Web-based Air pollution monitoring system

Under Graduate Final Year Project Report

Abdul Basit	FA17-BCE-062
Aisha Khan	FA17-BCE-081
Arooj Shafique	FA17-BCE-083
Nabia Imran	FA17-BCE-004

Project Advisor: Usman Rafique

Fall 2020

COMSATS UNIVERSITY ISLAMABAD, LAHORE CAMPUS, PAKISTAN

Submission Form for FYP Report

PROJECT ID		NUMBER OF MEMBERS	4

TITLE	Web-Based Air pollution monitoring system

SUPERVISOR NAME	Usman Rafique

MEMBER NAME	REG. NO.	EMAIL ADDRESS
Abdul Basit	FA17-BCE-062	Fa17-bce-062@cuilahore.edu.pk
Aisha Khan	FA17-BCE-081	Fa17-bce-081@cuilahore.edu.pk
Arooj Shafique	FA17-BCE-083	Fa17-bce-083@cuilahore.edu.pk
Nabia Imran	FA17-BCE-004	Fa17-bce-004@cuilahore.edu.pk

CHECKLIST:

Number of pages in this report	32
I/We have enclosed the soft-copy of this document along-with the codes and scripts created by myself/ourselves	**YES / NO**
My/Our supervisor has attested the attached document	**YES / NO**
I/We confirm to state that this project is free from any type of plagiarism and misuse of copyrighted material	**YES / NO**

MEMBERS' SIGNATURES
Abdul Basit
Aisha Khan
Arooj Shafique
Nabia Imran

Supervisor's Signature

Note 1: This paper must be signed by your supervisor
Note 2: The soft-copies of your project report, source codes, schematics, and executables should be delivered in a CD

This work, entitled "**Web-Based Air pollution monitoring system**"

has been approved to fulfil partial requirements for the award of

BS in Computer Engineering to

Abdul Basit	FA17-BCE-062
Aisha Khan	FA17-BCE-081
Arooj Shafique	FA17-BCE-083
Nabia Imran	FA17-BCE-004

Fall 2020

External Examiner:

Head of Department:

Department of Electrical and Computer Engineering

COMSATS UNIVERSITY ISLAMABAD

LAHORE CAMPUS– PAKISTAN

Declaration

"No portion of the work referred in this report has been submitted in fulfilment of another degree or qualification for any other institute or university".

MEMBERS' SIGNATURES
Abdul Basit
Aisha Khan
Arooj Shafique
Nabia Imran

Abstract

With the increase in industrialization and urbanization, there is an exceptional surge in air pollution. Currently, we are unaware of trends in air pollution in Pakistan and most importantly SMOG crises faced by Lahore (punjab, Pakistan) every winter for the last 4 years, the origin of SMOG has not yet been found. Air pollution has devastating effects on public health and the economy, especially in urban areas, according to WHO, killing more than 4.2 billion people each year. The objective is to implement an air quality monitoring system using a wireless sensor network (WSN) which can be deployed in multiple areas for the AQI measurements. The proposed system will help understand the magnitude of this issue, so people can become aware of the severity of the problem. We propose a monitoring system which is according to our country standards. The focus of this project is to develop modular devices that will help us to collect data regarding air quality. The proposed system will collect real-time data on air pollution, it will allow us to calculate the AQI and predict smog level with the help of appropriate machine learning model. Through this system we will be able to update the general public of the air quality conditions as well as the future smog levels for them to take necessary precautions.

Table of Contents

Table of Figures

Chapter 1

1 Introduction

1.1 Overview

One of the crucial environmental concerns of today's world is air pollution. Apart from degrading the natural resources and environment, it severely affects human health. The short and long-term effects range from respiratory diseases, headache, anxiety, fatigue, irritation of the eyes, nose, and throat – to permanent health effects like asthma, cancer, bronchitis, heart and lung diseases, pneumonia, nervous system damage. According to the estimates of the Global burden of Disease Study, approximately 4.2 million premature deaths caused from outdoor air pollution globally, in 2015.

Nowadays the air over the world has become contaminated with gases caused by industries and automobiles. The number of death rates due to their pollution is growing rapidly in metropolitan areas. In metropolitan areas the adverse health impacts of air pollution exposure are extremely condemning, which are population hubs, also flashpoints for pollution discharge.

With the rapid development of motor vehicle activities and industrializing operations, Pakistan is facing a high level of urban air pollution, as well as other adverse socio-economic, health, environmental, and welfare impacts. According to the estimates of WHO Global Health Observatory, in Pakistan, about 30 deaths per 100,000 occur due to indoor air pollution, while 25 deaths per 100,000 occur due to outdoor air pollution. [1] An average adult inhales and exhales approximately 7 to 8 liters (1/4th of a cubic foot) of air every minute. It totals 11,000 liters of air (approximately) per day. That is why it is salient to determine the standard of air an individual is inhaling, especially for a person who is facing some respiratory problems. [2]

1.2 Air Quality Index (AQI)

Air quality refers to the condition of the quality of air surrounding us. The release of a high level of chemical pollutants into the atmosphere due to the flaming of fossil fuels and the manufacturing processes has a drastic impact on air quality. Depending on human activities and some natural phenomena, there can be several air pollutants. The AQI is based on the measurement of these pollutants present in the air. The pollutants that are accounted for the majority of air pollution worldwide are particulate matter ($PM_{2.5}$ and PM_{10}), Ozone (O_3), Nitrogen Dioxide (NO_2), Sulfur Dioxide (SO_2), and Carbon Monoxide (CO) .

