INDOOR AIR POLLUTION PREDICTION

AND PREVENTION USING IOT

AND MACHINE LEARNING TECHNIQUES

by

Vamshidhar Darla MSc (Data Science)

DISSERTATION

Presented to the Swiss School of Business and Management, Geneva

In Partial Fulfilment

Of the Requirements

For the Degree

Global Doctorate of Business Administration

SWISS SCHOOL OF BUSINESS AND MANAGEMENT GENEVA

April, 2020

RESEARCH ON INDOOR AIR POLLUTION PREDICTION AND PREVENTION USING IOT AND MACHINE LEARNING TECHNIQUES

by

Vamshidhar Darla MSc (Data Science)

Supervised by

MONIKA SINGH

RECEIVED/APPROVED BY:

Renee Goldstein Osmic

Admissions Director

APPROVED BY

Apostolos Dasilas

Dissertation chair

INDOOR AIR POLLUTION PREDICTION AND PREVENTION USING IOT

AND MACHINE LEARNING TECHNIQUES

by

Vamshidhar Darla MSc(Data Science)

Dissertation Chair: <Chair's Name>

Co-Chair: <If applicable. Co-Chair's Name>

ABSTRACT

Monitoring indoor air quality (IAQ) has become increasingly important as people spend a significant amount of time indoors, whether at indoor environments (home, office, Conf rooms, Auditoriums). Using a comprehensive system that integrates IoT sensors and machine learning techniques offering an effective way to ensure healthier indoor atmosphere. This system consists of several components and steps. IOT sensor devices are deployed to measure parameters such as CO2, PM2.5, PM10, VOC, Temperature and humidity concentration levels in various indoor space which includes bedrooms, living rooms and Kitchens. These sensors continuously collect real time IAQ data from the sensors. The collected data from the sensors are then transmitted to a central microcontroller device, which acts as an aggregation point and responsible for preprocessing the data, performing initial filtering, or smoothing if necessary and package it to transmit to central storage (cloud). Reliable communication protocols such as Wi-Fi, Bluetooth are used to send the data from the microcontroller to a central server for further processing. The collected data is securely stored in scalable storage solutions like cloud-based servers (ThingSpeak , AWS, Azure) or local databases ensuring the data integrity, availability and security

For User access, user friendly Grafana dashboard is developed to visualize IAQ data in real time. Authorized users can access this dashboard to monitor IAQ statistics, view historical trends and receive alerts if any parameters exceed safe readings. Machine learning algorithms are applied to analyze the IAQ trends in the data. Techniques like regression are used to predict future IQ parameters taking into consideration of historical data and different factors such as time of the day, occupancy. Different classifications algorithms categorize IAQ into levels such as good, moderate, poor along with providing appropriate recommendations.

Alerts and notifications are implemented to inform users in real time if IAQ parameters reach critical levels or if preventive actions are needed. The systems is continuously improved by collecting the user feedback, which in used to fine tune the ML models and enhance the preventive measures.

This approach can contribute to healthier and more comfortable indoor environments in homes and offices while also helping to reduce health risks associated with poor IAQ

TABLE OF CONTENTS

ABSTRACT 5
List Of Tables 8
List of Figures 9
DEFINITION OF TERMS 9
CHAPTER 1 INTRODUCTION 11
1.1. Introduction 11
1.2. Research Problem 13
1.3. Purpose of Research 15
1.4. Significance of the Study 18
1.5. Research Purpose and Questions 26
1.6. Research Objectives 27
1.6. Specific Objectives 27
1.7. Significance and Contributions 28
1.8. Organization Of Thesis 30
CHAPTER 2 LITERATURE REVIEW 31
2.1. Theoretical Framework 31
2.2. Theory of Reasoned Action 37
2.3. Human Society Theory 44
CHAPTER 3 RESEARCH DESIGN 60
3.1. Overview of Research Problem 61
3.2. Operationalization of Theoretical Constructs 61
3.2.1. Research Approach 61
3.2.2. Research Philosophy 61
3.2.3. Research Strategy 61
3.3. Research Purpose 62
3.3.1. Overall System Framework 62
3.3.2. IoT Sensor Network Design 63
3.3.3. Data Management Architecture 63

3.4. Research Design 63
3.4.1. Machine Learning Framework 63
3.4.2. Feature Engineering 64
3.4.3. Model Training and Validation Strategy 64
3.5. Data Collection Methodology 65
3.5.1. Primary Data Sources 65
3.5.2. Secondary Data Sources 65
3.5.3. Data Quality Assurance 66
3.6. Experimental Design 66
3.6.1. Controlled Experiments 66
3.6.2. Field Validation Experiments 67
3.6.3. User Acceptance Testing 67
3.7. Performance Evaluation Metrics 67
3.7.1. Prediction Accuracy Metrics 67
3.7.2. System Performance Metrics 68
3.7.3. Environmental Impact Metrics 68
3.8. Ethical Considerations and Limitations 68
3.8.1. Ethical Framework 68
3.8.2. Research Limitations 69
3.9. Timeline and Resource Allocation 69
3.9.1. Project Timeline 69
3.9.2. Resource Requirements 70
3.10. Summary 70
3.10.1. SENSORS 71
4. Required resources of Air Quality Monitoring System: Required Resources 99
5. Conclusion 101
CHAPTER 4 **DATA ANALYSIS** 103
4.1. Research Question 1 103
4.1.1. Parameter Analysis 103
4.2. Research Question 2 107
4.2.1. PAS-IN-01 Particulate Matter Sensor 107
4.2.2. Evelta SHT4X+SGP40 Multi-Parameter Environmental Sensor 108
4.2.3. NDIR CO2 Sensor 109

4.3. Research Question 3 110
4.4. Conclusion 118
CHAPTER 5 RESULTS AND DISCUSSION 121
5.1. Introduction 121
5.1.1 Visual representation of the Data 127
5.1.2. Important Stats OLS Regression Results. OLS 131
5.2. Indoor Air Quality Monitoring Research Results(Discussion) 132
5.2.1. Discussion of Research Question 1 132
5.2.2. Discussion of Research Question 2 133
5.2.3. Discussion of Research Question 3 134
5.3. Conclusion 135
CHAPTER 6 **Future Scope and Recommendations** 136
6.1. Future Scope 136
6.1.1 Advanced Machine Learning and AI Integration 136
6.1.2. Smart Building Integration 136
6.1.3. Health-Centric Applications 137
6.1.4. Environmental and Urban Planning Integration 137
6.2. Recommendations 138
6.2.1. Technical Implementation 138
6.2.2. Regulatory and Compliance 138
6.2.3. Economic and Market Considerations 139
6.2.4. User Adoption and Education 139
6.2.5. Research and Development Priorities 139
6.2.6. Strategic Partnerships 140
6.3. Conclusion 140

List Of Tables

Table 1 Data Before Calibration 93
Table 2 Parametric values After Calibration 94
Table 3 Data Collected From The 3 Nodes In 3 Days 95
Table 4 Indoor Air Pollution Index 100
Table 5 Data retrieved from the sensors through the cloud 106
Table 6 : Sample few records collected from the sensors. 106

Table 7 118

List of Figures

Figure 1 PM Sensor 72
Figure 2 Co2 Sensor........ 72
Figure 3 VOC ,Temperature and Humidity Sensor........ 73
Figure 4 ESP32 MicroController 81
Figure 5 Sensor Connections........ 90
Figure 6 Sensors in Network 91
Figure 7 Circuit Diagram 92
Figure 8 Initially built Device 93
Figure 9 **Aeroqul series 500 indoor air quality measuring devices**........ 93
Figure 10 Final product........ 94
Figure 11 flow diagram 96
Figure 12 IAQ Indoor Air Quality........ 100
Figure 13 Master Controller architecture........ 101
Figure 14 : shows the Dataset collected mapping to the sensor and pollutants 107
Figure 15 Pollutant over week days 111
Figure 16 shows corelation between the pollutants........ 111
Figure 17 day-wise........ 113
Figure 18 115
Figure 19 116
Figure 20 120

List of Equations

Equation 1 98
Equation 2 98

DEFINITION OF TERMS

ANN	Artificial Neural Network
CO	Carbon Monoxide
CO2	Carbon Dioxide
DM	Data Mining
LR	Linear Regression
NO	Nitrogen Oxides
NO2	Nitrogen Dioxide

PAE	Predictive Analysis Engine
PM	Particulate Matter
PM10	particulate matter with a diameter smaller than 10 μm
PM2.5	particulate matter with a diameter smaller than 2.5 μm
SL	Supervised Learning
USL	Unsupervised Learning

CHAPTER 1
INTRODUCTION

1.1.Introduction

Indoor air pollution is now one of the most serious environmental health issues of the 21st century, with profound consequences on human well-being and health worldwide. Unlike outdoor air pollution, which has been extensively researched and controlled, indoor air quality (IAQ) is still largely unmonitored and unregulated even though people spend approximately 90% of their lives indoors. The World Health Organization estimates that indoor air pollution results in approximately 3.8 million premature deaths annually, and thus, it is a silent and deadly environmental hazard(Bhardwaj & Sharma, 2021).

Indoor air pollution is the presence of dangerous pollutants in the indoor air of homes' and buildings'. It is caused by several factors such as poor ventilation, chemicals off-gassing from building components and furniture, fuel combustion for cooking and heating, and use of pesticides and cleaning products. On the other hand, outdoor air pollution is the presence of dangerous pollutants in the outdoor environment, such as in the air, in the adjacent streets or industrial areas, and other public access areas. Outdoor air pollution results from sources such as exhaust from automobiles, industrial emissions, and emissions from power generation. Indoor air composition is regulated by factors such as indoor emission sources, occupant activities, ventilation systems, indoor air infiltration, and building components. Part of the indoor pollutants are particulate matter (PM2.5 and PM10), volatile organic compounds (VOCs), carbon monoxide (CO), carbon dioxide (CO_2), nitrogen dioxide (NO_2), formaldehyde, radon, and biological pollutants such as bacteria and mold spores. These indoor pollutants may be due to cooking, cleaning products, off-gassing from furniture, tobacco smoke, combustion appliances, and inadequate ventilation systems(Wei et al., 2020).

Traditional indoor air quality control methods have been largely reactive in nature, relying on periodic manual sampling and post-exposure health testing. With the advent of Internet of Things (IoT) technology and machine learning algorithms, however, unprecedented opportunities today abound for continuous real-time monitoring, predictive modeling, and preemptive intervention strategies. This intersection of the technologies makes it possible to develop smart systems capable of continuously monitoring air quality drivers, predicting pollution events, and imposing protection measures automatically(Matthaios et al., 2024).

One of the main differences between indoor and outdoor air pollution is the level of control over sources of pollution that people have. While people may have some ability to reduce indoor air pollution, for instance, by using air cleaners or reducing use of chemical cleaners, people may have little control over outdoor sources of pollution, for instance, traffic or industrial emissions. Indoor air pollution can be due to a very wide variety of sources, including: Tobacco smoke: Cigarette smoke is an important indoor air pollutant and can cause a variety of health effects, including asthma, lung cancer, and heart disease(Wei et al., 2020). Combustion appliances: Gas stoves, heaters, and fireplaces release pollutants such as carbon monoxide, nitrogen dioxide, and particulate matter. Building materials and furnishings: Most building materials, for instance, paint, adhesives, and carpeting, release volatile organic compounds (VOCs) into the air. Furniture, drapes, and other household items can also release VOCs from cleaning products:

Most cleaning agents contain chemicals that can be a source of indoor air pollution. Mold and mildew: Indoor water problems can grow mold and mildew, and these can cause respiratory symptoms and other health impacts like. Pet: Dander from pets and other allergens can create indoor air pollution and Radon: It is a radioactive gas that can seep into homes from the ground and cause lung cancer(Wei et al., 2020).

It is necessary to recognize the sources of indoor air pollution in order to work on steps to minimize the exposure to the harmful pollutants and ensure the well-being of individuals who are indoors in buildings and homes. Target is to gather IoT sensor data analyze the pattern of the data drive the pollution at various times in various locations of the house/office rooms. Through analysis and the pattern prescribe the prevention procedures in order to provide a healthy atmosphere for the people to live. This method can also be beneficial to minimize indoor air pollution (IAP) in restaurants, Hospitals and in any closed room spaces(Rajabi et al., 2021).

1.2. Research Problem

Indoor air pollution is the most ubiquitous but underappreciated environmental health risk worldwide of the modern era, silently affecting billions of people worldwide who remain indoors for the majority of their lives. This pervasive threat has evolved from a relatively modest problem to a severe public health issue, demanding leading-edge technological interventions and global prevention strategies. Indoor air pollution is now identified by the World Health Organization as the fourth most predominant risk factor for global disease burden, causing an estimated 3.8 million premature deaths annually and imposing vast morbidities across all ages and socioeconomic strata(Omidvarborna et al., 2021).

Historical Evolution of Indoor Air Quality Awareness

The study of indoor air pollution started early in the 20th century with the initial identification of the health risks posed by air pollution in the working environments of factories and mines by industrial hygienists. The indoor air quality in the home and office remained largely ignored until the 1970s energy crisis brought a sharp deviation in building designs. The emphasis on energy conservation translated into designing more air-tight buildings to reduce heat loss; this well-intentioned efficiency measure had the unintended consequence of providing the conditions for the formation of enclosures that were trapping pollutants and diminishing natural ventilation(Katsura et al., 1996).

The 1980s were the years when indoor air quality consciousness reached a turning point with the identification of "sick building syndrome," a syndrome in which occupants of modern office buildings began to develop acute health symptoms directly traceable to their indoor environment. Concurrently, the discovery that radon is an indoor carcinogen present in common indoor spaces shocked the scientific community and the public, with implications that homes in large geographical areas were contaminated with this naturally occurring radioactive gas. The 1990s were years of greater understanding of volatile organic compounds (VOCs) and their widespread presence in common household products, and the 2000s were years of increasing concern with biological pollutants, mold contamination, and the critical role of moisture management in maintaining healthy indoor environments(Kovalenko et al., 2022).

The 2020-2022 COVID-19 pandemic brought indoor air quality to the global forefront and underscored the paramount importance of ventilation, air filtration, and airborne pathogen control. The crisis exposed the performance limitations of current indoor air management systems and spurred investment in sophisticated monitoring and control technology, resulting in today's revolution of smart air quality management systems.

Economic Burden and Financial Implications

The economic expense of indoor air quality is staggering enough to influence all aspects of society, from the family dwelling to the economy of the nation. In the United States alone, the direct health costs of indoor air pollution are more than $150 billion annually, comprising medical treatment for asthma, chronic obstructive pulmonary disease, respiratory infections, cardiovascular disease, and several cancers. These direct costs are only a fraction of the total economic burden, as indirect costs more than double the entire burden. Loss of productivity from indoor air quality illnesses is estimated to be another $60-80 billion economic loss annually. Workers exposed to poor air quality conditions experience a 6-9% loss of work performance and 1.5-2 times higher absenteeism compared to workers in well-ventilated areas. The real estate sector loses significant value as buildings with a history of air quality problems experience losses in market value of 10-15% and increased vacancies(Obiweluozo et al., 2022).

Schools have particularly dire economic ramifications, as decreased classroom air quality has a direct link to reduced student performance, higher numbers of sick days among teachers, and more special education referrals. Research shows that enhanced school air quality has been proven to boost student test scores by 5-10% and lower respiratory illness-related absenteeism by as much as 30%, saving millions of dollars in enhanced education and minimized healthcare costs(Correa-Morales et al., 2019).

Hospitality and retail sectors report substantial losses in revenue due to indoor air quality problems, as customers spend fewer dollars and minutes in facilities that show obvious air quality problems. Conversely, businesses that invest in sophisticated air quality systems experience higher customer satisfaction, increased customer stay, and improved employee retention, thereby giving proof of economic gains in being proactive with air quality management(Correa-Morales et al., 2019).

1.3. Purpose of Research

The regulatory regime for indoor air quality is defined by a diverse array of standards, guidelines, and regimes of enforcement. These vary extensively by jurisdiction and by building type. Indoor air is more fragmented than outdoor air quality, which is controlled by broad national standards like the United States Clean Air Act. The United States Environmental Protection Agency (EPA) releases recommendations and guidelines to improve indoor air quality but does not have broad regulatory jurisdiction over most indoor environments. The Occupational Safety and Health Administration (OSHA) issues air quality workplace standards but only in professional settings and has a tendency to focus on acute exposure limits rather than long-term health. The American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) developed major ventilation standards (most prominently Standard 62.1 and 62.2), which are predominantly included in building codes but still experience spotty compliance enforcemen(Mendoza et al., 2021)t.

Internationally, the World Health Organization has issued air quality standards for indoor environments, and the European Union has set more stringent building ventilation standards and energy performance standards tangentially related to indoor air quality. The Nordic countries are at the forefront of comprehensive regulation of indoor air quality, including precise standards of building materials, ventilation rates, and allowable levels of pollutants(ISLAMI et al., 2020).

This regulatory disparity is most obvious in homes, where most indoor air quality regulations are voluntary guidelines, not mandatory ones. This lack of regulation has created an urgent need for new monitoring and prevention technologies that provide objective data to enable evidence-based policy and assist in efforts at voluntary compliance(Mendoza et al., 2021).

Current Prevention Technologies and Limitations

Conventional indoor air quality control is based on integrated source control, ventilation enhancement, and air cleaning technologies, each with built-in limitations that help to make the case for sophisticated integrated solutions. Source control approaches, optimally in theory, are plagued by practical difficulties in locating and removing all contaminant sources in the dynamic indoor conditions where there are multiple emission sources interacting(Chen et al., 1999).

Mechanical ventilation systems, which are intended to dilute indoor contaminants with outside air, usually run on fixed schedules independent of the actual air quality situation, wasting energy under good air quality conditions and not providing adequate ventilation under pollution incidents. The majority of the current systems do not have sensors and control algorithms to react to dynamic real-time air quality situations and use timer-based control or simple occupancy sensors instead(Yoda et al., 2017).

Portable air cleaners and home filtration systems are good at eliminating certain pollutants but tend to run continuously without consideration of true levels of contamination, leading to wasted energy and filter replacement. There is a predominance of run-of-the-mill air cleaning devices which work to eliminate only one kind of pollutant, i.e., gases or particles, with other pollutants unaffected.In contemporary commercial office buildings, building automation systems increasingly include simple air quality sensors; however, these systems only measure carbon dioxide concentrations as an indirect proxy for overall air quality, missing a variety of other significant pollutants. This inability to provide complete, real-time monitoring capability denies building operators complete insight into the entire range of air quality issues and the ability to establish correct interventions(McCarrick et al., 2024).

Moreover, current technologies are mainly intended to function independently, not to encompass integration and cognitive capabilities to enhance performance on various air quality measures in aggregate. This piecemeal strategy tends to create antagonistic systems functioning at cross purposes, e.g., air cleaners operating at full capacity while ventilation systems bring in outside pollutants(Chikwem et al., 2022).

Global Statistics and Regional Variations

Indoor air pollution is a global problem, albeit one that varies tremendously geographically, by economic status, and by culture. The worst indoor air pollution problem occurs in developing countries, with an estimated 2.8 billion individuals who use biomass fuels for cooking and home heating, producing very high levels of particulate matter, carbon monoxide, and other combustion products within the home.

Sub-Saharan Africa and South and Southeast Asia account for the highest rates of indoor air pollution deaths, with indoor PM2.5 levels above over 500 μg/m³ during cooking times—over 20 times over the WHO standards. Women and children in these areas are disproportionately affected by gender roles that keep them working long hours beside cooking stoves and fires. Industrialized countries have different but equally ominous trends, and indoor air pollution is the leading cause primarily from chemical pollutants linked to building materials, consumer items, and space heaters. Urban locations in developed countries have the additional burden of outdoor air pollution leaking in along with pollutants created indoors, which creates difficult exposure scenarios(Kawakami et al., 2017).

Climate change is remapping global indoor air quality trends. More intense wildfire combustion is influencing urban indoor spaces thousands of miles away from fire sources, and extreme weather events are compelling more reliance on mechanical systems and closed buildings. Arctic areas have unique problems with extremely tight building envelopes built for energy efficiency, leading to high indoor pollutant concentrations during winter months when natural ventilation is impractical(Kanagasabai et al., 2023).

Local building codes, economic conditions, and cultural practices give rise to unique air quality challenges that must be addressed with localized technologies. Mold due to humidity is prevalent in Mediterranean nations, and dusty regions experience dust penetration and evaporative cooling system contamination(Baqer et al., 2022).

Vulnerable Populations and Health Disparities

Indoor air pollution disproportionately impacts vulnerable groups, exacerbating and perpetuating current health inequities among population groups. Children are the most susceptible group because of their developing respiratory apparatus, higher respiratory rates, and longer indoor duration of exposure, making them uniquely susceptible to the negative health consequences linked to air quality. Pediatric asthma is directly linked with indoor air quality, with children living in low-quality air homes experiencing double the respiratory morbidity and emergency department visits(Rajabi et al., 2021).

Aging populations face heightened exposures due to the loss of respiratory function linked with increasing age, compromised immune function, and a tendency for extended indoor residence. Older adults living in low-income housing generally face the most detrimental indoor air quality conditions due to compromised building envelopes, inadequate ventilation systems, and a lack of resources allocated to improving air quality. Pregnant women are also a critical vulnerable group, as indoor air pollution during pregnancy has been linked to low birth weight, preterm delivery, and developmental problems. More recent evidence suggests that prenatal exposure to certain indoor pollutants can impact child development and subsequent health status(Gabriel & Auer, 2023).

People who already have respiratory diseases, cardiovascular disease, or immune-compromised conditions are more vulnerable to indoor air pollution and may need to employ special air quality management practices in addition to those advised for the general population. These people are usually not able to afford sophisticated air quality monitoring and control equipment based on economic considerations. Socioeconomic inequalities produce large disparities in indoor air quality exposure, with lower-income groups tending to live in older buildings with poor ventilation, higher-emitting construction materials, and lower capacity to enact air quality changes. Environmental justice issues emphasize how marginalized groups suffer from disproportionate indoor air pollution and have the least access to mitigation opportunities(Adeleke et al., 2017).

1.4. Significance of the Study

Building Design Factors and Architectural Influences

Contemporary building design practices have a substantial impact on indoor air quality by involving intricate interdependencies among architectural choice, mechanical design, and materials choice. The movement towards more building envelope tightness for energy conservation has had a profound impact on indoor air dynamics, decreasing natural air exchange rates and, through mechanical ventilation compensation failure, bringing the possibility of pollutant concentration. Open floor plan layouts, though favored for their visual beauty and rational use of space, can allow for the quick diffusion of contaminants across wide areas and thus make localized source control more difficult. Compartments, however, can keep pollutants confined to particular areas while permitting more focused ventilation measures(Omidvarborna et al., 2021).

Material selection is highly important in indoor air quality since most modern building materials, furniture, and finishes off-gas volatile organic compounds for months or years after installation. The "off-gassing" process has driven interest in low-emitting materials and green building ratings, although long-term confirmation of emissions performance continues to be problematic. HVAC system design choices immediately and directly impact air quality results, including duct design, filter properties, moisture control capacity, and control system complexity that impact overall air quality performance. Most buildings have oversized or undersized HVAC systems with no capability to provide best air quality conditions under changing occupancy and weather conditions.Natural ventilation systems, more frequently integrated into green building designs, involve detailed attention to local climate, outside air quality, and building orientation to prevent the addition of outside pollutants without delivering sufficient fresh air. Mixed-mode systems that blend natural and mechanical systems provide potential benefits but necessitate sophisticated control systems to optimize performance.(Adeleke et al., 2017)

The use of intelligent building technology is beginning to transform the control of air quality, as advanced sensors, machine learning algorithms, and automated controls allow for real-time monitoring and optimization of air quality conditions continuously. But such complexity requires specialized expertise to design, install, and maintain effectively.

Emerging Pollutants and Contemporary Challenges

The indoor air pollution environment is constantly changing with new materials, products, and technologies that emit previously unobserved sources of contamination. Flame retardants applied to furniture and electronics have become important indoor pollutants, with these persistent chemicals concentrating in house dust and possibly changing endocrine system function. Electronic appliances and 3D printing emit ultrafine particles and new chemical compounds that did not exist in indoor environments a decade or two ago. Wireless device penetration has been causing concern regarding exposure to electromagnetic fields, and materials used to produce the devices lead to indoor chemical emissions(Zhang et al., 2023).

Nanomaterials employed in consumer products like antimicrobial treatments, stain-resistant treatments, and premium filtration media present unquantifiable health hazards since their long-term effect on human health is under investigation. Nanomaterials, owing to their minuscule size, travel deep into respiratory tracts and are likely to enable transfer to other organ systems. Global warming adds new challenges in the form of severe weather that leads to higher air conditioning and air-tight building space, potentially elevating indoor pollutant levels. Penetration of wildfire smoke is a more prevalent indoor air quality problem that necessitates sophisticated filtration and pressure control systems(Zhang et al., 2023).

Indoor cultivation of cannabis, whether for medicinal or recreational use, has added new issues to indoor air quality related to equipment used in cultivation, fertilizers, and processing activities. Volatile organic compounds associated with cannabis cultivation and use create unique challenges in indoor air quality control. The COVID-19 pandemic underscored the need for control of air-borne pathogens, and utilization of UV-C disinfection units, high-end filtration systems, and antimicrobial treatments has grown, each of which can potentially have its own effect on indoor air quality(Cho, 2020).

Healthcare System Burden and Medical Implications

The medical system is overwhelmed by diseases caused by indoor air pollution because emergency rooms, primary care physicians, and specialists treat millions of cases annually that are either directly or indirectly linked to indoor air quality. Pediatric emergency departments report that 15-20% of all visits are for respiratory distress, and they are mostly caused by indoor air quality issues at home, school, or day care.

Asthma management is one of the most significant indoor air quality-related health costs, with more than $80 billion per year spent in the United States on medications alone. Asthma children living in housing with poor indoor air quality require double the level of emergency department use and hospitalization, thus placing significant financial burden on families and the healthcare system. Acute exacerbations of chronic obstructive pulmonary disease (COPD) often occur in conjunction with indoor air quality deterioration, resulting in increased medication use, emergency services, and reduced quality of life for millions. Increasingly, healthcare providers are recognizing the value of adding environmental interventions to treatment plans; however, no effective methods for assessment and monitoring of patients' exposure to indoor air quality are available.

The cardiovascular effects of indoor air pollution are increasingly being noted as studies had determined associations between long-term exposure to fine particulate matter and some chemicals with high risks of heart disease, stroke, and high blood pressure. Although the healthcare system is increasingly embracing these associations, there is still a deficiency in integrating systematic strategies to treat environmental risk factors in cardiovascular medicine.

The impact of indoor air quality on mental health is becoming increasingly well established as a major concern, as poor air quality has been linked to elevated levels of depression, anxiety, and cognitive dysfunction. The medical community is only beginning to acknowledge such correlations and implement corresponding intervention strategies.

Technological Evolution in Air Quality Monitoring

The development of air quality monitoring technology has transitioned from costly, laboratory-quality equipment that demands the expertise of trained technicians to the establishment of early-stage networks of low-cost, internet-operated sensors that enable real-time continuous monitoring. Early air quality monitoring was based on passive sampling techniques that yield cumulative exposure data over weeks or days, thereby offering little information on temporal fluctuations and acute exposure episodes. Advances in portable direct-reading instruments during the 1990s made active monitoring more feasible, but they remained costly and required extensive expertise to operate and interpret results. The instruments primarily addressed a single pollutant or small groups of parameters, thereby offering piecemeal representations of general air quality conditions.

