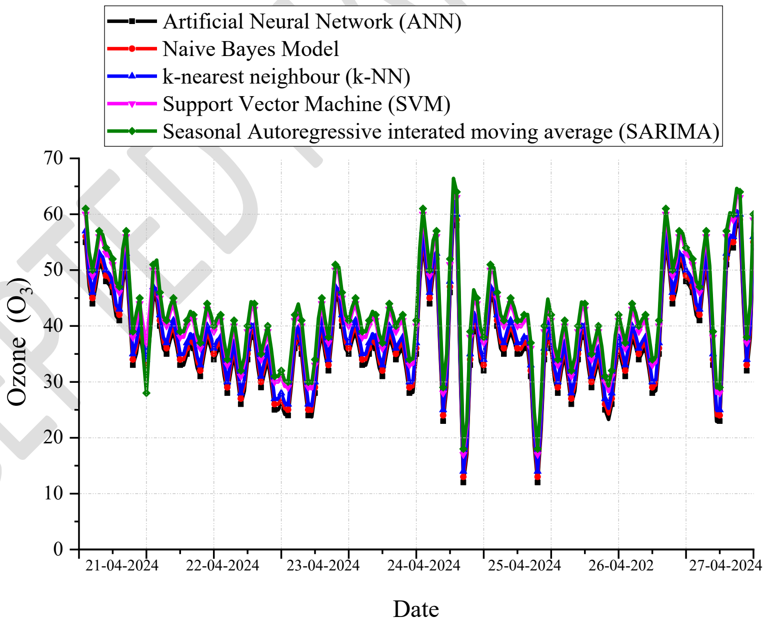


Figure 9 Prediction of O_3 (a) Training Data (b) Test Data

The collected data show that the SARIMA algorithm is more accurate in forecasting O_3. Table 4 displays the various models' performance metrics for the O_3 test data. The confidence interval is 95%.

Table 4 Performance metrics of O_3

Performance Indices	MAE	RMSE	Coefficient of determination (R^2)
ANN	1.5783	1.9759	2.3521
Naive Bayes Model	1.5045	1.8454	2.2758
kNN	1.4755	1.8075	2.2276
SVM	1.4755	1.7565	2.1875
Seasonal Autoregressive interated moving average (SARIMA)	1.3854	1.7784	2.1758

4.1 NO_2

The study compares the performance of AI-based algorithms in predicting NO_2 levels in air pollution during training and testing phases. These algorithms use machine learning and statistical techniques to gather historical data and analyze it to predict future ozone concentrations. The data is then analyzed to understand its distribution and relationships between variables. The most effective SARIMA model is selected for real-time NO_2 forecasts and integrated into an operational system for ongoing monitoring and decision-making, providing timely and accurate information for air quality management. The prediction of training data of NO_2 was shown in figure 10.

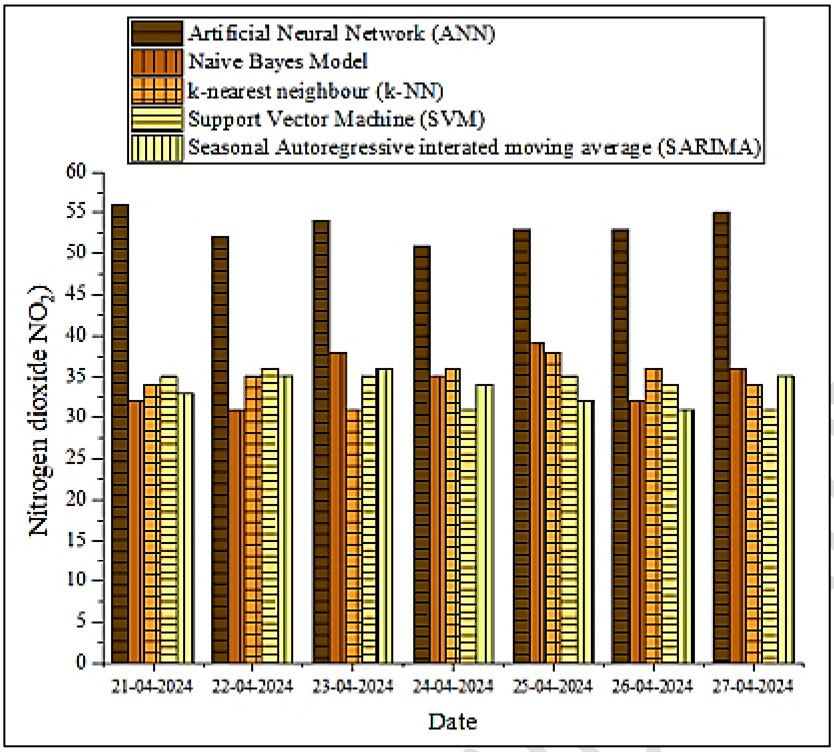


Figure 10 Prediction of training data of NO_2

The SARIMA algorithm is found to have superior accuracy in forecasting NO_2, with a confidence interval of 86%, as indicated by the performance indices in Table 5.