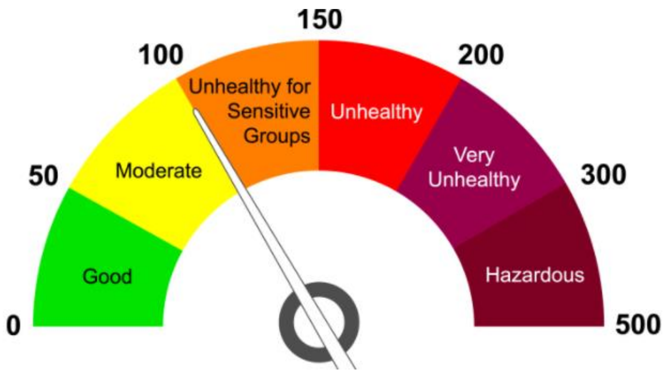

Figure 1: Standard Air Quality Index

- **Good:** Air quality is reasonable in this range.
- **Moderate:** Air quality is suitable although it may be distress for people who have a unique sensitivity to air pollution.
- **Unhealthy for Sensitive Group:** Adverse health effects may be experienced by people with lung or heart diseases. Children and elders may also be affected. The overall public may not be affected.
- **Unhealthy:** Adverse health effects maybe are experienced by everyone and members of the delicate group may face more severe health issues.
- **Very Unhealthy:** The complete population is likely to be disturbed. Any outdoor activities should be avoided.
- **Hazardous:** All outdoor activities must be avoided as severe health effects may be experienced by the entire population.

1.3 Health Hazards due to Air Pollution

One of the most severe urban air pollutions exists in Pakistan, capable of damaging human health significantly. According to WHO, air pollution is liable for an approximated 4.2 million deaths per year,

- 29% lung cancer issues,
- 43% of chronic obstructive pulmonary illness,
- 25% of ischemic cardiac disease,
- 17% acute respiratory issues,

- 24% of all deaths from sudden attacks [3].

Pakistan ranks 2nd in the world's AQI ranking. According to the report circulated by the medical journal, Lancet in 2015 around 22 percent or more than 310,000 deaths in Pakistan are produced mainly by air pollution [5]. Outdoor air pollution accounts for more than 22,000 deaths while indoor air pollution accounts for more than 40 million cases of acute respiratory infections and 28,000 deaths per year [4].

Figure 2: Pakistan ranking 2nd in World's AQI Ranking

1.4 Lahore Smog

Punjab is facing a public health emergency of smog Lahore being in the most critical situation. Lahore is the second-largest city and the most pullulated city in Pakistan. Recently it has been covered in a heavy blanket of smog [6]. Starting from October to February, it worsens the air quality in Lahore imposing a high threat to the lives of more than **12 million** people. Rapid shoot in multiple health-related problems due to Smog as it is hazardous to health is leading to the rise of concerns about the long-term devastating effects on public well-being.

An increase in vehicles, deforestation, urbanization, and industrialization has contributed to the rise of this emergency. [7]

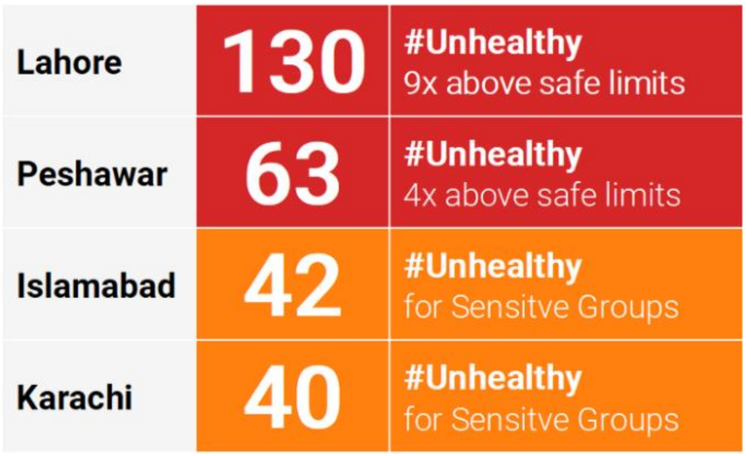

Figure 3: AQI in major cities of Pakistan

1.5 Problem Statement

Air pollution is one of the most common environmental problems that cannot be ignored. Inhaling polluted air for a long-time harm human health in several ways. The evaluation and computation of the air pollution level are set up on criteria that are present in every state in the world. Established air quality monitoring methods are costly. So, a cheap, remote, and real-time air pollution monitoring system is a need of the hour, that gives more accurate data results and requires less human interaction in hazardous regions.

According to the Environment Protection Department of the Government of Punjab (EPD) [8], the data is collected once a day with 4 devices that are recently working for pollution monitoring, but they do not give many sufficient and reliable results. This data cannot be analyzed to detect a problem. Our monitoring system will test the air quality of different areas with authentic results. This will allow users to realize which area of the town has more or less contamination so they can choose their paths accordingly.

1.6 Objectives & Goals

- To detect the level of Air pollution in a specific outdoor area and calculate the AQI of the area.
- To detect the smog in the area and its origin. And monitor it on a web application.
- To predict the level of smog in a particular area using the machine learning model.

1.7 Deliverables

- Web-based application for monitoring outdoor air pollution.
- Machine Learning model for prediction of Smog level.

- Modular devices that will help us to mould our devices according to the requirements of the area without affecting the rest of the system.

1.8 Engineering Design

We propose a solution in the form of a device that is capable of being interfaced with multiple sensors that can be utilized for observing the air pollutants such as CO, NO_x. and dust particles. The information gathered from the sensors is delivered to the microcontroller and then to the cloud with the support of a Wi-Fi module where the data is processed, stored on a database and shown on the website. Statistical graphs and trends observation will be updated based on data. This collected data is also used to predict the future level of the AQI of the area with the help of which the smog level will be predicted. The last section deals with the end-user layer i.e. website. The graphs and data will be updated on our website. The collected data will be visible to the public. This allows the citizens to take suitable precautions when needed.

1.8.1 Hardware Requirements

- **Arduino:** It is an open-source micro-controller based on the ATmega328P microcontroller. Arduino will be our module controller. It will be interfaced to the sensor. Arduino has 6 10-bits ADC channels through which the analog outputs from each of the sensors will be converted to digital values for processing.