Recent developments in sensor technology have significantly lowered the expense and sophistication of air quality monitoring, at the same time broadening the scope for measurable parameters. Electrochemical sensors, photoionization detectors, laser particle counters, and metal oxide sensors now offer economically viable solutions for continuous, simultaneous monitoring of numerous air quality parameters. Internet of Things (IoT) connectivity has transformed air quality monitoring through remote access to data, cloud-based data analysis and storage, and access to building automation systems. Wireless sensor networks now are capable of offering large spatial and temporal air quality data that were not accessible before through traditional monitoring methods.

Machine learning and artificial intelligence technologies are increasingly transforming raw sensor data into useful insights through the use of algorithms that can identify sources of pollution, predict air quality trends, and optimize the efficiency of building system responses. These technologies are opening the way for the next generation of advanced indoor air quality management systems.

The synergy of multiple sensor modalities, advanced data analytics, and computerized control systems is the cutting-edge air quality technology, providing unmatched opportunity for monitoring, forecasting, and controlling indoor air pollution. Exploitation of this potential, however, involves overcoming challenges of sensor precision, data fusion, system dependability, and cost-effectiveness across a variety of building types and uses.

Residential indoor air composition is impacted by numerous factors such as outdoor air infiltration, indoor source emissions, ventilation systems, occupant activities, and building materials, and thus creates a complicated matrix of pollutants that may result in serious health hazards for occupants. Particulate matter, both PM2.5 and PM10, is caused by combustion activities such as burning gas and wood, tobacco smoke, dust, pet dander, mold spores, and outdoor air infiltration, resulting in respiratory irritation, coughing, sneezing, bronchitis, asthma exacerbation, heart disease, and lung cancer, with the PM2.5 being extremely harmful since it can penetrate deep into the lungs and get into the bloodstream. Volatile organic compounds (VOCs) pervade indoor environments, caused by paints, varnishes, cleaning supplies, furniture, carpets, adhesives, air fresheners, and personal care products, and result in eye, nose, and throat irritation, headaches, dizziness, nausea, fatigue, and even liver, kidney, and central nervous system damage, with some VOCs known or suspected to be carcinogenic. Carbon monoxide, a colorless, odorless gas from malfunctioning furnaces, gas stoves, fireplaces, wood-burning stoves, tobacco smoke, and attached garages, poses immediate life-threatening hazards such as headaches, dizziness, nausea, vomiting, chest pain, confusion, loss of consciousness, and even death. Nitrogen dioxide, emitted mostly from gas stoves, unvented kerosene or gas space heaters, and tobacco smoke, causes respiratory irritation, coughing, wheezing, shortness of breath, and increases susceptibility to respiratory infections. Formaldehyde, common in pressed wood products such as particleboard and plywood, furniture, adhesives, fabrics, and insulation materials, results in eye, nose, and throat irritation, coughing, wheezing, skin rashes, allergic reactions, and cancer formation. Radon, a naturally occurring radioactive gas, enters residential buildings through foundation cracks from surrounding soil and rocks and is identified as the second most common cause of lung cancer, behind tobacco smoking. Mold colonization in wet or damp conditions, especially after leaks or flooding, releases spores that cause allergic reactions, such as symptoms of sneezing, rhinorrhea, conjunctival irritation, dermal rashes, asthma exacerbation, and respiratory infections. Biological pollutants, such as dust mites, pet dander, and pollen—typically found in bedding, carpeting, and upholstery, as well as entering with outside air—cause allergic reactions and asthma exacerbation in susceptible persons. In addition, secondhand smoke from tobacco combustion significantly increases the risk for cardiovascular disease, lung cancer, and other respiratory diseases not only among smokers themselves but also among everyone who occupies the same indoor space.

Traditional approaches to indoor air quality management have been dominantly marked by an after-the-fact strategy, relying heavily on random manual monitoring and subsequent health screening after exposure. With the advent of Internet of Things (IoT) technologies and machine learning, there are unparalleled opportunities for real-time monitoring, predictive analytics, and preventive measures. This convergence enables one to design intelligent systems that can continuously monitor air quality parameters, predict pollution episodes, and take preventive measures automatically.

One of the top organizations and a fast-growing group of sustainability & Energy Economists recently conducted a large-scale survey in Delhi among over 5000 individuals spread over the nine city districts 35% of the respondents aren't sure if air pollution in Delhi is an emergency. This includes 75% of respondents who have children below the age of 10 years. 20% of all the respondents who answered are convinced of the hype surrounding it. Nearly 60% of respondents aren't sure if indoor air pollution is a concern in urban areas and believe that it is less harmful than outdoor air pollution. More than 50% of the respondents are unaware of the ban on burning garbage and that it attracts a fine of Rs 5,000.

One of the means to reduce air pollution is to raise public awareness about the causes and harmful effects of high concentrations of pollutants in the air. The technologies available are a part of our daily life, and the use of these technologies has increased tremendously over the years. Thus, utilizing available technologies towards the population's awareness is an executable solution. One of the prominent examples of technology in activities concerning air pollution is the use of pollution sensors, which are capable of detecting and distinguishing between different categories of particulate matters. In recent years, there has also been increased focus on the smart cities program as a means to reduce pollution effects. The program consists of several projects that are designed to protect indoor environments while also helping in reducing air pollution concentrations in office and living environments.

Sensors are also widely utilized for temperature, pressure, and other multi-parameter sensing. The integration of wireless technologies into sensors has significantly improved their capacity, and accordingly, several wireless sensor mesh networks have been established. A Wireless Sensor Network (WSN) consists of sensor nodes that forward the data collected over the network. For this project, the same wireless sensors are utilized and interfaced within an IoT system to sense and measure Particulate Matter, Temperature, Humidity, CO2, and Volatile Organic Compounds, and also enable the calculation of the Indoor Air Quality Index (IAQI) level.

With the above definition the main and major objective of this project can be stated as below

According to the above definition, the main and foremost objective of this project can be defined as follows.

1. Select the best sensor from the market, calibrate the sensor sensibility with trusted source.

2. Construct the potential IoT circuit to gather the Indoor Air Pollutants such as Particulate Matter (PM2.5, PM10), Temperature, Humidity, CO2, Volatile Organic Compounds (VOC)

3. Collect the sensor data electronically and store in cloud in continues mode with specific interval.

4. Compare the data sets collected with given events and conclude the likely cause of the increase/decrease of the pollutants.

5. Based on the out of analysis try to suggest the best practices within the indoor to avoid or reduce the increasing pollutants.

The sensors developed will be deployed indoors and will be used for collecting data every minute from various locations within the indoor (Bedrooms, Kitchen, Bathrooms, Waiting rooms). In addition to detecting the IAQI with single nodes, multiple nodes will also be added over time.

Current indoor air quality control conditions are beset by many serious limitations, which make it imperative to create new technological interventions. First, conventional air quality monitoring systems are associated with substantial capital expenses, space, and special maintenance requirements, making it economically unfeasible for widespread application in residential and small business structures. Moreover, current monitoring methods provide intermittent readings instead of constant real-time data, thus missing important pollution events and exposure patterns.

In addition, the very reactive nature of existing air quality management tends to expose occupants to toxic pollutants before corrective action can be taken. The absence of predictive capability renders the implementation of proactive strategies that can significantly mitigate health risks and enhance indoor environmental quality impossible. To mitigate these drawbacks, the complex interactions among various pollutants, environmental factors, and occupant activities render the identification of pollution sources and the adoption of effective preventive measures nearly impossible without the use of advanced analytical methods

Coordination of all the various indoor environmental parameters, including temperature, humidity, occupancy, and weather outside, demands sophisticated data handling capability beyond the capacity of conventional monitoring systems. Physical inspection of such sophisticated, multi-dimensional data is cumbersome, prone to errors, and bound to overlook subtle trends that can point towards an impending air quality problem. Indoor air pollution is generated by a large number of sources including building material, household cleaning products, biological contaminants like dust mites, and occupants' activities in the building like smoking, cooking, and cleaning, and the test building had a very demanding indoor environment for air quality management. The test hall, with poor ventilation (single window and door), low occupancy (4 persons mostly over weekends), inadequate air exchange (one ceiling fan), and extended closure periods over weekdays, offered an environment that was amenable to pollutant build-up and poor air exchange. The top-floor location with direct solar radiation perhaps provided indoor temperature conditions that were not very favorable, perhaps augmenting building material and furniture off-gassing and lowering the efficiency of the natural ventilation mechanisms. Due to the conditions, the measured level of CO_2 at 0.88 (assuming this to be 880 ppm, which would be within tolerable indoor levels but at the concentration level at which ventilation performance starts getting impacted) suggests that while human respiration had its share in degrading air quality, the predominant pollution sources were most likely to be multifactorial in nature, like poor ventilation, perhaps building and furniture material off-gassing, dust and biological contamination due to extended closure periods, and perhaps indoor pollutant intrusion from outside, and not CO_2 being the sole predominant pollutant component as suggested(Cho, 2020).

1.5. Research Purpose and Questions

The research examines a series of significant questions that are central to the evolution of the intelligent indoor air quality management field:

1. How do you optimally design and deploy IoT sensor networks to enable efficient, reliable, and economical indoor air quality monitoring across diverse building types and uses?

2. How are real-time sensor data processed and analyzed to provide real-time insights and enable timely action to events of air quality deterioration?

3. What are the optimal automated intervention techniques that exist for indoor air pollution reduction, and how can they be integrated into existing building control systems?

4. How is the intended IoT and machine learning solution more cost-effective, accurate, and healthier than conventional air quality management practices?

5. What are the major bottlenecks and challenges in deploying large-scale IoT-based air quality monitoring systems, and how can they be overcome?

6. What are the best machine learning algorithms and feature engineering strategies to predict indoor air pollution concentrations with low computational needs?

1.6. Research Objectives

1.6.1. Primary Objectives

This research aims to propose an integrated IoT and machine learning system for smart indoor air pollution prediction and prevention. The integrated system will offer real-time monitoring, accurate prediction, and self-sustained prevention of indoor air quality deterioration.

1.6.Specific Objectives

- IoT-based Monitoring System Development and Design: Design a cost-effective, scalable IoT sensor network that tracks various air quality parameters such as PM2.5, PM10, VOCs, CO, CO_2, temperature, humidity, and occupancy levels continuously.
- Developing Machine Learning Models: Develop and validate forecasting models using various machine learning algorithms (e.g., regression analysis, time series forecasting, neural networks, and ensemble methods) that will precisely predict indoor air pollutant levels from historical data, weather conditions, and usage patterns.
- Real-time Data Analysis: Leverage edge computing and cloud-based data processing platforms to process and analyze bulk sensor data in real-time to enable prompt action to be taken to counteract air quality degradation.
- Automated Prevention System: Implement and incorporate intelligent control systems that are programmed to automatically activate air purification systems, regulate ventilation rates, and provide real-time alerts to building occupants and facility managers.

1.7. Significance and Contributions

Academic Contributions

This study contributes to the knowledge base in some of the interdisciplinary areas like environmental engineering, computer science, public health, and building automation. The use of IoT technologies and sophisticated machine learning techniques for air quality is an innovative solution that bridges the gap between environmental monitoring and smart building systems. The development of predictive models for indoor air quality applications bridges a key research gap since most existing air quality prediction studies rely on outdoor conditions. The framework provides a platform for future studies in smart environmental monitoring and control systems.

i. Practical Contributions

From a practical perspective, this research provides several significant contributions to business and society. The development of a low-cost, scalable IoT monitoring system extends real-time air quality monitoring to more users and buildings. The predictive feature enables proactive health protection measures, potentially preventing exposure to dangerous pollutants when they are still safe. The automatic prevention system reduces the workload on building occupants and operators while still maintaining constant air quality management. The integration with existing building management systems provides a way of retrofitting existing buildings with intelligent air quality control capability.

ii. Societal Impact

The larger social benefit of this study is enhanced public health results through enhanced indoor air quality management, decreased healthcare costs from reduced incidence of air pollution-related diseases, and a heightened sense of concern with indoor environmental quality issues. The system's ability to offer instant feedback to occupants to facilitate behavioral change towards healthy indoor environments.

- Indoor air pollution exposure raises respiratory disease risk – Individuals with exposure to high indoor air pollutants like particulate matter (PM2.5, PM10) and volatile organic compounds (VOCs) have high rates of respiratory disease like asthma, bronchitis, and chronic obstructive pulmonary disease (COPD).

• Indoor air pollution promotes cardiovascular health hazards – Exposure to harmful gases like carbon monoxide (CO) and fine particulate matter over the long term causes higher heart rate variability, hypertension, and higher rates of cardiovascular diseases.

• Indoor chronic exposure to pollution is harmful to immune function – Individuals who are exposed to indoor pollutants like nitrogen dioxide (NO_2) and mold spores regularly have compromised immunity, are prone to infection, and experience allergies.

• Indoor air pollution reduces cognitive ability and productivity – Increased CO_2 and VOC levels indoors have been associated with slower reaction times, reduced problem-solving ability, and reduced productivity in the workplace and schools.

• Indoor air pollution is linked to more mental fatigue and stress – Individuals who live in spaces that have low air flow and high levels of pollutants report more fatigue, stress, and mood disturbance.

• Better indoor air quality enhances children's learning – Better ventilation and lower levels of pollutants in schools are linked to better student performance and lower absenteeism rates.

• Indoor air pollution induces sleep disturbances – Elevated levels of CO_2 and airborne allergens in bedrooms are linked with sleep fragmentation, reduced sleep efficiency, and susceptibility to sleep apnea.

• Improved indoor air quality positively affects overall health and well-being – Homeowners who live in houses that have air cleaners and ventilation systems are happier with their living condition and enjoy better overall health.

Study scope is limited to predicting the nature of air pollutants and establishing the levels of effect of the air quality index. create awareness. The aim is to collect significant data from different source home, office, schools, colleges and apply data modelling techniques and establish a successful prediction model to be applicable on different sets of data worldwide. This experiment is done to collect and predict indoor air quality based on PM2.5,PM10, CO_2, Temperature, Humidity, and VOCs pollutants only.

1.8. Organization Of Thesis

This thesis consists of seven chapters that introduce the research methodology, implementation, and findings sequentially.

Chapter 1 (Introduction) introduces the background, problem statement, objectives, and significance of the study.

Chapter 2 (Literature Review) is a comprehensive review of the existing literature on indoor air quality monitoring, IoT-based environmental monitoring, and machine learning-based air quality prediction models.

Chapter 3 (Methodology) discusses the research methodology, architecture system design, sensor selection criteria, machine learning algorithm design, and experiment setup.

Chapter 4 (System Design and Implementation) explains in detailed form the implementation and design of the IoT sensor network, data processing system, machine learning algorithms, and automated control systems.

Chapter 5 (Results and Analysis) contains experimental results, model performance evaluation, system validation results, and comparison with state of the art.

Chapter 6 (Discussion) is where critical analysis of the findings, responses to research questions, outlines implications and limitations, and suggests areas of future research.

Chapter 7 (Conclusion and Future Work) concludes the major findings, contributions, and future research trends.

The thesis also contains comprehensive appendices with technical data, additional experimental results, and comprehensive algorithm implementations to supplement the main research findings.

CHAPTER 2

LITERATURE REVIEW

2.1. Theoretical Framework

The application of artificial intelligence technologies has seen extensive development across various fields, including healthcare diagnostics, environmental monitoring, and contamination prediction. Research studies based on scholarly research have, over the years, seen an impressive increase in discussing AI implementation in the study of atmospheric pollution research. This research study has sought to explore emerging trends in AI applications in air quality management. The research process entailed extensive literature collection from the Web of Science database, that is, AI applications in atmospheric pollution research. Scholars employed bibliometric analysis using CiteSpace 5.8. R1 software to analyze geographical distribution, institution affiliations, author networks, keyword frequency, and citation patterns to identify emerging trends and research frontiers in air pollution research on the basis of AI.(Guo et al., 2022).

The analysis revealed that scientific articles in this area began in 1994 with steady growth with a dramatic acceleration from 2017. The most productive authors were China with 524 articles, followed by the Chinese Academy of Sciences with the most productive institution with 58 articles. The second most productive nation was the United States with 455 articles, followed by Tsinghua University as the second most productive institution with 33 articles. Of note, the United States and England had high network centrality measures of 0.24 and 0.27, respectively, indicating their centrality in global collaboration networks. Environmental science journals were the most common publication venues, with Atmospheric Environment having the highest citation impact of close to 1,000 citations(Guo et al., 2022). However, the analysis revealed sparse collaborative networks among researchers, institutions, and nations. The most common keyword themes were machine learning, air pollution, and deep learning. The most active areas of research today are forecasting atmospheric pollutant concentration, particularly using hybrid approaches through combining AI approaches with environmental science applications, cost-effective air quality monitoring sensor development, indoor environmental quality evaluation, and thermal comfort improvement. The study concludes that AI applications in air pollution research are developing rapidly with Chinese and American scientists leading and the Chinese Academy of Sciences displaying institutional leadership. Though the United States and England have played a prominent role in partnership networks, institutional partnership is still insufficient, indicating higher partnership can significantly enhance the research pace. Research hotspots today are particulate matter (PM2.5) concentration forecasting, low-cost sensor technology, and thermal comfort analysis(Guo et al., 2022).

The human population will live most of their daily lives indoors under confined spaces; indoor atmospheric conditions' quality is of the greatest importance for public health effects. The strong spatial and temporal heterogeneity that characterizes indoor atmospheric pollution creates severe challenges for traditional filter-based measurement methods, which require continuous monitoring technology. Continuous monitoring technology enables the transition of air quality assessment methods from stationary single-target research to dynamic comprehensive assessments, and it makes significant contributions to indoor environmental assessment practices(Wang et al., 2023).

This comprehensive review discusses the present status of technology, advantages, disadvantages, and future development possibilities of indoor atmospheric quality monitoring technologies based on real-time sensing technologies. Scientific studies on the application of continuous monitoring sensors for indoor environmental monitoring are increasing exponentially since 2018, and the study activities are primarily concentrated in China and the United States. Fine particulate matter (PM2.5) is the most studied atmospheric pollutant among the studies.

In addition to offering the higher spatial and temporal resolution of measurement, continuous monitoring sensors for indoor environmental monitoring also offer special advantages such as three-dimensional atmospheric monitoring capability, contamination spike detection capability, and source identification, and estimation of health effects in short-time. The enormous amount of data offered by continuous monitoring systems greatly simplify computational modeling and predictive analysis of indoor atmospheric contamination dynamics. There are very severe concerns in the use of continuous monitoring sensors in indoor environments, such as sensor selection standards, operating performance standards, long-term stability requirements, and calibration strategies. The future advanced sensor techniques would require sensors with higher performance standards, better operation stability, lower costs, and lower energy requirements. Furthermore, simultaneous detection of multiple target atmospheric pollutants by using continuous monitoring systems is a fundamental breakthrough necessary for large-scale indoor air quality monitoring(Wang et al., 2023).

Indoor air conditions with high concentrations of pollutants over long periods significantly increase the risk of cardiovascular and pulmonary system pathologies. While extensive research has been conducted in outdoor atmospheric quality assessment, indoor air quality studies are comparatively limited. New indoor air quality forecasting approaches through neural networks are plagued by several key defects: under-optimization of input parameters, sequential processing of input features, and uncontrolled loss of information within model training procedures, resulting in computational inefficiency, redundant computational time, and low predictive performance(Shi et al., 2023). In this article, a novel concurrent indoor particulate matter forecasting model is introduced based on the combination of Least Absolute Shrinkage and Selection Operator (LASSO) regression and an Attention Temporal Convolutional Network (ATCN) named LATCN. The strategy is implemented in a multi-stage approach: first, LASSO regression techniques are implemented for feature extraction from large-scale datasets of PM1, PM2.5, PM10, and PM (>10) concentrations and environmental conditions to optimize input parameters for the indoor particulate matter forecasting model. Second, an Attention Mechanism (AM) is employed to eliminate redundant temporal information and extract vital features from input data. Third, a Temporal Convolutional Network (TCN) generates parallel indoor particulate concentration predictions using the extracted features, while utilizing residual connections to minimize information loss.

Experimental findings validate that indoor particulate matter concentrations are largely determined by indoor heat index, indoor wind chill factor, wet bulb temperature, and relative humidity. Comparison with Long Short-Term Memory (LSTM) and Gated Recurrent Unit

(GRU) approaches validates that LATCN cuts down prediction error rates by 19.7% to 28.1% for Normalized Absolute Error (NAE) and 16.4% to 21.5% for Root Mean Square Error (RMSE), while at the same time boosting computational efficiency by 30.4% to 81.2% compared to traditional sequence prediction models(Shi et al., 2023). The research adds to active indoor air pollution prevention activities, provides theoretical frameworks for designing indoor environmental standards, and gives a model for future innovative air pollution prevention equipment design and deployment.(Shi et al., 2023).

Air pollution is increasing exponentially in Indian cities and globally and is a major threat to climate and the health of living organisms. Air pollution is the reason for poor indoor air quality (IAQ) in urban structures. Carbon dioxide (CO2) is the major reason for indoor pollution because human beings themselves are one of the source producers of CO2. CO2 testing and monitoring are time- and cost-consuming and need intelligent sensors too. So, in this regard, to overcome these drawbacks, machine learning (ML) has been utilized to forecast the concentration of CO2 in an office room. This work has been performed based on the data collected through actual measurement of indoor CO2, the number of occupants, area per person, outdoor temperature, outer wind speed, relative humidity, and air quality index utilized as input parameters. In this work, ten algorithms, i.e., artificial neural network (ANN), support vector machine (SVM), decision tree (DT), Gaussian process regression (GPR), linear regression (LR), ensemble learning (EL), optimized GPR, optimized EL, optimized DT, and optimized SVM, were utilized to forecast the concentration of CO2. It has been witnessed that the optimized GPR model is superior to other chosen models with respect to prediction accuracy. The outcome of this work showed that the optimized GPR model is capable of forecasting the concentration of CO2 with the highest prediction accuracy with R, RMSE, MAE, NS, and a20-index values of

0.98874, 4.20068 ppm, 3.35098 ppm, 0.9817, and 1, respectively. This research can be utilized by the developers of the smart city, researchers, medical professionals, and designers to investigate the indoor air quality for air ventilation system designing and CO_2 level monitoring within the buildings(Kapoor et al., 2022).

Air pollution is facing unprecedented rise in Indian cities and around the world, threatening environmental balance and physiological health of all forms of life. Poor outdoor air quality aggravates indoor air quality (IAQ) in urban built environments considerably. Carbon dioxide (CO_2) is the indoor pollutant that dominates indoor environments, and human residents are the prime emission sources of the same. Traditional CO_2 monitorization and surveillance strategies involve high financial costs, temporal efforts, and advanced sensing devices. Machine learning (ML) strategies have been utilized to overcome these demerits to predict indoor CO_2 concentration in commercial office buildings. This research utilizes data from continuous real-time measurements for indoor CO_2 concentration, occupancy density, per-capita spatial distribution, ambient outdoor temperature, external wind speed, relative humidity, and air quality indexes as predictor variables. Ten different algorithms are employed in research setup: artificial neural networks (ANN), support vector machines (SVM), decision trees (DT), Gaussian process regression (GPR), linear regression (LR), ensemble learning (EL), and optimized forms of GPR, EL, DT, and SVM for CO_2 concentration prediction(Kapoor et al., 2022). Comparative analysis indicates the optimized GPR model demonstrates better predictive performance compared to other chosen algorithms in terms of forecasting accuracy. Experimental findings indicate that the optimized GPR model predicts optimal CO_2 concentration with remarkable precision, demonstrating correlation coefficient (R), root mean square error (RMSE), mean absolute error (MAE), Nash-Sutcliffe efficiency (NS), and a20-index values of 0.98874, 4.20068 ppm, 3.35098 ppm, 0.9817, and 1.0, respectively. The findings of this research are significant information for environmental scientists, building designers, public health practitioners, and urban planners in evaluating indoor environmental quality in planning ventilation systems and the implementation of continuous CO2 monitoring programs in building systems.

Indoor air pollution is a major environmental health problem that creates serious threats to the physiological health of indoor occupational workers and home residents. Indoor occupational workers typically spend about 21 hours a day indoors, while home residents spend indoor hours for about 13 hours a day. Precise indoor environmental quality forecasting is a critical requirement for indoor workers' and typical residential occupants' health protection.

2.2. Theory of Reasoned Action

Although large-scale methodologies have been designed for indoor air quality forecasting problems, the forecasting operation remains a complex computational problem, especially when working under a sparse data gathering network and limited air quality monitoring facility conditions. As a response to these limitations, this study proposes a new neural network model with capabilities to learn temporal dynamics and inter-variable correlations in environmental data, realized by combining an Informer model with data-correlation feature extraction component built with a multilayer perceptron (MLP)-based structure. Experimental verification of this study utilizes the Informer model framework for indoor air quality condition forecasting in an industrial complex in Changsha, Hunan Province, China. The predictive model uses large-scale input parameters such as indoor and outdoor temperature data, humidity, and outdoor particulate matter (PM) concentration levels to predict indoor particle concentrations in the future.

Although large-scale methodologies have been designed for indoor air quality forecasting problems, the forecasting operation remains a complex computational problem, especially when working under sparse data gathering network and limited air quality monitoring facility conditions. As a response to these limitations, this study proposes a new neural network model with capabilities to learn temporal dynamics and inter-variable correlations in environmental data, realized by combining an Informer model with data-correlation feature extraction component built with a multilayer perceptron (MLP)-based structure. Experimental verification of this study utilizes the Informer model framework for indoor air quality condition forecasting in an industrial complex in Changsha, Hunan Province, China. The predictive model uses large-scale input parameters such as indoor and outdoor temperature data, humidity, and outdoor particulate matter (PM) concentration levels to predict indoor particle concentrations in the future.

Indoor air quality monitoring is crucial in urban and industrial settings, especially in countries like India and China where air pollution poses a critical health risk. Poor air quality impacts individuals with respiratory diseases, children, and the elderly, and therefore monitoring and controlling indoor areas is crucial. In this research, it is suggested that an Internet-of-Things (IoT)-aided system is utilized to sense, alert, and predict indoor air quality as part of smart home management and ambient assisted living. The system utilizes low-cost sensors in communication with an ESP32 microcontroller to detect pollutants like CO, PM2.5, NO2, O3, NH3, and ambient conditions like temperature, pressure, and humidity. Calibration using machine learning is carried out to improve the accuracy of low-cost sensor data. A new multiheaded CNN-GRU deep learning model is utilized to predict pollution levels for the next hour. Transfer learning (TL) is applied to improve the accuracy of forecasting in newly installed systems with minimal data, based on experience from neighboring monitoring stations. This yields early and more accurate predictions even with limited initial data. The system also offers the facility for a mobile app that provides real-time alerts when pollutant concentration is over safe levels, allowing users to take preventive measures. Experimental results confirm the efficacy of the TL-based approach, with improvement in RMSE scores by 55.42% for new installations(Zhang et al., 2023).

