IoT-based AQI monitoring: Evaluation of low-cost PM sensors and AQI Estimation using AQSense

Thesis submitted in partial fulfillment
of the requirements for the degree of

Master of Science
in
Electronics and Communication Engineering
by Research

by

Ishan Patwardhan
2020702018
ishan.patwardhan@research.iiit.ac.in

International Institute of Information Technology
Hyderabad - 500032, INDIA
June 2022

International Institute of Information Technology
Hyderabad, India

CERTIFICATE

It is certified that the work contained in this thesis, titled "IoT-based AQI monitoring: Evaluation of low-cost PM sensors and AQI Estimation using AQSense" by Ishan Patwardhan, has been carried out under my supervision and is not submitted elsewhere for a degree.

Date	Advisor: Dr. Sachin Chaudhari

To

My Family and Friends

Acknowledgments

I would first like to express my gratitude to my advisor Prof. Sachin Chaudhari for guiding me through the right paths throughout my master's course. His expertise, experience, and innovative ideas helped me explore new technical subjects. His constructive remarks and supervision motivated me continuously to complete the journey of my master's degree with joy. He persistently pushed me to get better every day, helped me to enhance my technical expertise, and learn many new things throughout the writing of this dissertation.

I would also like to thank my project partners, Spanddhana, Jayati, and Nitin, along with my colleagues and friends, for providing me assistance continuously at several instances. Finally, I would like to express gratitude to my family. It would have been impossible for me to complete my study without their tremendous understanding and encouragement over the past two years,

I would also like to thank I-Hub Data for providing a fellowship that supported me financially for this dissertation. This exclusive fellowship made me feel more confident with my technical skills and research.

Abstract

Air pollution poses a threat to the lives of all living beings. Government authorities generally monitor pollution levels using a high-grade setup. The high-grade instruments are expensive and also require space for setup. A few low-cost sensors have been introduced for monitoring air quality. But, those sensors also have a few shortcomings. This thesis mainly focuses on the performance evaluation of low-cost PM sensors and an image processing-based technique to estimate air quality.

Firstly, the performance of three new and popular low-cost particulate matter (PM) sensors, namely SDS011, Prana Air, and SPS30, for measuring $PM_{2.5}$ and PM_{10} levels is evaluated against a standard reference Aeroqual Series-500. The test setup was exposed to PM concentrations ranging from 30 μg/cm^3 to 600 μg/cm^3. The results were based on 1 min, 15 min, 30 min, and 1 hr average readings. The experiments were carried out in indoor as well as outdoor environments. The comparative evaluation was performed before and after calibration. The performance of these sensors is evaluated in terms of coefficient of determination (R^2), coefficient of variation (C_v), and root mean square error (RMSE).

A real-time Air Quality Index (AQI) estimation technique using images and weather sensors on Indian roads is also presented. A mixture of image features, i.e., traffic density, visibility, and sensor features, i.e., temperature and humidity, were used to predict the AQI. Object detection and localization-based Deep Learning (DL) method and image processing techniques were used to extract image features. At the same time, an ML model was trained on those features to estimate the AQI. For this experiment, a dataset containing 5048 images was collected over four months from September-December 2021 using AQSense device that was developed in International Institute of Information Technology (IIIT), Hyderabad, and co-located AQI values across different seasons were collected by driving on the roads of Hyderabad city in India. The experimental results report an overall accuracy of 82% for AQI prediction. A few challenges faced during the measurement campaign are also discussed.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Internet of Things (IoT) refers to multiple things that can communicate with each other. Previously, this ability was limited to only big machines like computers and servers. With the advancement in the semiconductor industry, we can see billions of devices connected to the internet at any instant. The data that is being generated by these devices is enormous [3]. According to a study [4], it has been projected that the total number of devices that are connected to the internet will be around 30 billion by 2023. It is also reported that the share of IoT devices among all the internet-connected devices was 33 percent in 2018, which will grow to 50 percent by 2023. This means that we will witness a flurry of smart devices in the future. These devices include various products today, such as smartwatches, smart homes, industrial automation, home appliances, HVAC solutions, etc. The main characteristic of intelligent machines is their ability to sense a few parameters such as temperature, humidity, pressure, and gases. The sensor data is sent to the servers that have the AI engines running in the background [5, 3]. These engines analyze the data in real-time and send the feedback to the IoT nodes, triggering the appropriate actuator mechanism. The IoT devices, in general, aim to revolutionize the method of working of the existing technologies, improve the efficiency of machines and businesses, and make the life quality of the individuals better.

Air pollution is one particular domain that has been a topic of interest for IoT applications recently. According to a report [6], air pollution is a cause of respiratory diseases and premature deaths. The government currently does pollution monitoring in India. The pollution and weather monitoring systems are installed at a few points in a city. But, the number of these stations is generally low concerning the area of coverage. For example, in Hyderabad, there are six monitoring stations against the landmass of 217 sq-km. The main reason for this scarcity is the bulkiness of the measuring instruments. Hence, the spatial coverage of pollution data is less. One possible solution to improve the spatial resolution is deploying a low-cost sensing network of IoT-based nodes. Such networks can help enhance the spatial resolution of the pollution data and assist in identifying the areas that have remained unchecked by the sparse deployment. A few attempts have been made to deploy IoT- based air pollution monitoring networks that monitor the concentration of harmful gases and particulate matter (PM) along with the meteorological parameters [7, 8].

Although low-cost sensing networks are a promising approach for improving the spatial resolution of the data related to the concentration of pollutants such as PM, a few limitations need to be checked before using these sensors. The readings indicated by the low-cost sensors may not be the best approximation and might require calibration. It is essential to have a basic background in the performance of low-cost sensors to get the best results. There is a lack of literature on the performance evaluation of low-cost PM sensors. This is also the primary motivation for this work. This thesis mainly focuses on two issues. First, a comparative assessment of the three most popular low-cost PM sensors is presented for the readers. This analysis is presented to bring attention to the issues associated with the low-cost PM sensors, such as inaccuracies and inter-unit variation. Second, a novel methodology to monitor air pollution without PM sensors is presented. This method further tries to address the problems posed by using low-cost PM sensors apart from accuracy and inter-unit variation.

1.2 Summary of Contributions

The main contributions from this thesis are presented in the chapters mentioned as follows -

- **Chapter 4**
  - The performance of three most popular low-cost PM sensors namely SDS011, Prana, SPS30 is evaluated against Aeroqual.
  - Performance is evaluated in mobile as well as stationary mode in separate experiments.
  - Correlation and coefficient of variation are used as the evaluation metrics.
  - A calibration technique is presented using linear regression.
- **Chapter 5**
  - A newly developed sensor node named AQSense that was used to collect training data for image-based AQI estimation is presented.
  - A dataset of more than 5000 data points collected over six months are presented along with the details of the measurement campaign.
  - Results of different ML-based algorithms are discussed.
  - Normalized Difference Vegetation Index (NDVI) analysis of the col
  - A dataset of more than 5000 data points collected over six months are presented along with the details of the measurement campaign.
  - Results of different ML-based algorithms are discussed.

- Normalized Difference Vegetation Index (NDVI) analysis of the collected data is presented to show the effect of vegetation on PM level.lected data is presented to show the effect of vegetation on PM level.

Note: The object detection algorithm YOLOv5 and its training on Indian dataset is not discussed in this dissertation. Only results are shown. Credits belong to Nitin Nilesh and Jayati Narang.

1.3 Structure of Thesis

The rest of this thesis is organized as follows-

- **Chapter 2** provides a brief introduction to IoT, applications of IoT, and challenges involved.
- **Chapter 3** gives an overview of the related work and literature about air pollution monitoring networks and low-cost sensors.
- **Chapter 4** presents a performance evaluation of low-cost PM sensors against a standard reference in two different indoor and outdoor experiments. The linear regression-based calibration technique is also shown.
- **Chapter 5** presents the proposed methodology for predicting the AQI without using the PM sensor. Details of the dataset and measurement campaign are also put in focus. The challenges faced during the campaign are also discussed. The results obtained are based on the feature-rich dataset of 5000 data points.
- **Chapter 6** contains the conclusion of this thesis.

Chapter 2

An Overview on IoT

This chapter gives an introduction to IoT. The basic building blocks of IoT are discussed, followed by a brief overview of the applications and technologies that have evolved using IoT. Additionally, a few challenges that are related to IoT are discussed. This chapter is presented concisely, and more information about the concepts mentioned in this chapter can be found in the following sources for interested readers [9, 10, 3, 11, 12].

2.1 Introduction to IoT

IoT refers to the interconnected systems (things) that can transfer data through a suitable protocol. There is no formal definition of IoT yet. A few attempts have been made to define IoT. Two definitions of IoT are mentioned as follows-

- Gartner Research [13] defines it as *the network of physical objects that contain embedded technology to communicate and sense or interact with their internal states or the external environment.*

- United Nations International Telecommunication Union [14] defines it as *a global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies.*

Before the arrival of the IoT concept, the connectivity among the devices or internet access was restricted to high-end computers, servers, and other big machines. Such machines primarily perform computations, information transfer, storage, etc. Due to the advancement of semiconductor technology over the past few decades, new-age chips can accommodate more features in smaller surface areas, giving rise to Wi-Fi and Bluetooth Low Energy (BLE) enabled microcontroller units (MCU) in the market. Such modern MCUs are present in smaller electronic appliances such as fans, bulbs, etc. All such devices/things can gain communication with each other, and therefore, such networks are termed the "Internet Of Things." Fig. 2.1 shows an illustration of devices connected representing an IoT network. These networks are now commonly found in smart homes, factories, and other automation-based indus-

Figure 2.1 IoT network [15]

tries. IoT helps us connect computational, mechanical, and virtual machines by enabling the feature to exchange data through the internet or some different mode of communication.

2.2 Architecture

Fig. 2.2 shows the architecture of an IoT system. To this date, there has not been any standard IoT architecture. A few models are described in [16]. There are a few models that consider five layers, and there are several others that include four or three stages for an IoT project. But, three major stages are common to all IoT-based projects [17]. For that reason, this thesis presents a three-layer architecture of IoT. The three layers are- sensing and actuation, communication, cloud storage and data analysis. The information flow can occur from one end to another in both ways, as shown in Fig. 2.2 by green and red arrows. Each of these building blocks of the IoT system is described briefly in the following subsections.

2.2.1 Sensing

The sensors and actuators are the embedded system devices with one or more sensors to capture and quantify parameters, such as temperature, pressure, humidity, current, wind speed, PM, etc. In an extensive IoT network, multiple units of the same hardware devices can be present. Each machine is referred to as a node in the network. Apart from having the sensors, nodes may have a feedback system and contain actuators that react based on the input received by the device from the sensors. We

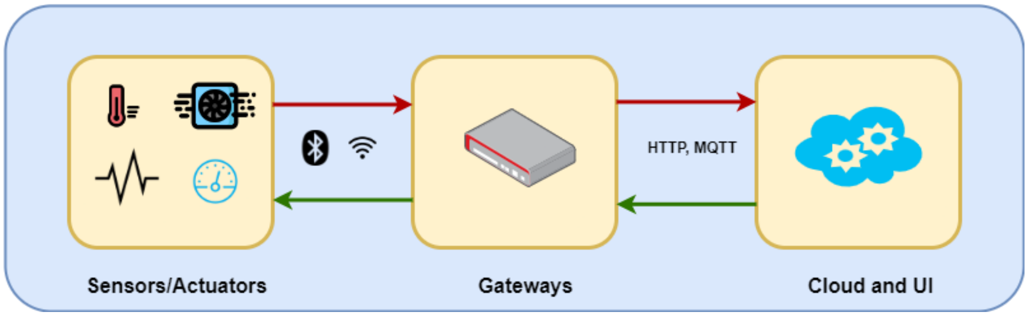


Figure 2.2 Three layer IoT architecture

see these kinds of devices every day. Examples include vending machines, alarm systems, etc. In a vending machine, the screen or buttons of the machine sense or capture the inputs from a user, and actuators dispense the desired item in a specific quantity asked by the user. The inventory report is sent continuously to the server in modern machines, which helps the owner refill the machine in good time. Also, the payment sent digitally by the user is verified through the server before triggering any actuator. Node devices use a protocol such as BLE, Wi-Fi, and ZigBee to communicate with the gateway. The information is relayed to the cloud by the gateway using protocols like HTTP, MQTT, DDS, and AQMP.

2.2.2 Communication

The data collected by the sensor nodes need a medium to move forward towards its destination. The gateways generally serve this purpose. A gateway is responsible for the movement of data in both directions. It may be a wireless or a wired device and might use different protocols such as Wi-Fi, LoRa, etc. Many communication chips are available in the market that can be embedded with the low-end MCU or interfaced externally using serial wired protocols such as UART, SPI, and I2C. These chips allow the low-end devices to establish wireless communication with the gateway. Few examples of wireless technologies include WiFi (802.11b/n/g/ah), BLE, ZigBee (IEEE 802.15.4), Z-wave, LoRa, NB-IoT and LTE-M [5]. Each protocol has some unique characteristics. The selection of the protocol depends on the designers. There are tradeoffs between power, range, cost, communication speed, accessibility, and security.

2.2.3 Cloud and User Interface

2.2.3.1 Cloud

A cloud can be considered as a virtual machine that stores data from all the sensing elements of an IoT network. The IoT devices have limited processing power due to design constraints. In a complex network, there is a lot of data generated by the machines. The cloud generally contains a big processing unit and supercomputers to make an intelligent decisions or regulate coordination between the nodes.