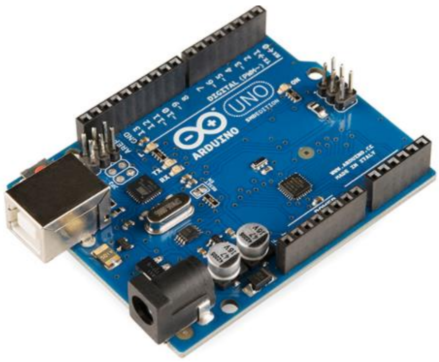

Figure 4: Arduino UNO

- **ESP32:** It is a low-cost SoC that works with the Arduino IDE. It has a built-in Wi-Fi and Bluetooth module. It will be used as the main MCU to collect data from the all modules attached and send data to the cloud for processing.

Figure 5: ESP32 Module

- **MQ-7:** It is used as a Carbon monoxide (CO) sensor. It works on 5 volts and the output is the analog voltage which changes in accordance to the concentration of the gas. Some processing is done through software for the calculation of gas concentration in PPM. Before starting the gas concentration measurement, the sensor needs to be calibrated to work efficiently.

Figure 6: MQ-7 Gas Sensor Module

- **MQ-131:** It is used as an Ozone(O_3) sensor. It works on 5 volts and the output is the analog voltage which changes in accordance to the concentration of the gas. Some processing is done through software for the calculation of gas concentration in PPM. Before starting the gas concentration measurement, the sensor needs to be calibrated to work efficiently.

Figure 7: MQ-131 Gas Sensor Module

- **MQ-135:** It is used as a Nitrogen oxide (NO_x) sensor. It works on 5 volts and the output is the analog voltage which changes in accordance to the concentration of the gas.

Some processing is done through software for the calculation of gas concentration in PPM. Before starting the gas concentration measurement, the sensor needs to be calibrated to work efficiently.

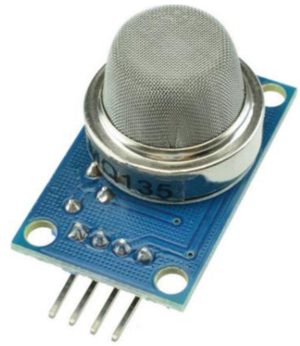

Figure 8: MQ-135 Gas Sensor Module

- **DHT11:** It is a combined temperature and humidity sensor with a digital serial output through a single wire. The operating voltage ranges from 3.5 volts-5.5 volts. Temperature ranging from 0 °C to 50 °C and humidity ranging from 20% to 50% can be measured using this sensor with an accuracy of ±1°C/±1%. The data is output in the form of 40 bits in a serial manner with the first byte being the high 8 bits of humidity value, second as the low 8 bits of humidity value, third as the high 8 bits of temperature value, fourth as low 8 bits of temperature value and the fifth as the 8 parity bits. To get output from the module the I/O pin of the interfaced microcontroller is first held low and then high.

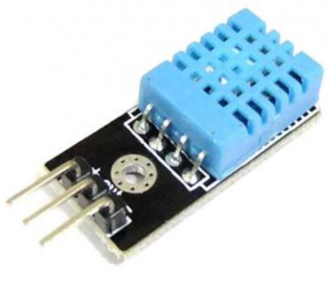

Figure 9: DHT11 Sensor Module.

- **PMS3003:** It is a Particle Concentration Sensor that is used to find the number of suspended materials in the air, i.e., the concentration of the particles, and gives output in digital interface**.** It works at 5 volts in contains fan which is used to suck air into the chamber where it is analyzed using an optical sensor and a diode.

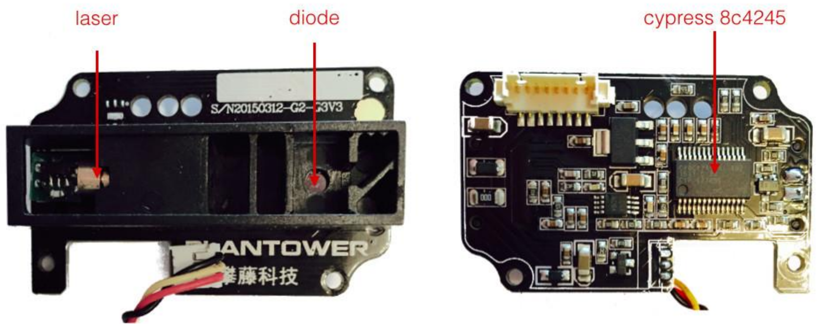

Figure 10: PMS3003 sensor Module.

1.8.1.1 Hardware Components Price in Pakistan

Hardware Components	Price (Rs)
Arduino	730
ESP32	850
MQ-7	230
MQ-131	N/A
MQ-135	230
DHT11	180
PMS3003	6,000

Table 1: Hardware Components price in Pakistan

1.8.2 Software Requirements

- **AWS S3 for cloud storage:** It is an expansible, fast, web-based cloud storage service. It can be used to store and revive any amount of data by users of all sizes and industries**,** with multiple use cases such as websites, mobile apps, data lakes, online backup and archiving of data and applications, business applications, big data analytics, and IoT devices. It also offers scalability, security, and efficient performance. In our project, the database will be stored in AWS S3.
- **AWS EC2 for Cloud Computing:** It is a web service that gives cloud computing ability that is steady and resizable. The AWS E2 includes the following components: Pricing, Security, Operating System Support, Migration, and Fault tolerance. EC2 is reliable, secure, inexpensive, and contains flexible tools. The computing of the database will be done through AWS EC2.

- **AWS Route for DNS protocol:** Amazon route 53 will be used as a DNS service for the domain. Route 53 DNS service, routes internet traffic to the website by translating domain names to numeric IP addresses, that computers use to link. This is a web-service with a highly accessible and expansible cloud Domain Name System (DNS). The computed results of the operating system will be routed through the AWS route to the website.
- **Arduino IDE for coding:** The primary text editing software used for Arduino programming in Arduino IDE (Integrated Development Environment). The code is typed on this before uploading it to the board. In our project, it will be used as a coding platform.
- **MySQL:** It is an open-source relational database management system. All the data will be stored in the structural database implemented in MySQL.
- **Operating System:** Linux based operating system will be used for computing results based on data prediction and analysis.