Table 5 Performance metrics of NO_2

Performance Indices	MAE	RMSE	Coefficient of determination (R^2)
Artificial Neural Network (ANN)	2.6546	1.9356	3.2654
Naive Bayes Model	2.2184	1.8645	3.0645
k-nearest neighbour (kNN)	2.1545	1.7466	2.9892
Support Vector Machine (SVM)	2.0548	1.6659	2.8464
Seasonal Autoregressive interated moving average (SARIMA)	2.0259	1.5268	2.5165

4.2 NOx,

Figures 11 demonstrate the performance analysis of NO_X forecasting during training phases. Nitrogen oxides (NO_x) are reactive gases from combustion processes in vehicles, industrial facilities, and power plants, detected using SSRIMA and AI algorithms. Controlling NO_x emissions is crucial for improving air quality.

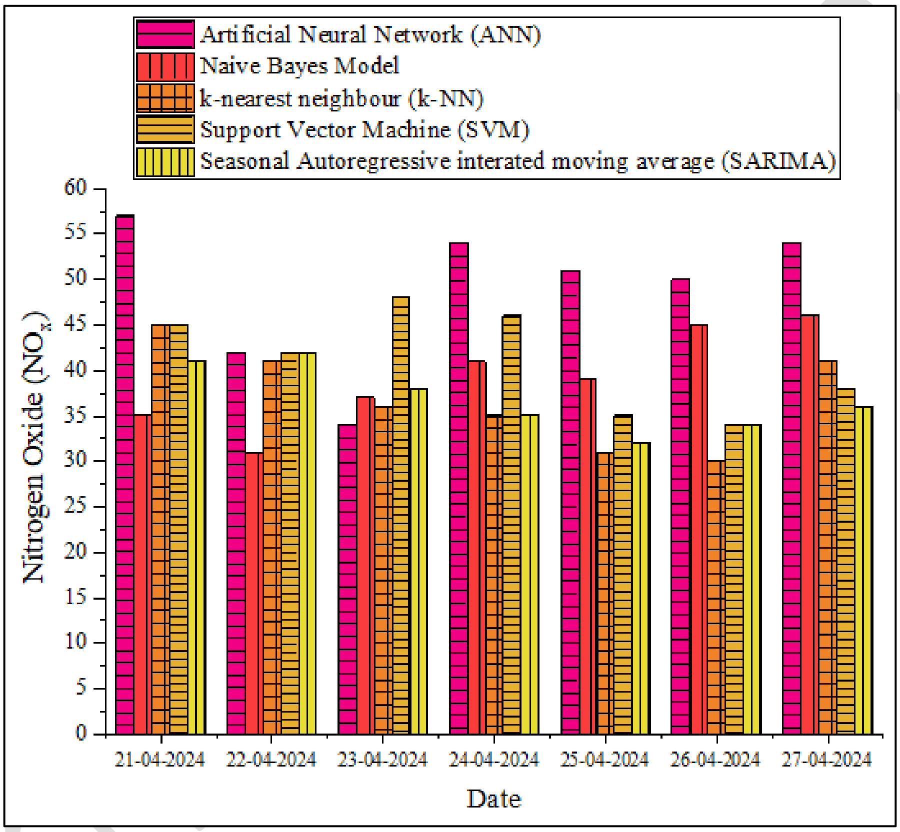


Figure 11 Prediction of NO_X Test Data

The SARIMA algorithm is found to have superior accuracy in forecasting NO_X, with a confidence interval of 87%, as indicated by the performance indices in Table 6.