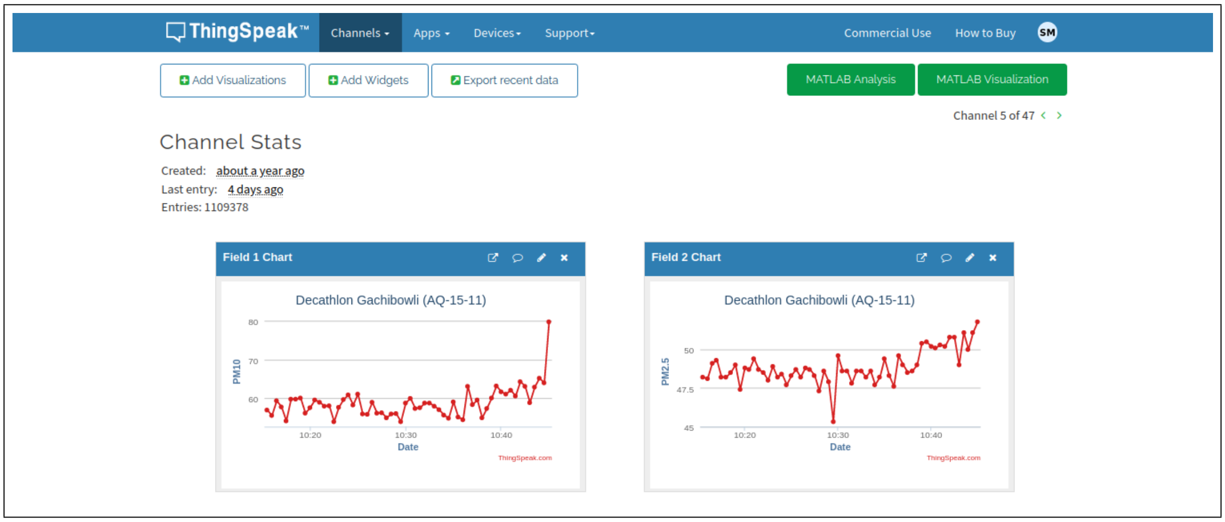


Figure 2.3 ThingSpeak User Interface

These network developers put all their decision-making algorithms into this computer. In other words, the cloud is the master of the IoT network, and it can also analyse the collected data according to the conditions specified by the user. It can also dictate the job to be performed by the nodes based upon the output of decision-making algorithms.

Another necessary functionality of the cloud is that it works as global storage. Any node can request data anytime from the cloud. Also, developers make use of the data that is stored in the cloud to develop algorithms. We also get many insights from this data, which leads to new ideas. Some examples of cloud storage services are Amazon Web Services IoT, IBM Watson IoT Platform, Microsoft Azure IoT Hub, Google Cloud IoT, etc.

2.2.3.2 User Interface

The user interface in IoT systems serves the purpose of human-device interaction. It helps us understand the system in a better and more intuitive way. IoT devices may generate a lot of information at short time intervals which is very difficult for a human to understand in a raw manner. The user interfaces conveniently present the data and information to the user. One such interface used throughout writing this dissertation is ThingSpeak. Fig. 2.3 shows a screenshot of this interface. It becomes easy to keep track of devices using this interface. All the progress is visible in real-time on the screen. We can also monitor the data patterns and perform mathematical operations on the time-series data generated by the IoT devices.

2.3 Applications

The advent of IoT has widened the scope of several industries. It is a breakthrough in technology because the earlier products and devices did not have the luxury of storing a large amount of sensor data. They also lacked the feature of edge computing. There have been many innovations in the global industries recently. A few of them are listed below in brief. Interested readers can refer to the following sources for detailed information [18, 19, 20, 21, 22, 23, 24, 25, 26, 27].

- **Smart cities:** The main idea behind smart cities is to improve the lifestyle and address the problems related to pollution, traffic, logistics, education, public transport, etc., for sustainable development. The term "smart" refers to an intelligent system that can make its own decision without any manual intervention up to a certain extent. For example, a traffic light that adjusts the duration of red and green signals in real-time according to the vehicle density at the junctions. The smart Heating Ventilation and Air Conditioning (HVAC) systems in the smart buildings ensure good ventilation. The Government of India launched the smart city mission in 2015, where a hundred cities were selected for smart development to improve the existing infrastructure and make the systems more efficient by using technology that will benefit the common public. This includes smart street lamps, free internet, smart public transport, intelligent parking logistics, etc.
- **Healthcare:** IoT has revolutionized this industry. Several wearable low-power devices have been introduced to provide detailed insights to the consumer [21, 22]. This helps the patients, doctors, analysts, and researchers to progress faster than ever. The data generated by the new-age IoT devices are accessible globally and just a few clicks away. This has helped improve the research and diagnosis of diseases, including COVID-19. During the COVID-19 pandemic, many doctors and analysts could diagnose and monitor the patients from long distances.
- **Consumer electronics:** Several consumer-grade electronic products have been introduced to the market that has helped in improving and monitoring the lifestyle of individuals. These products range from smart homes to wearable devices. A smart home may contain automatic lights/fan control, fire alarm systems, water leakage warning, electric breakdown handling systems, anti-theft systems, etc. This has improved the sense of security for families and made the lifestyle easy. Smart-home systems can inform the owner about any possible problem before it happens. Many smart wearable products can monitor steps, heart rate, blood pressure, oxygen level, etc. [23, 24]. Such devices can also generate a detailed report over time and inform the user about the possible health conditions that an individual might be suffering from.
- **Smart industries:** The concept of warehousing and logistics has been modified using the IoT-based chains. The industrial-grade approaches have seen a significant disruption after the introduction of IoT. A lot of products have been introduced that improve the safety of workers and can prevent substantial mishaps in the factories. The heavy machinery have become interconnected and can communicate with each other, leading to the reduction of manual intervention [18, 19].

These industries are seeing a major improvement in power efficiency, turn-around time, etc. Other examples include fully automated supermarkets such as AmazonGo. These are the modern supermarkets that do not require human presence. This has enabled the companies to set up 24-hour open supermarkets and beneficial facilities. This store can also be set up in places with scarce human resources.

- **Automobiles:** This technology has been present for a very long time. But, recently automobile industry has seen a significant technology breakout in the electric vehicle sector. Electric vehicles are being projected as the next big thing for the automobile industry [25, 26, 27]. They will lead us towards a cleaner environment in the future and support sustainable development. Another promising aspect in this sector is autonomous driving technology. IoT plays a pivotal role in facilitating this transformation. A fully electric and autonomous car can be seen as a collection of many sensors and components that have been interfaced together as a system. IoT has also facilitated other new features to the traditional cars such as tracking of vehicle from anywhere in the world using smartphone in case of loss or theft. The self-driving rental automobile services use an IoT-based device in their vehicles to track and monitor the driving performance of the customer.

2.4 Challenges

We have discussed several applications of IoT in the previous section. But, IoT is a relatively new research area and suffers from a few challenges that need to be addressed to utilize the resources best. A few of such challenges are discussed below -

- **Reliability of sensors:** As discussed in the previous sections, a few IoT systems rely on low-cost and small form-factor devices. These tiny devices are not the ideal substitutes for high-grade instruments. The low-cost devices suffer from certain inaccuracies that need to be considered before deployment. A few fixes, such as intelligent sensing and calibration, help to a certain extent but cannot be regarded as a long-term solution. This problem is discussed in detail in Chapter 5.

- **Power consumption:** Low-power IoT devices need to be more efficient in terms of power consumption. The battery's capacity is limited, and all the devices cannot be powered through the regular supply line because some devices are intended to be installed in places where wires and cables do not reach, or the appliance cannot be made entirely dependent on the main supply for some particular application such as security systems. To increase the battery life, developers are forced to compromise on the features that are present on the device. This problem can also be solved if the battery technology progresses and we can get higher capacity batteries in a smaller area, just like the rate at which semiconductor technology has moved ahead.

- **Security:** As mentioned in this chapter previously, standard protocols such as Wi-Fi and BLE have been embedded in the small low-end MCU. But, these chips do not implement the protocol with all the features. Sometimes, the data sensed by the IoT nodes are susceptible and should be kept private and secure at any cost. Although the name of the protocols used by the low-end chip is the same, the encryption methods used on these chips are primarily old and vulnerable. This concern is on the device level. Other problems include transferring the data in a secured fashion from the gateway to the cloud. It is also necessary to secure the cloud as it contains data from all the network nodes. Thankfully, a few players like Amazon, Google, and Microsoft offer secure cloud services for the users. So, an individual does not need to worry about security concerns on the gateway and cloud levels. But, the low-end devices still have a long way to go.

- **Privacy:** The IoT devices are becoming omnipresent. There are so many IoT-enabled devices available in the market that people use daily in their lives, such as television, air conditioner, toys, health trackers, security systems, etc. Although these devices might be used for leisure or for some utility that makes the lifestyle more comfortable, some people might not like their day-to-day life data to be stored somewhere on the cloud. The data has a lot to say about a person's life. Many insights about an individual can be inferred from the data, such as sleeping habits, medical conditions, work routine, cooking schedule, etc. Hence, it is a severe concern for the privacy of the individuals if this data reaches the wrong hands.

- **Deployment:** The idea of IoT may be popular in the research community, but it still has a long way to go to win the trust of ordinary people and authorities, with a weak technical background. As mentioned in the preceding paragraph, the individuals or organizations concerned with privacy do not absorb the idea of IoT readily. It is sometimes difficult to convince people to install a new IoT device near them. All the hard work and planning for developing the IoT devices might go in vain if the deployment is impossible.

Chapter 3

Overview of AQI Monitoring Networks

This chapter briefly describes the motivation for working on monitoring air pollution. A thorough literature review of various approaches followed previously is also presented. A few low-cost sensor networks and other IoT and image processing-based air quality estimation ideas that scholars worldwide have brought forward are also introduced briefly.

3.1 Motivation

Air pollution is a matter of grave concern and is the cause of many airborne diseases and premature deaths of many individuals belonging to all age groups [6]. Particulate matter (PM) is one of the air pollutants that can reach the lungs and prove fatal if an individual is exposed to it for a long term. Apart from this, it can cause many respiration and heart-related diseases [28]. Due to rapid and unplanned constructions in the urban areas, this problem is more prevalent in developing nations with high population and traffic density. The process of urbanization leads to the degradation of vegetation, and also, the closely packed houses of the metropolitan areas have poor ventilation. The individuals living in such conditions are mainly at a higher risk of getting affected by the air pollution [29]. Therefore, it is essential to monitor the air quality and take suitable actions. This thesis mainly focuses on the performance evaluation of low-cost PM sensors and an image processing-based technique to estimate air quality.

3.2 Related Work

There have been many attempts to monitor air pollution around the world. The experiments include static, mobile, and modern image processing-based ideas to estimate air quality. A few relevant approaches are discussed in the following subsections.

3.2.1 Stationary Networks

These kinds of networks are used mainly by the Governmental authorities to keep the public aware of the AQI. The high-grade PM monitoring devices are usually bulky and expensive, such as instruments based on Tapered Element Oscillating Microbalance (TEOM) technology and Beta Attenuation Monitor (BAM) [30]. They also require frequent servicing to get the best performance. Only a few PM-level monitoring systems are deployed in a city for the aforementioned reasons. Therefore, the data captured from these devices represent only a particular location and does not cover all the areas.

In the existing literature, it has been shown that the PM has a very high spatial variance, i.e., PM concentration might vary at places as close as a few meters [31]. To address this problem, there have been a few modern low-cost IoT-based approaches that help increase the spatial resolution of the PM data. Most of them involve hardware devices that can be easily mounted on poles, walls, traffic lights, etc. Out of many examples, two such networks include London Air Quality Network (LAQN) [32], CitySense [33].

In India, Central Pollution Control Board (CPCB) is the official authority that takes care of air and water pollution-related matters [1]. They have launched a National Air Monitoring Programme (NAMP), under which three pollutants - sulphur dioxide, nitrogen dioxide, and PM have been identified for regular monitoring. Besides pollutants, meteorological parameters such as temperature, humidity, wind speed, and direction have also been included. CPCB is carrying out the monitoring, State Pollution Control Boards, Pollution Control Committees; National Environmental Engineering Research Institute (NEERI), Nagpur. CPCB coordinates with the other agencies to ensure the uniformity and consistency of air quality data and provides technical and financial support for operating the monitoring station. There are 804 operating stations in 344 cities/towns in 28 states and 6 Union Territories. Fig. 3.2.1 shows the monitoring stations across India. The air pollution and weather data is recorded and is available to the public in real-time.

In LAQN [32], the network consists of 33 measurement units spread across London city. A few more nodes that the local authorities have deployed contribute additional data to the network. This data is available in the public domain, and anyone can check the air quality in real-time on the website.

The CitySense [33] project aims to develop a wireless network of sensors on the scale of an urban area in the city of Cambridge. The sensors also include air quality monitoring sensors. A total of a hundred Linux-based PCs are deployed at various places like streetlamp posts and poles. The nodes are equipped with radios that work as a mesh. The data is continuously pushed to the servers and made available to the public through a web app.