1.9 Block Diagram

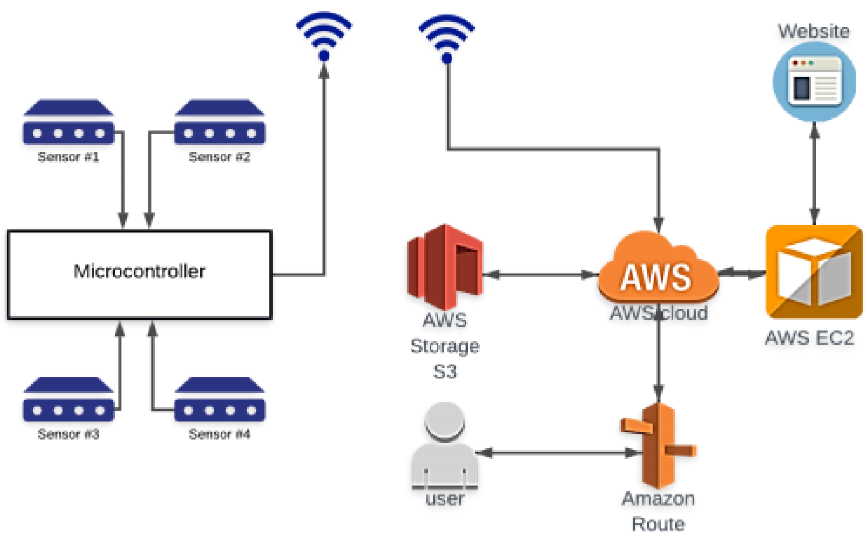

Figure 11: System block diagram.

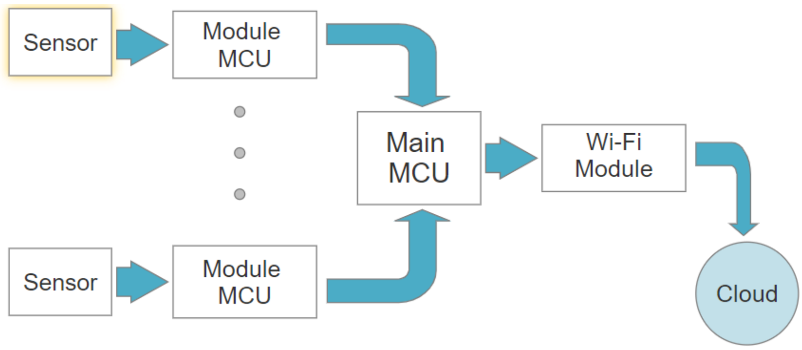

Figure 12: Block Diagrams of the Hardware.

1.10 Flowchart

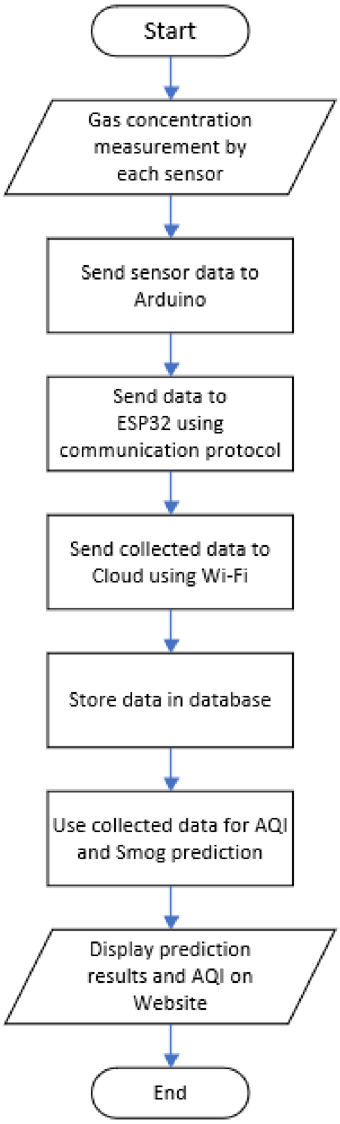

Figure 13: Flowchart of the System

Chapter 2

2 Literature Review

2.1 Literature review related to Machine Learning

2.1.1 Machine learning algorithms for predicting air pollutants

In this paper, ML models are proposed for the prognostication of $PM_{2.5}$ evaluated at different altitudes above the ground level. The data set is collected from the air monitoring station in Bangkok, Thailand. The 4 classification models implemented in this paper are Logistic Regression, Neural network, Naive Bay's, and Random Forest. To estimate the performance of these models, traditional matrices are computed like F1-score and confusion matrix while F1-score is particularly considered. Different hyperparameters are calibrated to obtain the best model's performance. To overcome the imbalance of the data set and enhance the model's performances, a re-sampling technique SMOTE (Synthetic Monitory Over-Sampling Technique) is used. The F1-score achieved by these classification models is reasonably good ranging from 0.6 to 0.8. The Random Forest model provides the best performance with an F1-score of 0.82 [5].

2.1.2 A Machine Learning Approach to Predict Air Quality in California

In this paper, a system of forecasting is presented. This prediction system is based on the concentration of individual contaminants that can forecast the air quality on an hourly basis. Two prediction models based on SVR (Support Vector Regression) are proposed, the PCA SVR-RBF model and SVR-RBF model. With the State of California observations, the data set is collected from the US EPA's Air Quality. To increase the accuracy of the prediction models, the approximation of 2^{nd} order polynomial is used to manipulate the missing data, the outliners are removed and Yeo-Johnson's transformation method is implemented for the transformation of the data set. To judge the performance of these models, confusion matrix invalidating and training data sets obtained with both models are reported. Both models achieved almost similar results with SVR-RBF slightly more accurate in predicting the AQI but both were unable to predict irregular $PM_{2.5}$ values [6].