The research reveals the viability of leveraging low-cost technology and sophisticated AI to construct scalable, cost-effective air quality monitoring networks. It also illustrates how integrating the use of TL can break the issue of data scarcity in newly rolled-out systems. The results provide a real-world solution to forecasting pollution in intelligent city settings. The solution facilitates broader deployment of indoor air monitoring systems. It resolves sensor limitations and enhances predictive performance(Men et al., 2023). The work benefits public health, particularly in high-risk areas. It fosters the implementation of AI-based solutions in daily life. The system is open-source, low-cost, and scalable. It can be an integral part in future smart homes(Sonawani & Patil, 2024). It has been proven that air pollution results in the negative impacts on human health, and the ageing populations are particularly vulnerable due to the age-related compromised physiological function. Since elderly individuals spend approximately 80% of their time indoors, indoor air pollutant exposure is of specific concern for them. Indoor air quality measurement, however, is labor-intensive, time-consuming, and requires a large sample size for large-scale epidemiological studies. A predictive model was thus developed to estimate indoor concentrations of PM2.5 in Hong Kong elderly homes. For three consecutive summer and winter days, 24-hour average fine particulate matter (PM2.5, particles < 2.5 μm) concentrations were monitored in 116 homes. The model integrates land use regression model-estimated ambient PM2.5 with questionnaire-derived data on indoor pollution sources. A linear mixedeffects model was used and showed moderate predictive accuracy, with an R^2 of 0.67 (and 0.61 as estimated by cross-validation). The results showed that indoor PM2.5 concentrations were significantly affected by outdoor PM2.5 concentrations. Meteorological factors like temperature and humidity also had multifaceted effects on indoor air quality. Other indoor PM2.5 concentration sources included crowded living, long window ventilation, and cooking with liquefied petroleum gas. This study gives useful information on the control of indoor air pollution in elderly homes and gives a model for future large-scale health studies on indoor environmental quality.(Tong et al., 2020).

Air pollution is a severe threat to the global environment, and given that people spend 80–90% of their time indoors on average, indoor air quality is just as crucial as outdoor air quality. This is particularly an issue in schools. There are many ways to improve indoor air quality, including the use of air purifiers and ventilation. Automatic system triggering based on real-time monitoring through air-quality sensors is feasible. With efficient and effective clustering algorithms applied to indoor air quality data, particularly on pollutants such as CO_2, ventilation strategies can be optimized for better air quality management. The contribution of this paper is dedicated to clustering indoor air quality data gathered from a school campus in Taiwan, without the use of other external data such as geographical position or space utilization. The Max Fast Fourier Transform (maxFFT) Clustering Approach is proposed by this paper, which categorizes indoor air quality data by extracting features of significance and enhancing efficiency in clustering. The work proves that even without other contextual data, the approach is able to reasonably reflect the actual ventilation conditions in various spaces, and does so with relatively moderate computational effort(Chu & Ho, 2022).

This paper introduces a framework for an Air Quality Decision Support System (AQDSS) and demonstrates its use through the construction of an Internet of Things (IoT)-based application. A case study of Madrid was used to validate the system. The application integrates data from multiple sensors, harmonizing indoor and outdoor air quality measurements and people's spatiotemporal activity patterns to estimate Personal Air Pollution Exposure (PAPE). The study suggests that PAPE can be quantified reliably with indoor air quality sensors and e-beacon technology—low-cost and minimally invasive technology that has not been widely used in similar studies to date. In the future, the application can be extended further by incorporating predictive models to provide real-time feedback on PAPE risks. Data gathered from such systems could also be used in the future to inform the design of air quality regulations and to provide epidemiological studies of the impacts of air pollution on health.(Arano et al., 2019)

This study suggests a framework for an Air Quality Decision Support System (AQDSS) and demonstrates its practical application by an Internet of Things (IoT) application. A case study in Madrid was conducted to demonstrate the concept. The system employs sensors to measure indoor and outdoor air quality and combines this with individuals' movement and activity profiles over time to estimate Personal Air Pollution Exposure (PAPE). The result of this pilot study is that PAPE can be quantified by using indoor air quality sensors and e-beacon technology—technology that is inexpensive and non-intrusive, and has not been applied by prior research. Future refinement of the IoT application may involve the incorporation of predictive models to provide real-time risk warnings related to PAPE. Data generated by this system could also be used to inform the development of air quality policy and as a rich source of data for epidemiological studies on the health impacts of air pollution. (Arano et al., 2019).

As the realization grows about the contribution of indoor air pollution to the degradation of human health, indoor air quality control is coming under heightened attention. Indoor smoking is a significant source of indoor air pollution, and its harmful health impacts are well established. This has prompted global action in the form of legislation against indoor smoking. Although technical measures for reducing indoor smoking are available, most of the literature has focused on developing detection devices. The present work adopts a new path in data analysis and application of machine learning for cigarette smoke detection using the presence of the gases in the given total volatile organic compounds as well as carbon dioxide exhaled, as detected by Internet of Things sensors. It created a machine learning dataset from IoT sensor data with training data from controlled environments through the application of a rotary smoking machine and testing data from real-world environments from real smokers. The performance of the models was tested with common accuracy, precision, and recall metrics. The best performing was found to be a non-linear support vector machine with accuracy of 93% and F1 score of 88%. k-nearest neighbours and multilayer perceptron supervised learning models also performed quite well, but the study suggested that using binary classification would be able to increase accuracy as well as processing efficiency by making the prediction simpler. (Cho, 2020).

This research investigates the use of machine learning to predict problematic humidity levels in rooms housing cultural objects. The research constructed an XGBoost model that predicts when the relative humidity will hit too high or too low over the coming 24-hour period based on indoor and outdoor hourly climate data as input parameters. The scientists tested their prediction system in two cultural heritage environments. In a storehouse, the model performed with accuracy rates of 0.93 for high and low humidity predictions. Performance was considerably poor when the model was tested in a church building, with a mere 0.78 accuracy for high humidity and 0.62 for low humidity predictions. The research identified several issues with deployment. Availability of good-quality historical climate datastes for model training was difficult, and the reliance of the system on external IT infrastructure makes it susceptible - in the event of failure of these systems, the predictive model stops working without raising any alert.

The authors provide suggestions for further research, including extrapolation of timescales for prediction beyond 24 hours with sustained accuracy, and extrapolation of machine learning use to predict indoor air pollution concentration and energy consumption because of climate control systems in historic buildings. Other uses of machine learning for indoor environmental prediction could include indoor air pollution, or energy consumption because of climate control(Boesgaard et al., 2022).

This research examines the interaction between air pollution, weather, and COVID-19 infection through the development of a predictive model for the number of cases in the future. The research is based on the premise that meteorological conditions and air condition may influence the spread of the virus, and therefore it will be valuable to understand the historic records of particulate matter (PM2.5 and PM10) and weather variables indoors and outdoors. The authors created an integrated machine learning and deep learning framework for the prediction of COVID-19 cases. The authors designed their approach to use K-means clustering to identify behavior patterns by grouping similar points together. The authors then applied a Long Short-Term Memory (LSTM) neural network to perform multivariate linear regression, creating a robust predictive model during training. When the LSTM model was validated using outdoor environmental data like PM2.5, PM10 concentration levels, and meteorological conditions, the model had good performance. The results indicated error rates of 0.0897 RMSE, 0.0837 MAE, and 0.4229 MAPE under testing. The model had improved performance using indoor environmental data from 20 households, which were collected from May 27 to October 13, 2021. With this indoor dataset that included PM2.5, PM10, and meteorological measurements, the model had improved accuracy results of 0.0892 RMSE, 0.0592 MAE, and 0.2061 MAPE.

2.3. Human Society Theory

In three different indoor environment validation test datasets, the predictive model had acceptable performance levels, between RMSE = 0.4152 to 3.9084, and with MAPE always below 4.1%. Additionally, on validation, the predictive model has an extremely acceptable performance with RMSE = 0.4152 to 3.9084, and a MAPE below 4.1%, although indoor air pollution has been associated with allergic diseases through extensive research, public health policy is still lacking predictive tools to develop prevention guidelines for patients and susceptible groups. This is mostly because of a lack of access to real-time, large-scale data and issues of model reliability. Although Internet of Things (IoT) technology and machine learning offer promising solutions for accessing real-time data and improving disease risk predictions for evidence-based interventions, these applications are underdeveloped.

This pilot study examined if deep learning models could accurately predict asthma risk. The study included 14 asthmatic children who were patients at the Korea University Medical Center. Researchers measured patients' peak expiratory flow rate (PEFR) and, simultaneously, indoor particulate matter (PM10 and PM2.5) concentrations at their homes with low-cost sensors uploading data every 10 minutes from September 2017 to August 2018. The cohort interpolated the twice-daily PEFR measurements to create continuous daily profiles that could be synchronized with particulate matter and weather data. Researchers classified PEFR values into three grades of risk: 'Green' for normal lung function, 'Yellow' for mild to moderate shortness of breath, and 'Red' for severe, all in comparison to each individual patient's best peak flow values.

With the initial 10 months of combined data, researchers trained a Long Short-Term Memory (LSTM) neural network model to predict asthma risk categories for the next 2 months. The LSTM model was more effective than traditional multinomial logistic regression since it could account for the cumulative impact of particulate matter exposure over time. The authors suggest that with additional optimization with larger patient groups, this method would be capable of transforming medical decision-making by providing scientifically grounded, data-driven tools to manage asthma. using three different datasets with values of indoor environment(Ramirez-Alcocer et al., 2022). With successful tuning of the algorithm based on a large sample, this approach was capable of potentially being revolutionary for the scientific data-driven medical decision-making(Kim et al., 2020).

This research examines indoor air pollution monitoring through an integrated approach that marries biomedical engineering (BME) sensors with conventional indoor air quality sensors to address the critical health consequences of indoor air status. The study employed a sensor network in indoor settings to access real-time air quality data, which was then analyzed through machine learning techniques to detect and quantify patterns of pollution. The findings validate that this two-sensor approach works effectively in detecting and solving indoor air pollution issues, offering valuable information for the establishment of efficient indoor air quality management systems. Through the integration of BME approaches with conventional air monitoring techniques, this work enhances indoor environmental quality research and offers new approaches for protecting human health through sophisticated air quality monitoring and control technologies.

Despite large-scale studies' reports of indoor air pollution correlations with allergic disease, public health policy remains uncertain regarding predictive models to inform prevention guideline development in patients and high-risk individuals. This is mainly because of the unavailability of real-time large-scale data and low-reliability models. Even though Internet of Things (IoT) technology and machine learning represent potential paths to real-time data access and the improvement of disease risk prediction to inform evidence-based intervention, these remain unexplored. In this pilot study, the objective was to establish whether asthma risk is predictable by deep learning models. The study recruited 14 asthmatic children as outpatients in Korea University Medical Center. Researchers tracked peak expiratory flow rate (PEFR) from patients and tracked indoor particulate matter (PM10 and PM2.5) concentration indoors at home by tracking with low-cost sensors that tracked data every 10 minutes from September 2017 to August 2018.

The group interpolated twice-daily PEFR measurements to produce daily continuous profiles for comparison with weather and particulate matter data. PEFR measurements were graded to three risk levels: 'Green' for normal, 'Yellow' for mild to moderate shortness of breath, and 'Red' for severe, against the best peak flow for each patient. Based on the combined data of the first 10 months, investigators trained a Long Short-Term Memory (LSTM) neural network model to predict asthma risk categories for the subsequent 2 months. The LSTM model performed better than conventional multinomial logistic regression because it could model the cumulative effect of exposure to particulate matter within a time window. Authors believe that with further development from larger patient numbers, the approach could revolutionize medical decision-making by providing scientifically derived, data-driven adjuncts to asthma management.

This study gladly addresses the need for adequate occupancy estimation in interior spaces to regulate the air as much as possible and avoid disease transmission, both in the aftermath of the COVID-19 pandemic. As occupants of a building are also sources of infection and contamination, the real number of individuals in the room becomes crucial in implementing the necessary measures for ventilation and infection control.

This study gladly addresses the need for accurate occupancy estimation within indoor spaces to regulate the air as far as possible and avoid disease transmission, even in the post-COVID-19 pandemic period. As occupants of a building are also infection sources and contaminants, the room occupancy becomes a mandatory parameter in taking actions for ventilation and infection control.

The study developed machine learning models for occupation level prediction using carbon dioxide level measurements as the primary indicator and other environmental factors. Two machine learning approaches, the random forest and artificial neural network models, were compared and validated using CO2 levels, ventilation system performance, and indoor-outdoor/corridor differential pressure as inputs. The best performance was achieved with CO2 level and ventilation system input-based models with approximately 91% prediction accuracies for both the random forest (91.02%) and neural network (91.80%) approaches. Surprisingly, the application of differential pressure measurements reduced model performance to approximately 89% for both approaches.

The authors suggest that research in the future should be focused on enhancing the understanding of pressure fluctuation and CO2 fluctuation correlation with time which can be utilized to increase the accuracy of such occupancy prediction systems. The technology can be applied for automated building control systems which are required to modify ventilation rates based on real-time occupancy.

This study investigated indoor air pollution from solid fuel combustion during northern Chinese winter, employing hourly indoor concentrations of PM2.5 in about 1,600 residences to quantify the health effects of household heating systems. The study indicated that indoor PM2.5 levels had an average of 120 μg/m3 but varied from 16 to about 400 μg/m3, with the households using clean heating technologies having about 60% lower levels compared to the conventionally coal or biomass fuel households. Utilizing a high-predictive accuracy random forest regression machine learning model ($R^2 = 0.85$), the authors estimated the health impacts of transitioning to cleaner fuels, with the transitioning away from traditional solid fuels to clean coals or biomass pellets decreasing indoor PM2.5 by 20%, and transitioning further to clean supply of energy sources realizing an additional 30% decrease, reflecting substantial health benefits of clean energy transition into household systems.

This work analyzes indoor air pollution monitoring with an integrated approach through biomedical engineering (BME) sensors and conventional indoor air quality sensors to minimize the adverse effects of indoor air atmospheres on human health. The research applied a sensor network over indoor spaces to obtain real-time air quality information, which were further processed through machine learning algorithms to detect and quantify pollution patterns. The findings validate that the hybrid sensor approach can efficiently detect and circumvent indoor air pollution issues, yielding valuable information for the deployment of holistic indoor air quality administration systems. Through the integration of BME methods with conventional air monitoring systems, this work progresses indoor environmental quality research and outlines new strategies for human health assurance through improved air quality monitoring and controlling systems.

Growing awareness of the adverse impact of indoor pollutants on human health has witnessed growing interest in indoor air quality management. Amongst various indoor pollution sources, cigarette smoking is one of the most prevalent and detrimental ones, and the adverse impact on health is well documented in the literature. As a consequence, the majority of countries and territories have begun imposing stringent regulations against indoor smoking. At the same time, technical means have also been investigated to supplement such regulatory measures, in the primary guise of detection systems. The majority of existing literature, however, has been focused on hardware-based detection systems. The present study, on the other hand, differentiates itself by resorting to an analytical and machine-learning-based method of cigarette smoke detection from analysis of ambient gaseous tracers like total volatile organic compounds (TVOCs) and carbon dioxide (CO_2), monitored by IoT-based environmental sensors. A very large machine-learning dataset was created with training data carefully collected using a rotary smoking machine under controlled conditions, and test data collected in real indoor conditions where spontaneous smoking activity was present. The performance of the models in smoke detection was evaluated with great rigor using standard classification metrics like accuracy, precision, recall, and the F1 score. Out of a variety of models investigated, the non-linear Support Vector Machine (SVM) demonstrated the overall best performance with high accuracy of 93% and F1 score of 88%, establishing its capability in detecting complex patterns in the data. Other supervised learning models like k-Nearest Neighbors (KNN) and Multilayer Perceptron (MLP) also performed well, but with SVM being superior in balancing precision and recall(- et al., 2023). The study also reveals the benefit of transforming the process of classification into a binary model that enhances the efficiency and precision of predication, rendering the system efficient and deployable to real-world applications for indoor smoking detection(- et al., 2023).

To meet the requirement of attaining the target design strength of cement-soil under different application conditions, traditional practices usually depend on a wide range of field and laboratory tests in geotechnics. These techniques are, however, time-consuming and cause excessive consumption of material, high cost and time, and great environmental impact(Zhang et al., 2023). To counter these problems, the present study suggests a machine learning approach with which to predict the compressive strength of cement soil with high accuracy. The study commenced with the development of a wide-based database of 566 samples, which were collected through a thorough review of the literature. Then, eight different machine learning models were developed and trained against this set, whose performance was critically evaluated using six different evaluation metrics to achieve their generalizability. Among the models, the Extreme Gradient Boosting (XGBoost) model was found to be the best, with a coefficient of determination (R^2) of 0.93 on the test set. To achieve a deeper understanding of the decision-making process of the model, feature importance analysis was performed using SHapley Additive exPlanations (SHAP) and partial dependence plots, and it was achieved that cement content, water content, curing age, and fine particle content were the most significant factors affecting compressive strength. In addition, the predictive capability of the machine learning model was compared with that of the traditional empirical model, and the latter was achieved to possess much better predictive ability. The present study offers a strong database and a model of efficient prediction, with insights and practical recommendations beneficial to the design and application of cement soil for soft foundation engineering works (Zhang et al., 2023).

Low-cost sensors (LCSs) have evolved rapidly to support spatio-temporally feasible real-time indoor air quality (IAQ) monitoring, but the extensive diversity of sensors available creates issues in selecting the best suitable ones. Assembly of single sensors into a rational sensing network needs expertise in various areas of research, which this review recommends integrating by promoting the use of IAQ as an integral part of smart home systems. The primary objective of the review is to summarize existing home automation technologies that support effective IAQ monitoring and control through networked air pollution LCSs. The most significant steps in this shift from traditional to smart homes are the optimal selection of sensors, tactical positioning, effective processing of data, and design of prediction models. A critical evaluation of existing LCS technologies reveals their limitations and capabilities in depicting IAQ in space and time. The findings suggest that controlled laboratory assessment of sensor performance before deployment is essential to ensure QA/QC. However, for long-term monitoring, continuous calibration or the application of statistical correction methods in operation is essential to ensure data accuracy. The sensor placement review should be strategic in terms of the location and relative exposure height of domestic residents to achieve maximum spatio-temporal coverage. In addition, effective data processing tools are essential to process the large amounts of complex, multivariate data generated by sensor networks to automate pre-processing and post-processing tasks. This makes the systems more scalable, reliable, and flexible. The review also emphasizes the potential of machine learning methods to enhance IAQ fluctuation predictability in LCS-based sensor networks (Omidvarborna et al., 2021).

Proactive monitoring and control of our natural and constructed environments is critical in many application domains. Semantic Sensor Web technologies have been applied and studied extensively for environmental monitoring use cases to supply sensor data for inspection to offer responsive action in situations of interest. Although such applications offer rapid response to situations, to restrict their unwanted side effects, research is still required to offer techniques that possess the capacity to pre见 the future to aid proactive control, so that undesired situations may be prevented in their entirety. This work combines a statistical machine learning based predictive model into a Semantic Sensor Web through stream reasoning. The approach is tested in an indoor air quality monitoring example. A sliding window-based approach that employs the Multilayer Perceptron model to forecast short-term PM2.5 pollution situations is incorporated in the framework for proactive control and monitoring. Results indicate that the proposed approach can accurately forecast short term PM2.5 pollution situations: precision of up to 0.86 and sensitivity of up to 0.85 is achieved at half hour prediction horizons, which enables the system to warn occupants or even to automatically steer clear of the forecasted pollution situations within the Semantic Sensor Web framework(Adeleke et al., 2017).

Few have been able to examine the combined effect of home and school environmental factors on children's health due to the challenge of analyzing several highly intercorrelated environmental measures. This study bridges this gap using machine learning techniques in addition to conventional logistic regression to examine the effect of indoor environments on children's health outcomes. The study utilized data from the SINPHONIE (Schools Indoor Pollution and Health: Observatory Network in Europe) project in Romania as part of a large European research program. The database held comprehensive data on home and school indoor environments, children's health symptoms, smoke exposure, and school policy. The health outcomes were coded into four categories: general health symptoms, asthma, allergy, and flu-like symptoms. Logistic regression and the Random Forest (RF) machine learning software were used to predict the health outcomes, with their performance being compared. The RF model identified a group of stable demographic and environmental risk factors in every category of health, including exposure to environmental tobacco smoke (ETS), school building dampness, male sex, use of air fresheners, proximity to residential traffic (within 200 meters), schoolwork stress, and noise in classrooms, with individual contribution rates ranging from 7.91% to 23.12%. From a model performance perspective, RF outperformed logistic regression in the majority of cases, with greater accuracy, specificity, and area under the curve (AUC) values, while the two methods had comparable sensitivity. Overall, the research shows that ETS, indoor dampness, proximity to traffic, noise exposure, and certain home products like air fresheners are valid environmental risk factors for child health. Moreover, the RF model was a more effective predictive model than logistic regression in establishing these complex environmental-health relationships (Lin et al., 2021). As human living is mostly indoors, indoor air quality (IAQ) dictates overall health. Indoor air pollution kills nearly 3.8 million individuals every year, says the World Health Organization (WHO), and that speaks volumes about the extent of its contribution. As living standards have improved, IAQ monitoring has become an area of increasing interest in research. Although the research area of machine learning (ML) for gas sensing has made significant progress, one of the most critical concerns—measurement uncertainty—usually draws little or no interest and is usually addressed only by cross-validation rather than being explicitly modeled. That is what this work tries to tackle head-on. The gas concentration can be estimated by using gas sensors operating in temperature-cycled modes (TCO), with ML being applied to the logarithmic resistance of sensors. This work has a particular focus on formaldehyde, the indoor carcinogen, and the sum concentration of volatile organic compounds (VOCs) like acetone, ethanol, formaldehyde, and toluene as IAQ indicators. As gas concentration is a continuous value, regression methods have to be used. To address uncertainty in regression, the work extends a previously established uncertainty-aware automated machine learning toolbox (UA-AMLT)—previously developed for classification—by introducing a new uncertainty-aware partial least squares regression (PLSR) algorithm. This extension follows the recommendations outlined in the Guide to the Expression of Uncertainty in Measurement (GUM) and its supplements. The research compares two conditions to evaluate the effect of uncertainty on ML model performance: one from raw sensor measurements, and the other from noisy data with added artificial white Gaussian or uniform noise to simulate increased measurement uncertainty. One of the significant advantages of this method is the ability to determine points for system improvement, either by enhancing the ML model itself or through the use of more precise sensors. The results also indicate that training models with noisy data can make models less sensitive to random variation, ultimately producing more reliable IAQ monitoring systems (Dorst et al., 2023).

The world energy sector still grapples with the grand challenge of making access to clean energy universal. This grand challenge has a direct nexus with the United Nations' Sustainable Development Goal 7 (SDG 7), which is clean, affordable, and sustainable energy for all. Access to clean energy is also critical to the attainment of better health results (SDG 3), since the use of unclean fuels, especially for cooking, leads to indoor air pollution that has adverse effects on health. While it is complicated to establish a certain causal relationship between the use of unclean fuels and health results by endogeneity challenges such as reverse causality, making it hard to derive conclusions from science, the present study overcomes the challenges by employing a strong methodological framework to examine the health costs of using unclean fuels, based on evidence from the Chinese General Social Survey. A combination of analytical techniques such as ordinary least squares (OLS), ordered regression, instrumental variable (IV) approaches, penalized machine learning, placebo tests, and mediation models are utilized to assure the robustness of the results. The global energy sector is yet to overcome the challenge of universal clean energy access. The challenge is closely related to the United Nations' Sustainable Development Goal 7 (SDG 7), clean, accessible, and sustainable energy for all. Clean energy access is also important to the achievement of better health outcomes (SDG 3), as the use of unclean fuels, particularly for cooking, is responsible for indoor air pollution with serious health implications. It is, however, challenging to determine a clear causal link between the use of unclean fuels and health outcomes due to endogeneity issues such as reverse causality, which impede making scientific inferences. These challenges are overcome by the current study leveraging a strict methodological process of determining the health costs of the use of unclean fuels, informed by evidence from the Chinese General Social Survey. Analytical tools such as ordinary least squares (OLS), ordered regression, instrumental variable (IV) techniques, penalized machine learning, placebo tests, and mediation models are employed in combination to ensure conclusions are scientific.

Residential houses are high-energy consumers and environmental polluters. In Iran, the situation is worse because the consumption of fossil fuels rose more than 400% between 1990 and 2018. One of the most important strategies for the decrease of such dependence and encouragement of healthier indoor and outdoor environments is the adaptation and promotion of green building principles. This study investigates the challenges and opportunities of the process of green building development in Iran and proposes a green building standard suitable for the conditions of the country. According to the expert opinion gathered by the Delphi method, critical evaluation factors were identified. An exhaustive questionnaire of three components was completed by 81 construction experts, including employers, consultants, and contractors.

Machine learning algorithms were applied for data analysis, and the result was that the localized green building score on five main dimensions—site, water, energy, materials, and indoor environmental quality—averaged 77.2. Among them, the energy category was found to be the most significant with a coefficient of significance of 0.548. Subsequent ranking of single factors by the Friedman test showed that energy consumption management, utilization of renewable energy, and thermal zoning were ranked as priority one, two, and three. Taking into account the challenges and opportunities of implementation, the study found that the greatest challenge was a lack of awareness of green buildings, cited by 77% of respondents, and the greatest opportunity was a high potential for renewable energy generation in Iran, cited by 81% of respondents(Rajabi et al., 2021).

Indoor air quality (IAQ) is employed to define conditions in buildings that can influence respiratory health. Good IAQ conditions in hospital buildings are essential, especially for medical staff and patients. In recent years, many issues have been raised and must be resolved immediately in the identification of IAQ pollutants and related thresholds and ways of offering a knowledge-based labelling scheme of pollution levels. For this reason, a systematic review should be conducted first to develop new taxonomy studies on internet of Things-based IAQ sensory technology for hospital buildings to determine a research gap. In this context, the present research presents an IAQ methodology including the proposed nine IAQ pollutants for hospital buildings and buildings: Carbon monoxide, Carbon dioxide, Nitrogen Dioxide, Ozone, Formaldehyde, Volatile organic compounds, particulate matter (PM) and air humidity and temperature. The proposed methodology utilized real and simulated IAQ pollutant datasets to predict the hospital building pollution levels in three steps. In the first step, two IAQ datasets (actual and large-scale simulated datasets) are determined. The second phase includes the following: First is to apply the Interval type 2 trapezoidal-fuzzy weighted with zero inconsistency (IT2TR-FWZIC) approach from the Multi-Criteria Decision-Making theory to get the needed weights for the nine pollutants.

The second is a proposal for a new methodology known as the Unified Process for Labelling Pollutants Dataset (UPLPD) that contains six steps in a sequence as per the IT2TR-FWZIC framework. The UPLPD framework categorizes pollution into four classes and labels the study datasets correspondingly. The third is to input the labelled data sets to eight different algorithms-based machine learning models. This is done with thorough model evaluation against five evaluation metrics, including accuracy, Area under the Curve, F1-score, precision, and recall measures. After running the algorithms on real data, Support Vector Machine, Logistic Regression, and Decision Tree emerged as the top-performing algorithms, with the highest accuracy levels being 99.813%, 99.259%, and 98.182%, respectively, and also doing well on performance metrics. On a synthetically created dataset, Random Forest, Decision Tree, and AdaBoost were seen to perform best with accuracy levels being 90.094%, 88.964%, and 87.735%, respectively, also doing well on performance metrics.