3.2.2 Mobile Networks

In this kind of network, the hardware devices are smaller than the stationary deployment. As the name suggests, these devices can be carried from one place to another, providing a better spatial resolution. These look like small gadgets that are easily portable and battery-powered. They also contain

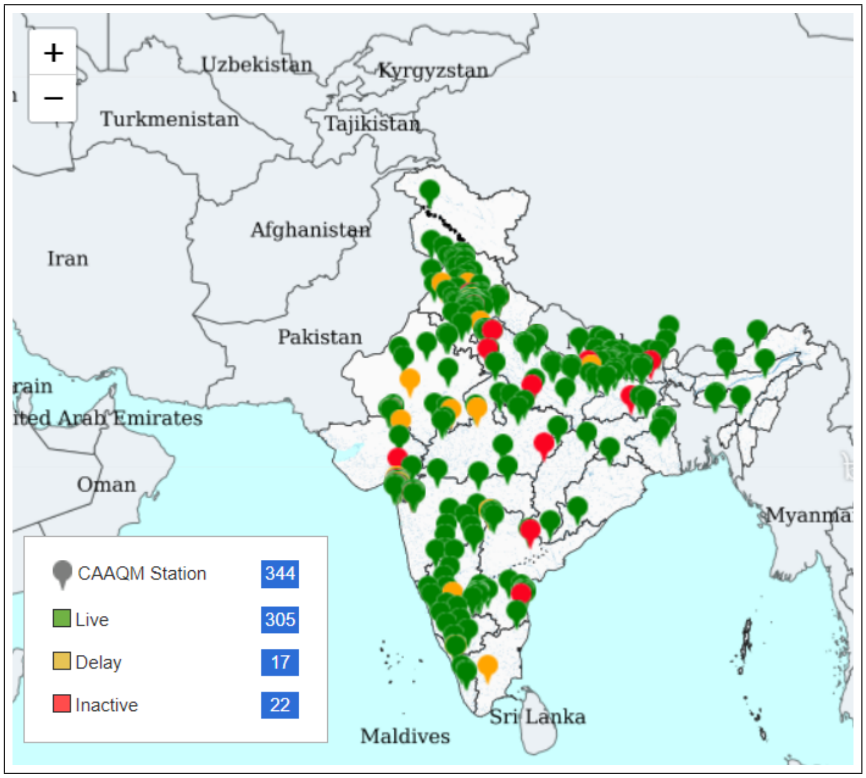


Figure 3.1 CPCB monitoring stations under NAMP [1]

wireless interfaces like Wi-Fi, BLE, 4G, etc. Data can be transferred through an auxiliary device like a smartphone to the cloud. There are generally two variations of a network of mobile nodes. The main difference is how devices are ported from one place to another.

In the community-based monitoring networks, the devices are generally given to the common public, researchers, professionals, etc. They need to recharge the devices and connect them to their smartphones. The hardware's form factor is like a smartphone and can be easily carried in a bag or pocket. A few examples of such kinds of networks include MegaSense [34], and GasMobile [35]. These projects include a small sensing node and an interfacing protocol like USB, BLE, to connect the devices to the smartphone. The basic idea here is that people can carry these nodes along with themselves and check the air quality continuously. The data also gets transferred to a central server with the help of their smartphones, which helps the researchers and authorities collect meaningful insights and make decisions over a long period.

MegaSense [34] gadget is a part of the HOPE project in Helsinki, Finland. It aims to monitor air quality with the help of public support. MegaSense is a battery-powered mobile device that is equipped with sensors. Individuals can carry this device to the places they commute to every day. They can monitor the air quality in real-time with the help of an android application that provides the UI for MegaSense. In this way, users can check the air quality around them, and the data is also sent to the servers, which can further be analyzed to extract more information on the developer's end. A district-level map of pollution data is also created using the data obtained from various users.

In GasMobile [35], researchers have developed a device that can monitor outdoor air pollution and directly connect to the smartphone via a USB interface. Users need to connect their smartphones to the device. The data is also sent to the server using the user's cellular network and is available to the public. In this way, the pollution data can also be accessed by the people who do not own the device.

The other kind of mobile network is the vehicular sensing network. The sensing nodes are mounted on the public transport vehicles in this arrangement. These systems work on either a cellular network or in offline mode. The data does not automatically get transmitted to the server offline. It needs to be manually transferred to a device using cables. This information is then relayed to the internet with the help of an auxiliary device. A few examples of such kinds of networks include City Scanner [36, 37], and Google street view car [38].

3.2.3 Machine Learning and Image Processing based AQI Estimation

So far, we have discussed the methods that include some hardware that can sense and report the AQI. As mentioned at the beginning of this thesis, hardware-based solutions require frequent maintenance and cleaning. For that, researchers worldwide have introduced a few novel approaches that can report the AQI without using any PM sensor to avoid this hassle. They predict the AQI based on the images and extra parameters such as temperature and humidity. These are machine-learning-based approaches and use historical data to predict the AQI.

A few of them [39, 40, 41] rely on the hybrid supervised deep learning models that can predict air quality using historical data. In some of them, features are extracted from the images automatically to predict the AQI. The dataset generally consists of photos taken from traffic cameras, CCTV footage, and other public camera sources. The PM and weather data are aggregated from the nearest weather station.

In [39], a CNN-LSTM (Convolutional Neural Network- Long Short-Term Memory) based supervised hybrid deep learning model is trained to predict the air quality. Two approaches were used in this experiment - univariate and multivariate. In the univariate process, the model contained only one pollutant, whereas, in the multivariate model, information about multiple pollutants and meteorological data were stored in the model. The data was taken from the pollution monitoring facilities installed by the Municipal office in Barcelona, Spain, from June'18 to January'19. The final weights obtained after training were used to predict the AQI data for Barcelona and other cities with similar meteorological conditions. This approach suffers from a major shortcoming. The authors have used a model trained for one city for other cities with similar meteorological conditions. The pollution data depends on the weather parameters and the traffic, construction, and population density. If there is a rich source of pollution, such as chimneys or mining, this model will miss the actual AQI values for such places.

In another approach, [42], CNN architecture along with transfer learning on a manually created dataset. The model is trained on 591 images obtained from public cameras in Beijing, China, and the corresponding PM data was taken from the nearest air quality monitoring station. A CNN was trained on

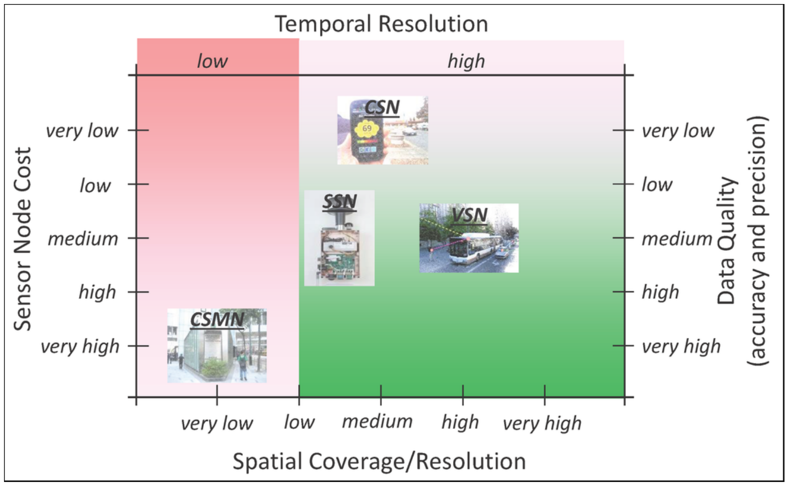


Figure 3.2 Trade off between different types sensor networks w.r.t cost, spatial coverage/resolution, data quality (accuracy and precision) [43].

the images, and the AQI was classified into three categories- Good, Moderate, and Severe. The authors report a maximum accuracy of 68.74% for their CNN-based model. However, a few things should be noted about this approach. This model works like a black box with no explicit features to catch in every image sample. Also, the authors have captured data from only one place and do not consider the spatio-temporal relationship of PM data. The PM data can vary at several places in the same city. The authors have tried to use a model based on just 591 samples of the same place captured for a concise time frame. This kind of model can report PM values on any image if it can provide the features expected by the model unknowingly.

Apart from this, all the image-based approaches suffer from a common issue with their dataset. The data points are aggregated from various sources. Images are taken from public cameras, PM, and weather data from the nearest weather station. In this sense, the PM and other parameters annotated with the sample image may not represent the correct values, and the dataset might be faulty. This problem is addressed later in this thesis, where we discuss a similar approach with our dataset, which has been collected with specially designed hardware.

This chapter covered different types of initiatives taken by individuals worldwide to monitor air quality. Selecting any network as the universal solution for monitoring air quality is challenging. There are a few tradeoffs related to each method discussed in this chapter. Fig. 3.2 shows the tradeoffs involved with each type of technology in brief. It is up to the user to select the appropriate solution as per their application and constraints.

Chapter 4

Comparative Evaluation of New Low-Cost Particulate Matter Sensors

The weather and pollution stations use bulky and expensive standard instruments to measure PM concentration accurately. In recent times, a few new low-cost sensors have been introduced to the global market for monitoring particulate matter (PM). These sensors usually come in a small factor to compact the overall device. They enable the dense deployment of compact devices to increase the spatial resolution of the PM data. But, there is one major drawback of using these sensors, i.e., the inaccuracy of their measurements. In this chapter, the performance of three such low-cost PM sensors, namely SDS011, Prana Air, and SPS30, for measuring $PM_{2.5}$ and PM_{10} levels is evaluated against Aeroqual Series-500 PM monitor as the reference.

4.1 Introduction to Low-cost PM Sensors

The low-cost PM sensors were introduced to the global market a few years ago. Before that, only standard, bulky, high-grade, and expensive instruments (approx. USD 10,000) were available to measure the PM concentration. The availability of low-cost PM sensors (USD20-30) has enabled developing and emerging nations to afford the measurement of PM levels in their countries. These sensors also help increase the spatial and temporal resolution of the PM data.

4.1.1 Sensor Design

The commercially available low-cost PM sensors work on the principle of scattering of light as shown in Fig. 4.2(a). Four major electronic components are used in the system - MCU, a light source (LED), a photodiode (PTD), and a focusing lens. The light source continuously emits photons directed towards the PTD using a lens. In normal conditions, the light coming from the source reaches the PTD, and it works in the ON condition. The sensor also has two more components- inlet pipe and exhaust fan. The exhaust fan has a rotation direction such that it pulls the air from the inlet vent and pushes it out of the sensor from the other end. When the air enters the sensor, its particles scatter the light, and the PTD turns OFF for that duration. The PM concentration is estimated internally by the MCU present in the

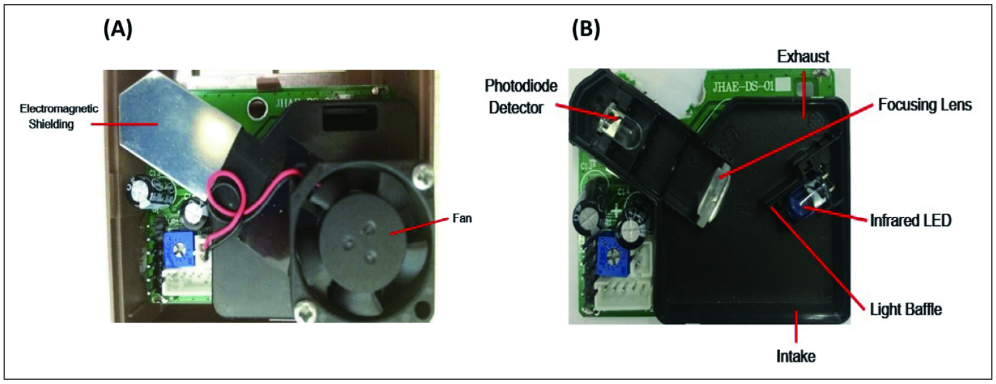
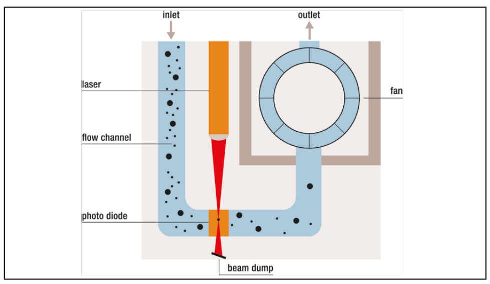
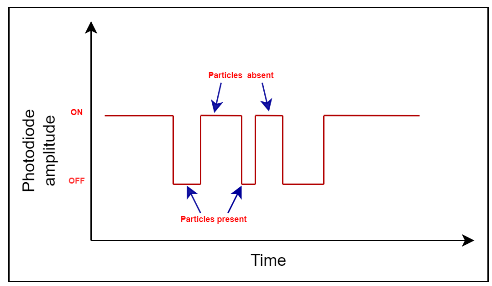


Figure 4.1 Cross-section of a low-cost PM sensor [2]



(a) Light scattering mechanism [44]



(b) Electrical response

Figure 4.2 Light scattering principle and electrical response of a low-cost PM sensor

sensor based on the number of these ON/OFF transitions of the PTD. Fig. 4.1 shows the actual internal setup of a low-cost PM sensor. These sensors are calibrated in the factory by the manufacturer. The data can be retrieved from these sensors using the suitable protocol (UART/ I2C/ PWM). Fig. 4.2(b) conveys the method discussed above in the graphical form.

4.1.2 Limitations

Although the low-cost PM sensors have a large number of benefits, there are a few drawbacks of using these sensors, such as -

- The accuracy of these sensors might suffer from place to place. The manufacturer calibrates the sensor in the factory. The type of particle used to calibrate the sensor may be different from the particle that the sensor is exposed to in the field.