2.1.3 Comparative Analysis of Machine Learning Techniques for Predicting Air Quality in Smart Cities

In this paper, machine learning models are proposed for the prediction of air pollution. Four machine learning models implemented in this work are Random Forest Regression, Decision

Tree Regression, Gradient Boosting Regression, and Multi-Layer Perceptron Regression. To find the best performance model, these models were compared based on error rate and processing time. According to the results achieved by the simulation, Random Forest Regression showed the best performance in predicting the air pollution for data sets of varying characteristics collected at varying locations. Random Forest Regression achieved the lowest error rate among the four models and processing time lower than gradient boosting and multi-layer perceptron algorithms. The processing time of the Decision Tree Regression was found to be the least but its error rate was higher than the others. Hence, it can be concluded that Random Forest Regression is the finest technique among the four models for predicting air pollution while Gradient Boosting Regression is the worst, with the highest processing time and high error rate [7].

2.1.4 Machine learning algorithms in air quality modeling

This research focuses on the development of ML techniques and their role in improving the accuracy of prediction models for air pollution prediction. It consists of a review of the 38 most important and relevant studies in the last 6 years that have applied machine learning methods. The major techniques for air pollutant concentration forecasting based on Linear Regression, Neural Network, Support Vector Machine, and Ensemble Learning algorithms are discussed in this paper. According to this paper, Ensemble Learning is the most up-to-date machine learning approach followed by Support Vector Machine and Neural Networks in predicting air contaminant concentration. The research indicates that Ensemble Learning and Regression approaches are preferred for air pollution estimation while Support Vector Machine and Neural Network-based approaches are preferred for air pollution prediction. Furthermore, the performance results indicate the superiority of ensemble learning and regression approaches over neural networks and SVR approaches because of their low variability [8].

2.1.5 Forecasting smog-related health hazard based on social media and physical sensor

In this paper, a predictive analytic framework is proposed that utilizes information data from social media and physical sensors to forecast health hazards associated with smog. It is based on two major parts: (1) model health risks and dangers associated with smog by extracting raw microblogging content and information. (2) develop an artificial neural network (ANN) based model to predict the next day's smog-related public health risks and smog severity monitoring. In this analytic model, smog intensity index (SSI), public health index (PHI), public observation, and social media records are used in decision making which leads to early emergency warning in case of excessive smog level and real-time smog concentration monitoring. The ANN model is built by using the ELM algorithm, BP algorithm, and other regression methods. This analytic model displays prediction results for eight cities in China when they are struck by the high concentration of smog by using both social media and physical sensor data [9].

2.1.6 Prophet forecasting model: a machine learning approach to predict the concentration of air pollutants ($PM_{2.5}$, PM_{10}, O_3, NO_2, SO_2, CO) in Seoul, South Korea

In this paper, a Prophet forecasting model (PFM) is presented. This model is used to forecast air pollution in Seoul, South Korea, over a period of 1 year. The air pollutants that are observed through this model are: $PM_{2.5}$, PM_{10}, O_3, NO_2, SO_2, and CO. Machine learning models are implemented to measure the meteorological parameters. The data set is collected from the Seoul Open Data Plaza in Seoul that contains the hourly air quality measurements of three years. To check the accuracy of the PFM, statistical indicators are used that include: Mean Squared Error, Root Mean Squared Error, and Mean Absolute Error. This model is more successful than ARIMA (Autoregressive Integrated Moving Average) and SARIMA (Seasonal Autoregressive Integrated Moving Average Model). PMA gives accurate results even if the dataset has minimum values. This model takes 10 times less to operate than ARIMA [10].

2.1.7 Detection and Prediction of Air Pollution using Machine Learning Models

In this paper, a prediction model is proposed for air pollution especially caused by PM2.5. This work logistic regression is used to detects the contamination level of data samples, Autoregression is applied for the prediction of $PM_{2.5}$ concentration based on the previous concentration level. This model is designed to reduce the level of $PM_{2.5}$ in nearing years. The system aims to forecast the concentration level of $PM_{2.5}$ by containing the data set of daily meteorological conditions in a certain city. The two phases that are followed in the system are the training phase and the testing phase. The data set to trace the air quality was acquired from the UCI repository, it includes temperature, wind speed, dewpoint, pressure, $PM_{2.5}$ concentration, and result. From both of the machine learning models, Logistic regression was considered best on the data set with more precise mean accuracy and standard deviation accuracy. Autoregression predicted the value of $PM_{2.5}$, 7 days before the present-day. The results successfully show that logistic regression and autoregression efficiently detect and predict the meteorological conditions and $PM_{2.5}$ concentration. It consists of a review of the 17 most relevant studies that have applied machine learning techniques in the previous years [11].

2.1.8 Observation in Machine Learning

From the above reviews, it has been concluded that the Random forest algorithm for machine learning is the one with the highest accuracy rate because most of the papers concluded that it has so far the best results regarding Logistic regression, linear regression, and neural networks.

2.1.9 Proposed Machine Learning Model for the System

The above reviewed papers indicate that the most frequently used machine learning technique for the air quality and smog prediction is Random Forest with highest accuracy rate. Therefore, it has been decided to use Random Forest Regression for our machine learning model in order predict air quality and smog.

2.2 Literature review related to Hardware

2.2.1 Arduino-Based Real-Time Air Quality and Pollution Monitoring System

In this project, a real-time air pollution monitoring system is developed based on the Arduino microcontroller. It consists of a small, affordable gas sensor, MQ135, interfaced with an Arduino microcontroller unit. It is a hand-held air quality measuring system, small in size and low in cost. Ozone, carbon monoxide, carbon dioxide, and nitrogen dioxide can be monitored through this system. The real-time data for this air quality monitoring system is collected from different environmental pollutions such as vehicle smoke, coil smoke, etc. According to this research publication, the functionality and results achieved of this system are similar to the existing high-cost air quality detectors. It is a portable, user-friendly, and proficient system for air quality measurement [12].