Table 6 Performance metrics of NO_X

Performance Indices	MAE	RMSE	Coefficient of determination (R^2)
Artificial Neural Network (ANN)	3.9496	3.0651	3.0549
Naive Bayes Model	3.5164	2.8161	2.9169
k-nearest neighbour (k-NN)	2.6292	2.6469	2.6695
Support Vector Machine (SVM)	2.15489	2.4265	2.1989
Seasonal Autoregressive interated moving average (SARIMA)	1.9469	1.9466	1.9466

4.3 SO_2

SO_2 data metrics are essential for assessing air quality, identifying pollution sources, and evaluating regulatory compliance. Key metrics include concentration (ppb or µg/m³), temporal trends (changes over time), spatial distribution (locations within a region), and pollution episodes (short-term spikes above regulatory standards). These metrics help identify hotspots, pollution sources, and areas with elevated levels, guiding mitigation efforts and regulatory enforcement. Monitoring these metrics helps assess the severity of air pollution events and issue public health advisories or alerts. The training phase of SO_2 prediction was shown in figure 12.

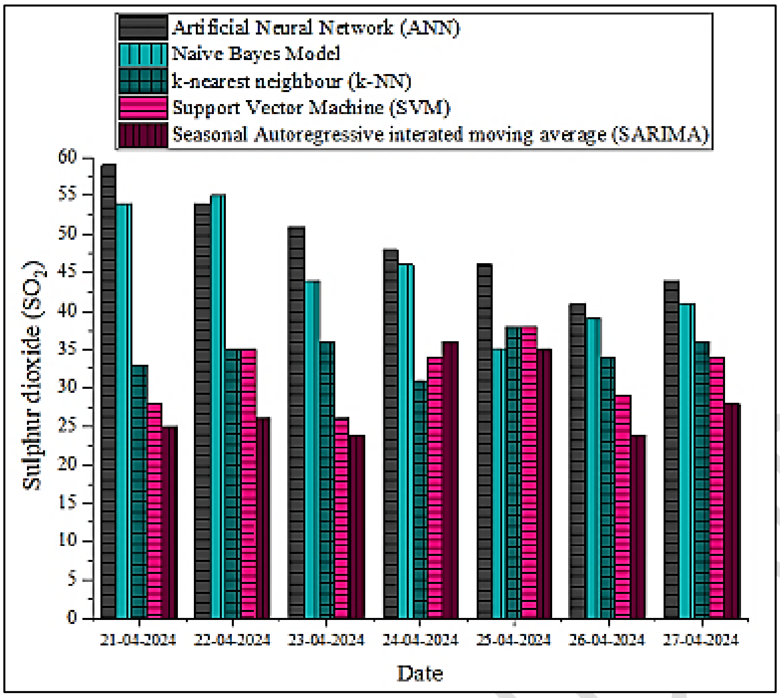


Figure 12 Prediction of SO_2 Test Data

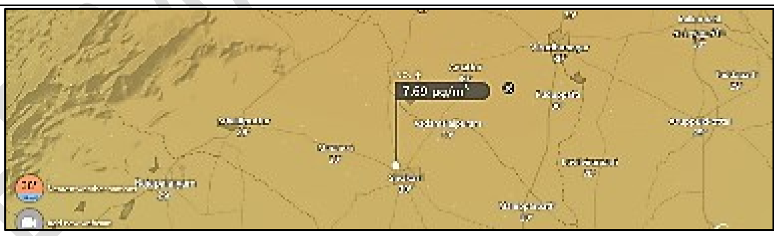
The SARIMA algorithm is found to have superior accuracy in forecasting SO_2, as indicated by the performance indices in Table 7 with a confidence interval of 80%.

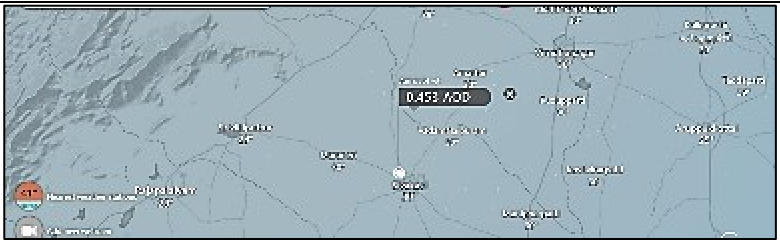
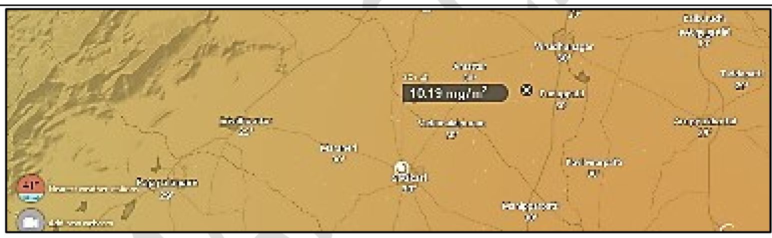
Table 7 Performance metrics of SO_2