- These sensors use a light source such as LED or LASER, which have a limited lifetime. Also, the performance might be affected as the light source's lifetime nears its end. The typical lifetime is up to 8000 hours.

- When we buy multiple copies of the same sensor, we expect all of them to give similar readings when placed together in an identical environment. But, there is always some variation in the readings of all the copies.

4.2 Evaluation Methods

Previous works that have been done to compare and evaluate the performance of the low-cost PM sensors use the weather station as the reference. In [45], the sensors were kept near the weather station, and the data was collected for a long time. In [46], the sensors were placed in an artificial wind tunnel along with the reference instrument. The particle concentration was controlled by varying the speed of the wind. However, these methods have a few limitations. In [45], authors evaluate the performance in stationary mode. The weather station can not be moved. Hence, the sensors were exposed to the particles reaching the monitoring facility only. As the number of these stations is significantly less, the setup might not capture a wide range of PM values. Also, different kinds of particles can come from various sources that have the same dimension. It is essential to capture data from all sorts of sources possible to remark on the performance of the sensors in different natural conditions. In [46], a single type of particle was used for the evaluation of sensors. But, it is impossible to determine which kind of particle is suspended in the air in the actual deployment field. It can be a mixture of several particles with the same dimension.

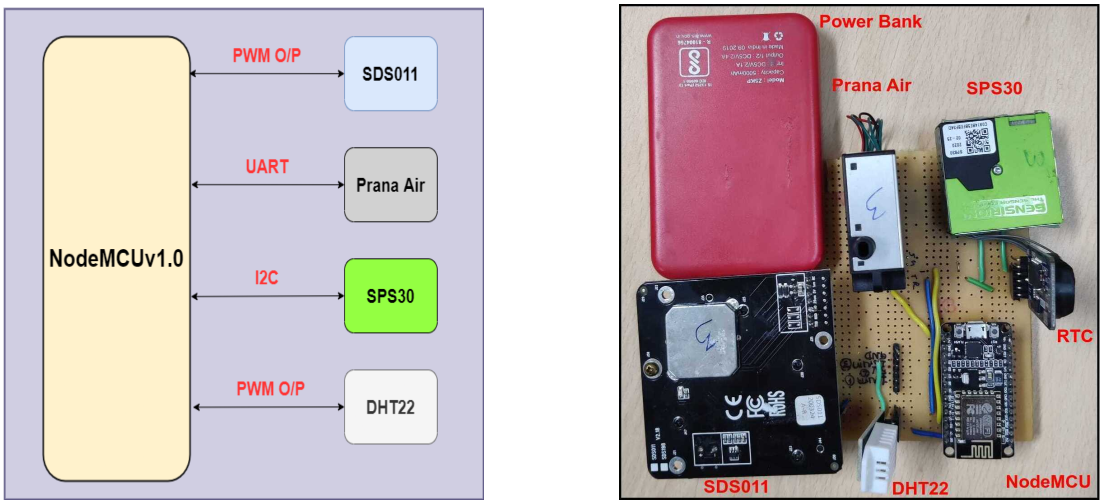
In this chapter, the evaluation of three low-cost PM sensors is presented. The performance was analyzed in the indoor as well as outdoor environment. Unlike the existing literature, which primarily used a stationary setup, we used a portable setup to test the PM sensors in a few different parts of the city. This way, sensors are exposed to different particles from sources like factories, vehicles, restaurants by taking the test setup and the reference instrument near such places. In addition, this study evaluated two new sensors (Sensirion and Prana) along with one well-studied sensor (SDS011).

4.3 Experimental Setup

This section briefly describes the architecture of the hardware setup and the reference device used for this study. A single node consisted of three different PM sensors, and three such nodes were used for this experiment. Aeroqual Series-500 was the reference instrument. More discussion about the hardware setup is presented in the following subsections.

4.3.1 Reference Instrument

Series-500 manufactured by Aeroqual [47] was the reference instrument used for this experiment. It is a monitoring device with a separate “head” for monitoring PM and gases. It can measure only one pollutant at a time. Aeroqual is a portable air pollution monitoring device that can measure $PM_{2.5}$ and



(a) Block architecture (b) Actual hardware

Figure 4.3 Hardware setup

PM_{10} levels simultaneously with a minimum time resolution of 1 min. It works on laser particle counter (LPC) technology and is factory calibrated. A case study [48] reports a very high correlation between this portable monitor and higher specification environmental monitors.

4.3.2 Hardware Design

Three identical test nodes were created for the experiment. Fig. 4.3 shows the schematic view and actual view, respectively, for each such node, which consists of one unit of SDS011, Prana Air, and SPS30 each. ESP8266 based Wi-Fi enabled NodeMCU v1.0 microcontroller module was used to interface these sensors. Samples were collected at 2 s intervals, and all data was pushed to Thingspeak, an MQTT-based IoT platform. The Wi-Fi access point was created using a smartphone, and a 4G cellular network was employed to access the internet. All information was downloaded in CSV format from the Thingspeak platform for all three nodes. We also downloaded the data from our standard reference Series-500. The data was processed and analyzed using the python programming language for 1 min, 15 min, 30 min, and 1 hr averaged readings.

4.4 Measurement Campaigns

In this experiment, the performance evaluation was done in indoor and outdoor environments. The main reason to perform these two experiments separately was to understand the ability of these sensors when exposed to the static and dynamic environment. The following two sub-sections briefly describe the process of both experiments.

Figure 4.4 Open street map of the chosen places for data collection

4.4.1 Indoor Experiment

In this method, the test setup was stationary and placed near an open window to record the PM concentration continuously for one week. The room windows remained available for the whole duration of this experiment to keep good ventilation. The PM level inside the room does not change by an enormous amount during any time of the day. This happens because the indoor environment is primarily free of any pollution source except for a few places such as the kitchen. The primary PM source is the dust coming from nearby roads or areas with a pollution source. Hence, the range of PM levels captured during the day is not very dynamic. The objective of this experiment was to examine the sensor in such static condition.

4.4.2 Outdoor Experiment

All three nodes and the reference instrument Series-500 were placed inside a vehicle with open windows in the outdoor experiment. All devices were powered using a 5000 mAh power bank. The vehicle was taken to several parts of the city to cover industrial, residential, and commercial areas and places with high and low traffic movement. A few halts for short intervals of 30-45 minutes were made to collect the readings of that particular area. The main objective was to evaluate the ability of sensors to record the change in PM levels. Fig. 4.4 displays the $PM_{2.5}$ levels at the locations chosen for this experiment.

4.5 Data Pre-processing and Performance Evaluation Parameters

The data obtained from the test setup was filtered for outliers before performing any analysis. All sensors work in a specific humidity range, so all the readings obtained beyond the operating humidity range were removed.

The performance parameters used in this paper are R^2, C_v, and RMSE. The coefficient of determination R^2 analyzes the ability to report the changes in PM levels in comparison to the reference instrument. It is given in [49] by

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}, \tag{4.1}$$

where y_i denotes the observations, $\bar{y}$ denotes the average of the observations and $\hat{y}_i$ is the prediction of y_i using the fitted model. The R^2 values were calculated with raw sensor output and calibrated values. The calibration of the test sensors with respect to the reference sensor is done using simple linear regression.

Coefficient of variability C_v measures the reproducibility or the variance across multiple units of the same sensor and is given by

$$C_v = \frac{1}{N} \sum_{i=1}^{N} \frac{\hat{\sigma}_i}{\hat{\mu}_i} \times 100\%, \tag{4.2}$$

where N is the total number of samples, $\hat{\mu_i}$ denotes the sample mean of readings of all three units of a particular sensor at one moment of time and $\hat{\sigma_i}$ is the average standard deviation of all the copies of a sensor [50]. C_v was calculated only on the raw data in this paper.

RMSE denoted by E_{rms} is a metric representing the average of the square root of the sum of squares of errors.

$$E_{rms} = \frac{1}{N} \sqrt{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}, \tag{4.3}$$

where $\hat{y}_i$ is the prediction of y_i using the fitted model. RMSE was only calculated for the calibrated data in this paper.

4.5.1 Calibration using Linear Regression

Linear regression technique was used in this experiment for correcting the offset of the low-cost sensors. The manufacturers of these sensors claim that the PM values reported by the sensors are the true values. Also, the sensors are calibrated in the factory environment. Hence, a simple regression should be good enough ideally if the sensors are of a good quality. For this reason, this technique was selected in this experiment for the calibration among the several other advanced ML techniques that are available with us.

The offset in the low-cost sensors measurements is compensated by doing linear regression of the form

$$y = mx + c, \tag{4.4}$$

where x and y are the reading from the test sensors and the reference instrument, respectively, while m and c are constants that represent the slope and intercept of the fitting curve, respectively. This method gives us a model to calculate the actual PM concentration using raw values from test sensors.

4.6 Results and Discussion

4.6.1 Indoor experiment

Based on approximately 200,000 data points, the performance of the sensors is analyzed. All the sensors and reference instruments data in the indoor setup plotted against time are shown in Figs. 4.5 and 4.6 for $PM_{2.5}$ and PM_{10}, respectively. The graphs in Figs. 4.5 and 4.6 show that the sensors' data and reference instruments' data follow a similar trend. The sensors underestimate the PM values most of the time compared to Aeroqual. It can also be observed that the bias between the sensors and the reference instrument is lower at lower PM values and higher at higher PM values. The matching trend indicates that the sensors can comfortably capture the PM values change.

Tables 4.1 and 4.2 contain the average R^2 values of all three copies of the same sensor for $PM_{2.5}$ and PM_{10} before performing the calibration. A very high correlation value was observed for all the sensors in this setup. All sensors respond equally well while capturing the changes in $PM_{2.5}$ levels. A minor drop in R^2 values is observed in the estimation of PM_{10} levels for all the sensors.

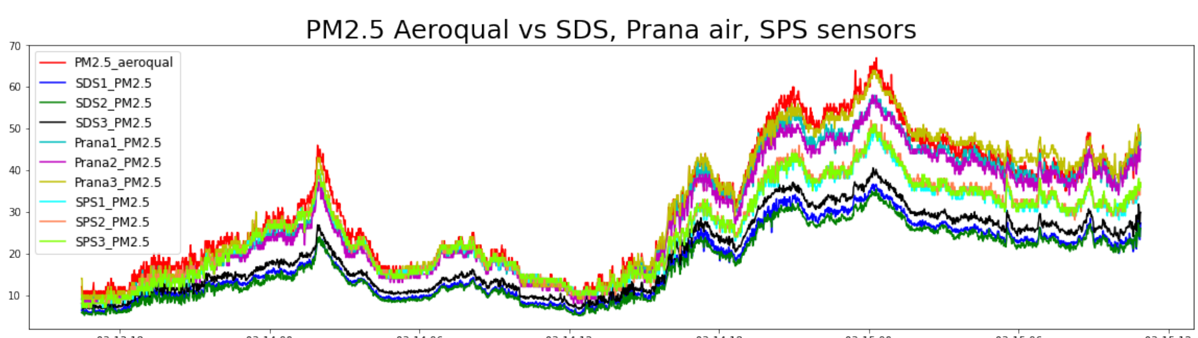


Figure 4.5 $PM_{2.5}$ trend for indoor experiment.

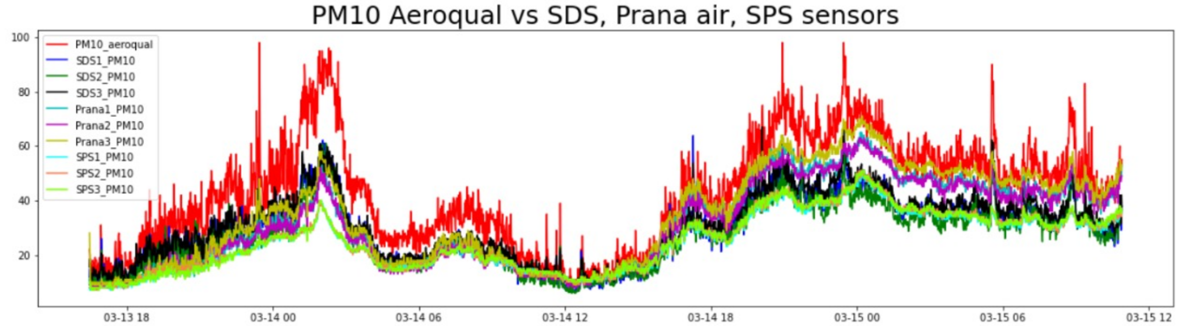


Figure 4.6 PM_{10} trend for indoor experiment.

Particle Dimension	**Sensor Name**		
	SDS011	***Prana Air***	***SPS30***
$PM_{2.5}$	9.6	5.73	2.75
PM_{10}	9.05	6.64	2.95

Table 4.3 C_v values for $PM_{2.5}$ and PM_{10} in % for indoor experiment

Averaging Interval	**Sensor Name**		
	SDS011	***Prana Air***	***SPS30***
1 min	0.99	0.99	0.98
15 min	0.99	0.98	0.98
30 min	0.98	0.98	0.97
1 hr	0.97	0.97	0.96

Table 4.1 R^2 values for $PM_{2.5}$ for indoor experiment

Averaging Interval	**Sensor Name**		
	SDS011	***Prana Air***	***SPS30***
1 min	0.87	0.98	0.98
15 min	0.98	0.90	0.90
30 min	0.98	0.90	0.90
1 hr	0.97	0.89	0.90

Table 4.2 R^2 values for PM_{10} for indoor experiment

Table 4.3 indicates the C_v values for estimation of $PM_{2.5}$ and PM_{10}. It can be observed that this parameter is not more than 10% for any sensor. SDS011 is found to have the most variation of approximately 9% for both kinds of particles, followed by Prana Air (≈ 5-7%) and SPS30 (≈ 3%).