2.2.2 Low-Cost IoT Based Air Quality Monitoring Setup Using Arduino and MQ Series Sensors with Dataset Analysis

This conference paper includes a discussion of a low-price IoT-based air quality monitoring system with a setup including Arduino and MQ sensors. The MQ gas sensors included are MQ135 and MQ7. Ammonia, alcohol, carbon dioxide, and smoke can be detected using the MQ135 gas sensor while the MQ7 gas sensor is used for the sole detection of carbon monoxide. The setup is installed at different locations all over VIT University, Vellore. The system can be used for air quality measurement both indoors and outdoors. ESP-01 Wi-Fi module is used so that the collected data from gas sensors can be sent to an IoT supporting platform and can be visualized since Arduino UNO does not support Wi-Fi. Many machine learning models were examined to find the best suitable model including the Vector Auto Regression model (VAR) and Long Short Term Memory model but the results indicated best performance was given by the Vector Auto Regression model (VAR) and thus concluded to be the most suitable model for air quality measurement. While using Arduino and MQ2 gas sensors results that the system is highly cost-efficient [13].

2.2.3 Arduino and Sensor-Based Air Pollution Monitoring System Using IoT

In this research journal, an air quality monitoring system is proposed. It is a three-phase pollution monitoring system based on IoT. It consists of IoT kits, each consisting of gas sensors, an Arduino microcontroller, and a Wi-Fi module. The IoT kits are installed at various locations in different cities with the help of which the air quality data is collected. This collected data is sent to Arduino IDE by the gas sensors and then to the cloud by Arduino IDE and stored on a database. The collected data is accessible through a mobile phone or a computer. The gas sensor used is the MQ2 gas sensor for the measurement of LPG, alcohol, propane, carbon monoxide, and methane while the DHT11 Humidity sensor is used for measuring temperature and humidity. ESP8266 is used as a Wi-Fi module as Arduino UNO does not support Wi-Fi. According to the paper, the most important aspect of this research is the use of Arduino UNO as it controls the entire process [14].

2.2.4 An Innovative Air Quality Monitoring System based on Drone and IoT Enabling Technologies

Air pollution has been one of the big problems faced by the world over the past few years. So, there is a need for a monitoring system to detect air pollutants not only in urban areas but also in remote areas. In this article, a low-cost and low-energy consuming tracking system is proposed based on a drone. The pollutants to be measured in this project are PM_1, $PM_{2.5}$, PM_4, PM_{10}, CO, CO2, NHX as well as humidity, temperature, pressure, and NOX. The drone is sent to the area where the pollutants have to be measured. The drone has LoRa, a low-power wireless device used for covering large distances attached to it. STM32L073RZ is used as a microcontroller to process data received from sensors. MQTT sensor is further used to send data to a web server, where correct and real-time data be accessed by anyone through the internet [15].

2.2.5 A Modular Plug-And-Play Sensor System for Urban Air Pollution Monitoring: Design, Implementation, and Evaluation

In today's world urban air pollution has caused adverse effects on the environment and human health. Modern monitoring systems provide pollution concentration details to identify human exposures. The hardware requirements and adjustments are generally adjusted according to the applications. These systems are difficult to maintain, rearrange, and expand in terms of sensing capabilities. This paper intends to address the above issues through a proposed Modular Sensor System (MSS) designing and Universal Sensor Interface (USI), as well as modular design in the sensor node. The prototype of the MSS sensor node was designed and tested, it has expandible sensor modules with the plug-and-play characteristic and supports numerous Wireless Sensor Networks (WSNs). Analysis of the outcomes shows that MSS sensor nodes can comfortably be accessed in a variety of situations, adapting and rearranged with power, and trace low-level air pollution with high power and accuracy. The sensor modules for THP, NO_2, and CO were successfully regulating [16].

2.2.6 Air Pollution Monitoring System based on Geosensor Network

In this paper, a method for tracking sir pollution based on geo sensors is implemented. To design the model, 24 sensors and 12 routers are placed in different areas. Various kinds of sensors such as humidity, temperature, illumination, altitude, dust, air pressure, carbon dioxide, wind speed, ultraviolet, and wind direction are mounted. Geo-sensor is a network employed to measure geospatial data. Geospatial data is the data that is collected from a specific location on earth. In this project, the context model involves understanding the condition of air pollution in distant areas. The context model helps the flexible sampling interval change for an efficient trade-off between sampling levels and the lifetime of the battery. As the context models reduce the transmission of data it can save the geo sensor's limited batteries [17].

2.2.7 Urban Air Pollution Monitoring System with Forecasting Models

In this paper, urban air quality monitoring and forecasting system is proposed. This system forecasts the concentration level of pollutants before 24 hours. The concentration level of pollutants is collected wirelessly from tracking nodes that are attached with an arrangement of meteorological and gaseous sensors. This concentration information is analyzed using machine learning (ML) algorithms. The algorithms that are used to construct an efficient forecasting model for O_3, NO_2, and SO_2 are M5P model trees, artificial neural networks (ANN), and support vector machines (SVM). In this system, univariate and multivariate modeling is followed, and root means square error (RMSE) and prediction trend accuracy are followed for performance estimation. The result analysis of the system shows that the M5P algorithm displays efficient and accurate predictions for all gases in all perspectives according to NRMSE and PTA. ANN gained very bad results due to its poor production capacity, while SVM is better than ANN for this system due to its high flexibility for geographical data [18].

2.2.8 Hardware related Observation

The above-reviewed papers indicate that the results obtained using Arduino microcontroller are comparable to the results obtained by high-cost microcontrollers. It is a good choice for the proposed system as it is efficient as well as cost-effective, in addition to that ESP variants are used with Arduino microcontroller since it does not have a Wi-Fi module required to deliver a wireless network. The sensors used for the air quality measurement are MQ-series gas sensors which are low cost, reliable, and readily available. All the papers propose the AQI measurement based on the concentration of harmful gases or materials present in the atmosphere. Some of them neglected $PM_{2.5}$ and PM_{10} which are a fine particulate matter of dust contributing to a major part in air pollution.