Performance Indices	MAE	RMSE	Coefficient of determination (R^2)
Artificial Neural Network (ANN)	1.5654	1.9416	2.9652
Naive Bayes Model	1.4265	1.5164	2.5165
k-nearest neighbour (k-NN)	1.2194	1.0265	2.4649
Support Vector Machine (SVM)	1.0564	0.9846	2.0659
Seasonal Autoregressive interated moving average (SARIMA)	0.9126	0.9466	1.8465

4.4 Air pollutants monitoring in online

Online monitoring of air pollutants involves real-time measurement and analysis of air quality parameters using a network of stations equipped with sophisticated sensors and instruments. These stations are strategically located in urban, industrial, and residential areas, and each station is equipped with sensors for specific pollutants. Data collected is transmitted to a central database or monitoring center, allowing immediate access to information and timely response to pollution events. Regular calibration and maintenance of monitoring equipment ensure data accuracy and reliability, displayed on the website for online monitoring of air pollutants such as NO_2, PM2.5, aerosol, ozone layer, SO_2, CO and dust mass from anywhere in the world and it is clearly illustrated in figure 13.

NO_2	
PM 2.5	

Aerosol	
Ozone Layer	
SO_2	
Surface Ozone	
CO Concentration	

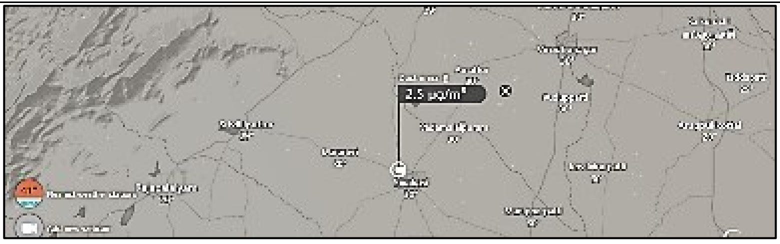
Dust Mass	

Figure 13. Website that monitors air quality

Data analysis and visualization tools are used to present the data in an easily understandable format. Online monitoring systems can be integrated with alert systems to notify authorities and the public about air quality issues. Figure 14 described about the air pollution level in Sivakasi district, TamilNadu, India.