4.6.2 Outdoor Experiment

The results obtained in this section are based on an analysis of approximately 30,000 data points, collected at 2 s intervals, before calibration. Figs. 4.7 and 4.8 show the plots for $PM_{2.5}$ and PM_{10}, respectively. The graphs show that the bias between the sensors and reference instrument is high, and

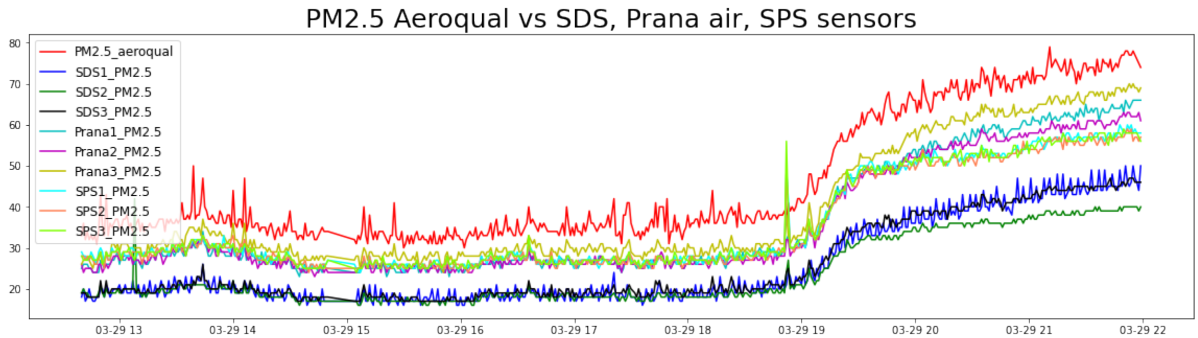


Figure 4.7 $PM_{2.5}$ trend for outdoor experiment.

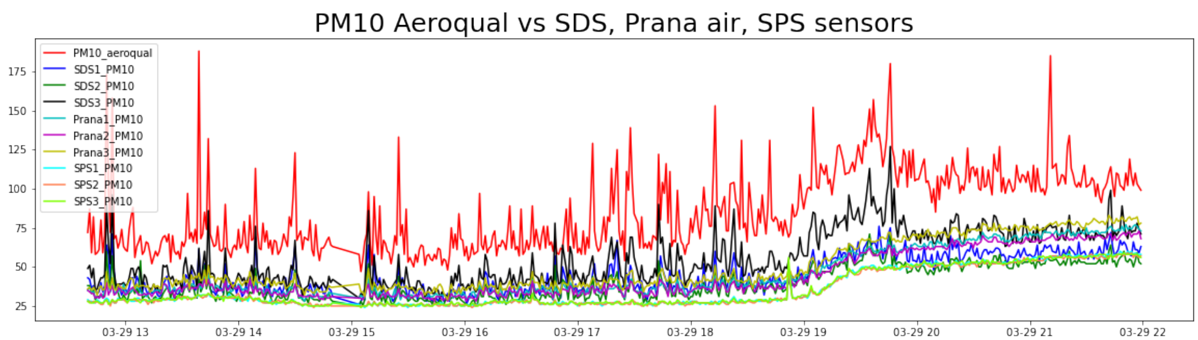


Figure 4.8 PM_{10} trend for outdoor experiment.

sensors sometimes do not follow the same trend. This might be due to the sudden increase in the PM concentration and the change in the type of particle.

Tables 4.4 and 4.5 contain the R^2 values for $PM_{2.5}$ and PM_{10} respectively. The R^2 and C_v values for mobile setup were not calculated for 30 min and 1 hr intervals due to insufficient data points. The overall response is less accurate than the indoor experiment results from these tables. It is observed that SDS011 has the highest R^2 value for both kinds of particles. Prana Air and SPS30 have an almost similar response for $PM_{2.5}$ and PM_{10}.

Table 4.6 shows the C_v values obtained from the outdoor experiment. There is a minor change in C_v values for all sensors except SDS011 compared to the indoor experiment. The C_v values were observed to be in the range of 3-9% approximately for $PM_{2.5}$ and 3-20% for PM_{10}. It is seen that the SDS011 has the highest C_v value for this experiment.

It can be observed from Figs. 4.5-4.8 that the performance of the sensors is better in the indoor environment. This difference in performance can be justified by looking at the range of the PM values that the sensors were exposed to in both experiments. In the indoor experiment, the PM values were restricted to a range of 20-70μg/cm^3 while in the outdoor experiment, the observed range was up to 200 μg/cm^3. Apart from this, there were many sudden spikes in the outdoor experiment, which indicates

Averaging Interval	Sensor Name		
	SDS011	***Prana Air***	***SPS30***
1 min	0.89	0.60	0.70
15 min	0.89	0.74	0.71

Table 4.4 R^2 values for $PM_{2.5}$ for outdoor experiment

Averaging Interval	Sensor Name		
	SDS011	***Prana Air***	***SPS30***
1 min	0.73	0.61	0.70
15 min	0.87	0.68	0.65

Table 4.5 R^2 values for PM_{10} for outdoor experiment

Particle Dimension	Sensor Name		
	SDS011	***Prana Air***	***SPS30***
$PM_{2.5}$	9.29	6.86	3.88
PM_{10}	20.13	7.89	3.38

Table 4.6 C_v values for $PM_{2.5}$ and PM_{10} in % for outdoor experiment

the environment was very dynamic. On the other hand, the PM levels in the indoor experiment changed gradually, and there were no abrupt changes in the PM level. Hence, the sensors were able to follow the trend of the reference instrument.

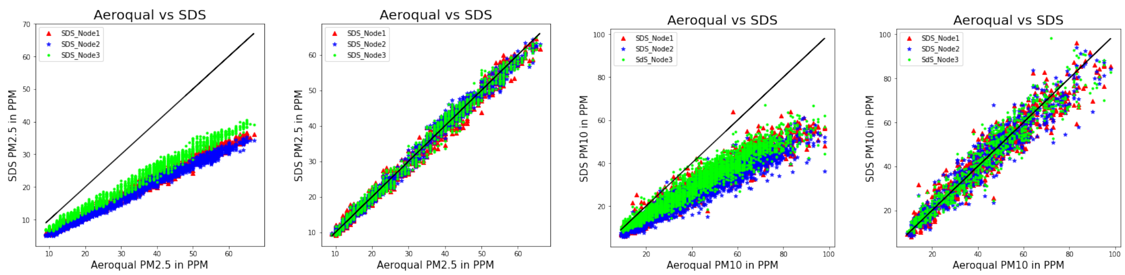
4.6.3 Calibration

This section contains results obtained after performing the calibration for the indoor experiment only. Data points obtained from the outdoor experiment are fewer, so they have been excluded from this section. The R^2 values for 1 hr averaged samples, after performing calibration, were calculated as 0.92, 0.91, 0.86 for SDS011, Prana Air, and SPS30, respectively. Table 4.7 indicates the E_{rms} values for 1 hr averaged samples for all test sensors. It was also observed to be reasonable for the collected data.

Figs. 4.9-4.11 present the scatter plots of all units represented by different colors before and after performing the calibration. These figures further explain the E_{rms} values. For the sensors having the least E_{rms} values, Prana Air for $PM_{2.5}$ and SDS011 for PM_{10}, the data points align very well around the average value of the reference instrument.

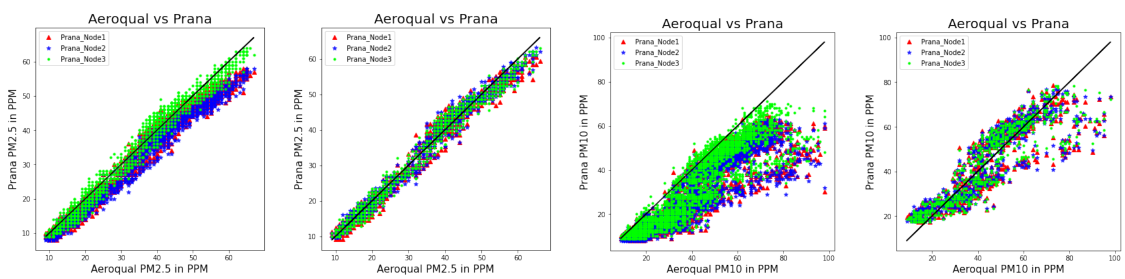
Particle Dimension	Sensor Name		
	SDS011	***Prana Air***	***SPS30***
$PM_{2.5}$	3.40	1.80	2.63
PM_{10}	2.42	8.3	8.8

Table 4.7 E_{rms} values for 1 hr averaged readings



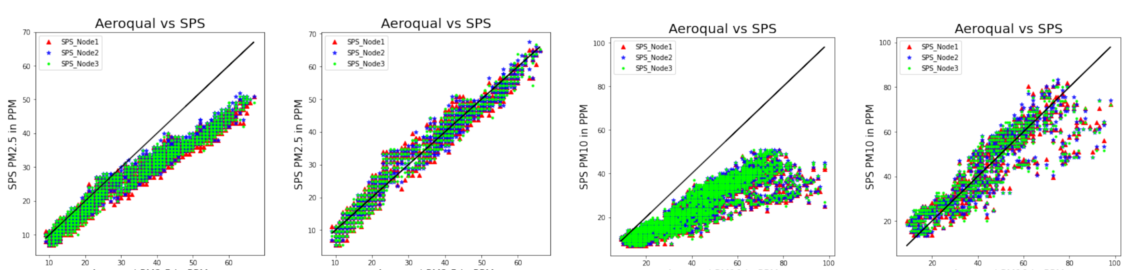
(a) $PM_{2.5}$ before calibration (b) $PM_{2.5}$ after calibration (c) PM_{10} before calibration (d) PM_{10} after calibration

Figure 4.9 SDS011 scatter plots



(a) $PM_{2.5}$ before calibration (b) $PM_{2.5}$ after calibration (c) PM_{10} before calibration (d) PM_{10} after calibration

Figure 4.10 Prana Air scatter plots



(a) $PM_{2.5}$ before calibration (b) $PM_{2.5}$ after calibration (c) PM_{10} before calibration (d) PM_{10} after calibration

Figure 4.11 SPS30 scatter plots

Chapter 5

AQSense: Image Processing and IoT based AQI Estimation Device

The AQI estimation techniques used in the dense IoT-based networks or weather stations generally use the $PM_{2.5}$ and PM_{10} readings from the PM sensor. But, PM sensors also come with their limitations. Standard instruments are not easily portable for mobile measurements, and on the other hand, the low-cost sensors may have inaccuracies and a short lifetime. This chapter presents an image processing-based method that can report AQI in real-time without using the PM sensor. Image-based features like traffic density, visibility, along with temperature, and humidity were used to determine the AQI. Supervised learning algorithms and deep learning-based object detection models were used to predict the AQI. The data for training and testing the models was collected by driving through the roads of Hyderabad city in India. This technique can predict the AQI with an accuracy of up to 90%, according to the obtained experimental results.

5.1 AQI Estimation using ML-based Techniques

Predicting the AQI using natural images and machine learning-based models has been a recent approach in the research community. In [39], a combination of a convolutional neural network (CNN) and long short-term memory (LSTM) deep neural network model was proposed to predict the concentration of air pollutants in multiple locations of a city by using spatial-temporal relationships of PM and weather data. The already available meteorological data was used to make the model, and the final weights were used to predict the AQI of the city with similar weather conditions. But in this method, the model is based on the dataset generated by the weather stations, which has a low spatial resolution, as mentioned in the previous chapter. Other image processing-based techniques use the fact that the PM values affect the visibility. This change in visibility can be extracted from the images as a feature and can further predict AQI. In [42], authors use CNN architecture along with transfer learning on a manually created dataset. The dataset contains a total of 591 images collected across different seasons from a Beijing tourist website with their respective $PM_{2.5}$ values. The air quality is categorized into three levels i.e., good, moderate, and severe. Overall, the proposed method achieved 68.74% classification accuracy. However, this method has a significantly less number of data points to build, and it classifies the data into only three categories. The authors have trained a CNN on the images and do not use other features. In the other method by [51], various numerical and categorical features from an image are extracted using image processing algorithms. A support vector regression (SVR) model was trained on these features to predict the PM values. Their dataset contains images collected from a fixed scene with respective $PM_{2.5}$ value, weather data, and geographic location. It has a total of 6587 images spanning three cities in China. The SVR model performs well for the two cities but fails in the third city because of the narrow range of $PM_{2.5}$ values. The main shortcoming of these methods is that images and PM values were not taken at the same place and time. They are aggregated from different sources. Prior literature [31] shows that PM data has spatial and temporal variation. The previous methods mentioned above may have inaccuracies because they have aggregated data from several sources.