2.2.9 Proposed Hardware for the System

The above papers have been analyzed thoroughly to choose the right sensors and microcontroller for our project. The microcontroller used for the gas sensor modules will be an Arduino and the main microcontroller will be ESP32 because it comes with the built-in Wi-Fi module. In most of the above-reviewed papers, ESP8266 is used for Wi-Fi but we will use EPS32 as it is a successor of ESP8266 and comes with enhanced capabilities. ESP32 will be used to transfer data to the cloud using internet connectivity. Following is the list of sensors and their use for the proposed system:

- **DHT11**: Temperature and humidity Sensor.
- **MQ-7**: CO sensor.
- **MQ-131**: Ozone sensor.
- **MQ-135**: NO_x sensor.
- **PMS 3003**: $PM_{2.5}$ & PM_{10} Sensor.

2.3 Literature review related to WSN

2.3.1 Design of Air Quality Monitoring System Based on Web Using Wireless Sensor Network

This paper is addressing the problem of Air pollution by designing a monitoring system to identify the concentration of air pollutants such as SO_2, O_3, CO, NO, PM, and Pb. It has been done by designing a Wireless Sensor Network system. The methodology used in this paper is to place 3 sensor nodes on the pole of some height. Each sensor node has a node MCU which is connected to the power supply, Wi-Fi, and five sensors, each sensor for measuring different pollutants. Then data is collected on each node and it will be sent to the webserver, where it is stored and can be accessed by smartphones or an online system. For interface web server PHP programming with My SQL data engine is used [19].

2.3.2 IoT Enabled Air Quality Monitoring System (AQMS) using Raspberry Pi

In this paper air quality monitoring system has been designed based on IoT (Internet Of Things). This paper points out the use of SBC to integrate IoT and WSN Air Quality Monitoring System (AQMS), to improve the speed and to reduce complexity they used SBC. Integration of cloud services with SBC provides the alert process in real-time. Web server nodes having commercial sensors in them are used for sensing gases such as CO, CO2, NH3, and NOx to analyze both indoor and outdoor air quality. Then the results are determined using an open-source Thing Speak IoT platform. For purpose of sending and receiving data HTTP application protocol is used. In this project, only four nodes are used [20].

2.3.3 Industrial Air Pollution Monitoring System Based on Wireless Sensor Networks

In Tanzania, the air pollution monitoring system based on sensor nodes is proposed to help the regulatory agency to overcome industrial pollution. This system allows sensors to transmit data within time constraints. So that whenever pollution occurs regulatory agency will act on time. Sensors can detect six types of air pollutants- particulate matter, SO_2, O_3, NO, CO, and lead. By using a routing algorithm sensors self-align themselves to form a radio network after this gas level is measured and data is transmitted to a central node (that gathers data from all the sensors). The focal point of future work is to place many central nodes. Also, to store data SD card and to monitor pollution at exact location GPS module may be interfaced with the web page by which authorities can regulate the industries [21].

2.3.4 Environmental Monitoring Using Wireless Sensor Networks (WSN) based on IoT.

The environmental monitoring system in this paper has been designed using sensors for temperature, Humidity, CO2, Vibration, and pollution. This project aims to interface Raspberry Pi with sensors such as MQ2 and MQ7. The concept of wireless sensor nodes is also used. For collecting data from the environment raspberry pi and sensors are used. Data is transmitted to the webserver, from where the public can access it through the internet. Raspberry pi connects different sensor nodes using the Zigbee protocol. In wireless sensor nodes, data is collected by sensors and sent to sensor nodes, routers receive data and transmit this data to the gateway (Raspberry pi), where data is stored and then sent to a web server. From a web server, anyone can access it. The concept of IoT is also used to connect different devices [22].

2.3.5 Development of WSN Based Air Pollution Monitoring using NS2

In Yangon, the wireless mesh network of sensor nodes is developed to sense air pollutants such as NO, CO, O_3, SO_2, PM2.5, and PM10. This project aims to design a system at a low cost. The project design has four-phase, first phase has sensors to detect air pollutants, second consists of ADC to convert analog information to digital. The third phase has three components: RAM to store data, Microprocessor to perform the different operations, and a web server. The last phase consists of a Wi-Fi card and an IPv6 stack. In air pollution monitoring system NS2 (Network Simulator 2) is used as a simulator for simulators. By analyzing the AODV protocol using NS2 it is concluded that by increasing sensor nodes PDR also improves [23].

2.3.6 A Wireless Sensor Network Air Pollution Monitoring System

The Wireless Sensor Network (WSN) is discussed in this paper about the air quality monitoring system in Mauritius. The "Wireless Sensor Network Air Pollution Monitoring System (WAPMS)" was the name of the system; a large number of wireless sensors were installed on the island of Mauritius to monitor air quality by using the AQI of the area. The pollutants that are observed through this system are O_3, fine particulate matter, NO_2, CO, SO_2, and total reduced sulfur compounds. In this system, the efficiency of the system is enhanced by implementing and designing Recursive Converging Quartiles (RCQ) algorithm and to enhance the power management Hierarchical routing protocol is used. The Dynamic Source Routing protocol (DSR) is used for data transmission. WAPMS conveys the state of the air quality to users efficiently through colored line graphs, assesses the level of health issues for a certain area, overcomes the problem of high-power consumption in WSN, and delivers accurate results. Moreover, it shows the Port Louis map, displaying the locations of wireless sensors deployed in several areas and data readings collected by each senor. This system displays highly efficient and accurate results through sensors [24].