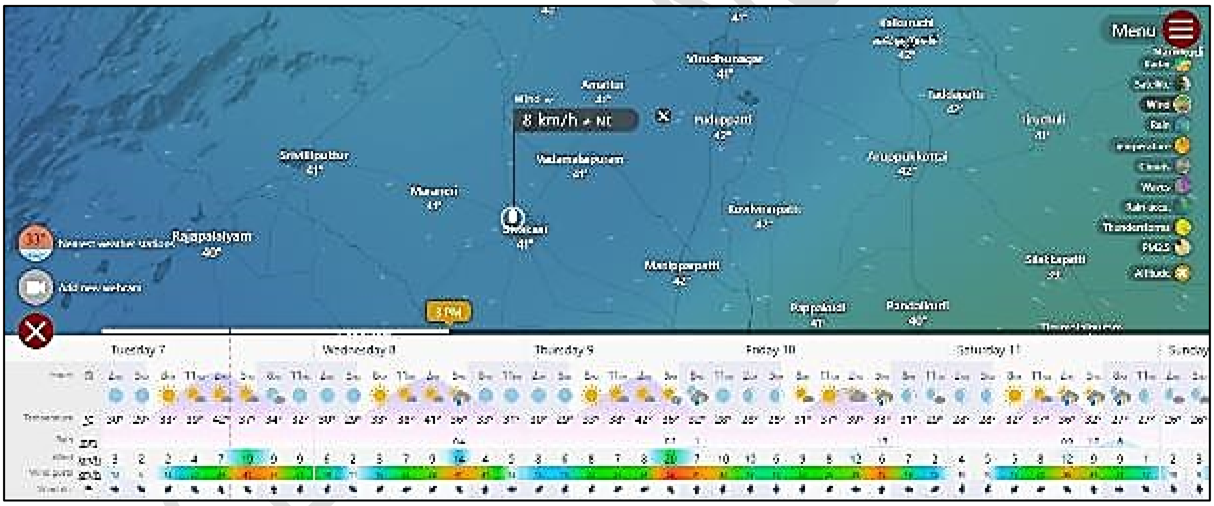


Figure 14. View of location selection

5. Conclusion

An IoT based AI technique was developed to provide realistic real-time air quality tracking and monitoring from any location. An IoT-based hardware prototype was created to test the functioning of the suggested approach. The acquired data is evaluated and measured using the built AI based and time series models. The numerous time-series models are developed and used to forecast particular air pollutants such as NO_2, NO_x, NH_3, CO, SO_2, and O_3. it was collected in Sivakasi, Tamilnadu India. First, the models are trained using training data and then tested using previously unknown test data. The values for the next four hours were predicted with an 87% confidence interval by all of the AI methods that were used, including Artificial Neural Network (ANN), Naive Bayes Model, k-nearest neighbour (k-NN), and Support Vector Machine (SVM). Using test data, it was found that the performance indices of the selected models were sufficient. The findings show that, with a 95% accuracy level, SARIMA is more accurate than the other three techniques in all case studies and has the lowest MAE, coefficient of determination, and RMSE values. As a result, the SARIMA model was identified as a more appropriate forecasting approach for predicting future air pollution values.

Evaluating the Air Pollution Index (API) with the aid of technology represents a significant advancement in understanding and managing environmental health. Utilizing sophisticated technologies such as Internet of Things (IoT) sensor arrays and artificial intelligence (AI), modern systems can monitor a wide range of pollutants in real-time. These sensors collect continuous data on air quality parameters, which are then transmitted to centralized servers for comprehensive analysis. AI algorithms, such as neural networks and machine learning models, process this data to generate accurate and timely assessments of the API. By leveraging technology, API evaluations become more precise and responsive to dynamic environmental conditions. This capability enables authorities and communities to make informed decisions regarding public health interventions, urban planning, and pollution control measures. Moreover, the integration of IoT and AI reduces reliance on traditional monitoring methods that may be slower or less adaptable to rapid changes in air quality. As a result, technology-driven API evaluations not only enhance our understanding of air pollution but also empower proactive strategies to mitigate its adverse effects on human health and the environment.

Abbreviations and Nomenclature

ANN	-	Artificial Neural Network	R^2	-	Coefficient Of Determination
k-NN	-	k-nearest neighbour	O_3	-	Ozone
SVM	-	Support Vector Machine	NO_2	-	Nitrogen Dioxide
SARIMA	-	Seasonal Autoregressive Interated Moving Average	CO	-	Carbon Monoxide
MAE	-	Mean Absolute Error	NH_3	-	Ammonia
RMSE	-	Root Mean Square Error	NO_x	-	Nitrogen Oxide
WHO	-	World Health Organisation	SO_2	-	Sulphur dioxide
LSTM	-	Long Short Term Memory			
IOT	-	Internet of Things			
ENN	-	Elman Neural Network			
AAA	-	Artificial Algae Algorithms			
PM	-	Particulate Matter			
AQI	-	Air Quality Index			

Competing interests

The authors declare no conflicts of interest.

Authors contribution

R.Nirmalan conceptualized the study, designed the research methodology, and implementation of the proposed work. **P. Thendral** conducted data collection, preprocessing, and model training. **S.Karukuzhali** performed data analysis, interpretation and validation of results. All authors contributed to the drafting and revision of the manuscript and approved the final version for submission.

Funding

This research study is sponsored by the institution

REFERENCES

Thambavani, D. S., Rajeswari, G., & Vijaya, V. (2009). Effect of air pollution on plant metabolism at Sivakasi, Tamil Nadu. *Journal of Ecobiology*, ***24***, 153-157.

Manikandan, M., & Abdullah, S. S. (2016). Causes and effects of air pollution-A study with special reference to Sivakasi Taluk of Virudhunagar district, Tamil Nadu. *Wide Spectrum*, ***4***, 34-48.

Raj, V. A., Priya, R. M. P., & Meenakshi, V. (2017). Air pollution monitoring in urban area. *Int. J. Electron. Commun. Eng*.

Xiaojun, C., Xianpeng, L., & Peng, X. (2015, January). IOT-based air pollution monitoring and forecasting system. In *2015 international conference on computer and computational sciences (ICCCS)* (pp. 257-260). IEEE.

Ayele, T. W., & Mehta, R. (2018, April). Air pollution monitoring and prediction using IoT. In *2018 second international conference on inventive communication and computational technologies (ICICCT)* (pp. 1741-1745). IEEE.