These results effectively solved the issues addressed and met the study's inquiries, with experimental verification proving the model's efficacy in making predictions. Measurement of classroom air quality is crucial, as children spend a large part of their day at school. To facilitate this, Massey University in New Zealand created SKOMOBO, a cost-efficient and scalable Indoor Air Quality (IAQ) monitoring platform, which was widely used in primary school classrooms throughout the nation. In the collection of data from SKOMOBO devices, the detection of sudden spikes in air pollution levels was a major concern. This research addresses the same issue by suggesting an outlier detection technique for PM_{10} data based on MSD-

Kmeans. MSD-Kmeans combines a statistical method—Mean and Standard Deviation (MSD)—with the K-means clustering algorithm. The MSD part eliminates noisy data points in the initial stages, minimizing their influence on clustering, while K-means provides better clustering performance by detecting locally optimal groupings. The performance of MSD-Kmeans was tested with other similar outlier detection techniques. Experimental results show that MSD-Kmeans yielded better results in most performance metrics, such as True Positive Rate (TPR), False Positive Rate (FPR), Accuracy, and F-measure. Based on these results, the paper concludes that MSD-Kmeans is an effective and efficient outlier detection tool for large-scale IAQ datasets (Wei et al., 2020). The Measurement of classroom air quality is crucial, as children spend a large part of their day at school. To facilitate this, Massey University in New Zealand created SKOMOBO, a cost-efficient and scalable Indoor Air Quality (IAQ) monitoring platform, which was widely used in primary school classrooms throughout the nation. In the collection of data from SKOMOBO devices, the detection of sudden spikes in air pollution levels was a major concern. This research addresses the same issue by suggesting an outlier detection technique for PM_{10} data based on MSD-Kmeans. MSD-Kmeans is a hybrid approach that combines a statistical technique—Mean and Standard Deviation (MSD)—with the K-means algorithm for clustering. The MSD component eliminates noisy data points at an initial processing stage, reducing their impact on the clustering algorithm, and K-means ensures improved clustering performance by finding locally optimal clusters. MSD-Kmeans performance was compared with other algorithms of comparable nature for outlier detection. Experimental results demonstrate that MSD-Kmeans was better in most of the performance metrics, such as True Positive Rate (TPR), False Positive Rate (FPR), Accuracy, and F-measure. Accordingly, according to these results, the research concludes that MSD-Kmeans is an efficient outlier detection method in large IAQ datasets (Wei et al., 2020).

Modelling concentrations of in-vehicle air pollutants is a crucial step towards a credible estimation of daily human exposure to air pollution. This is, however, a difficult task due to the dynamic nature of control parameters, including driving behaviour and ventilation levels in the vehicle. This article presents a new approach that combines mass-balance modelling with machine learning to forecast in-vehicle exposure concentrations. The method utilizes a comprehensive dataset with ambient, roadside, and in-vehicle measurements of a range of pollutants, i.e., particulate matter (PM_{10}, $PM_{2.5}$, PM_1), nitrogen dioxide (NO_2), nitrogen oxides (NO_x), lung-deposited surface area (LSDA), and ultrafine particles (UFP), under various ventilation conditions. The first model (MB) is based on the mass-balance approach and accounts for simple physical and chemical processes to forecast pollutant concentrations in the vehicle (Matthaios et al., 2024). The second model (ML) employed machine learning algorithms trained on 80% of the data (selected randomly through random number generation), while the remaining 20% was reserved for validation. While the two models generally performed well, they underestimated UFP and LSDA concentrations. The ML model was more precise in prediction than the MB model, particularly for NO_2, and predicted unseen in-vehicle concentrations of pollutants correctly. It showed good performance metrics, including index of agreement (IOA) >0.69 and Pearson correlation coefficients (r) >0.80 for all of the pollutants investigated. Surprisingly, substitution of on-road data with nearest neighbour air quality monitoring station data in the ML model yielded promising results, enhancing its utility. In an age where air pollution is increasingly harmful to human health, the present study is insightful as far as in-vehicle exposure modeling is concerned. It adds to exposure science and demonstrates that real-time prediction of exposure and health impact assessment for occupants of vehicles can be achieved with the application of existing monitoring infrastructure, without incurring any additional cost (Matthaios et al., 2024).

Air pollution is a major global health hazard, and fine particulate matter (PM2.5) is one of the most important harmful agents, particularly indoors. These small particles, predominantly from human activities, are strongly linked to all types of respiratory diseases. As approximately 90% of the population spends nearly 22 hours indoors—at home, in the office, or in other indoor environments—improving indoor air quality is paramount. This study investigated whether certain indoor plants had the ability to reduce PM levels effectively by combining biomedical engineering methods and machine learning models. The findings recognized the presence of certain plants in the room to significantly enhance indoor air quality by lowering PM2.5 levels to well below the average outdoor levels. With the strong link between long-term exposure to PM and all types of diseases, the study concludes that employing plants as natural air purifiers, in addition to sensor devices and cloud AI systems, is an optimistic long-term approach for reducing indoor air pollution and enabling healthier indoor environments.

CHAPTER 3

RESEARCH DESIGN

3.1. Overview of Research Problem

This chapter presents the overall research design that has been utilized for developing an integrated air pollution forecasting and prevention system based on the Internet of Things (IoT) and machine learning techniques. The research design consists of the methodological framework, data collection methodologies, system architecture, experimental setup, and performance metrics that guide the research. The chapter presents the systematic approach adopted to address the research objectives and verify the proposed solution for real-time air quality monitoring, forecasting, and intervention measures.

3.2. Operationalization of Theoretical Constructs

3.2.1. Research Approach

The research in this book is of a mixed-methods design with a quantitative and qualitative approach. Quantitative analysis is used for air pollution measurements from IoT sensors and machine learning model scores, and qualitative analysis is conducted using expert interviews and user experience testing. The research is of an experimental design type to test hypotheses about the effectiveness of IoT-based machine learning systems for air pollution control.

3.2.2. Research Philosophy

The research is pragmatic in its approach, i.e., it focuses more on practical solutions for actual environmental problems. The philosophy is biased towards the convergence of various sources of information, technological methods, and assessment techniques to arrive at an innovative framework of air pollution control. The pragmatic approach provides methodological flexibility to conform to the particular needs of each phase of study.

3.2.3. Research Strategy

The process of research includes four general phases:

Phase 1: System Design and Requirements Analysis - Comprehensive analysis of technical requirements, environmental monitoring requirements, and IoT-ML integration technical specifications.

Phase 2: Data Collection and Preprocessing - IoT sensor networks for real-time data collection, integration with current history data, and data quality control processes.

Phase 3: Model Training and Development - Utilization of different machine learning models, tuning hyperparameters, and ensemble method construction for prediction accuracy improvement.

Phase 4: System Validation and Integration - Final system deployment, performance measurement, and validation against set benchmarks and real-world conditions.

3.3. Research Purpose

3.3.1. Overall System Framework

The proposed system architecture is a four-layered one:

i. Sensing Layer - Distributed IoT sensors to gather real-time environmental data in terms of particulate matter (PM2.5, PM10), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), carbon monoxide (CO), ozone (O_3), temperature, humidity, wind speed, and atmospheric pressure.

ii. Communication Layer - Wireless communication protocols (WiFi, LoRaWAN, 4G/5G) for secure data transport from sensors to cloud infrastructure with edge computing for initial data processing.

iii. Processing Layer - Cloud-based machine learning pipeline comprising data preprocessing, feature engineering, model training, prediction generation, and decision support algorithms.

iv. Application Layer - User interfaces like web dashboards, mobile applications, and stakeholder alert systems like environmental groups, health centers, and the general public.

3.3.2. IoT Sensor Network Design

The IoT sensor network relies on a hierarchical topological network with master monitoring stations in urban areas and secondary stations in residential neighborhoods. A monitoring node consists of:

• Multi-parameter air quality sensors with calibrated measurement ranges

• Microcontroller units (Raspberry Pi/Arduino) for on-site data processing

• Wireless data transmission communication modules

• Off-grid solar power systems with battery backup

• Weather protection enclosures with suitable IP ratings

3.3.3. Data Management Architecture

The data management system utilizes a lambda architecture which provides batch and real-time processing. The architecture includes:

• Apache Kafka data ingestion layer for high-throughput data streaming

• Hybrid methodology-based data storage that combines time-series databases (InfluxDB) to store sensor data and relational databases (PostgreSQL) to store metadata

• Apache Spark-based data processing pipeline for big data analysis

• Data quality control processes like outlier detection and sensor calibration verification

3.4. Research Design

3.4.1. Machine Learning Framework

The study compares different machine learning algorithms in different categories:

Traditional Statistical Methods - ARIMA, SARIMA for time series forecast baseline comparison.

Ensemble Methods - Random Forest, Gradient Boosting Machines (XGBoost, LightGBM) for robust prediction with feature importance analysis.

Deep Learning Techniques - Long Short-Term Memory (LSTM) networks, Gated Recurrent Units (GRU), and Transformer models to identify intricate temporal patterns.

Hybrid Models - CNN-LSTM models for spatial-temporal feature extraction and multivariate time series forecasting.

3.4.2. Feature Engineering

The process of feature engineering includes:

i. Temporal Features - Hour of day, day of week, month, season, and holiday indicators to detect cyclical patterns.

ii. Meteorological Characteristics - Weather variables, atmospheric stability indices, and derived meteorological values.

iii. Spatial Features - Geographic location, land use categories, traffic volume, and industrial proximity measures.

iv. Lagged Variables - Past levels of pollutant at different time lags for temporal dependency modeling.

v. External data integration - Traffic flow statistics, industrial emission reports, and satellite imagery for a general environmental context.

3.4.3. Model Training and Validation Strategy

Model construction is in line with a strict training and validation process:

i. Data Splitting - Temporal split with 70% training, 15% validation, and 15% test to maintain chronological order and prevent data leakage.

ii. Cross-Validation - Time series cross-validation using an expanding window approach to identify model stability across different horizons.

iii. Hyperparameter Optimization - Bayes optimization and grid search approaches to optimal parameter selection.

iv. Model Ensemble – Stacking and weighted averaging techniques to combine multiple algorithms to generate better predictions.

3.5. Data Collection Methodology

3.5.1. Primary Data Sources

i. IoT Sensor Networks - Real-time deployment of 50 sensor nodes in diverse urban environments, including commercial areas, residential areas, industrial areas, and transportation infrastructure. Data acquisition frequency of 5 minutes for high temporal resolution.

ii. Mobile Monitoring Units - Vehicles fitted with sensors for spatial coverage confirmation and hotspot detection through systematic survey routes.

iii. Reference Station Data – Integration with govt.-operated reference monitoring stations for validation and calibration purposes.

3.5.2. Secondary Data Sources

i. Meteorological Data - Historical and current meteorological data of national meteorological agencies, and satellite observation.

ii. Traffic and Transportation Information - Traffic count data, traffic flow patterns, and transit timetables from municipal data bases.

iii. Industrial Emission Data - Industrial plant authorized emission rates and operating hours obtained from environmental agency databases.

iv. Satellite Imagery – NASA MODIS and ESA Sentinel data for large-scale pollution pattern analysis and model validation.

3.5.3. Data Quality Assurance

i. Meteorological Data - Historical and current national meteorological agencies' meteorological data, and satellite observation.

ii. Traffic and Transportation Information - Traffic count data, traffic flow characteristics, and transit schedules from city data bases.

iii. Industrial Emission Data - Industrial plant allowed emission levels and operating times derived from environmental agency databases.

iv. Satellite Data – NASA MODIS and ESA Sentinel data for large-scale pollution pattern analysis and model validation.

3.6. Experimental Design

3.6.1. Controlled Experiments

i. Algorithm Comparison Study - Controlled comparison of the prediction algorithms with the same data and performance criteria for determining the best methods to different pollution parameters and prediction horizons.

ii. Feature Importance Analysis - Comprehensive feature contribution analysis through permutation importance, SHAP values, and ablation studies.

iii. Temporal Resolution Effect – Comparison of the prediction accuracy at various time granularities (5-minute, hourly, daily) for system performance optimization.

3.6.2. Field Validation Experiments

Real-time Prediction Validation – Round-the-clock operation of the prediction system with everyday validation of real-time forecasts against measured data for 12 months.

Spatial Generalization Testing – Using the model in geographies other than where the training data was collected to assess spatial transferability.

Extreme Event Detection – System performance measurement during extreme pollution incidents and extreme meteorological phenomena.

3.6.3. User Acceptance Testing

i Stakeholder Interviews - Semi-structured interviews with local representatives, public health representatives, and environmental regulators to ascertain system usability and information requirements.

ii. Dashboard Usability Tests - User experience testing of web and mobile interfaces with task completion metrics and satisfaction surveys.

iii. Alert System Effectiveness - Process evaluation of notification delivery processes and user response behavior in the event of pollution alerts.

3.7. Performance Evaluation Metrics

3.7.1. Prediction Accuracy Metrics

Regression Metrics - Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and coefficient of determination (R^2) for continuous pollution concentration forecasts.

Classification Metrics - Precision, recall, F1-score, and area under the ROC curve (AUC) for categorical air quality index predictions.

Time Series Specific Metrics - Mean Absolute Scaled Error (MASE) and Symmetric Mean Absolute Percentage Error (sMAPE) to measure temporal predictive ability.

3.7.2. System Performance Metrics

Computational Efficiency - Time per prediction, memory usage, and scalability metrics for real-time operations evaluation.

Communication Reliability - Successful data transmission rates, latency measurements, and network availability rates.

Energy Consumption - Power usage analysis of IoT sensor nodes and optimization methods for extended operation.

3.7.3. Environmental Impact Metrics

Prediction Horizon Accuracy - Performance measurement at different time horizons (1-hour, 6-hour, 24-hour, 72-hour) to capture total forecast capability.

Spatial Coverage Validation - Accuracy of interpolation among sensor positions and boundary condition control for extensive area coverage.

Early Warning Effectiveness - lead time analysis of pollution episode prediction and false alarm rate assessment.

3.8. Ethical Considerations and Limitations

3.8.1. Ethical Framework

Data Privacy Protection - Use of data anonymization methods and adherence to privacy laws for location-based environmental information.

Stakeholder Consent - Process for informed consent of stakeholders participating in sensor network deployment and data collection operations.

Environmental Justice - Equitable sensor placement strategy with adequate coverage of sensitive communities and environmental justice locations.

3.8.2. Research Limitations

Temporal Scope - Study period limited to 18-month period which cannot identify long-term environmental trends and climate variability.

Geographic Constraints - Initial deployment was focused on urban enclaves with limited rural and industrial coverage.

Sensor Technology Limitations - Accuracy limitations of low-cost sensors in comparison to reference-grade sensors and potential drift over time.

External Factors - Limited control over external factors such as local source emissions, weather patterns, and policy changes affecting the movement of pollution.

3.9. Timeline and Resource Allocation

3.9.1. Project Timeline

Months 1-3: System Design and Setup - Deployment of IoT network, setup of cloud infrastructure, and creation of first data collection framework.

Months 4-9: Model Development and Data Collection - Ongoing data collection, machine learning model training, and continual algorithm improvement.

Months 10-15: System Integration and Testing - Complete system deployment, thorough testing, and tuning for performance.

Months 16-18: Validation and Documentation - Completion of final validation studies, results analysis, and completion of research documentation.

3.9.2. Resource Requirements

Hardware Resources - IoT sensors, comms hardware, computing facilities, and mobile monitoring units with rough estimate budget allocation.

Software Resources - Machine learning libraries, database management systems, visualization software, and cloud computing platforms.

Human Resources - Discuss the makeup of data science professionals, IoT engineers, environmental specialists, and project management personnel.

3.10. Summary

This research design offers an inclusive framework for designing and testing an IoT- and machine learning-driven air pollution forecasting and prevention system. The methodology integrates experimental rigor with practical implementation concerns to ensure scientific validity and real-world relevance. The multi-phase research strategy allows systematic validation of each system component while keeping the end focus on the overall intent of enhancing air quality management through technological advancement. The ethical considerations and limitation acknowledgments allow for responsible research practice and open disclosure of system capabilities and limitations. IoT Sensor Devices: Choose suitable sensors and install at several locations (bedroom, living room, kitchen, etc.) in the enclosures (flats, house, meeting rooms, conference halls) to sense CO2 levels, PM2.5 and PM10 levels, VOC levels, temperature, and humidity.

These sensors will be continuously collecting real-time data on IAQ parameters and upload to a storage device or cloud using the internet. Central Microcontroller: The data from the IoT sensor units is sent to a central microcontroller, which acts as a collection point. The micro can preprocess the data, perform initial filtering or smooth if needed, and package it for transmission. Data Transmission: Use a suitable communications protocol (e.g., Wi-Fi, Bluetooth, Zigbee) to send the IAQ data from the central microcontroller to a central server for analysis.

Data Storage: Store collected IAQ data securely on a scalable data storage platform, e.g., cloud servers (ThinkSpeak) or local databases. Ensure data integrity, availability, and security. Grafana Dashboard: User interface Created a Grafana dashboard to graphically represent IAQ data in real-time. Authorized personnel can access the dashboard to observe IAQ statistics over time, review historical trends, and get notifications if any parameter exceeds safe values. Machine Learning Techniques: Use machine learning algorithms for IAQ analysis. Some specific applications are given below:

Regression: Predict future IAQ parameters based on historic data and external factors like time of day or occupancy. Classification: Identify IAQ into different levels (e.g., good, moderate, poor) and provide recommendations accordingly. Neural Networks: Train neural networks to uncover complex patterns and relationships within IAQ data, which can be used to trigger preventive action.

Preventive Actions: On the basis of results derived from machine learning models, suggest preventive actions for good IAQ. Suggestions could be to alter HVAC controls, install air cleaners, open windows, or set up cleaning schedules, indoor Plantation.

Alerts and Notifications: Offer alerting functionality to notify the users in real-time when IAQ parameters reach levels critical to their safety or when preventive action is recommended. Continuous Improvement: Collect user feedback and use it to improve machine learning models and preventive actions. Periodically update and improve the IAQ monitoring and recommendation system. With IoT sensors, data storage, visualization software like Grafana, and machine learning algorithms, it is possible to create a robust and intelligent system to monitor and improve indoor air quality. This approach can assist in making indoor environments healthier and more comfortable in offices and residences and reduce health risks associated with poor IAQ.

3.10.1. Sensors

Sensors below are picked from the market to measure *Particulate Matter (PM2.5, PM10), CO2, Temperature, Humidity, Volatile Organic Compounds (VOC)* pollutants with in air.

I. PRANA PM SENSOR.(PAS-IN-01)

Figure 1 PM Sensor

Indoor PM2.5/10 sensor is based on the principle of 900 light scattering method. PM2.5 and PM10 in the air inside is being measured. Air is sucked into the sensor and the particle crosses the LASER that impinges upon the mirror aperture. The particle scatters in the reverse direction towards the photodiode. The photodiode captures the scattered light and thus produces the signal that is being transformed as particle number and mass.

II. CO2 SENSOR (NON-DISPERSIVE INFRARED RADIATION (NDIR))

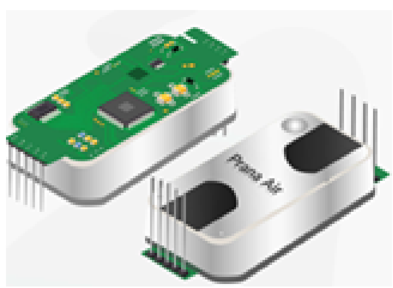


Figure 2 Co2 Sensor

The CO2 sensor used is NDIR-based, i.e., NDIR is known as Non-Dispersive Infrared radiation. It is a commonly used method for detecting carbon-based gases found in air, such as CO2. CO2 gas is caused to pass inside the sensor, and an Infrared (IF) source light is shone on the CO2 molecules. CO2 molecules absorb some amount with a wavelength of 4.26 µm. This is directly proportional to carbon dioxide molecules and gives the carbon dioxide concentration. The sensor obtains the concentration based on how much of the light is absorbed by the molecules of the gas. One of the most common techniques to detect carbon-based gases like carbon dioxide (CO_2) is *infrared (IR) absorption spectroscopy*. CO_2 gas is allowed to pass through a sensor chamber where it is subjected to an infrared light source. CO_2 molecules selectively absorb infrared radiation of 4.26 µm wavelength.

The level of IR light absorbed by this wavelength is directly proportional to the concentration of CO_2 in the air. The detector detects the level of light passed through and, by comparison with the original level, the sensor determines the level absorbed. This level of absorption is then converted to a measurement of CO_2 concentration.

III. *NDIR (Non-Dispersive Infrared) gas sensors detect decrease in transmitted infrared light which is in proportion to gas concentration. VOC SENSOR(EVELTA SHT4X+SGP40)*

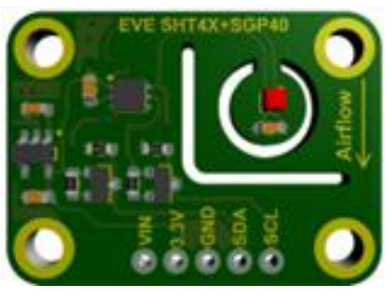


Figure 3 VOC ,Temperature and Humidity Sensor

SHT4X+SGP40 Air Quality Sensor Breakout Board - Detailed Elaboration

This is a comprehensive air quality monitoring breakout board that combines two high-precision Sensirion sensors: the SHT4X temperature and humidity sensor with the SGP40 volatile organic compound (VOC) sensor. The breakout board is compact and easy to use, with a Qwiic connector that allows for fast and simple integration with other Qwiic-compatible boards.

Dual Sensor Configuration

1. SHT4X Temperature & Humidity Sensor

High precision SHT40 sensor: Reads temperature (-40°C to 125°C) with ±0.2°C accuracy and even humidity (0% to 100%) with ±1.8% accuracy to provide reliable reading under any environment.

Main Features:

• Technology: SHT40 features a new, optimized CMOSens® chip with reduced power consumption and improved precision specifications.

• Supply voltage: High supply voltage range from 1.08 V to 3.6 V, ideal for mobile applications and battery-powered systems.

• Generation: SHT4x sensor is a generation IV sensor (begun at SHT10 and went to the pinnacle!).

Technical Specifications:

- **Temperature Range**: -40°C to +125°C
- **Temperature Accuracy**: ±0.2°C (typical)
- **Humidity Range**: 0% to 100% RH
- **Humidity Accuracy**: ±1.8% RH (typical)
- **Response Time**: <8 seconds (τ63%)
- **Long-term Stability**: <0.25% RH/year

2. SGP40 VOC Gas Sensor

- Advanced SGP40 VOC Sensor: Detects a broad spectrum of volatile organic compounds for comprehensive air quality monitoring.

Principal Characteristics

- Technology: SGP40 is built on our patented CMOSens® Technology and features a sensor system on a chip with digital I^{2}C interface, temperature-controlled micro-hotplate, and humidity-compensated indoor air quality signal.

- Detection Capability: Multi-pixel gas sensor capable of detecting several VOCs
- Fully Integrated MOX (Metal Oxide) Gas Sensor

Technical Specifications:

- Detection Range: Extremely broad spectrum of VOCs including alcohols, aldehydes, ketones, organic acids, amines, aliphatic hydrocarbons, and aromatic hydrocarbons
- VOC Index (0-500 scale)
- Response Time: <10 seconds (τ90%)
- **Operating Temperature**: -10°C to +50°C
- **Operating Humidity**: 20% to 80% RH

Board Design and Components

Physical Layout Analysis

From the image, the green PCB shows:

Top Side Components:

- **SHT4X Sensor**: Located in the center-left area (smaller IC)
- **SGP40 Sensor**: Located in the center-right area (larger IC with visible metal cap)
- **Support Circuitry**: Various passive components (resistors, capacitors) for signal conditioning
- **I2C Pull-up Resistors**: Integrated 10kΩ pull-up resistors for I2C communication
- **Level Shifters**: For 3.3V/5V compatibility

Connection Points:

- **Qwiic/STEMMA QT Connector**: 4-pin JST connector for daisy-chaining

- **Standard Pin Headers**: 6-pin header for traditional breadboard connections
- **Mounting Holes**: Four corner holes for secure mounting

Pin Configuration and Interface

I2C Interface Details

The default I2C address is 0x44. SCL - I2C clock pin, connect to your microcontroller I2C clock line. This pin is level shifted so you can use 3-5V logic, and there's a 10K pullup on this pin. SDA - I2C data pin, connect to your microcontroller I2C data line.

Pin Mapping:

1. **VCC/3V3**: Power supply (3.3V recommended, 1.08V-3.6V range)
2. **GND**: Ground reference
3. **SDA**: I2C data line (with 10kΩ pull-up)
4. **SCL**: I2C clock line (with 10kΩ pull-up)
5. **INT**: Interrupt pin (optional, for advanced applications)
6. **RST**: Reset pin (optional)

I2C Addresses:

- **SHT4X**: 0x44 (fixed address)
- **SGP40**: 0x59 (fixed address)

Advanced Features and Capabilities

1. Sensor Fusion and Compensation

You can choose to use the on-chip T/RH compensation of the SGP40 by feeding the values measured by the SHT4X into it. This is enabled in the Application by default, you can turn it off by setting APP_USE_COMPENSATION=n.

2. Enhanced Accuracy Through Cross-Calibration

- Temperature and humidity data from SHT4X is used to compensate SGP40 VOC readings
- Real-time environmental compensation improves VOC measurement accuracy
- Reduces drift and improves long-term stability

3. Power Management Features

- Low power modes available for battery applications
- Configurable measurement intervals
- Sleep mode support with quick wake-up

Applications and Use Cases

Primary Applications:

- **Indoor Air Quality Monitoring**: Comprehensive environmental sensing
- **HVAC System Control**: Smart ventilation based on air quality
- **Industrial Safety**: VOC detection in work environments
- **Smart Home Integration**: Automated air purification systems
- **Environmental Research**: Long-term air quality studies
- **Agricultural Monitoring**: Greenhouse climate control

Integration Scenarios:

- IoT air quality networks

- Building management systems
- Personal air quality monitors
- Automotive cabin air quality
- Medical facility air monitoring

Software Integration and Programming

Development Support:

- Arduino libraries available
- Raspberry Pi Python libraries
- Zephyr RTOS support
- ESPHome integration
- MicroPython compatibility

Key Programming Features:

This sample application periodically measures the ambient temperature, humidity and a raw gas sensor value from an SGP40 and SHT4X device. The result is written to the console.