5.2 AQSense: IoT based Hardware Device for AQI Estimation

5.2.1 Why AQSense?

This section presents the hardware device AQSense used to collect the data throughout the measurement campaign. It is important to mention that the purpose of this device is only to collect the PM and weather data with the best estimation of the ground truth. This data is used only to train the ML-based model discussed later in this chapter. AQSense is not meant to be used as a portable AQI meter or a gadget for the general public. The ultimate objective of this work is to predict the AQI using the picture of traffic. Once the model is fully trained, the inputs required are mentioned as follows -

- Picture of traffic (can be captured with any camera-enabled device).
- Temperature and humidity values (can be obtained through nearest weather station data).

Although, the best practice would be to have a temperature and humidity sensor at the point where AQI needs to be predicted to get the most accurate results. But, the weather station data can also help provide a rough idea.

5.2.2 Technical Specifications

The main idea behind this device is to estimate the AQI using the features that affect the particulate matter. Previous works show that PM values depend on meteorological parameters. In addition to that, this approach includes traffic and visibility as two more features. Fig. 5.1 shows the hardware and block diagram of AQSense. This solution has three parts: an embedded hardware device with PM, temperature, humidity sensors, a camera, embedded firmware for sampling and sending data to the cloud, and ML-based algorithms to estimate the AQI in real-time.

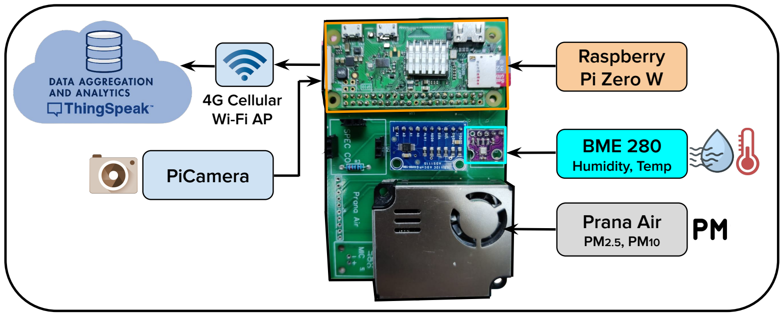


Figure 5.1 Block diagram and hardware setup.

Fig. 5.1 shows the block architecture of the hardware developed for this experiment. A Rpi0 MCU and a PiCamera are connected to capture and process the vehicle images. The other sensors interfaced with the MCU include BME280 for temperature and humidity. A *Prana Air* [52] sensor was used for measuring the $PM_{2.5}$ and PM_{10} concentrations. It is a reliable PM sensor, as shown in the study[53]. The data collected from *Prana* sensor was used to calculate the AQI, which also served as the ground truth for the ML algorithm developed for this experiment. The dimensions of this device are 100x80 mm making it compact and suitable for mobile use.

The accuracy and resolution of all the sensors are listed in Table 5.1. The hardware setup can send the processed data into a remote server, suitable for edge computing. A sample from each sensor was collected once in every 30s by the MCU. The sample is processed using the methodology defined in the upcoming sections. A cellular 4G-based Wi-Fi access point sent the data to the remote server.

Sensor	**Parameter**	**Range**	**Tolerance**
BME 280	Temperature	-40 to 85 °C	$\pm 1\%$
	Humidity	0 to 100%	$\pm 3\%$
Prana Air	$PM_{2.5}$	$\leq 1000 \mu g/m^3$	$\pm 10\%$
	PM10PM_{10}	$\leq 1000 \mu g/m^3$	$\pm 10\%$

Table 5.1 Specifications of sensors

5.3 Data Measurement Campaign

With the help of the hardware setup mentioned in the previous section, a traffic dataset was collected, containing images of traffic and the measurement of pollution levels. The device was placed on top of a car, as shown in Fig. 5.2. The car was driven during the daytime and captured variations, including

Figure 5.2 AQSense mounted on top of a car

different scenarios (urban and sub-urban areas), traffic conditions, and pollution levels.

The dataset was captured across Sep 21-Dec'21, comprising two seasons, monsoon and winter. The attempt was to get a diverse dataset. A total of 5048 samples were collected in this duration. Datapoints collected between Sep'21 and Oct'21 were considered for the monsoon season. The rest of the data collected from Nov 21- Dec'21 was accounted for for the winter season. Fig.5.3 shows the routes traveled during this campaign in the metro city of Hyderabad, India. Each captured image is associated with co-located respective sensor values, i.e., temperature, humidity, $PM_{2.5}$, and PM_{10} measurement. The AQI level is computed using the $PM_{2.5}$ and PM_{10} values as per the Central Pollution Control Board, India [1], and categorized into five classes which are as follows: 1. **Good** (0 - 50) 2. **Satisfactory** (51-100) 3. **Moderate** (101-200) 4. **Poor** (201-300), and 5. **Severe** (>300). The distribution of the collected data in terms of the AQI level and month is shown in Fig. 5.4.

5.3.1 Dataset

Bike	**Truck**	**Car**	**Bus**	**Rickshaw**	**Temperature**	**Humidity**	**AQI**
1	1	3	0	3	28	62	**Good**
4	2	3	2	3	32	42	**Moderate**

Table 5.2 Specifications of sensors

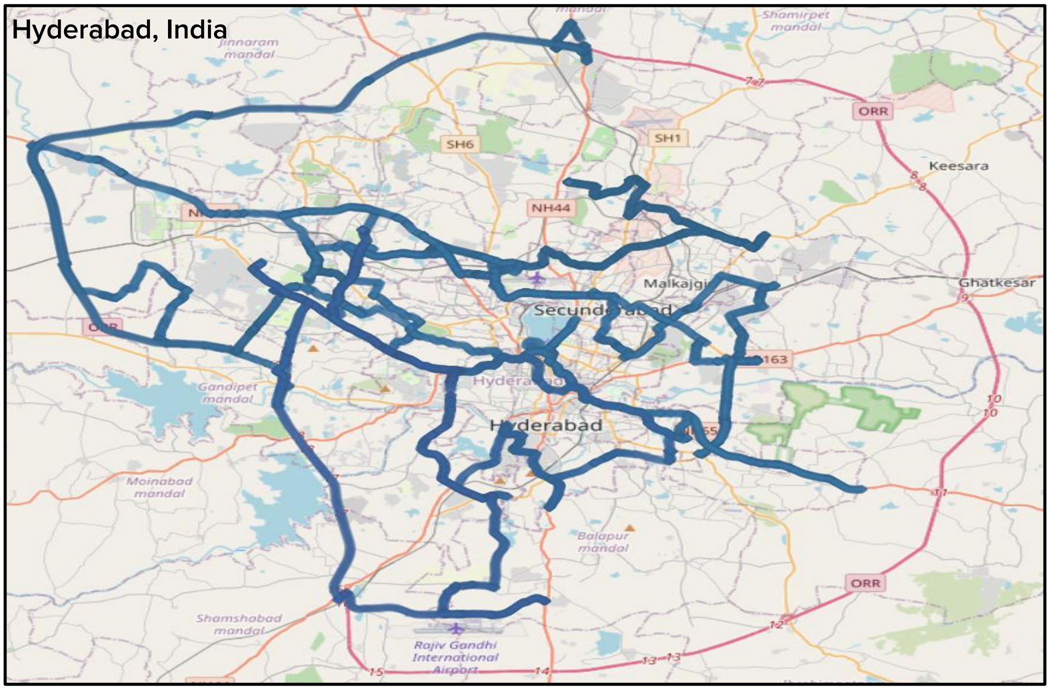


Figure 5.3 Street view of routes traveled during measurement campaign (Total distance = 1000 km).

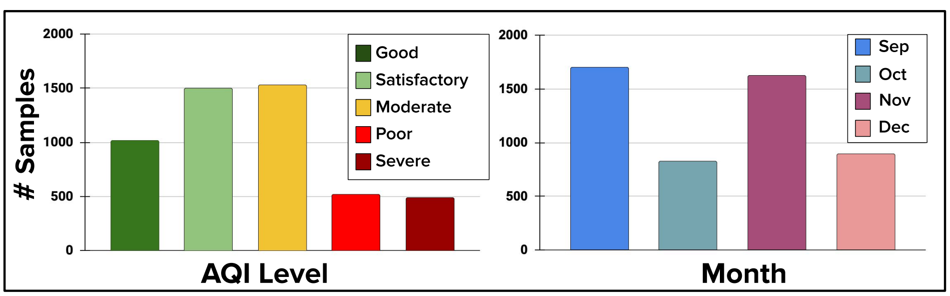


Figure 5.4 Left: Frequency of the AQI levels in the collected dataset. Right: Frequency of samples collected across months. (Best viewed in color).

The final dataset obtained for this experiment has a total of 5048 samples. Table 5.2 shows the sample entries of the dataset. It has eight features, namely motorcycle, car, bus, truck, autorickshaw, temperature, humidity, and visibility, and the target is the AQI category. The AQI categories are decided by using the standard Indian chart as shown in Fig. 5.5. For this experiment, the last two categories, i.e., "Very poor" and "Severe," were clubbed in the "Severe" category due to the lack of data points in those two categories. Fig. 5.4 shows the spread of data with the help of a histogram that was collected during the measurement campaign. Most samples (around 80 %) belong to the first three categories based on the AQI level.

AQI Category	PM2.5 (ug/m3)	PM10 (ug/m3)	Health Impact
Good (0-50)	0-30	0-50	Minimal
Satisfactory (51-100)	31-60	51-100	Minor Breathing discomfort to sensitive people.
Moderately polluted (101-200)	61-90	101-250	Breathing discomfort to asthma patents, elderly and children.
Poor (201-300)	91-120	251-350	Breathing discomfort to all
Very poor (301-400)	121-250	351-430	Respiratory illness on prolonged exposure.
Severe (401-500)	250+	430+	Health impact even on light physical work. Serious impact on people with heart/lung disease.

Figure 5.5 Standard PM reference chart according to Indian standards

5.3.2 Issues and Challenges

The measurement campaign was conducted over four months and included two seasons - monsoon and winter. AQSense was tested thoroughly regarding embedded hardware and firmware credibility before taking it to the field. Still, plenty of challenges was faced during the campaign that resulted in corrupt/ redundant data or, sometimes, hardware failure. A few trivial routines need to be followed on-field to capture the data without compromising the quality. Most of the significant issues were faced during the monsoon season. Also, after the monsoon season, we addressed a few problems by making suitable changes and faced more minor challenges during the winter season. A few of such challenges and issues are listed from personal experience, which might be insightful to this thesis's on-field researchers and readers.

- **Hardware reset** - Each hardware component has a fixed humidity range for proper functioning. As we were collecting PM data, we had to maintain proper ventilation. Hence, we did not have the freedom to use an air-tight container for the AQSense to prevent the effect of outdoor humidity on our device. Consequently, there were a few instances when the AQSense experienced hardware resets and malfunction due to suspected short circuits during the monsoon season.

- **Unreliable sensor data** - This issue was particularly faced during the monsoon season. The operating range of the BME280 is 0-100% relative humidity (RH), while that of Prana Air is 0-70%. Sometimes, during the rainy season, we faced a lot of variation (from 60% to 100%) in the humidity as we moved in the vehicle. When the rain was in patches, the PM sensor became unreliable, and its readings started to fluctuate in a big range even if we halted the vehicle at a place with RH less than 70%. We can easily filter the data by putting a condition on 70% RH, but, in this case, data points below 70% were also incorrect. Such data points need to be removed manually.

- **Dirt on glass case** - We put the AQSense in a ventilated IP65 rated plastic box with a glass cover for the measurement campaign. But sometimes, due to a lot of pollution and humidity, the glass cover requires cleaning so that the image quality does not get affected. Also, the images captured with dusty surfaces incorrectly calculated the visibility feature that was used to predict the AQI.

5.4 Methodology

This AQI estimation technique involves the following three steps -

- Initially, the object detection algorithm You Only Look Once (YOLOv5) was trained on Indian Driving Dataset (IDD).
- The weights obtained by training YOLOv5 were then used to do feature extraction using AQSense on edge. A dataset was collected using AQSense in the city of Hyderabad from Sep-Dec'21.
- Using the collected dataset, supervised learning algorithms were used for AQI classification. The weights obtained from this step are the final solution and can be used directly for estimating AQI from the image. The PM sensor can be removed from the AQSense hardware setup.

The first step of the methodology that involves training the YOLOv5 algorithm on IDD is beyond the scope of this thesis, and the subsequent steps are discussed in the following subsections.

5.4.1 Feature Extraction using AQSense

The AQI estimation algorithm relies on the features such as traffic, visibility, temperature, and humidity. A trained YOLOv5 algorithm was deployed on the AQSense to quantify the traffic. The total number of vehicles that occur in the scope of the picture captured by the camera are quantified into their respective categories using the YOLOv5 algorithm. There were a total of five categories of vehicles chosen for this experiment that are mentioned as follows: 1. **Motorcycle** 2. **Car** 3. **Truck** 4. **Bus** 5. **Autorickshaw**. These are the most commonly found vehicle categories in the urban areas of India. The vehicle count of each category is used as a feature to predict the AQI category. The general intuition of using each category as a feature is that pollution levels might partially depend on the vehicles' engine capacity. For example, a motorcycle has a small engine. Hence, pollution content released from its exhaust would be generally lower than a car or a truck. However, this might not be necessarily true because there can be exceptional cases where vehicles might not be in good health. In that case, smaller vehicles might be emitting more pollutants. This can contribute to a drop in the accuracy of the final model to a small extent.