2.3.7 WSN related Observation

Most wireless sensor nodes are implemented using Arduino variant MCU's and ESP variants like ESP8266 and ESP32 but some are also implemented using System On Chip (SOC) like Raspberry Pi. All sensors are connected to MCU or SOC directly which limits the sensor interfacing on the ADC of the chip. The data is collected on the central node or the cloud. The most important fact observed is the limitation on the number of gas sensors that can be interfaced with a single MCU.

2.3.8 Proposed WSN for the System

The WSM for this system will come with an innovation of pushing the limitation on the number of gas sensors that can be interfaced with a single MCU. The collected data will be used for the prediction of air pollutants so that the level of smog in the atmosphere can be predicted with high accuracy.

3 References

[1] J. K. C. &. Boonphun and P. Wongchaisuwat, “Machine learning algorithms for predicting air pollutants,” 2019.

[2] Castelli, . M. C. F. M. P. A. S. S. . V. and L. , “A Machine Learning Approach to Predict Air Quality in California,” 2020.

[3] B. J. K. C. W. and P. , “Machine learning algorithms for predicting air pollutants,” 2019.

[4] C. M. C. F. M. P. A. S. S. V. and L. , “A Machine Learning Approach to Predict Air Quality in California,” 2020.

[5] Ameer, S. S. M. A. K. A. S. H. M. C. I. S. U. A. and M. N. , “Comparative Analysis of Machine Learning Techniques for Predicting Air Quality in Smart Cities,” 2019.

[6] Halsana and S. , “Machine learning algorithms in air quality modeling,” *Algorithms. International Journal of Scientific Research in Computer Science, Engineering and Information Technology,* 2020.

[7] Chen, J. C. H. W. Z. H. D. P. and J. Z. , “Forecasting smog-related health hazard based on social media and physical sensor,” *Information Systems,* 2017.

[8] Shen, J. V. D. M. and S. , “Prophet forecasting model: a machine learning approach to predict the concentration of air pollutants (PM2.5, PM10, O3, NO2, SO2, CO) in Seoul, South Korea,” 2020.

[9] R, A. C. D. C. R. K. N. D. G. P. A. and V., “Detection and Prediction of Air Pollution using Machine Learning Models,” *International Journal of Engineering Trends and Technology,* 2018.

[10] Ahasan, M. A. R. S. S. A. H. A. R. H. and M. Z. , “Arduino-Based Real-Time Air Quality and Pollution Monitoring System,” *International Journal of Innovative Research in Computer Science,* 2018.

[11] Sai, K. B. M. S. S. and H. P. , “Low-Cost IoT Based Air Quality Monitoring Setup Using Arduino and MQ Series Sensors With Dataset Analysis,” *Procedia Computer Science,* 2019.

[12] Kumar, S. J. and A. , “Arduino and Sensor-Based Air Pollution Monitoring System Using IoT,” in *International Conference on Computing, Communication and Automation (ICCCA)*, 2017.

[13] L. A. e. al, “An Innovative Air Quality Monitoring System based on Drone and IoT Enabling Technologies,” in *IEEE International Workshop on Metrology for Agriculture and Forestry*,

Portici, 2019.

[14] Yi, W. L. K. L. and Y. , “A Modular Plug-And-Play Sensor System for Urban Air Pollution Monitoring: Design, Implementation, and Evaluation,” 2017.

[15] Jung, Y. J. L. Y. K. L. D. G. R. K. H. N. and S. , “Air Pollution Monitoring System based on Geosensor Network,” *IEEE International Geoscience and Remote Sensing Symposium,* 2008.

[16] K. B. Shaban, A. K. and E. R. , “Urban Air Pollution Monitoring System with Forecasting Models,” *IEEE Sensors Journal,* 2016.

[17] Purwanto, P. S. S. S. and S. , “Design of Air Quality Monitoring System Based on Web Using Wireless Sensor Network,” *Journal of Physics: Conference Series,* 2019.

[18] Balasubramaniyan, C. M. and D. , “IoT Enabled Air Quality Monitoring System (AQMS) using Raspberry Pi,” *ndian Journal of Science and Technology,* 2016.

[19] D. S. Simbeye, “Industrial Air Pollution Monitoring System Based on Wireless Sensor Networks,” *Journal of Information Science and Computing Technologies,* 2017.

[20] Jaladi, A. R. and e. a. , “Environmental Monitoring Using Wireless Sensor Networks (WSN) based on IoT,” *Int. Res. J. Eng.,* 2017.

[21] MAW, K. K. C. M. N. and H. T. , “Development of WSN Based Air Pollution Monitoring using NS2,” 2017.

[22] Hejlová, V. V. and V. , “Wireless Sensor Network Components for Air Pollution Monitoring in the Urban Environment: Criteria and Analysis for their Selection,” *Wireless Sensor Network,* 2013.

[23] T. . Kjellstrom, M. . Lodh, A. J. McMichael, G. . Ranmuthugala, R. N. Shrestha and S. . Kingsland, “Air and Water Pollution: Burden and Strategies for Control,” , 2006. [Online]. Available: https://ncbi.nlm.nih.gov/books/nbk11769. [Accessed 2 1 2021].

[24] “WHO - 7 million premature deaths annually linked to air pollution,” , . [Online]. Available: http://www.who.int/mediacentre/news/releases/2014/air-pollution/en/. [Accessed 2 1 2021].

Appendix D: Project Timeline

DATE	2^{nd} Jan 2021

PROJECT ID		TOTAL NUMBER OF WEEKS IN PLAN	

TITLE	Web-Based Air Quality Monitoring System

No.	STARTING WEEK	DESCRIPTION OF MILESTONE	DURATION
1		Project idea	15 days
2		Literature review	40 days
3		Hardware and software selection	10 days
4		Research on selected hardware	10 days
5		Setting up gas sensors with MCU	8 days
6		Research on selected software	7 days
7		Setting up ESP32 with gas modules	5-7 days
8		Training ML prediction model	15-20 days
9		Setting up cloud and database	12-15 days
10		Creating front-end and back-end	12-15 days
11		Result analysis	12-15 days
12		Documentation	20-25 days
			166-187 days