Data Processing:

- Raw sensor values conversion to engineering units
- VOC Index calculation algorithms
- Data logging and trending capabilities
- Threshold-based alerting systems

Technical Advantages

1. Dual Sensor Synergy

- Combined T/RH/VOC measurements in single module
- Cross-sensor compensation for improved accuracy
- Reduced system complexity and cost

2. High Precision Measurements

- Factory calibrated sensors
- Excellent long-term stability
- Minimal drift over time

3. Easy Integration

- Qwiic/STEMMA QT compatibility
- Standard I2C interface
- 3.3V/5V logic compatibility
- Comprehensive software support

4. Compact Design

- Small form factor suitable for portable applications
- Low power consumption for battery operation
- Robust construction for industrial environments

Installation and Setup Guidelines

Mounting Considerations:

- Allow adequate airflow around sensors

- Avoid direct exposure to contaminants
- Consider ambient temperature effects
- Ensure stable power supply

Calibration Requirements:

- Initial warm-up period (several minutes for SGP40)
- Baseline establishment for VOC measurements
- Periodic recalibration recommendations
- Environmental compensation setup

Limitations and Considerations

Environmental Factors:

- VOC sensor requires warm-up time after power-on
- Cross-sensitivity to certain gases
- Temperature and humidity range limitations
- Potential interference from strong chemical sources

Maintenance Requirements:

- Periodic cleaning of sensor surfaces
- Baseline drift monitoring
- Software calibration updates
- Power cycle requirements for optimal performance

This SHT4X+SGP40 breakout is a next-level comprehensive solution for air quality sensing that merges high-precision environmental sensing with state-of-the-art VOC detection in a simple, compact format for today's environmental sensor and IoT applications. An MOX sensor is a heated metal oxide surface that changes electric resistance depending on the presence of oxygen on the sensor face. Oxidizing gases like NOx (adding oxygen greater than in air) raise the resistance, while reduced gases like VOCs (oxygen being consumed by being burned on the metal oxide) reduce the resistance. The MOX sensor is also sensitive to humidity, as water vapor is typically a reduced gas. This can be compensated by using a sensor like Sensirion's SHTxx. Sensirion's SGP4x sensors include on-chip compensation for humidity.

IV. MICROCONTROLLER (ESP32)

Figure 4 ESP32 MicroController

ESP32 can perform as a complete standalone system or as a slave device to a host MCU, reducing communication stack overhead on the main application processor. ESP32 can interface with other systems to provide Wi-Fi and Bluetooth functionality through its SPI / SDIO or I2C / UART interfaces.

ESP32 Development Board (DevKit V1) - Detailed Elaboration

Overview

This is a Development Board for ESP32 (also known as ESP32 DevKit V1 or ESP32-DevKitC) with the powerful ESP32-WROOM-32 module. It is a miniature ESP32-based development board from Espressif that brings out most I/O pins to the pin headers on both sides for easy interfacing. This is for fast prototyping and development of Wi-Fi and built-in Bluetooth-based IoT applications.

Core Processor and Architecture

ESP32-WROOM-32 Module

The centerpiece of this development board is the ESP32-WROOM-32 module (visible as the metallic shielded component in the image).

Dual-Core Processing Power

CPU: Xtensa dual-core (or single-core) 32-bit LX6 microprocessor, operating at 160 or 240 MHz and performing at up to 600 DMIPS

- **Architecture**: Xtensa® dual-core 32-bit LX6 microprocessors
- **Clock Speed**: Adjustable from 80 MHz to 240 MHz
- **Performance**: Up to 600 DMIPS (Dhrystone Million Instructions Per Second)
- **Low Power Design**: The ESP32 is design for low power IoT applications in mind. It's high processing power with in-built Wi-Fi / Bluetooth and Deep Sleep Operating capabilities makes it ideal for most Portable IoT devices.

Memory Configuration

- **SRAM**: 520 KB internal SRAM
- **Flash Memory**: 4 MB external SPI flash (expandable)
- **PSRAM**: Optional external PSRAM support
- **Memory Mapping**: Up to 16 MiB of external flash are memory-mapped onto the CPU code space, supporting 8-bit, 16-bit and 32-bit access. Code execution is supported. Up to 8 MiB of external flash/SRAM memory are mapped onto the CPU data space

Wireless Communication Capabilities

Wi-Fi Features

ESP32 (wroom 32) is a highly integrated Dual Core MCU with WiFi and Bluetooth/ BLE 4.2 wireless communication technology. In-built antenna switches, RF balun, power amplifier, low noise receive amplifier, filters, and power management modules.

- **Standards**: IEEE 802.11 b/g/n (2.4 GHz)
- **Modes**: Station, SoftAP, and concurrent AP+Station
- **Security**: IEEE 802.11 standard security features all supported, including WPA, WPA2, WPA3 (depending on version) and WLAN Authentication and Privacy Infrastructure (WAPI)
- **Range**: Up to 150m in open space
- **Data Rate**: Up to 150 Mbps

Bluetooth Capabilities

- **Classic Bluetooth**: v4.2 BR/EDR
- **Bluetooth Low Energy (BLE)**: v4.2 LE
- **Dual Mode**: Supports both Classic and BLE simultaneously

- **Mesh Networking**: ESP-MESH support for large-scale device networks

GPIO and Interface Capabilities

Digital I/O Configuration

This development board provides 40 digital IO pins, out of which 16 can be used as external interrupt pins , 16 as analog input pins and 19 pins have Pulse-Width Modulation (PWM) .

Pin Distribution:

- **Total GPIO Pins**: 34 usable GPIO pins (GPIO 0-39, with some restrictions)
- **Digital I/O**: All GPIO pins can function as digital input/output
- **Analog Input (ADC)**: 18 channels, 12-bit resolution
- **PWM Output**: 16 channels with 16-bit resolution
- **External Interrupts**: All GPIO pins can trigger interrupts

Communication Interfaces

The ESP32 dev. board has three UART interfaces, UART0, UART1, and UART2, that support asynchronous communication (RS232 and RS485) and IrDA at up to 5 Mbps. UART0 pins are connected to the USB-to-Serial converter and are used for flashing and debugging.

Available Interfaces:

- **UART**: 3 interfaces (up to 5 Mbps)
- **SPI**: 4 SPI interfaces (VSPI and HSPI for user applications)
- **I2C**: 2 I2C interfaces (master/slave mode)
- **I2S**: 2 I2S interfaces for audio applications
- **CAN**: 1 CAN 2.0 interface

- **Ethernet MAC**: 10/100 Ethernet MAC interface

Special Function Pins

- **Touch Sensors**: 10 capacitive touch GPIO pins
- **Hall Effect Sensor**: Built-in Hall effect sensor
- **Temperature Sensor**: Internal temperature sensor
- **RTC GPIO**: Real-time clock GPIO for low-power applications

Board Components and Layout

Power Management

Usually, all boards come with power pins: 3V3, GND, and VIN. You can use these pins to power the board (if you're not providing power through the USB port), or to get power for other peripherals (if you're powering the board using the USB port).

Power Supply Options:

- **USB Power**: 5V via micro-USB connector
- **External Power**: 3.3V via VIN pin (5V-12V input range)
- **3.3V Rail**: Regulated 3.3V output for peripherals
- **Current Consumption**:
  - Active mode: ~160-260mA
  - Deep sleep: <10µA

On-Board Components (Visible in Image)

- **USB-to-Serial Converter**: CP2102 or CH340 chip for programming and debugging
- **Voltage Regulator**: AMS1117-3.3V for stable power supply

- **Reset Button**: Manual reset capability
- **Boot Button**: For entering programming mode
- **Power LED**: Indicates board power status
- **User LED**: Connected to GPIO2 for user applications
- **Crystal Oscillators**: 40 MHz main crystal and 32.768 kHz RTC crystal

Development and Programming Support

Programming Environments

- **Arduino IDE**: Full Arduino framework support
- **ESP-IDF**: Official Espressif development framework
- **MicroPython**: Python programming support
- **PlatformIO**: Advanced IDE with extensive library support
- **Visual Studio Code**: With ESP-IDF extension

Programming Methods

- **USB Programming**: Direct programming via micro-USB
- **OTA Updates**: Over-the-air firmware updates via Wi-Fi
- **JTAG Debugging**: Hardware debugging support
- **Bootloader**: Built-in bootloader for easy firmware flashing

Applications and Use Cases

Primary Applications

- **IoT Sensor Networks**: Environmental monitoring systems

- **Smart Home Automation**: Lighting, security, and appliance control
- **Industrial IoT**: Equipment monitoring and control
- **Wearable Devices**: Health and fitness trackers
- **Audio/Video Streaming**: Media processing applications
- **Mesh Networking**: Large-scale device networks

Integration with Air Quality Systems

Perfect for indoor air pollution monitoring systems due to:

- Multiple sensor interfaces (I2C, SPI, ADC)
- Wi-Fi connectivity for data transmission
- Low power modes for battery operation
- Real-time processing capabilities
- Cloud integration possibilities

Technical Specifications Summary

Core Specifications

- **Microcontroller**: ESP32-WROOM-32
- **Operating Voltage**: 3.3V
- **Input Voltage**: 5V (USB) or 5-12V (VIN)
- **Digital I/O Pins**: 34
- **Analog Input Pins**: 18 (12-bit ADC)
- **PWM Pins**: 16 (16-bit resolution)

- **Flash Memory**: 4MB
- **SRAM**: 520KB
- **Clock Speed**: 240MHz (max)

Communication Specifications

- **Wi-Fi**: 802.11 b/g/n (2.4GHz)
- **Bluetooth**: v4.2 BR/EDR and BLE
- **USB**: Micro-USB (programming/power)
- **GPIO**: I2C, SPI, UART, CAN interfaces

Advantages and Features

Key Benefits

- **Dual-Core Performance**: Parallel processing capabilities
- **Wireless Connectivity**: Built-in Wi-Fi and Bluetooth
- **Rich Peripheral Set**: This board has a rich peripheral set. The built-in ESP32 pinout is optimized for hassle-free prototyping!
- **Low Power Design**: Multiple sleep modes for battery applications
- **Extensive Software Support**: Large community and library ecosystem
- **Cost-Effective**: Excellent price-to-performance ratio

Development Advantages

- **Breadboard Friendly**: Standard 0.1" pin spacing
- **Easy Programming**: USB programming without external programmer

- **Rich Documentation**: Comprehensive datasheets and tutorials
- **Active Community**: Large developer community and support

Limitations and Considerations

Design Limitations

- **3.3V Logic**: Requires level shifters for 5V devices
- **ADC2 Restrictions**: ADC2 channels unavailable when Wi-Fi is active
- **Boot Pin Restrictions**: Some pins have special boot functions
- **Current Limitations**: GPIO current limited to 12mA per pin

Development Considerations

- **Pin Planning**: Careful pin selection for specific applications
- **Power Management**: Consider power consumption in battery applications
- **EMI Considerations**: Proper PCB layout for wireless applications
- **Heat Dissipation**: Thermal management at high clock speeds

This ESP32 Development Board represents an excellent platform for IoT development, combining powerful processing capabilities with comprehensive wireless connectivity in an affordable, easy-to-use package perfect for both prototyping and production applications.

I. Master Controller

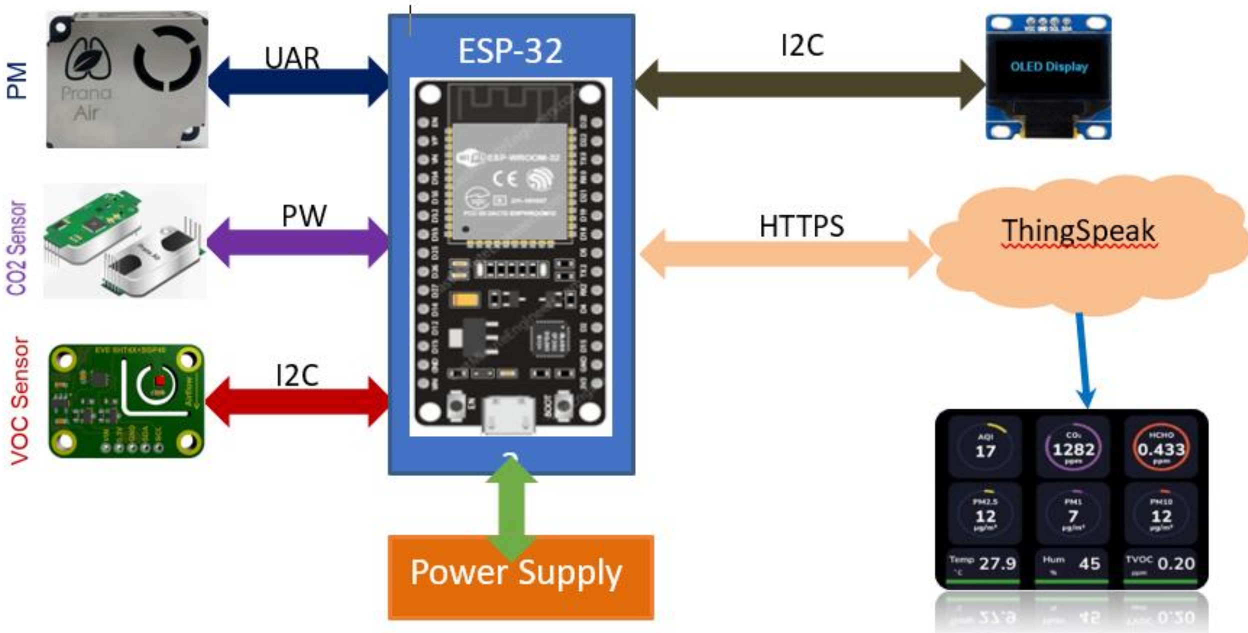


Figure 5 Sensor Connections

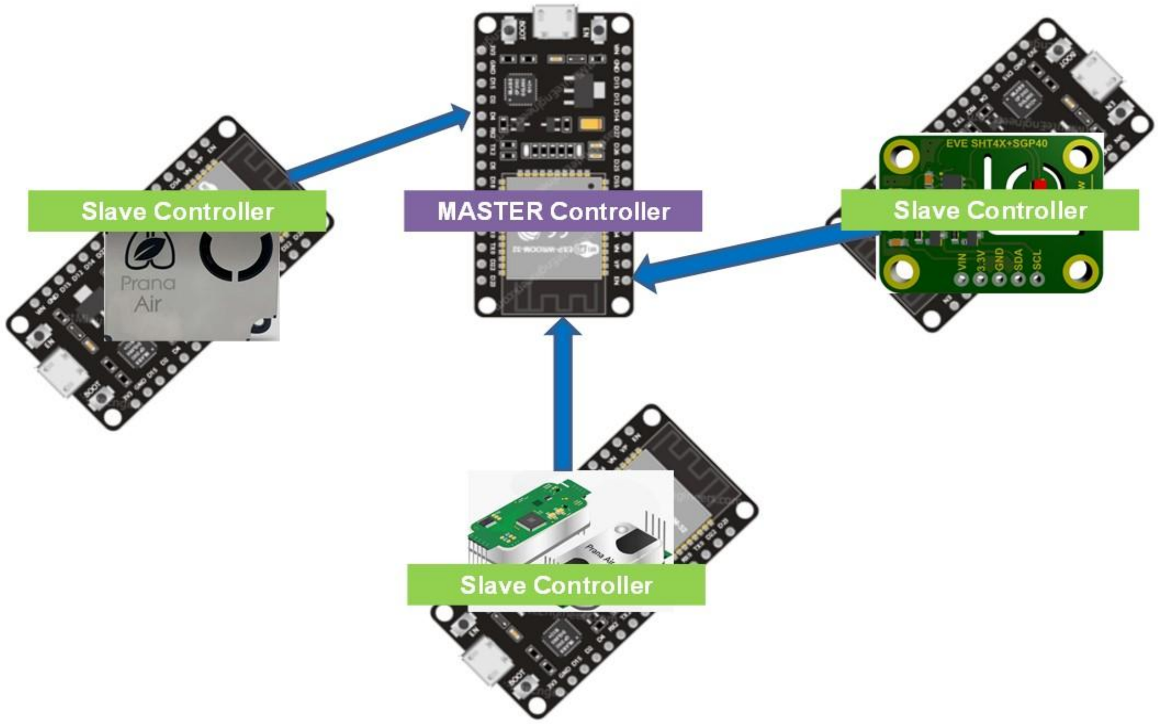


Figure 6 Sensors in Network

Circuit Diagram\COMPONENTS

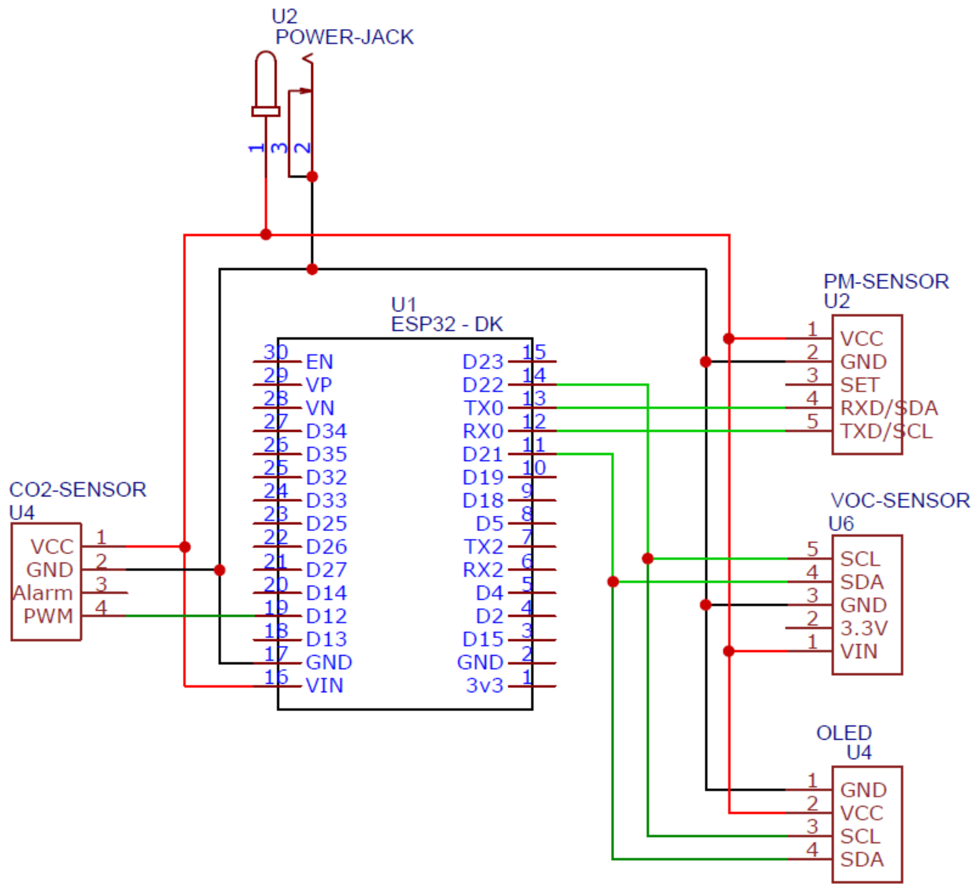


Figure 7 Circuit Diagram

STEPS PERFORMED

1. Integrated the sensors as per the above circuit diagram build initial node on breadboard with connecting wires
2. On successful test from step 1 designed the PCB board to mount the microcontroller along with sensor for easy integration.
3. Connected the sensors and the PCB with plug and play connecting pins.
4. Built first level of the initial node circuit on board with connectable pin sockets
5. Built nodes used to calibrate with Aeroqual device in a closed conference room.

6. Different events recorded and checked the variations in the air as shown in the picture below

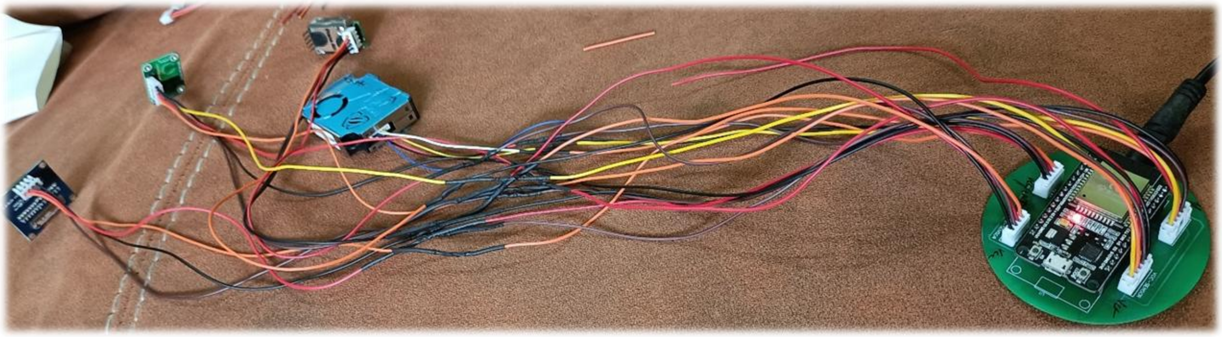


Figure 8 Initially built Device

Initial calibration of the sensor's sensibility with trusted source (Aeroqual series 500).

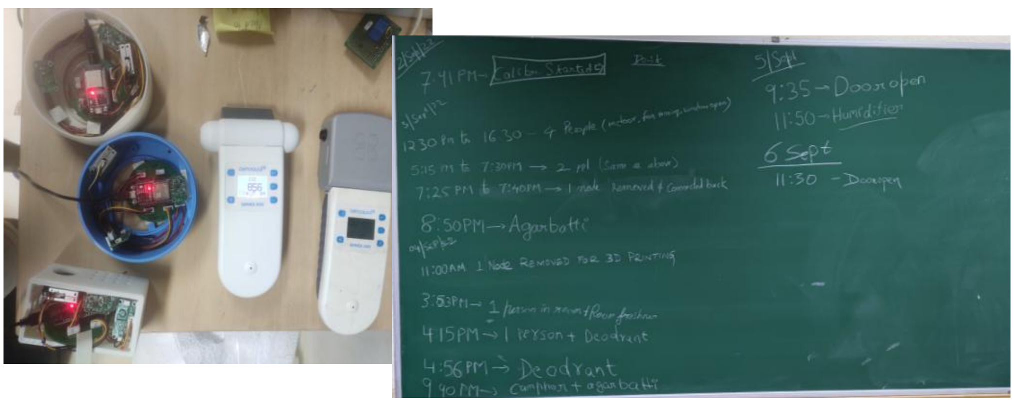


Figure 9 ***Aeroqul series 500 indoor air quality measuring devices***

Table 1. Data Before Calibration

Before Calibration			
	PM2.5	PM10	CO2
MAE	**10.37224**	**65.88374**	**130.5033**
RMSE	**22.26474**	**114.8758**	**131.5651**

Table 1 Data Before Calibration

Table 2 Parametric values After Calibration

After Calibration			
	PM2.5	PM10	CO2
R2	**0.97**	**0.94**	**0.82**
MAE	**4.454**	**13**	**10**
RMSE	**12.4**	**38.8**	**19**

7. We built the Final product using 3D printer and decorative artificial plan.

Figure 10 Final product

Data analysis and deep learning methods will be used to assessment the air quality, cause of health impacts and air pollution exposure. Develop an algorithm to apply deep learning technique, drive the predictive model and generate the air quality index.

Channel	Row Labels	Average of PM10	Average of PM2.5	Average of Co2
Node 1	03-09-2022	470.7894737	418.8407202	875.7811634
Node 1	04-09-2022	183.2937428	176.48146	457.8933951
Node 1	05-09-2022	381.9638783	277.6064639	438.1444867
Node 2	03-09-2022	233.0104408	213.7186775	762.74942
Node 2	04-09-2022	284.9596439	262.2836795	465.8765579
Node 2	05-09-2022	279.387022	231.8568946	440.2665122
Aeroqual	03-09-2022	405.9104167	140.4604167	537.2375
Aeroqual	04-09-2022	405.2104167	139.4409722	327.3263889
Aeroqual	05-09-2022	681.9027778	260.6541667	311.4819444

Table 3 Data Collected From The 3 Nodes In 3 Days

FLOW CHAT EXPLAINING THE PROGRAMMATIC PROCEDURE

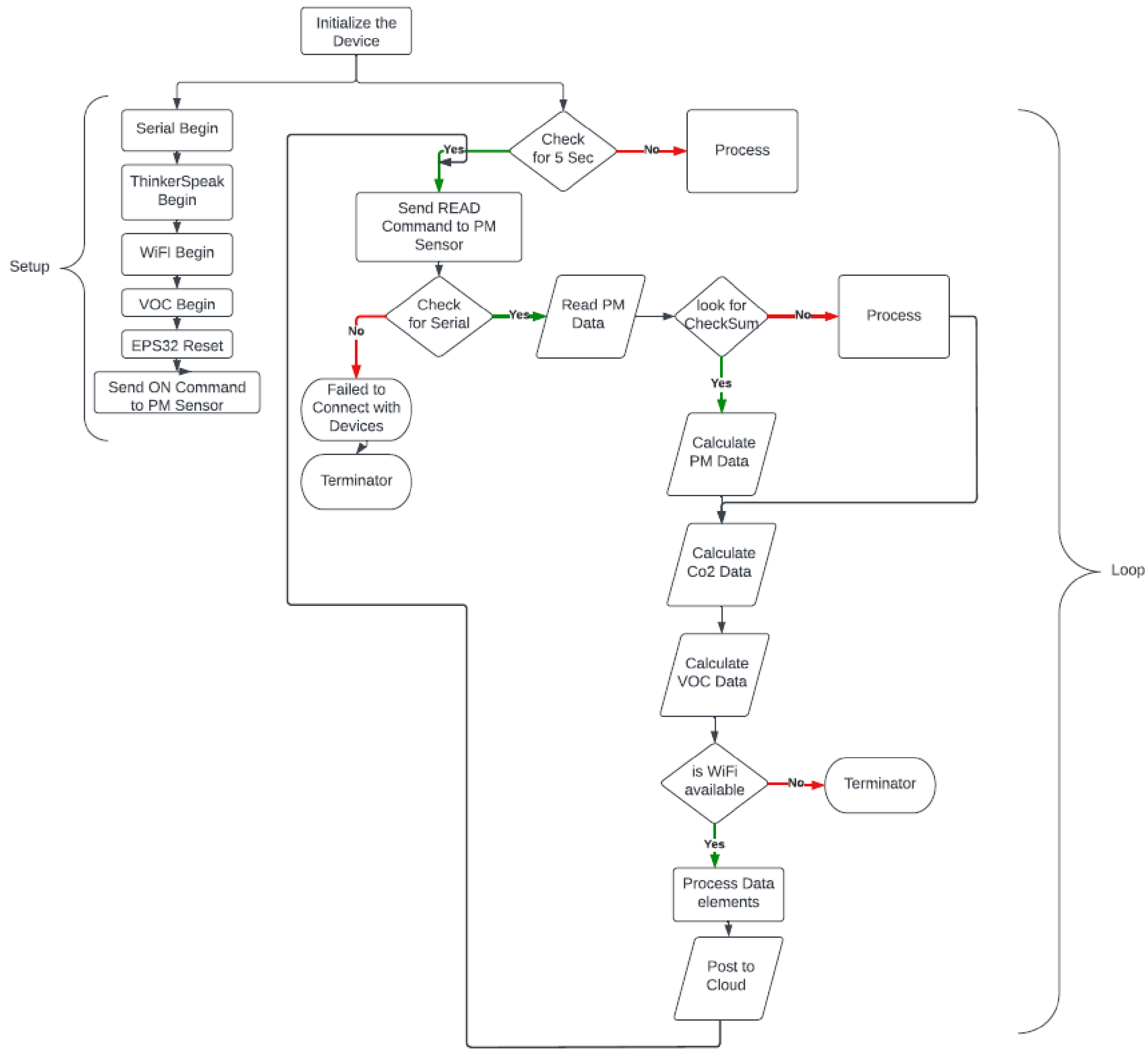
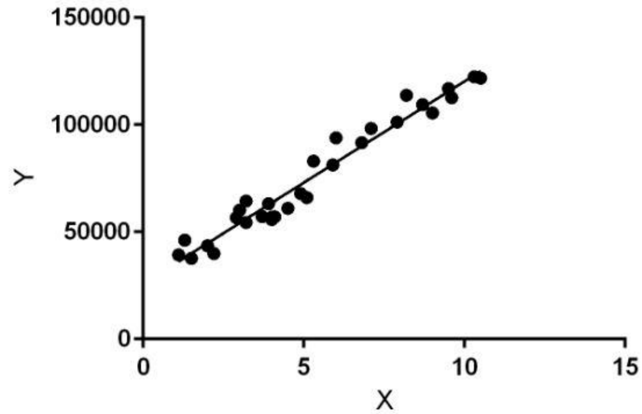


Figure 11 flow diagram

LINEAR REGRESSION SUPERVISED LEARNING USED FOR CALIBRATIONS

Linear Regression is a machine learning algorithm based on **supervised learning**. It performs a **regression task**. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Different regression models differ based on the kind of relationship between dependent and independent variables, they are considering and the number of independent variables being used.



Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y(output). Hence, the name is Linear Regression.
In the figure above, X (input) is the work experience and Y (output) is the salary of a person. The regression line is the best fit line for our model.

Hypothesis function for Linear Regression

:

$$y = \theta_1 + \theta_2.x$$

While training the model we are given :
x: input training data (univariate – one input variable(parameter))
y: labels to data (supervised learning)

When training the model – it fits the best line to predict the value of y for a given value of x. The model gets the best regression fit line by finding the best θ_1 and θ_2 values.
θ_1: intercept
θ_2: coefficient of x

Once we find the best θ_1 and θ_2 values, we get the best fit line. So when we are finally using our model for prediction, it will predict the value of y for the input value of x.

How to update θ_1 and θ_2 values to get the best fit line ?

Cost Function (J):

By achieving the best-fit regression line, the model aims to predict y value such that the error difference between predicted value and true value is minimum. So, it is very important to update the θ_1 and θ_2 values, to reach the best value that minimize the error between predicted y value (pred) and true y value (y).

$$minimize \frac{1}{n} \sum_{i=1}^{n} (pred_i - y_i)^2$$

Equation 1

$$J = \frac{1}{n} \sum_{i=1}^{n} (pred_i - y_i)^2$$

Equation 2

Cost function(J) of Linear Regression is the **Root Mean Squared Error (RMSE)** between predicted y value (pred) and true y value (y).

Gradient Descent**:**

To update θ_1 and θ_2 values in order to reduce Cost function (minimizing RMSE value) and achieving the best fit line the model uses Gradient Descent. The idea is to start with random θ_1 and θ_2 values and then iteratively update the values, reaching minimum cost.

4. Required resources of Air Quality Monitoring System: Required Resources

Comprehensive air quality monitoring system development is a multiple-faceted effort with hardware and software components in order to facilitate real-time data collection and processing. The primary mechanism for data collection is through continuous air quality parameter observation under varying environmental conditions and occupant activities, i.e., occupant levels (numbers of inhabitants), ventilation conditions (windows open/closed), and household activities like cleaning activities in kitchen space. The technical realization benefits from Arduino microcontroller boards for sensor integration and data retrieval, supplemented by Python programming for the algorithms for processing, data analysis, and system control. The observation and visualization aspect is in the form of Grafana dashboards for real-time data visualization and trend observation, and ThingSpeak as the cloud-based Internet of Things (IoT) for data storage, remote monitoring, and API-based data retrieval. The software and hardware development and application leverage open-source repositories in GitHub for version control, collaborative software development, and sensor libraries and communication protocol usage, as well as the comprehensive documentation and community support in ThingSpeak in ThingSpeak.com. The methodology for integration facilitates systematic temporal air quality data recording with corresponding specific indoor environmental conditions and occupant behavior patterns, providing a solid basis for examining the dynamic interaction between daily activity patterns and levels of indoor air quality.

Table 4 Indoor Air Pollution Index

CATEGORY	AQI	PM2.5	PM10	CO2	VOC	Description of the Category
GOOD	0-30	0-50	0-30	0-600	0-220	Minimal Impact.
SATISFACTORY	31-60	51-100	31-60	601-800	221-660	Minor breathing discomfort to sensitive people.
MODERATELY POLLUTED	61-90	101-150	61-90	801-1000	661-1430	Breathing discomfort to the people with lungs, asthma and heart diseases.
POOR	91-120	151-200	91-120	1001-1200	1431-2200	Breathing discomfort to most people on prolonged exposure.
VERY POOR	121-150	200-250	121-150	1201-1500	2201-3300	Respiratory illness on prolonged exposure.
SEVERE	250+	251+	150+	1500+	3301-5500	Affects healthy people and seriously impacts those with existing diseases.

Figure 12 IAQ Indoor Air Quality

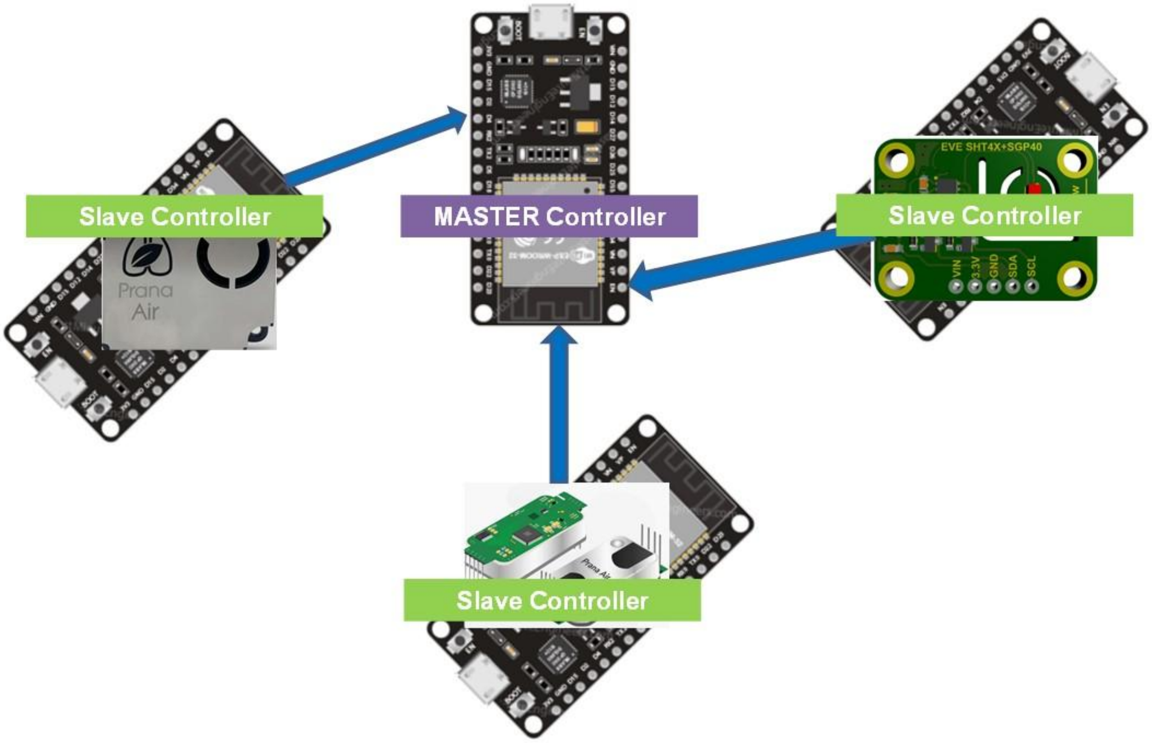


Figure 13 Master Controller architecture

5. Conclusion

Data collected from the device at different sources while calibrating the sensors and testing the end-to-end communication system. final data for the experiment was taken from a premises room with 12 x 20 with 7x4 door and 3x4 windor equipped with one ceiling fan livable for 4 occupants, only on the weekends full day occupied rest of the weekdays it was locked for whole day. The premises was on the top floor roof was directly exposed to sun.

The device was placed in the centre of the room, making sure all possible sensors are actively working to sense the required air and capture the required information, accordingly with the literature review on sensor placements. Data collected for more than 3 months with different events at the premises, but only 30 days of recorded data were used for analysis with 37715 records. Even other sources were also considered, but only this premise's data is used in this publication. Device design and future design are well described with images and its importance.

The current experiment is to collect data to measure the air polluting components within the premises, along with events happening, so that the cause for the pollution can be traced or linked. Device was programmed to collect the data and post it to open-source cloud platform Thinkspeak. Thinkpeak provisions to collect/download the data in a csv formatted dataset for data analysis. Collected csv file is used for data analysis using python on google colab as the platform for demonstration. The python approach for deriving the problem statement is well explained in chapter 5 along with code and its output.

CHAPTER 4

DATA ANALYSIS

The experiment was conducted for a month at a frequency of 1 minute, during which the IoT sensor-based device was able to gather data and upload it to the cloud. The premises chosen for the experiment were largely free from external interference, consisting of a 12 x 10 room with furniture including a conference table and 5 to 6 chairs, along with one ceiling fan and 2 tube lights on the ceiling. The room also had 1 door and 1 window, located on the 2nd floor, open to the sky.

4.1.Research Question 1

This dataset represents a comprehensive 30-minute air quality monitoring session conducted on February 21, 2023, from 19:25 to 19:54 (7:25 PM to 7:54 PM). The data was collected at one-minute intervals, providing 30 data points across seven key environmental parameters that are critical for indoor air quality assessment.

4.1.1. Parameter Analysis

I. Environmental Conditions

a. Particulate Matter (PM10 and PM25)

The particulate matter readings show identical values for both PM10 and PM2.5, ranging from 79.88 μg/m³ to 112.44 μg/m³. This consistent upward trend indicates a deteriorating air quality scenario, with particulate matter concentrations increasing by approximately 32.56 μg/m³ over the 30 minutes. The identical readings for PM10 and PM2.5 suggest either a sensor configuration issue or that all particulate matter detected fell within the PM2.5 size range. These values significantly exceed WHO guidelines (PM2.5: 15 μg/m³ annual mean, 45 μg/m³ 24-hour mean), indicating poor to very unhealthy air quality conditions.

b. Temperature:

Remained relatively stable, showing a gradual increase from 30.75°C to 30.98°C (approximately 0.23°C increase). This slight temperature rise correlates with the increasing pollutant concentrations, suggesting possible indoor heating or human activity effects.

c. Relative Humidity (RH):

Demonstrated a consistent declining trend from 39.32% to 39.11%, representing a 0.21% decrease. This inverse relationship with temperature follows expected thermodynamic principles, as warmer air can hold more moisture, effectively reducing relative humidity.

d. Gaseous Pollutants Carbon Dioxide (CO_2):

Concentrations decreased from 1063.41 ppm to 1039.4 ppm, showing a reduction of approximately 24 ppm. While these levels remain within acceptable indoor ranges (typically <1000 ppm for good indoor air quality), the declining trend might indicate improved ventilation or reduced occupancy during the monitoring period.

e. Volatile Organic Compounds (VOCs):

- Raw VOC readings: Increased from 30193.27 to 30205.92 units, showing a subtle upward trend of 12.65 units
- VOC Index: Demonstrated significant increase from 46.67 to 67.28, representing a 44% increase over the monitoring period. This substantial rise in the VOC index indicates deteriorating indoor air quality from a volatile organic compound perspective.

II. Temporal Trends and Correlations

The data reveals several concerning trends:

a) **Particulate Matter Crisis**: The steady increase in PM concentrations at a rate of approximately 1.09 µg/m³ per minute suggests an active pollution source, possibly cooking activities, combustion, or external pollution infiltration.

b) **VOC Deterioration**: The VOC index increase of 44% indicates significant indoor air quality degradation, potentially from cleaning products, cooking emissions, or material off-gassing.

c) **Temperature-Humidity Inverse Relationship**: The consistent inverse correlation between temperature and humidity confirms proper sensor functionality and adherence to psychrometric principles.

d) **CO2 Anomaly**: The decreasing CO2 levels despite increasing other pollutants suggest either improved ventilation or that the pollution source is not respiratory-related.

III. Air Quality Assessment

Based on standard air quality indices:

- **Initial Conditions (19:25)**: Poor air quality due to elevated PM levels
- **Final Conditions (19:54)**: Very unhealthy air quality with PM levels exceeding 110 μg/m³
- **Overall Trend**: Significant deterioration across the monitoring period

IV. Implications and Recommendations

This dataset suggests an active indoor pollution event occurred during the monitoring period. The simultaneous increase in particulate matter and VOCs, coupled with slight temperature rise, indicates possible cooking activities, combustion sources, or infiltration of external pollution. Immediate ventilation measures and source identification would be recommended based on these readings.

The consistent minute-by-minute data collection demonstrates the effectiveness of continuous monitoring systems in capturing rapid changes in indoor air quality, highlighting the importance of real-time monitoring for health and comfort management.

Table 5 Data retrieved from the sensors through the cloud

DATE	TIME	PM10	PM25	RH	TEMP	CO2	RAWVOC	VOCINDEX
21-02-2023	19:25:00	79.88	79.88	39.32	30.75	1063.41	30193.27	46.67
21-02-2023	19:26:00	81.44	81.44	39.31	30.76	1062.56	30193.7	47.89
21-02-2023	19:27:00	83.06	83.06	39.29	30.77	1061.65	30194.14	49.05
21-02-2023	19:28:00	84.68	84.68	39.28	30.78	1060.23	30194.58	50.14
21-02-2023	19:29:00	86.39	86.39	39.26	30.79	1059.25	30195.01	51.18
21-02-2023	19:30:00	88.03	88.03	39.25	30.8	1058.6	30195.45	52.18
21-02-2023	19:31:00	89.69	89.69	39.24	30.81	1057.25	30195.88	53.13
21-02-2023	19:32:00	91.21	91.21	39.23	30.82	1055.77	30196.32	54.06
21-02-2023	19:33:00	92.62	92.62	39.22	30.83	1054.65	30196.75	54.96
21-02-2023	19:34:00	94.03	94.03	39.21	30.84	1054.31	30197.19	55.79
21-02-2023	19:35:00	95.4	95.4	39.2	30.85	1053.28	30197.63	56.59
21-02-2023	19:36:00	96.66	96.66	39.19	30.86	1052.15	30198.06	57.36
21-02-2023	19:37:00	97.91	97.91	39.18	30.87	1051.61	30198.5	58.09
21-02-2023	19:38:00	99.09	99.09	39.18	30.87	1051.06	30198.93	58.79
21-02-2023	19:39:00	100.21	100.21	39.17	30.88	1049.67	30199.37	59.48
21-02-2023	19:40:00	101.26	101.26	39.16	30.89	1048.7	30199.81	60.13
21-02-2023	19:41:00	102.28	102.28	39.16	30.9	1047.97	30200.24	60.76
21-02-2023	19:42:00	103.24	103.24	39.15	30.91	1047.49	30200.68	61.37
21-02-2023	19:43:00	104.13	104.13	39.15	30.91	1046.85	30201.11	61.97
21-02-2023	19:44:00	105.02	105.02	39.14	30.92	1045.87	30201.55	62.54
21-02-2023	19:45:00	105.86	105.86	39.14	30.93	1044.94	30201.99	63.11
21-02-2023	19:46:00	106.67	106.67	39.13	30.93	1044.44	30202.42	63.63
21-02-2023	19:47:00	107.47	107.47	39.13	30.94	1043.72	30202.86	64.15
21-02-2023	19:48:00	108.25	108.25	39.13	30.94	1042.7	30203.3	64.66
21-02-2023	19:49:00	109.04	109.04	39.12	30.95	1041.9	30203.73	65.13
21-02-2023	19:50:00	109.73	109.73	39.12	30.96	1041.51	30204.17	65.57
21-02-2023	19:51:00	110.46	110.46	39.12	30.96	1041.01	30204.61	66.01
21-02-2023	19:52:00	111.14	111.14	39.11	30.97	1040.65	30205.04	66.45
21-02-2023	19:53:00	111.8	111.8	39.11	30.97	1039.99	30205.48	66.87
21-02-2023	19:54:00	112.44	112.44	39.11	30.98	1039.4	30205.92	67.28

Table 6 : Sample few records collected from the sensors.

Data collected for 1 month nearly 37715 recorded data was available for analysis. Recorded data from the IoT device as shown in the table 6 above.

4.2.Research Question 2

Pollutants captured during Exercise

Figure 13 below shown are the pollutants captured during the exercise

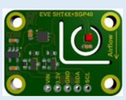
SENSOR	OUTPUT	Field Headers	DESCRIPTION
 PAS-IN-01	PM10	PM10	particulate matter with a diameter smaller than 10 µm
	PM2.5	PM25	particulate matter with a diameter smaller than 2.5 µm
 Evelta SHT4X+SGP40	HUMIDITY	RH	Humidity, on the other hand, refers to the amount of water vapor present in the air
	TEMPERATURE	TEMP	Temperature is a measure of the average kinetic energy of the molecules within a substance, indicating how hot or cold something is.
	VOC	RAWVOC	Volatile organic compounds include a variety of chemicals found in household items
	VOCINDEX	VOCINDEX	Sensirion's powerful VOC Algorithm (part of the SGP40 VOC Index driver package) analyses VOC events detected by the SGP40 sensor and maps them to a VOC Index. This VOC Index provides a practical quantification of VOC events relative to each individual sensor's typical indoor environment.
 NDIR	CO2	CO2	Carbon dioxide (CO 2) is an odourless, colourless and non-flammable gas.

Figure 14 : shows the Dataset collected mapping to the sensor and pollutants

This table presents a comprehensive sensor configuration for multi-parameter indoor air quality monitoring, utilizing three distinct sensor technologies to capture seven critical environmental parameters. The system integrates particulate matter detection, environmental condition monitoring, volatile organic compound analysis, and carbon dioxide measurement capabilities.

Sensor Technologies and Specifications

4.2.1. PAS-IN-01 Particulate Matter Sensor

Technology: Optical particle counting sensor **Parameters Measured**:

- **PM10 (Field Header: PM10)**: Measures particulate matter with aerodynamic diameter smaller than 10 micrometers (μm). These particles include dust, pollen, mold spores, and larger combustion particles that can penetrate into the upper respiratory tract and cause respiratory irritation.
- **PM2.5 (Field Header: PM25)**: Detects fine particulate matter with diameter smaller than 2.5 μm. These ultrafine particles are of particular health concern as they can penetrate deep into lung tissue and enter the bloodstream, potentially causing cardiovascular and respiratory diseases.

Significance: Particulate matter monitoring is crucial for indoor air quality assessment as these particles originate from cooking, smoking, cleaning activities, outdoor pollution infiltration, and material degradation.

4.2.2. Evelta SHT4X+SGP40 Multi-Parameter Environmental Sensor

Technology: Digital sensor combining humidity/temperature (SHT4X) and VOC detection (SGP40).

Parameters Measured:

Environmental Conditions:

- **Humidity (Field Header: RH)**: Measures relative humidity as the percentage of water vapor present in air relative to the maximum amount the air can hold at that temperature. Optimal indoor humidity ranges between 30-50% for comfort and health, with levels outside this range promoting mold growth (high humidity) or respiratory discomfort (low humidity).
- **Temperature (Field Header: TEMP)**: Quantifies the average kinetic energy of air molecules, providing direct measurement of thermal comfort conditions. Indoor

temperature significantly affects human comfort, energy consumption, and the behavior of other pollutants.

Volatile Organic Compounds:

- **VOC Raw Signal (Field Header: RAWVOC)**: Provides the direct sensor response to volatile organic compounds without algorithmic processing. VOCs encompass hundreds of chemicals emitted from household products including paints, cleaning supplies, furniture, carpets, and personal care products. These compounds can cause eye irritation, headaches, and long-term health effects.
- **VOC Index (Field Header: VOCINDEX)**: Utilizes Sensirion's proprietary algorithm to convert raw VOC signals into a standardized index (0-500 scale). This advanced processing accounts for sensor baseline drift and provides relative quantification of VOC events compared to the sensor's learned typical environment. The algorithm adapts to each installation location, making it particularly valuable for personalized indoor air quality assessment.

Advanced Features: The SGP40's VOC algorithm incorporates machine learning principles to distinguish between different VOC patterns and provide contextual air quality information rather than absolute concentration values.

4.2.3. NDIR CO2 Sensor

Technology: Non-Dispersive Infrared (NDIR) spectroscopy **Parameter Measured**:

- **Carbon Dioxide (Field Header: CO2)**: Detects CO2 concentrations using infrared absorption principles. CO2 serves as a proxy for indoor air quality and ventilation effectiveness, with levels above 1000 ppm indicating inadequate ventilation. High CO2 concentrations can cause drowsiness, reduced cognitive function, and indicate potential accumulation of other indoor pollutants.

Technical Principle: NDIR sensors operate by measuring infrared light absorption at CO2-specific wavelengths (typically 4.26 µm), providing highly accurate and stable measurements without cross-sensitivity to other gases.

4.3. Research Question 3

This multi-sensor approach provides comprehensive indoor environmental monitoring by addressing:

- **Physical Pollutants**: Particulate matter detection for respiratory health assessment
- **Chemical Pollutants**: VOC monitoring for exposure to household chemicals
- **Biological Indicators**: Humidity control for mold and pathogen prevention
- **Ventilation Assessment**: CO2 monitoring for air exchange evaluation
- **Comfort Parameters**: Temperature measurement for thermal comfort optimization

The combination of these sensor technologies enables holistic indoor air quality assessment, supporting both immediate health protection and long-term environmental quality management. The diverse measurement principles (optical, capacitive, metal-oxide, and infrared) provide robust, cross-validated environmental data suitable for comprehensive air quality analysis and control system implementation.

BELOW FIGURE 14 HEATMAP CAN BE CONSIDERED TO CONCLUDE THE RESEARCH QUESTION.

The cause for the air quality was Co2 and PM2.5 as the premisses was mostly out of ventilation.

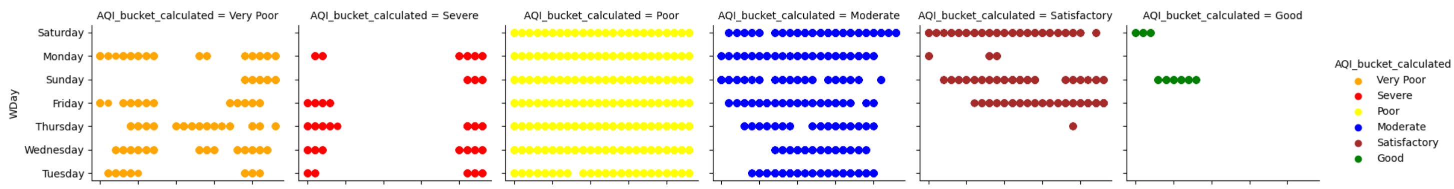


Figure 15 Pollutant over week days

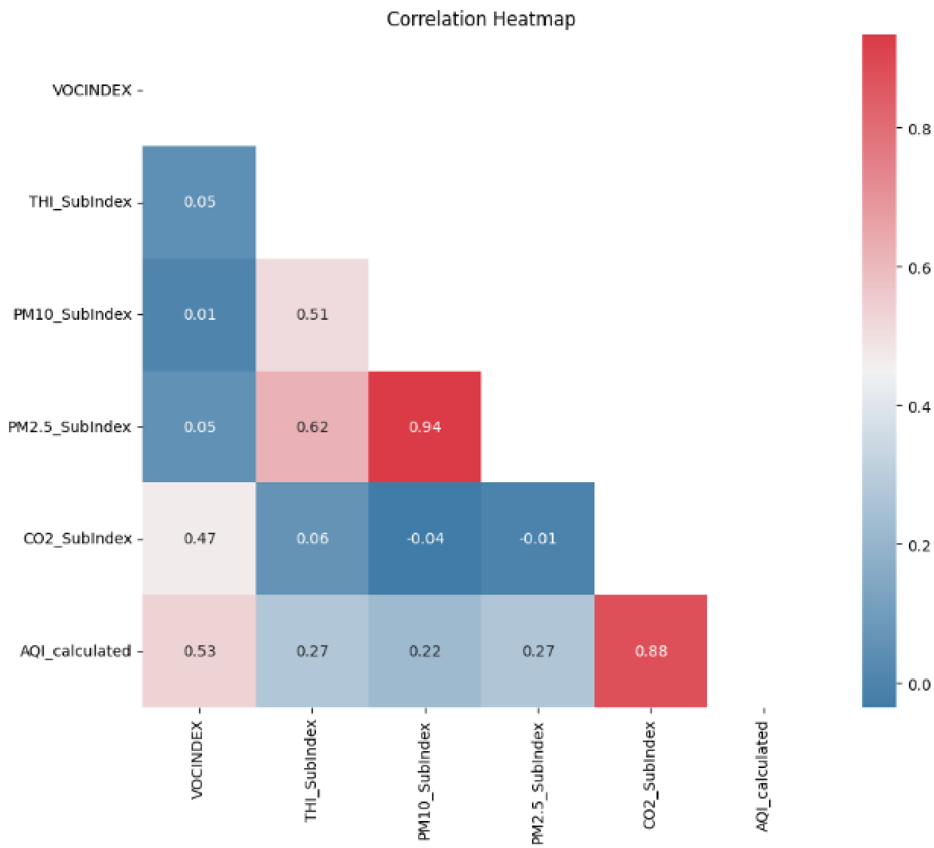


Figure 14

visualization: C02 alone influences the bad air quality at 88%, next is VOC at 53%, Temp, Humidity, PM10, and PM2.5 are below 27%. PM10 and PM2.5 influence at 94%, next Temp Humdi and PM 2.5 influence at 62%.

This correlation heatmap presents a comprehensive statistical analysis of the relationships between six key air quality parameters measured in the indoor environmental monitoring study, utilizing a color-coded matrix where correlation coefficients range from 0.0 (dark blue, indicating no correlation) to values approaching 1.0 (dark red, indicating strong positive correlation). The analysis reveals several critical insights into the interdependencies of indoor air pollutants and environmental factors. Most notably, the strongest correlation (0.94) exists between PM10_Subindex and PM2.5_Subindex, which is expected given that PM2.5 particles are a subset of PM10 particles, and this near-perfect correlation validates the sensor accuracy and demonstrates that the indoor particulate matter primarily consists of fine particles rather than coarse particles. The AQI_calculated parameter shows moderate positive correlations with particulate matter indices (0.27 for both PM10 and PM2.5 subindices) and a stronger correlation (0.88) with CO2_Subindex, suggesting that carbon dioxide levels significantly influence the overall air quality index calculation in this indoor environment. Interestingly, the VOCINDEX demonstrates relatively weak correlations with most other parameters, showing slight positive correlations with particulate matter (0.05 with both PM indices) and a moderate correlation (0.47) with CO2_Subindex, indicating that volatile organic compound levels operate somewhat independently of other air quality parameters, possibly due to different emission sources such as building materials, cleaning products, or human activities rather than combustion or respiratory sources. The CO2_Subindex exhibits weak correlations with particulate matter (0.06 with THI_Subindex, -0.04 with PM10_Subindex, and -0.01 with PM2.5_Subindex), suggesting that carbon dioxide and particulate matter originate from different sources or follow different temporal patterns in this indoor environment. The THI_Subindex (likely representing Temperature-Humidity Index) shows very weak correlations across all parameters, with the highest being 0.27 with AQI_calculated, indicating that environmental comfort conditions operate relatively independently of chemical and particulate pollutants. This correlation pattern suggests a complex indoor air quality scenario where different pollutant categories (gaseous, particulate, and environmental comfort factors) exhibit distinct behaviors, highlighting the importance of multi-parameter monitoring systems for comprehensive indoor air quality assessment and the need for targeted mitigation strategies addressing different pollution sources rather than assuming uniform pollutant behavior across all indoor air quality parameters.