The visibility feature is calculated using the BRISQUE algorithm on the captured images. As the camera can only take pictures of the traffic and road condition in front of the test vehicle, the information received by the image is limited in terms of scope. It is essential to capture the air pollution generated by the other elements, e.g., constructions (roads, buildings, etc.), fire, etc. To capture the essence of

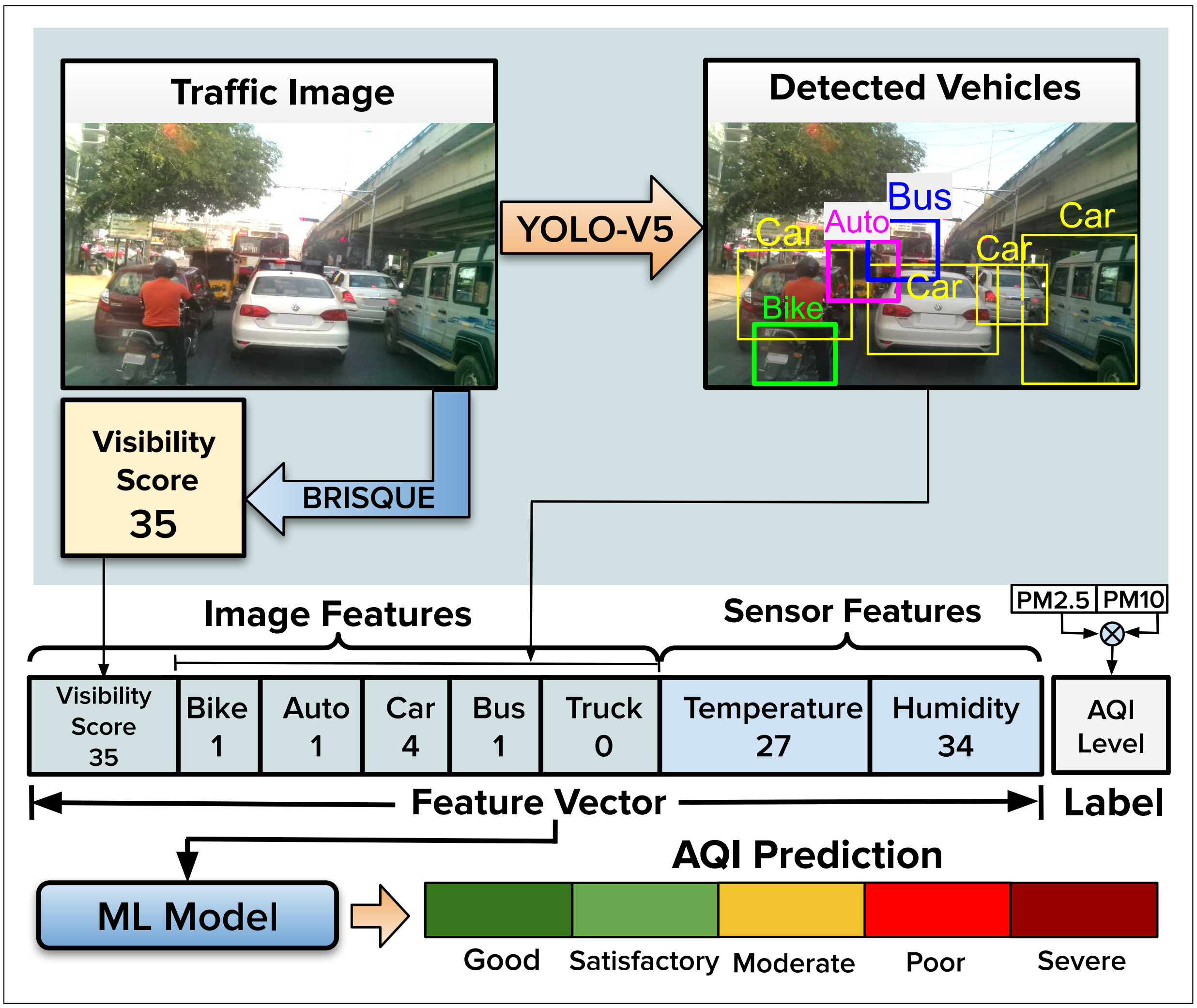


Figure 5.6 Algorithmic pipeline of the proposed method

pollution caused by other sources, the visibility of the image is computed using BRISQUE [54] which is a no-reference Image Quality Assessment (IQA) metric. As visibility is a subjective matter, a human evaluated dataset named TID2008 [55] which has 1700 images and their respective quality scores, is used as an image visibility score reference. The output of the BRISQUE algorithm for a given image is a number between 0 to 100, where 0 signifies the best and 100 signifies the worst visibility. An example of image visibility metric calculation can be seen in Fig. 5.6. The importance of this feature can be explained more clearly by the following example - There can be two instances such that the vehicle count is similar. Still, there can be a significant change in the PM concentration due to an external source such as a chimney, construction site, etc. In such cases, the visibility will get affected severely, which can be captured using the visibility score as a feature. This will help the model differentiate between the situations when the traffic and the meteorological conditions are the same.

5.4.2 AQI Estimation using Supervised Learning

This is the final step of the proposed methodology. Till this point, we have obtained a dataset with all the features as shown in Section 5.3.1. Here, a supervised learning-based model is trained on the dataset to predict the AQI category given the following features: temperature, humidity, traffic, and visibility. When the model is fully trained, the obtained weights can be saved. After this, the PM sensor can be removed from the AQSense. The device can predict the AQI by collecting the images and the weather parameters.

5.5 Experiments & Results

Table 5.3 shows the R^2 score of the features with respect to the AQI. It can be seen that all features have a positive correlation with the AQI in the range of 0.2-0.4. This indicates the partial dependence of the AQI on these features and, therefore, justifies the intuition of dependence of the AQI on traffic and visibility as discussed in the Sections 5.4.1. It is a well-known fact that weather conditions have an impact on PM values from the previous findings [56]. As the humidity increases, the dust tends to settle down, which reduces the number of impurities suspended in the air. For the same reason, most of the data points collected during the monsoon season (around 90%) belong to the first two categories.

The ML models were trained and validated for the dataset mentioned in Section 5.3.1. As there are seasonal variations in PM values [56], we trained three different models each for 1. Monsoon dataset (samples collected between Sep'21 - Oct'21) 2. Winter dataset (samples collected between Nov'21 -

	Vehicles	**Temperature**	**Humidity**	**Visibility**
AQI	0.33	0.29	0.31	0.41

Table 5.3 R^2 score of features w.r.t AQI

Method	Monsoon		Winter		Overall	
	Acc	F1	Acc	F1	Acc	F1
SVM	0.86	0.85	0.74	0.72	0.77	0.76
MLP	0.90	0.89	0.78	0.74	0.79	0.78
RF	**0.91**	**0.90**	**0.80**	**0.78**	**0.82**	**0.81**
CNN [42]	0.71	0.70	0.61	0.61	0.67	0.65

Table 5.4 Performance of various methods on overall and season-wise data.

Dec'21), and 3. Overall dataset (combining Monsoon and Winter dataset). The results obtained for all three datasets are presented in Table 5.4. Note that the ML methods (SVM, MLP, RF) were trained on features extracted using the method described in section 5.4.

For the overall dataset, the RF model achieved an accuracy of 82% and an F1 score of 81%. For the data points of monsoon season, it is observed that the RF classifier performs the best with an accuracy of 90.32%. The main reason for this relatively high accuracy is better model training as there is a significantly increased number of data points with low AQI values in monsoon season. Hence, most of the data points belong to the first two categories. This is more evident from Fig. 5.4 that shows the category-wise distribution of the dataset. Classes named "Good" and "Moderate" account for 50% of the collected data points, which leaves significantly less room for the misclassification of the samples. On the other hand, RF is the best performing model for the Winter dataset, with an accuracy of 80.14%. It is relatively low compared to the monsoon season as the data is spread over all categories of AQI, which also increases the chance of misclassification.

To compare our proposed method with the existing work, we applied the method proposed in [42] on each of the three datasets mentioned above and observed an overall improvement of 15%. The main reason behind this improvement is how features from images are extracted. As a traffic image can have different objects, our work emphasizes focusing only on those responsible for air pollution. On the other hand, applying plain CNN on straight-forward images fails to identify this paradigm.

5.6 NDVI Analysis

5.6.1 What is NDVI?

Normalized Difference Vegetation Index (NDVI) is a graphical indicator that indicates the presence of vegetation in a particular area. It is a technique used to classify the land as green, barren, etc. NDVI is calculated on the satellite images of the earth. Vegetation areas absorb red light and reflect the light falling in the near-infra-red region. Based upon the amount of light absorbed/reflected, these two parameters are quantified as Red, and NIR (Near Infra-red), and their normalized difference is taken to obtain the NDVI. This value ranges between -1 to $+1$. If the value is close to -1, there is a high probability that it is a water body, and if the value is close to $+1$, there is a high probability of vegetation in that area. The mathematical formula for calculating NDVI is given in the equation 5.1 -

$$NDVI = \frac{(NIR - Red)}{(NIR + Red)} \tag{5.1}$$

5.6.2 Why NDVI?

It has been found in the previous studies [57, 58] that vegetation is a sound-absorbent of PM. It helps settle the dust and acts as a natural bio-filter against the PM. In this experiment, we collected the GPS coordinates in our dataset and the PM values. Using NDVI, we try to locate the vegetation areas in Hyderabad city and then relate the PM values captured by the sensor. This will give us a better idea of the importance of vegetation in reducing the PM level. Ideally, let us suppose green spaces surround a busy road. In that case, the PM values should be lower than the site with the same level of traffic but without any vegetation land in the vicinity.

5.6.3 NDVI Analysis

5.6.3.1 Observations

Fig. 5.7 shows the plot of NDVI with PM values. The NDVI values range from 0.1 to 0.3 because the GPS coordinates collected are from the main road, and hence, the NDVI values are close to 0. As the data is collected on the streets, it cannot collect the vegetation part. Moreover, it is observed from the graph that the PM values are above 100. This is because the device is present in areas with traffic movement. One more thing that can be observed from Fig. 5.7 is that most of the data points that have high AQI are present in the low NDVI value range. This indicates that in the low vegetation zones, the PM settles less than the places with some traces of vegetation.

5.6.3.2 NDVI as a Feature

We tried to use NDVI as a feature in our dataset, which was shown earlier in section 5.3.1 in this chapter. Since vegetation helps in settling the dust particles, we had the intuition that including NDVI values in the model might help us in training a better model as it will be able to understand the difference in the cases where all other features, namely temperature, humidity, traffic, visibility are same. This is a very realistic scenario where the traffic and weather parameters could be the same. Still, the average PM values might differ due to the magnitude of vegetation cover in the subject areas. The maximum accuracy that we achieved with the help of this additional feature was 65% which is less than the models that do not consider this feature. The results obtained by using various algorithms are shown in Table 5.5

The NDVI-based model provides less accuracy because the data was collected in a moving vehicle on the road. The main road is a source of pollution at every instant time. The car also moves on the

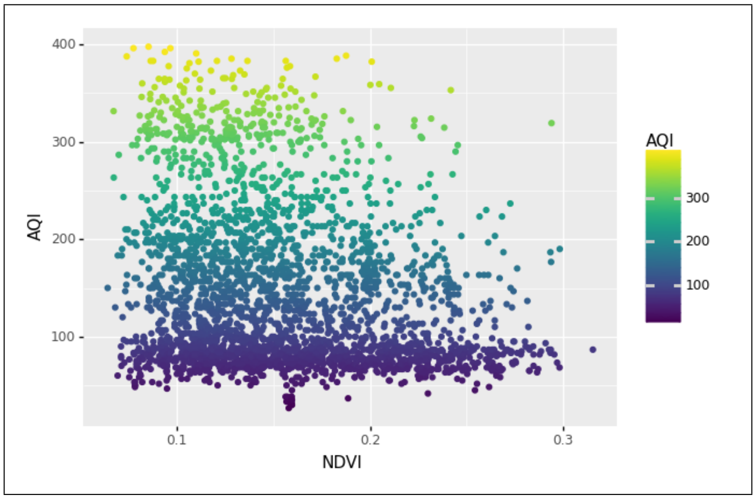


Figure 5.7 NDVI analysis of dataset

Method	Monsoon		Winter		Overall	
	Acc	F1	Acc	F1	Acc	F1
SVM	0.52	0.51	0.48	0.44	0.54	0.57
MLP	0.59	0.59	0.53	0.51	0.53	0.54
RF	**0.65**	**0.64**	**0.62**	**0.62**	**0.60**	**0.61**

Table 5.5 Performance of various methods on overall and season-wise data with NDVI as a feature.

main road all the time. Hence, each observation in the dataset corresponds to the average PM values of the previous 30 s before the observation was recorded, which can be expected to be the new emissions from the vehicles and other sources present around the sensor. Settling of dust requires some time which seems to be insufficient in this case.

The NDVI values would make more sense if the data were collected using a stationary node for a few days. The time series plot of such an experiment would have shown a good trend of settling particles with time. But in our case, we do not have multiple observation points of the exact or approximate GPS coordinates. Hence, we could not get the desired results for AQI prediction using NDVI as a feature.

Chapter 6

Concluding Remarks

6.1 Conclusion

This dissertation discussed the challenges associated with traditional air pollution monitoring systems. A few modern approaches for monitoring AQI include stationery, community, vehicular, and mobile-based IoT solutions. We found that all the modern techniques rely upon low-cost PM sensors, which have some limitations. Image-based AQI determination systems that do not use low-cost PM sensors were also discussed. The two main contributions from this thesis have been presented that address the concerns related to low-cost PM sensors and the current image-based AQI monitoring systems.