Zhao, Yu-Lin, Jiali Tang, Han-Pang Huang, Ze Wang, Tse-Lun Chen, Chih-Wei Chiang, and Pen-Chi Chiang. (2020) "Development of iot technologies for air pollution prevention and improvement." *Aerosol and Air Quality Research* **20**, 2874-2888.

Asha, P., L. B. T. J. R. R. G. S. Natrayan, B. T. Geetha, J. Rene Beulah, R. Sumathy, G. Varalakshmi, and S. Neelakandan. (2022) "IoT enabled environmental toxicology for air pollution monitoring using AI techniques." *Environmental research* **205**: 112574.

Dhingra, Swati, Rajasekhara Babu Madda, Amir H. Gandomi, Rizwan Patan, and Mahmoud Daneshmand. (2019) "Internet of Things mobile–air pollution monitoring system (IoT-Mobair)." *IEEE Internet of Things Journal* **6**: 5577-5584.

Idrees, Z., Zou, Z., & Zheng, L. (2018). Edge computing based IoT architecture for low cost air pollution monitoring systems: a comprehensive system analysis, design considerations & development. *Sensors*, ***18***, 3021.

Geetha Mani, Joshi Kumar Viswanadhapalli, P Sriramalakshmi. "AI powered IoT based RealTime Air Pollution Monitoring and Forecasting" , Journal of Physics: Conference Series, 2021

Subramaniam, Shankar, Naveenkumar Raju, Abbas Ganesan, Nithyaprakash Rajavel, Maheswari Chenniappan, Chander Prakash, Alokesh Pramanik, Animesh Kumar Basak, and Saurav Dixit. (2022 "Artificial intelligence technologies for forecasting air pollution and human health: A narrative review." *Sustainability* **14:** 9951.

Kothandaraman, D., N. Praveena, K. Varadarajkumar, B. Madhav Rao, Dharmesh Dhabliya, Shivaprasad Satla, and Worku Abera. (2022) "Intelligent forecasting of air quality and pollution prediction using machine learning." *Adsorption Science & Technology* 2022.

Cordova, Chardin Hoyos, Manuel Niño Lopez Portocarrero, Rodrigo Salas, Romina Torres, Paulo Canas Rodrigues, and Javier Linkolk López-Gonzales. (2021) "Air quality assessment and pollution forecasting using artificial neural networks in Metropolitan Lima-Peru." *Scientific Reports* **11**: 24232.

Schürholz, Daniel, Sylvain Kubler, and Arkady Zaslavsky. (2020) "Artificial intelligence-enabled context-aware air quality prediction for smart cities." *Journal of Cleaner Production* **271**: 121941.

Abd Rahman, Nur Haizum, Muhammad Hisyam Lee, Mohd Talib Latif, and S. J. J. T. Suhartono. (2013) "Forecasting of air pollution index with artificial neural network." *Jurnal Teknologi* **63**.

Asghari, M., and H. Nematzadeh. "Predicting air pollution in Tehran: Genetic algorithm and back propagation neural network." (2016) *Journal of AI and Data Mining* **4**: 49-54.

Yarragunta, SriramKrishna, and Mohammed Abdul Nabi. "Prediction of air pollutants using supervised machine learning. (2021)" In *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 1633-1640. IEEE.

Ren, Lulu, Farun An, Meng Su, and Jiying Liu. (2022) "Exposure assessment of traffic-related air pollution based on CFD and BP neural network and Artificial Intelligence prediction of optimal route in an urban area." *Buildings* **12**: 1227.

Li, Victor OK, Jacqueline CK Lam, Yang Han, and Kenyon Chow. (2021) "A big data and artificial intelligence framework for smart and personalized air pollution monitoring and health management in Hong Kong." *Environmental Science & Policy* **124**: 441-450.

Saini, J., Dutta, M., Marques, G., & Halgamuge, M. N. (Eds.). (2022). *Integrating IoT and AI for Indoor Air Quality Assessment*. Springer International Publishing.

Tiwary, A., & Williams, I. (2018). *Air pollution: measurement, modelling and mitigation*. Crc Press.

Gupta, A. K., Shukla, A. K., & Srivastava, R. (2020). Seasonal ARIMA models in forecasting of air pollution time series data. ***International Journal of Environmental Science and Technology*****, 17(8), 3695-3706.**

Roy, S., Tran, T. A., & Natarajan, K. (Eds.). (2023). *Recent Advancement of IoT Devices in Pollution Control and Health Applications*. Elsevier.