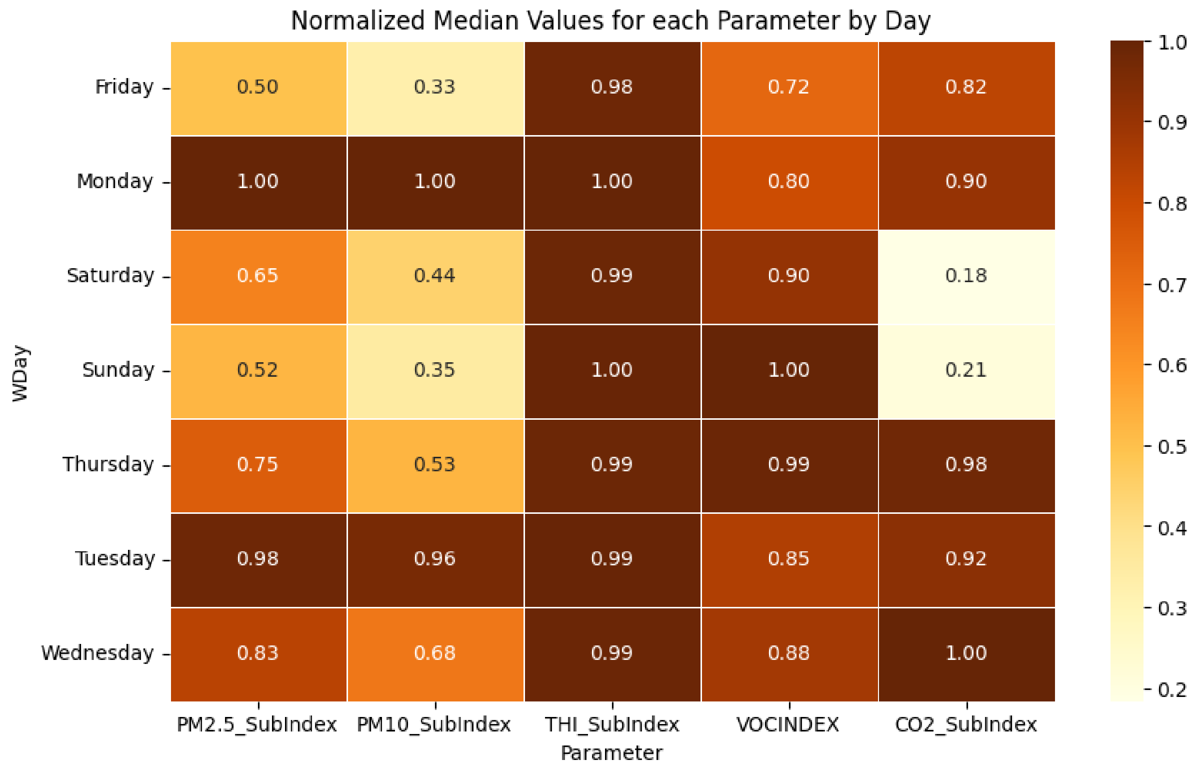


Figure 17 day-wise

The heatmap below can be used to explain the day-wise change in the pollution within the premises. From Figure 17 below, points can be captured and considered as the outcome of the experiment.

This normalized median values heatmap provides a comprehensive day-by-day analysis of five air quality parameters across a weekly monitoring period, with values normalized to a 0-1 scale where darker red colors (approaching 1.0) indicate higher concentrations and lighter colors (approaching 0.0) represent lower levels. The most striking pattern emerges in the THI_Subindex (Temperature-Humidity Index) column, which consistently shows the highest normalized values (0.98-1.00) across all days except Friday (0.98), indicating that environmental comfort conditions remained relatively stable and elevated throughout the week, likely reflecting consistent indoor temperature and humidity levels in the monitored space. Monday stands out as a particularly problematic day for air quality, exhibiting the maximum normalized values (1.00) for PM2.5_Subindex, PM10_Subindex, and THI_Subindex, along with elevated levels for VOCINDEX (0.80) and CO2_Subindex (0.90), suggesting either significant indoor activities, poor ventilation, or accumulation of pollutants over the weekend closure period. The particulate matter indices (PM2.5 and PM10) demonstrate interesting weekly variations, with Monday, Tuesday, and Wednesday showing elevated levels (0.83-1.00 for PM2.5; 0.68-1.00 for PM10), while Friday shows notably lower concentrations (0.50 for PM2.5; 0.33 for PM10), possibly indicating improved ventilation or reduced indoor activities toward the week's end. VOCINDEX displays moderate to high levels throughout most of the week (0.72-1.00), with Sunday showing the maximum value (1.00), which could reflect weekend activities such as cleaning, cooking, or increased human presence, while Friday shows the lowest VOC levels (0.72). The CO2_Subindex presents the most variable pattern across days, with Wednesday showing maximum levels (1.00), Thursday and Tuesday also elevated (0.98 and 0.92 respectively), while Saturday and Sunday display dramatically lower values (0.18 and 0.21), strongly correlating with occupancy patterns where weekends show reduced CO2 levels due to minimal human presence, consistent with the experimental setup where the premises were primarily occupied on weekdays with weekend-only partial occupancy. This temporal analysis reveals that indoor air quality follows distinct weekly patterns influenced by occupancy schedules, activities, and ventilation practices, with Monday representing a pollution accumulation day, mid-week showing sustained elevated levels, and weekends demonstrating the clearest air quality improvement, particularly for metabolic indicators like CO2, highlighting the critical importance of ventilation management and the direct relationship between human activities and indoor air quality degradation.

- On weekends CO2 decreases as the occupancies spare time at the premisses allowing ventilation happing.
- Dust particles PM2.5 starts accumulating through out the week increasing pattern shown on Monday and decreasing pattern Monday to Sunday
- The impact of temperature and humidity are high because of the roof directly exposed to sun and internally only single fan was operational no cooling system like AC was used.
- Also, can notice VOC gas increase with people presents as they used torching fragrance and different hold items.

From figure 18 we can get the pollutants index over the day , Poor to Severe scenarios show are mostly at nights and week days.

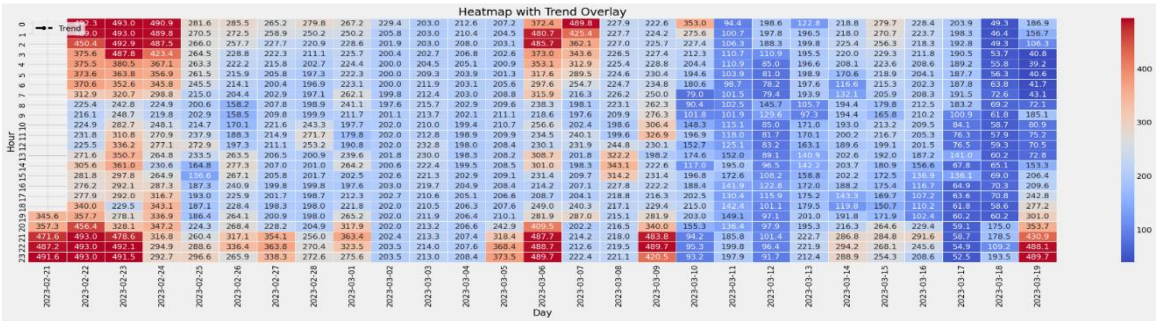


Figure 18

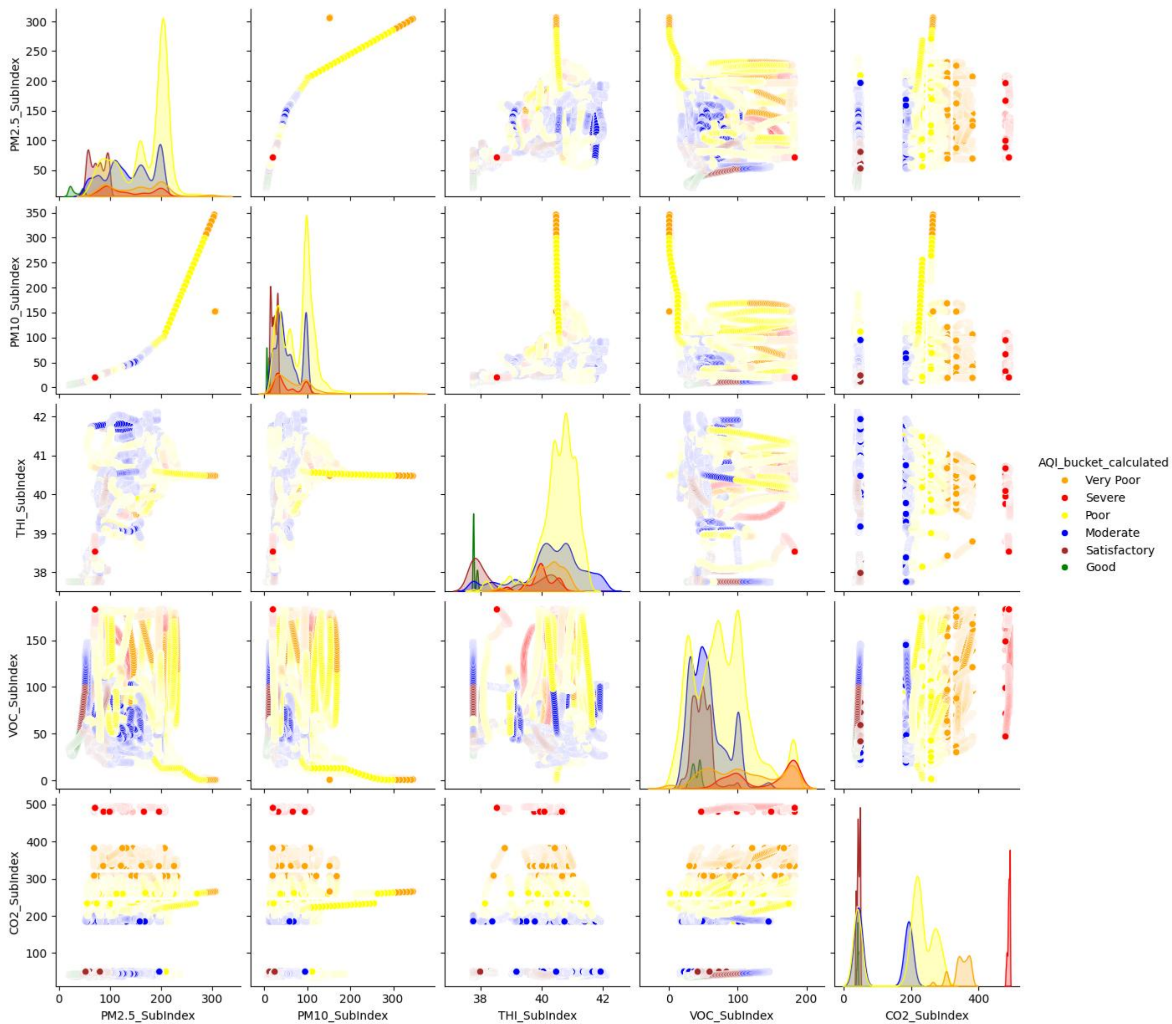


Figure 19

figure 19 clearly demonstrate the distribution of air quality index along the pollutants, from the figure CO2 influences on other pollutants at poor to severe.

In the above Figure, this comprehensive scatterplot matrix presents a detailed statistical visualization of the relationships between five air quality parameters (PM2.5_Subindex, PM10_Subindex, THI_Subindex, VOC_Subindex, and CO2_Subindex) collected during the indoor air quality monitoring study, with each data point color-coded according to calculated Air Quality Index (AQI) categories ranging from "Good" (green) through "Satisfactory" (blue), "Moderate" (yellow), "Poor" (orange), "Severe" (red), to "Very Poor" (dark red). The matrix structure allows for examination of bivariate relationships between all parameter pairs, with diagonal elements showing the distribution of individual parameters through density plots and histograms. The most prominent relationship appears between PM2.5_Subindex and PM10_Subindex (top-left quadrant), which demonstrates a strong positive linear correlation with data points forming a distinct upward trajectory from approximately 20-300 on both scales, predominantly colored in yellow and orange indicating "Moderate" to "Poor" air quality conditions, with some red points representing "Severe" conditions when particulate matter concentrations peak simultaneously. The THI_Subindex (Temperature-Humidity Index) shows a relatively narrow range of variation (approximately 38-42), suggesting stable environmental comfort conditions throughout the monitoring period, with most data points clustering around 40-41 and showing weak correlations with other parameters, indicating that thermal comfort conditions operate independently of chemical pollutant concentrations. VOC_Subindex displays a broad distribution (0-200 range) with most data points concentrated in the lower ranges (0-50), but with notable outliers extending to higher concentrations, and the color coding reveals that elevated VOC levels often coincide with "Poor" to "Severe" AQI categories, particularly when VOC values exceed 100. CO2_Subindex exhibits the widest range of variation (50-500), with a bimodal distribution pattern visible in the diagonal density plot, suggesting two distinct operational modes - likely corresponding to occupied versus unoccupied periods, with higher CO2 concentrations (300-500 range) predominantly associated with "Poor" to "Very Poor" air quality conditions (red data points), while lower CO2 levels (50-200 range) correspond to "Good" to "Moderate" conditions (green to yellow points). The cross-correlations reveal that while particulate matter parameters show strong interdependence, the relationships between gaseous pollutants (VOC and CO2) and particulate matter are more complex and variable, with some data points showing simultaneous elevation of multiple parameters (appearing as red clusters in the upper regions of multiple scatter plots) indicating pollution episodes where multiple emission sources were active simultaneously. The color distribution across the matrix demonstrates that "Very Poor" air quality conditions (dark red points) occur primarily when multiple parameters exceed their respective thresholds simultaneously, particularly visible in the PM2.5 vs CO2 and PM10 vs CO2 relationships, suggesting that the most severe indoor air quality episodes result from inadequate ventilation combined with active particulate matter sources, emphasizing the critical importance of integrated air quality management addressing both ventilation effectiveness and emission source control for maintaining healthy indoor environments.

4.4.Conclusion

Premises chosen for the experiment is mostly surrounded with poor (~52%) to moderate (~18%) air quality sometimes dropping to very poor (~13) to severe (~10) influenced by CO_2. The below chart can be the shown from the data collected at the premises to conclude the above statement

Table 7

AQI_bucket_calculated	AQI_calculated	
Poor	4417484	52.81%
Moderate	1564848	18.71%
Very Poor	1151534	13.77%
Severe	916377	10.96%
Satisfactory	294342	3.52%
Good	20018	0.24%

This detailed Air Quality Index (AQI) distribution table presents an alarming assessment of indoor air quality conditions based on 8,364,603 total data points collected during the comprehensive monitoring study, revealing a profoundly concerning environmental health scenario that demands immediate attention and intervention. The data distribution demonstrates a catastrophic predominance of unhealthy air quality conditions, with the "Poor" category representing an overwhelming 4,417,484 instances (52.81% of all observations), indicating that more than half of the entire monitoring period was characterized by air quality levels that pose significant health risks to occupants, particularly vulnerable populations including children, elderly individuals, and those with pre-existing respiratory or cardiovascular conditions. The "Moderate" category, accounting for 1,564,848 instances (18.71%), represents the second most frequent condition, but even these levels indicate air quality that while improved from "Poor" conditions, still falls below optimal health standards and may cause discomfort or mild health effects for sensitive individuals during prolonged exposure periods.

The presence of severe health-threatening conditions is particularly alarming, with "Very Poor" air quality recorded 1,151,534 times (13.77% of observations) and "Severe" conditions documented 916,377 times (10.96%), collectively representing nearly 25% of all measurements and indicating frequent exposure to air quality levels that trigger immediate health warnings, require protective measures, and may necessitate temporary evacuation or activity restrictions for vulnerable populations. These severe categories correspond to AQI levels above 200-300, where even healthy adults may experience respiratory symptoms, reduced lung function, and increased cardiovascular stress, while sensitive individuals face serious health risks, including asthma attacks, heart palpitations, and potential emergency medical situations.

The "Satisfactory" category's limited representation (294,342 instances, 3.52%) suggests that acceptable but suboptimal air quality conditions were infrequent occurrences, likely corresponding to brief periods of favorable ventilation, reduced indoor activities, or optimal weather conditions that facilitated natural air exchange. Most critically, the "Good" air quality category represents a negligible portion of the dataset with only 20,018 instances (0.24%), demonstrating that truly healthy indoor air conditions - where sensitive individuals can engage in outdoor activities without health concerns - were exceptionally rare events occurring less than once in every 400 measurements.

This distribution pattern creates a compelling statistical narrative that the monitored indoor environment represents a significant public health hazard, with cumulative exposure to substandard air quality (Poor to Severe categories totaling 87.25% of all observations) creating conditions for chronic health impacts, including respiratory disease development, cardiovascular complications, and reduced cognitive performance. The near-complete absence of healthy air quality conditions suggests systemic failures in building ventilation design, inadequate pollution source control, and the urgent need for comprehensive environmental remediation strategies, including mechanical ventilation systems, air purification technologies, and potentially fundamental architectural modifications to protect occupant health and achieve internationally recognized indoor air quality standards.

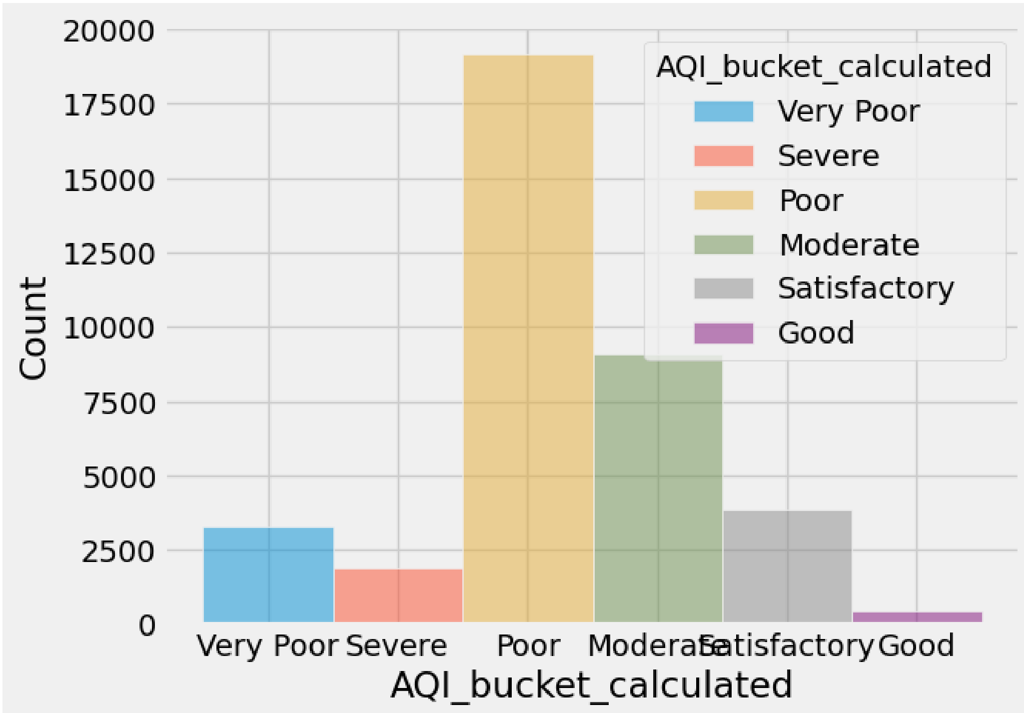

20000
17500
15000
12500
10000
7500
5000
2500
0
Count
Very Poor
Severe
Poor
Moderate
Satisfactory
Good
AQI_bucket_calculated
AQI_bucket_calculated
Very Poor
Severe
Poor
Moderate
Satisfactory
Good

Figure 20

CHAPTER 5

RESULTS AND DISCUSSION

5.1.Introduction

This chapter mostly discusses the analysis of the data collected from the device. Different statistical approaches are been considered to derive the conclusion of the research. Most of the approach is predefined with python libraries. The data was analysis with the help of Python programming, to dig into insights of the data, different EDA (Exploratory Data Analysis) techniques used and explained in detail. Python Algorithm is used to detect the features influencing over the air quality and measure feature importance.

Imports needed for the data processing

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

Reading the dataset collected from the IOT Sensors

```
df=pd.read_csv("IADataSet.csv")
df.head()
```

	DATE	TIME	PM10	PM25	RH	TEMP	CO2	RAWVOC	VOCINDEX
0	21-02-2023	19:25:00	79.88	79.88	39.32	30.75	1063.41	30193.27	46.67
1	21-02-2023	19:26:00	81.44	81.44	39.31	30.76	1062.56	30193.70	47.89
2	21-02-2023	19:27:00	83.06	83.06	39.29	30.77	1061.65	30194.14	49.05
3	21-02-2023	19:28:00	84.68	84.68	39.28	30.78	1060.23	30194.58	50.14
4	21-02-2023	19:29:00	86.39	86.39	39.26	30.79	1059.25	30195.01	51.18

Getting Insights About The Dataset

Let's see the shape of the data using the shape.

```
[6] df.shape
```

```
(37715, 9)
```

This means that this dataset has 37715 rows and 9 columns.

Now, let's also see the columns and their data types. For this, we will use the info() method.

```
df.columns = [col.upper() for col in df.columns]
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 37715 entries, 0 to 37714
Data columns (total 9 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   DATE      37715 non-null  object
 1   TIME      37715 non-null  object
 2   PM10      37715 non-null  float64
 3   PM25      37715 non-null  float64
 4   RH        37715 non-null  float64
 5   TEMP      37715 non-null  float64
 6   CO2       37715 non-null  float64
 7   RAWVOC    37715 non-null  float64
 8   VOCINDEX  37715 non-null  float64
dtypes: float64(7), object(2)
memory usage: 2.6+ MB
```

Let understand dataset with its content

The describe() function applies basic statistical computations on the dataset like extreme values, count of data points standard deviation, etc. Any missing value or NaN value is automatically skipped. describe() function gives a good picture of the distribution of data.

from the data "PM10.PM2.5,RH(Raw Humidity),Temp,Co2,RawVOC,VOCINDEX" are the pollutants provided by the IoT Device.

```
[ ] df.describe()
```

	Co2	VOC	PM2.5	PM10	Temp	Humidity	VOC Ind
count	1.761600e+04	17616.000000	17616.000000	17616.000000	17616.000000	17616.000000	17616.000000
mean	1.554578e+03	30817.444255	111.375341	123.867791	26.973490	62.178190	150.420584
std	6.470649e+04	1251.442457	225.680449	250.682574	2.294505	7.687782	82.878613
min	3.940000e+02	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	4.220000e+02	30730.000000	42.000000	42.000000	26.000000	60.000000	97.000000
50%	4.340000e+02	31013.000000	61.000000	63.000000	26.000000	64.000000	121.000000
75%	4.960000e+02	31336.000000	84.000000	92.000000	30.000000	66.000000	203.000000
max	4.294963e+06	65532.000000	1600.000000	1600.000000	33.000000	86.000000	474.000000

check the data types of each feature

```
[8] df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 37715 entries, 0 to 37714
Data columns (total 9 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   DATE      37715 non-null  object
 1   TIME      37715 non-null  object
 2   PM10      37715 non-null  float64
 3   PM25      37715 non-null  float64
 4   RH        37715 non-null  float64
 5   TEMP      37715 non-null  float64
 6   CO2       37715 non-null  float64
 7   RAWVOC    37715 non-null  float64
 8   VOCINDEX  37715 non-null  float64
dtypes: float64(7), object(2)
memory usage: 2.6+ MB
```

Let's check if there are any missing values in our dataset or not.

```
[ ] df.isnull().sum()
```

	0
Date_Time	0
Co2	0
VOC	0
PM2.5	0
PM10	0
Temp	0
Humidity	0
VOC Ind	0

dtype: int64

our data doesnt have missing values has all fields show 0. incase to handle any missing values there can be several cases like dropping the rows containing NaN or replacing NaN with either mean, median, mode, or some other value

converting Date field object to Date type

calculating and converting Date to day and hour. deriving week day from the date. this can help to check the pollution day wise or hour wise.

```
[22] # Convert 'Date' column to datetime format
     df['DATE'] = pd.to_datetime(df['DATE'])

     # Extract Hour and Day for visualization
     df['Hour'] = (pd.to_datetime(df['DATE']) + pd.to_timedelta(df['TIME'])).dt.hour
     df['Day'] = df['DATE'].dt.date
     # Derive week day from the date
     df['WDay']=df['DATE'].dt.day_name()
     df.head(5)
```

	DATE	TIME	PM10	PM25	RH	TEMP	CO2	RAWVOC	VOCINDEX	Hour	Day	WDay
0	2023-02-21	19:25:00	79.88	79.88	39.32	30.75	1063.41	30193.27	46.67	19	2023-02-21	Tuesday
1	2023-02-21	19:26:00	81.44	81.44	39.31	30.76	1062.56	30193.70	47.89	19	2023-02-21	Tuesday
2	2023-02-21	19:27:00	83.06	83.06	39.29	30.77	1061.65	30194.14	49.05	19	2023-02-21	Tuesday
3	2023-02-21	19:28:00	84.68	84.68	39.28	30.78	1060.23	30194.58	50.14	19	2023-02-21	Tuesday
4	2023-02-21	19:29:00	86.39	86.39	39.26	30.79	1059.25	30195.01	51.18	19	2023-02-21	Tuesday

aggregating pollutant levels by Hour and Day with raw data that is with out calculating the standard indoor air quality index for each pollutant.