First, a comprehensive study was performed to evaluate the low-cost PM sensors. The low-cost PM sensors were able to follow the trend of the reference instrument most of the time with reasonably correlated values. The sensors underestimated the PM values, which can be corrected to some extent by performing calibration. We were able to achieve low E_{rms} values after doing the calibration using a simple linear regression for the indoor experiment. Also, different copies of the same sensor output the PM values in a very close range, indicating a low inter-unit variability for the sensors examined in this paper. In general, it can be concluded that the new low-cost PM sensors can be used for measuring the PM levels, but calibration is required to get a better output. Furthermore, the low-cost PM sensors evaluated in this study were also able to provide reasonably correlated raw values for $PM_{2.5}$ and PM_{10} when exposed to the various types of particles during the outdoor experiment.

Next, a simple yet efficient image-based AQI classification technique on an IoT device (AQSense) using a mixture of ML and DL-based supervised learning algorithms was also discussed. This technique uses low-cost PM sensors only for the training period. Therefore, this model does not rely heavily on the PM sensor for its entire lifetime, and hence, it does not require any special maintenance frequently. Also, the users do not need to carry a separate hardware device for checking the AQI. The only required input from a user's side is an image of the traffic. Experimental results show accuracy up to 90% for the AQI classification. Additionally, a feature-rich dataset was created to be released in the public domain to promote further research.

Related Publications

Conference Papers:

- Ishan Patwardhan, Spanddhana Sara, Sachin Chaudhari"Comparative Evaluation of New Low-Cost Particulate Matter Sensors," *IEEE International Conference on Future Internet of Things and Cloud (FiCloud)*, 2021.
- Nitin Nilesh, Ishan Patwardhan, Jayati Narang, Sachin Chaudhari, "IoT-based AQI Estimation using Image Processing and Learning Methods" *IEEE International Conference on Image Processing (ICIP)*, 2022 *(under review)*.

Patent:

- Ishan Patwardhan, Sachin Chaudhari, Jayati Narang, Nitin Nilesh "Portable Air Quality and Traffic Monitoring Device" India Patent Appl. Num. 202 241 004 682, February, 2022.

Bibliography

[1] *National Air Quality Index*, accessed 23 July 2021, https://app.cpcbccr.com/AQI_India/.

[2] J. S. et al., "Validation of a light-scattering pm2.5 sensor monitor based on the long-term gravimetric measurements in field tests," *PLoS ONE 12(11): e0185700*, 2017.

[3] C. C. Sobin, "A Survey on Architecture, Protocols and Challenges in IoT," *Wireless Personal Communications*, vol. 112, 2020.

[4] "Cisco Annual Internet Report (2018–2023) White Paper," March 2020. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

[5] A. Al-Fuqaha et al., "Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications," *IEEE Communications Surveys Tutorials*, vol. 17, no. 4, pp. 2347–2376, 2015.

[6] P. Landrigan et al., "The Lancet Commission on Pollution and Health," *Lancet*, vol. 391, pp. 464–512, 2018.

[7] M. Badura, P. Batog, A. Drzeniecka-Osiadacz, and P. Modzel, "Evaluation of Low-Cost Sensors for Ambient PM2.5 Monitoring," *Journal of Sensors*, vol. 2018, no. 5096540, pp. 1–16, 2018.

[8] S. Brienza et al., "A Low Cost Sensing System for Coopertaive Air Quality Monitoring in Urban Areas," *Sensors*, vol. 15, pp. 12 242–12 259, 2015.

[9] R. Kamal, *Internet of Things: Architecture and Design Principles*. Mc Graw Hill Education, 2017.

[10] P. Lea, *Internet of Things for Architects*. Birmingham, UK: Packt Publishing Ltd, 2018.

[11] D. Norris, *The Internet of Things*. Mc Graw Hill Education, 2015.

[12] A. Bahgya and V. Madisetti, *Internet of Things: A Hands-on Approach*. Universities Press, 2015.

[13] "Internet Of Things (IoT)." [Online]. Available: https://www.gartner.com/en/information-technology/glossary/internet-of-things

[14] "Internet of Things Global Standards Initiative." [Online]. Available: https://www.itu.int/en/ITU-T/gsi/iot/Pages/default.aspx

[15] R. R. Chinthalapani, "Iot-based air pollution monitoring: Algorithms and implementation," https://web2py.iiit.ac.in/research_centres/publications/view_publication/mastersthesis/1020, 2021.

[16] N. M. Kumar and P. K. Mallick, "The internet of things: Insights into the building blocks, component interactions, and architecture layers," *Procedia Computer Science*, 2018, international Conference on Computational Intelligence and Data Science. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050918309049

[17] *IoT 101 An Introduction to Internet of Things*. Leverage, 2018.

[18] J. Markendahl, S. Lundberg, O. Kordas, and S. Movin, "On the role and potential of IoT in different industries: Analysis of actor cooperation and challenges for introduction of new technology," in *Internet of Things Business Models, Users, and Networks*, 2017, pp. 1–8.

[19] L. D. Xu, W. He, and S. Li, "Internet of Things in Industries: A Survey," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 4, pp. 2233–2243, 2014.

[20] "20 Examples of Wearables and IoT Disrupting Healthcare," Dec 2020. [Online]. Available: https://www.avenga.com/magazine/wearables-iot-healthcare/

[21] R. K. Kodali, G. Swamy, and B. Lakshmi, "An implementation of IoT for healthcare," in *IEEE Recent Advances in Intelligent Computational Systems (RAICS)*, 2015, pp. 411–416.

[22] M. Haghi Kashani, M. Madanipour, M. Nikravan, P. Asghari, and E. Mahdipour, "A systematic review of IoT in healthcare: Applications, techniques, and trends," *Journal of Network and Computer Applications*, vol. 192, p. 103164, 2021.

[23] D. Metcalf, S. T. Milliard, M. Gomez, and M. Schwartz, "Wearables and the Internet of Things for Health: Wearable, Interconnected Devices Promise More Efficient and Comprehensive Health Care," *IEEE Pulse*, vol. 7, no. 5, pp. 35–39, 2016.

[24] F. John Dian, R. Vahidnia, and A. Rahmati, "Wearables and the Internet of Things (IoT), Applications, Opportunities, and Challenges: A Survey," *IEEE Access*, vol. 8, pp. 69 200–69 211, 2020.

[25] X.Krasniqi and E.Hajriz, "Use of IoT Technology to Drive the Automotive Industry from Connected to Full Autonomous Vehicles," *17th IFAC Conference on International Stability, Technology and Culture TECIS 2016*, vol. 49, no. 29, pp. 269–274, 2016.

[26] B. V. Philip, T. Alpcan, J. Jin, and M. Palaniswami, "Distributed Real-Time IoT for Autonomous Vehicles," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 2, pp. 1131–1140, 2019.

[27] D. Minovski, C. Åhlund, and K. Mitra, "Modeling Quality of IoT Experience in Autonomous Vehicles," *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 3833–3849, 2020.

[28] M. Tagle et al., "Field performance of a low-cost sensor in the monitoring of particulate matter in Santiago, Chile," *Environmental Monitoring and Assessment*, vol. 192, no. 171, p. 18, Feb. 2020.

[29] P. Pant et al., "Monitoring particulate matter in India: Recent trends and future outlook," *The Verge*, vol. 12, no. 1, pp. 45–58, Jan. 2019.

[30] J. Schwab, H. D. Felton, O. V. Rattigan, and K. L. Demerjian, "New york state urban and rural measurements of continuous pm2.5 mass by fdms, teom, and bam," *J. of Air & Waste Manage. Assoc.*, vol. 56, no. 4, pp. 372–383, 2006.

[31] C. Rajashekar, T. Mukku, A. Dwivedi, A. Rout, S. Chaudhari, K. Vemuri, K. S. Rajan, and A. M. Hussain, "Improving spatio-temporal understanding of particulate matter using low-cost iot sensors," in *Int. Symp. Personal, Indoor and Mobile Radio Commun.*, 2020, pp. 1–7.

[32] L. A. Q. Network, "London air quality network—summary report 2018," *Techinal Report Environmental Research Group, King's College London: London, UK*, 2018.

[33] R. N. Murty et al., "CitySense: An Urban-Scale Wireless Sensor Network and Testbed," in *IEEE Conference on Technologies for Homeland Security*, 2008, pp. 583–588.

[34] A. Rebeiro-Hargrave et al., "MegaSense: Cyber-Physical System for Real-time Urban Air Quality Monitoring," in *15th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, 2020, pp. 1–6.

[35] D. Hasenfratz, O. Saukh, S. Sturzenegger, L. Thiele, *et al.*, "Participatory air pollution monitoring using smartphones," *Mobile Sensing*, vol. 1, pp. 1–5, 2012.

[36] A. Anjomshoaa et al., "City Scanner: Building and Scheduling a Mobile Sensing Platform for Smart City Services," *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 4567–4579, 2018.

[37] S. Mora et al., "Towards Large-scale Drive-by Sensing with Multi-purpose City Scanner Nodes," in *IEEE 5th World Forum on Internet of Things (WF-IoT)*, 2019, pp. 743–748.

[38] J. S. Apte et al., "High-Resolution Air Pollution Mapping with Google Street View Cars: Exploiting Big Data," *Environmental Science & Technology*, vol. 51, no. 12, pp. 6999–7008, 2017.

[39] A. S. Gilik, A. and A. Ozmen, "Air quality prediction using cnn+ lstm-based hybrid deep learning architecture," *Environmental Science and Pollution Research*, 2021.

[40] X. Zou, J. Zhao, D. Zhao, B. Sun, Y. He, and S. Fuentes, "Air quality prediction based on a spatiotemporal attention mechanism," *Mobile Information Systems*, 2021.

[41] J. Ma, Z. Li, J. C. Cheng, Y. Ding, C. Lin, and Z. Xu, “Air quality prediction at new stations using spatially transferred bi-directional long short-term memory network,” *Science of The Total Environment*, 2020.

[42] A. Chakma, B. Vizena, T. Cao, J. Lin, and J. Zhang, “Image-based air quality analysis using deep convolutional neural network,” in *IEEE International Conference on Image Processing (ICIP)*, 2017.

[43] W. Y. Yi et al., “A Survey of Wireless Sensor Network Based Air Pollution Monitoring Systems,” *Sensors*, vol. 15, no. 12, pp. 31 392–31 427, 2015.

[44] “Working of a pm sensor,” https://sensirion.com/products/product-insights/specialist-articles/particulate-matter-sensing-for-air-quality-measurements/.

[45] A. Manikonda, N. Zíková, P. Hopke, and A. Ferro, “Laboratory assessment of low-cost pm monitors,” *J. Aerosol Sci.*, vol. 102, pp. 29–40, 2016.

[46] K. Kelly, J. Whitaker, A. Petty, C. Widmer, A. Dybwad, D. Sleeth, R. Martin, and A. Butterfield, “Ambient and laboratory evaluation of a low-cost particulate matter sensor,” *Environmental Pollution*, vol. 221, pp. 491–500, 2017.

[47] “Series- 500 brochure,” https://www.aeroqual.com/wp-content/uploads/Aeroqual-Portable-Monitors-Long-Brochure-LR.pdf, (Accessed on 03/19/2021).

[48] “Portable pm colocation study,” https://www.aeroqual.com/wp-content/uploads/Portable-PM-Colocation-Study.pdf, (Accessed on 03/19/2021).

[49] A. Bucchianico, “Coefficient of determination,” *Encyclopedia of Statistics in Quality and Reliability*, 03 2008.

[50] C. Brown, *Applied Multivariate Statistics in Geohydrology and Related Sciences*. Springer, Berlin, Heidelberg, 1998.

[51] C. Liu, F. Tsow, Y. Zou, and N. Tao, “Particle pollution estimation based on image analysis,” *PloS one*, 2016.

[52] “Prana air,” https://www.pranaair.com/air-quality-sensor/outdoor-pm-sensor/.

[53] I. Patwardhan, S. Sara, and S. Chaudhari, “Comparative evaluation of new low-cost particulate matter sensors,” in *8th International Conference on Future Internet of Things and Cloud (FiCloud)*, 2021.

[54] A. Mittal, A. Moorthy, and A. Bovik, “No-reference image quality assessment in the spatial domain,” *IEEE Transactions on Image Processing*, 2012.

[55] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, and F. Carli, M.and Battisti, "Tid2008-a database for evaluation of full-reference visual quality assessment metrics," *Advances of Modern Radioelectronics*, 2009.

[56] P. Chrysikou and A. Samara, "Seasonal variation of the size distribution of urban particulate matter and associated organic pollutants in the ambient air," *Atmospheric Environment*, 2009.

[57] A. Diener and P. Mudu, "How can vegetation protect us from air pollution? a critical review on green spaces' mitigation abilities for air-borne particles from a public health perspective - with implications for urban planning," *Science of The Total Environment*, vol. 796, p. 148605, 2021.

[58] L. Chen, C. Liu, R. Zou, and Z. Yang, M.and Zhang, "Experimental examination of effectiveness of vegetation as bio-filter of particulate matters in the urban environment," *Environmental Pollution*, vol. 208, pp. 198–208, 2016.