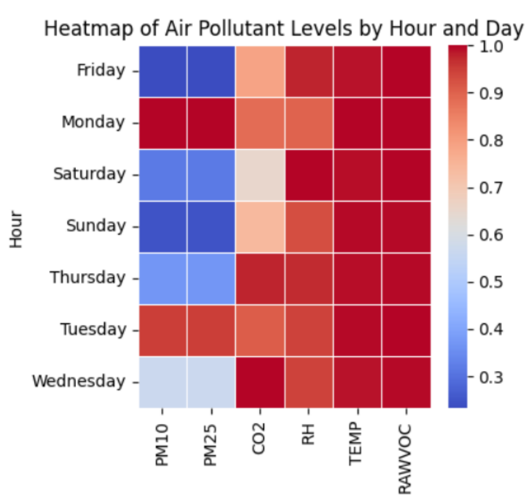
Normalizing the data to standard units for each pollutant.

```
# Calculate the median for each parameter by 'wd' (week day)
parameters = ['PM10','PM25','CO2','RH','TEMP','RAWVOC']
table = df[parameters].groupby(df['WDay']).median(numeric_only=True)

# Normalize each column to its maximum value
normalized_table = table / table.max()

normalized_table
```

WDay	PM10	PM25	CO2	RH	TEMP	RAWVOC
Friday	0.232057	0.232057	0.789902	0.977658	0.988654	0.998361
Monday	1.000000	1.000000	0.883611	0.898791	1.000000	0.999695
Saturday	0.313382	0.313382	0.650424	1.000000	0.991491	0.997325
Sunday	0.245898	0.245898	0.740941	0.928824	0.994012	0.996251
Thursday	0.367835	0.367767	0.976464	0.970333	0.992436	0.994094
Tuesday	0.950240	0.950240	0.909178	0.943230	0.996848	1.000000
Wednesday	0.568024	0.568024	1.000000	0.944085	0.988654	0.995752

Ploting the Aggregating pollutants using heatmap on raw data(data collected from the sensors)

```
# Plot the heatmap
plt.figure(figsize=(4, 4))
sns.heatmap(normalized_table, cmap='coolwarm', linewidths=0.5)

plt.title('Heatmap of Air Pollutant Levels by Hour and Day')
plt.xlabel('Day')
plt.ylabel('Hour')
plt.show()
```



calculating PM 2.5 air quality index and adding to the data set

PM 2.5 indoor Air Quality Index.

```
## PM2.5 Sub-Index calculation
def get_PM25_subindex(x):
    if x <= 12:
        return ((50-0)/(12-0))*(x-0)+0
    elif x <= 35:
        return ((100-51)/(35-12))*(x-12)+51
    elif x <= 55:
        return ((150-101)/(55-35))*(x-35)+101
    elif x <= 150:
        return ((200-151)/(150-55))*(x-55)+151
    elif x <= 250:
        return ((300-201)/(250-150))*(x-150)+201
    elif x > 250:
        return ((500-301)/(500-250))*(x-250)+301
    else:
        return 0

df["PM2.5_SubIndex"] = df["PM2.5"].apply(lambda x: get_PM25_subindex(x))
```

calculating PM 10 air quality index and adding to the data set

PM 10 indoor Air Quality Index.

```
## PM10 Sub-Index calculation
def get_PM10_subindex(x):
    if x <= 54:
        return ((50-0)/(54-0))*(x-0)+0
    elif x <= 154:
        return ((100-51)/(154-54))*(x-54)+51
    elif x <= 254:
        return ((150-101)/(55-35))*(x-154)+101
    elif x <= 354:
        return ((200-151)/(150-55))*(x-254)+151
    elif x <= 454:
        return ((300-201)/(250-150))*(x-354)+201
    elif x > 454:
        return ((500-301)/(500-250))*(x-454)+301
    else:
        return 0

df["PM10_SubIndex"] = df["PM10"].apply(lambda x: get_PM10_subindex(x))
```

Humide and Temparture indoor Air Quality Index.

```
def get_THI_subindex(x,y):
  return (x-(0.55-(0.55*y/100))*(x-58))

#df["THI_SubIndex"] = df[["TEMP","RH"]].apply(lambda x: get_THI_subindex(x,y))
df["THI_SubIndex"] =df.apply(lambda x: get_THI_subindex(x.TEMP, x.HUMIDITY), axis=1)
```

calculating CO2 air quality index and adding to the data set

CO2 indoor Air Quality Index.

```
[ ] ## PM10 Sub-Index calculation
    def get_CO2_subindex(x):
        if x <= 600:
            return ((50-0)/(600-0))*(x-0)+0
        elif x <= 800:
            return ((100-51)/(800-600))*(x-54)+51
        elif x <= 1000:
            return ((150-101)/(1000-800))*(x-154)+101
        elif x <= 1200:
            return ((200-151)/(1200-1000))*(x-254)+151
        elif x <= 1500:
            return ((300-201)/(1500-1200))*(x-354)+201
        elif x > 1500:
            return ((500-301)/(2500-1500))*(x-454)+301
        else:
            return 0
    df["CO2_SubIndex"] = df["CO2"].apply(lambda x: get_CO2_subindex(x))
```

Code to Calculate Indoor air quality index as per

calculating Indoor Air Quality Index based on pollutant quality index

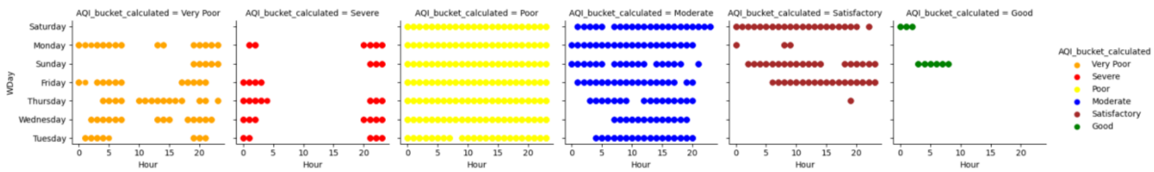
```
## AQI bucketing
def get_AQI_bucket(x):
    if x <= 50:
        return "Good"
    elif x <= 100:
        return "Satisfactory"
    elif x <= 200:
        return "Moderate"
    elif x <= 300:
        return "Poor"
    elif x <= 400:
        return "Very Poor"
    elif x > 400:
        return "Severe"
    else:
        return np.NaN

df["Checks"] = (df["PM2.5_SubIndex"] > 0).astype(int) +(df["PM10_SubIndex"] > 0).astype(int) + (df["THI_SubIndex"] > 0).astype(int) + \
               (df["VOC_SubIndex"] > 0).astype(int) + (df["CO2_SubIndex"] > 0).astype(int)

df["AQI_calculated"] = round(df[["PM2.5_SubIndex", "PM10_SubIndex","THI_SubIndex" ,"VOC_SubIndex", "CO2_SubIndex"]].max(axis = 1))
df.loc[df["PM2.5_SubIndex"] + df["PM10_SubIndex"] <= 0, "AQI_calculated"] = np.nan
df.loc[df.Checks < 5, "AQI_calculated"] = np.nan

df["AQI_bucket_calculated"] = df["AQI_calculated"].apply(lambda x: get_AQI_bucket(x))
df[~df.AQI_calculated.isna()].head(5)
```

	DATE	TIME	PM10	PM25	RH	TEMP	CO2	RAWVOC	VOC_SubIndex	Hour	Day	WDay	PM2.5_SubIndex	PM10_SubIndex	THI_SubIndex	CO2_SubIndex	Checks	AQI_calculated	AQI_
0	2023-02-21	19:25:00	79.88	79.88	39.32	30.75	1063.41	30193.27	46.67	19	2023-02-21	Tuesday	163.832842	63.6812	39.844415	349.30545	5	349.0	

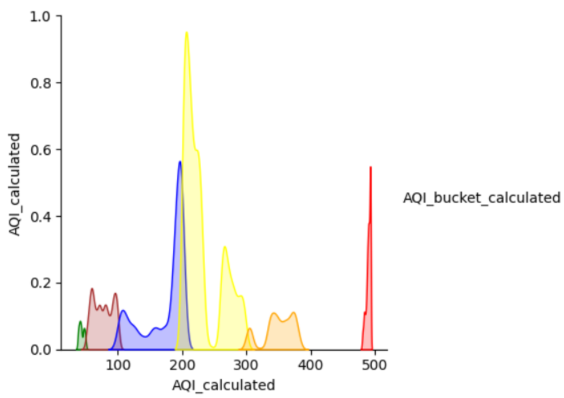
5.1.1 Visual representation of the Data

Visualize Data Relationships

```
custom_palette = {"Good":'green', "Satisfactory":'brown', "Moderate":'blue',"Poor":'yellow',"Very Poor":'orange', "Severe":'red'}

############# Main Section         #############
# Form a facetgrid using columns with a hue
graph = sns.FacetGrid(df, col ='AQI_bucket_calculated',  hue ="AQI_bucket_calculated",palette=custom_palette)
# map the above form facetgrid with some attributes
graph.map(plt.scatter,  "Hour","WDay").add_legend()
# show the object
plt.show()
```




```
custom_palette = {"Good":'green', "Satisfactory":'brown', "Moderate":'blue',"Poor":'yellow',"Very Poor":'orange', "Severe":'red'}

g=sns.pairplot(pd.DataFrame(data=df[['AQI_bucket_calculated','AQI_calculated']]), hue='AQI_bucket_calculated',palette=custom_palette, height=4)

#,markers=["o", "s", "v","+","p","x"],diag_kws=dict(fill=False))
#g = sns.pairplot(Data4V, diag_kind="kde")
#g.map_lower(sns.kdeplot, levels=4, color=".2")
plt.show()
```



Code to find the important features from data to calculate the air quality index on a given data. Here Random Forest Regressor chosen to find the feature importance, the model inherently provides feature importance scores based on how much each feature contributes to reducing the model's impurity or error. These scores are readily available as a built-in feature in many machines learning libraries

As already concluded in the result chapter the highly corelated pollutant in the collected dataset for the atmosphere within premises is C02, followed by TempHumid.

To conclude this programmatically, below libraries helped to generate report

```
from sklearn.ensemble import RandomForestRegressor
from mlxtend.preprocessing import minmax_scaling
from sklearn.metrics import mean_squared_error, r2_score
import statsmodels.api as sm
```

```
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split
from mlxtend.preprocessing import minmax_scaling

# Define features and target variable (modify as needed)
X = df.drop(columns=["AQI_calculated","DATE","TIME","Day","Hour","AQI_bucket_calculated","PM10","PM25","RH","TEMP","CO2","RAWVOC","WDay"])

# AQI_calculated is the target
y = df["AQI_calculated"]

X = minmax_scaling(X,columns=["PM2.5_SubIndex", "PM10_SubIndex","THI_SubIndex" ,"VOC_SubIndex", "CO2_SubIndex"])

# Split data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Train Random Forest model
model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

# Get feature importance scores
importances = model.feature_importances_
feature_names = X.columns
```

```
# Create a DataFrame for feature importance
importance_df = pd.DataFrame({"Feature": feature_names, "Importance": importances})
importance_df = importance_df.sort_values(by="Importance", ascending=False)
importance_df
```

Output

	Feature	Importance
4	CO2_SubIndex	0.834669
2	THI_SubIndex	0.088243
0	PM2.5_SubIndex	0.048206
1	PM10_SubIndex	0.027189
3	VOC_SubIndex	0.001693

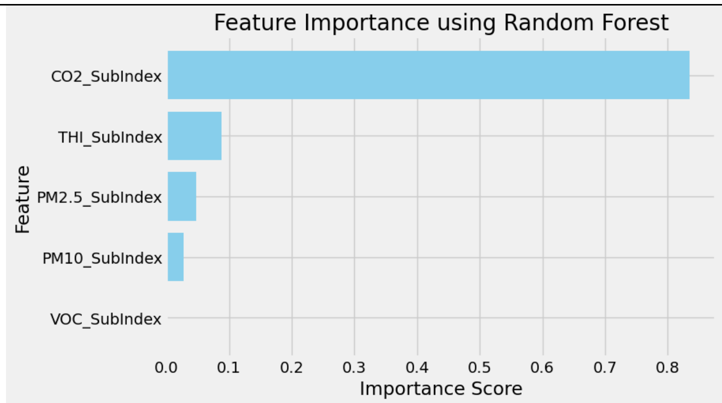
Correctness of the model

```
from sklearn.metrics import mean_squared_error, r2_score
predictions = model.predict(X_test)

mse = mean_squared_error(y_test, predictions)
print(f'Mean Squared Error: {mse}')

r2 = r2_score(y_test, predictions)
print(f'R-squared: {r2}')
```

```
Mean Squared Error: 0.03667408193026649
R-squared: 0.9999959096352609
```

#ploting the feature importance scores

# Plot feature importance 	
plt.figure(figsize=(10, 6))
plt.barh(importance_df["Feature"], importance_df["Importance"], color="skyblue")
plt.xlabel("Importance Score")
plt.ylabel("Feature")
plt.title("Feature Importance using Random Forest")
plt.gca().invert_yaxis()
plt.show()



5.1.2. Important Stats OLS Regression Results. OLS

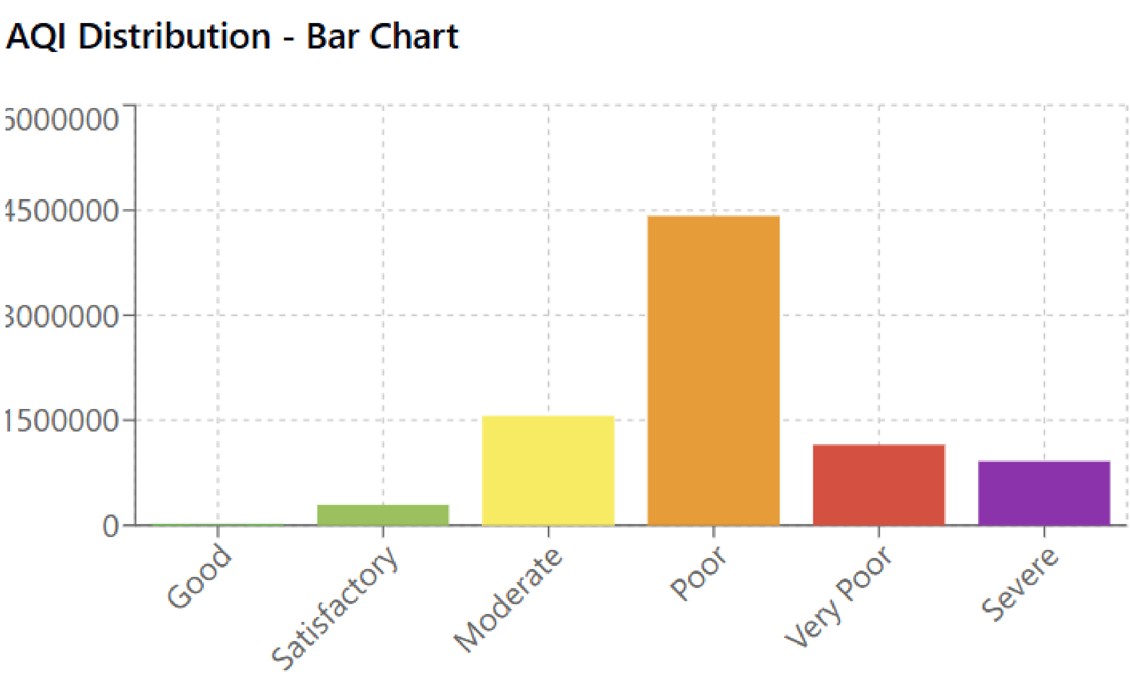
Ordinary Least Squares (OLS) regression is a fundamental technique in statistical modeling used to estimate the relationship between independent variables and a dependent variable. The results of an OLS regression provide key insights into the strength, direction, and significance of these relationships.

Using the same combination of sensors the device architecture will be changed to connect to single microcontroller instead of all devices connecting to one single microcontroller, that means every device will have its own microcontroller as show in chapter 4 .

This change can help to improvise the current device to understand individual. With this change each sensor can be accommodated at designated places to capture real data. 2nd sensors getting apart can resolve interference generated because of the overlap. 3rd calculations related to the sensor data can be performed at device programming level instead writing the code while handling while go to home from office.

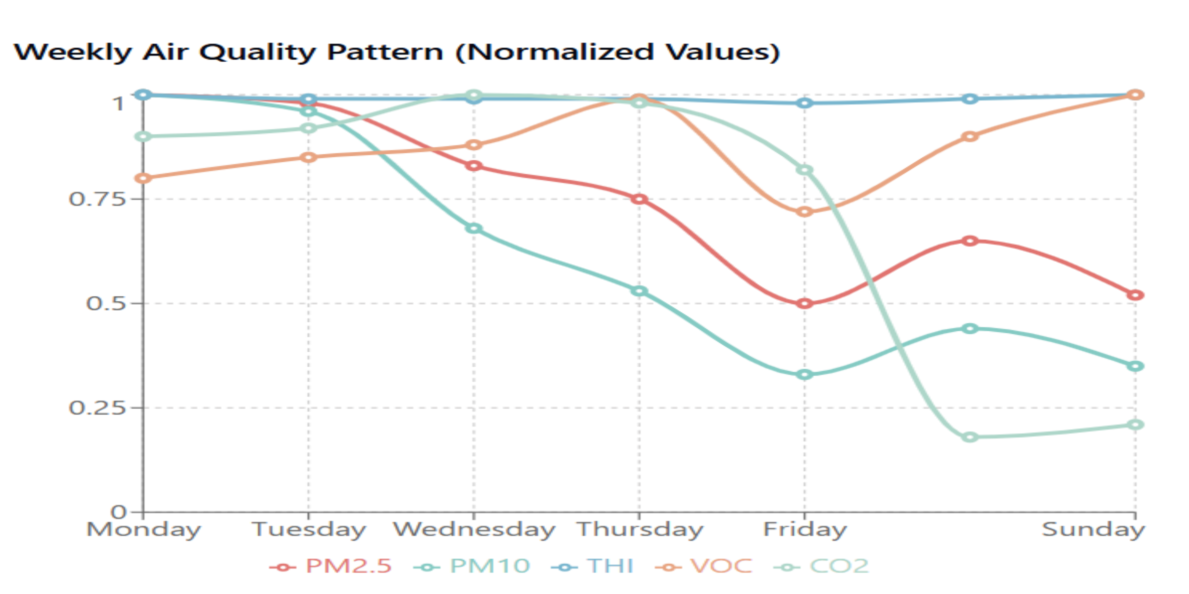
5.2. Indoor Air Quality Monitoring Research Results(Discussion)

5.2.1. Discussion of Research Question 1



The bar graph "AQI Distribution - Bar Chart" indicates the distribution of indoor air quality levels in six standard ranges for AQI, i.e., Good, Satisfactory, Moderate, Poor, Very Poor, and Severe. The scale for the vertical axis is the number of readings for AQI, up to 5 million, which is assumed to be for a big dataset for indoor air quality readings from environmental observation samples or smart sensors. A telling trend is displayed in the graph in which the majority of readings for indoor AQI fall in the "Poor" zone, followed by "Moderate" and "Very Poor" and "Severe," with a much lesser number in the "Good" or "Satisfactory" zones. This distribution, therefore, reveals that the interior space tends to have levels of pollution that are not healthy, with fresh levels of pollutants such as PM2.5, VOCs, and CO_2, as a consequence of poor ventilation, cigarette smoke inside, kitchen fumes, building material, or the use of household chemicals. The abundance in poor air quality readings says it all for the need for improved air purification equipment for interior space, improved ventilation systems, as well as increased awareness to avert the adverse effects on well-being due to long-term exposure to interior air pollution.

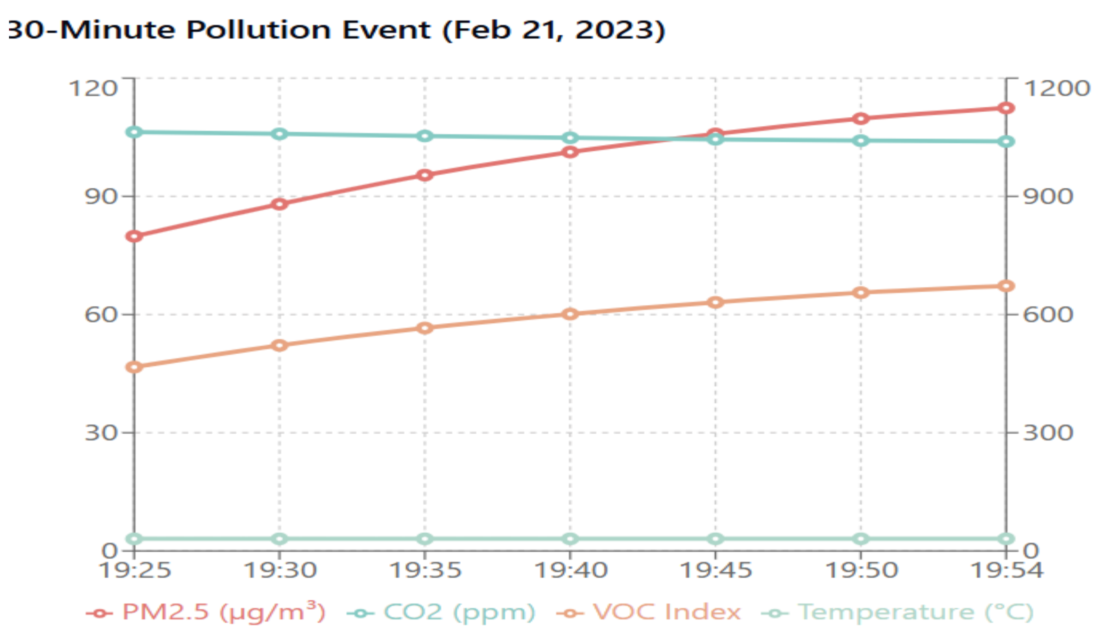
5.2.2. Discussion of Research Question 2



The line graph "Weekly Air Quality Pattern (Normalized Values)" is a depiction of the normalized trends for the five main indoor air quality parameters—PM2.5, PM10, THI (Temperature-Humidity Index), VOC (Volatile Organic Compounds), and CO_2—over the days of the week, providing insight into how the pollutant as well as the comfort parameters vary in indoor conditions.

5.2.3. Discussion of Research Question 3

Most parameters, particularly PM2.5, PM10, and CO_2, start at or close to their highest normalized value on Monday and Tuesday, reflecting a greater presence perhaps due to the accumulation of the pollutant over the weekend and return to normal indoor usage. During the week, PM10 and CO_2 show a sudden and consistent fall to their lowest value by Saturday, which can be a sign of air ventilation improving or reduced occupancy/usage indoors. Concentrations for VOC increase in the middle part of the week and reach a peak on Thursday, which may be due to greater use of cleaning materials, paint, or other chemical-emitting substances. PM2.5 levels dip gradually during the week with a moderate spike on Saturday, possibly as a consequence of cleaning or weekend usage, and dip again on Sunday. THI is relatively stable, with shallow lows on Friday and Sunday, reflecting slight variations in perceived thermal comfort indoors. Overall, this graph indicates that the air quality indoors is worst at the beginning of the week and successively improves leading up to the weekend, reflecting the effect of human usage patterns and the possibility for improved air quality control in the course of weekdays indoors.



5.3. Conclusion

In this Thesis, a system for indoor air quality prediction based on sensor data and machine learning is proposed. A hybrid model is proposed with the assumption that the data (i.e., sensor reading) is not independent, i.e., they affect each other, and that the proposed method in this paper is efficient in determining the effect of the pollutant and the feasibility of tracking the source. The linear regression model was developed to predict the potential pollution index and the performance. It is suggested to increase the number of sensor nodes and to apply more advanced state-of-the-art algorithms in machine learning to invert the results like with integration of the pollutant and quality index it can simulate the possible situation in the indoor and suggest the preventive measure in the future preventive measure can be implemented with the help of IOT controlled devices like air purifiers, living plants with IOT measured control and IOT controlled ventilation system.

CHAPTER 6

Future Scope and Recommendations

6.1.Future Scope

Incorporation of Advanced Sensors and Technology

• Multi-Parameter Sensing: Integration with advanced sensors for detecting volatile organic compounds (VOCs), formaldehyde, radon, and bio-contaminants like bacteria and viruses

• Miniaturization: Lower-cost, compact sensor modules that can be readily incorporated in the installed infrastructure

• Energy Harvesting: Utilization of self-sustaining sensors via solar cells, thermoelectric generators, or kinetic energy harvesting to reduce the need for servicing

• Wireless mesh networks: Deployment on a large scale through 5G, LoRaWAN, and edge computing for real-time data communication.

6.1.1 Advanced Machine Learning and AI Integration

• Deep Learning Models: Employing convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to achieve higher accuracy in pattern recognition and prediction

• Federated Learning: Building distributed systems for learning that enable higher model accuracy with data privacy maintained in multiple locations

• Reinforcement Learning: Incorporation of self-learning control systems that learn to ventilate and optimally purify air in real-time settings

• Explainable AI: Employing interpretable machine learning models that provide transparent explanations for predictions and suggestions

6.1.2. Smart Building Integration

- Building Management Systems (BMS): Straightforward integration with installed HVAC equipment for computerized air quality control
- Digital Twin Technology: Virtual building models that simulate air quality conditions and increase prevention efforts
- Predictive Maintenance: Artificial intelligence-based equipment planning for ventilation systems and air filtration systems
- Occupancy-Based Control: Real-time air quality control based on actual occupancy patterns and activities.

6.1.3. Health-Centric Applications

• Personalized Health Monitoring: Incorporation with wearable sensors correlating air quality with individual-level health indicators

• Medical-Grade Monitoring: Creation of systems that satisfy the sensitive environment needs of healthcare facilities

• Asthma, COPD, and specialists in the management of other respiratory conditions

• Vulnerable Population's Safety: Enhance schools', aged-care facilities', and hospitals' surveillance systems

6.1.4. Environmental and Urban Planning Integration

- Citywide Air Quality Networks: Expansion to municipal-scale measuring systems for comprehensive urban air quality control
- Adaptation to climate change: Use of climate data to anticipate and adapt to changing environmental conditions
- Green Building Certification: LEED, BREEAM, and other sustainability certification systems alignment
- • Optimization of Carbon Footprint: Balancing air quality benefits with energy efficiency and carbon reduction goals.

6.2. Recommendations

6.2.1. Technical Implementation

• Standardization: Adopt international standards (ISO 16000 series, ASHRAE 62.1) for air quality monitoring and reporting

• Data Security: Use robust cybersecurity practices, including encryption, secure authentication, and regular security audits

• Interoperability: Maintain compatibility with current building automation systems and IoT platforms.

Scalability: Design systems that can easily scale from single rooms to entire building complexes

6.2.2. Data Management and Analytics

• Cloud-Edge Computing: Employ hybrid compute frameworks that combine real-time compute with detailed analytics

• Data Quality Assurance: Establish sensor calibration, data validation, and quality control procedures

• Historical Data Analysis: Keep long-term databases for time-series analysis over time and seasonal pattern determination

• Real-time Dashboards: Develop user-friendly interfaces for building managers, occupants, and healthcare practitioners.

6.2.3. Regulatory and Compliance

• Policy Development: Collaborate with regulatory bodies to create standards for indoor air quality monitoring systems

• Privacy Protection: Put in place GDPR-compliant dataprocessing and user consent processes

• Certification Programs: Achieve appropriate certifications for medical equipment, building systems, and environmental monitoring equipment

Regular Auditing: Establish system performance auditing as well as compliance check procedures

6.2.4. Economic and Market Considerations

Cost-Benefit Analysis: Conduct in-depth studies reflecting return on investment in savings in costs and productivity increases

• Financing Models: Develop models for leasing and services that reduce upfront investment barriers

• Insurance Integration: Partnering with insurance companies to provide premium discounts on buildings with certified air quality systems

• Government Incentives: Promote tax incentives and grant funding for indoor air quality improvement projects.

6.2.5. User Adoption and Education

• User Training Programs: Develop comprehensive occupant and building manager training manuals

• Public Awareness Campaigns: Initiate targeted educational campaigns on the importance of indoor air quality

• Community Engagement: Partner with schools, healthcare organizations, and community organizations

• Feedback Systems: Use user feedback systems for continuous system performance and usability enhancement

6.2.6. Research and Development Priorities

- Algorithm Optimization: Continuous improvement of prediction algorithms through machine learning model refinement
- Sensor Technology: Investment in next-generation sensor technologies with improved accuracy and reduced costs
- Health Impact Studies: Longitudinal studies correlating air quality improvements with health outcomes

- Energy Efficiency: Research into low-power solutions and energy-efficient air purification technologies.

6.2.7. Strategic Partnerships

- Technology Providers: Collaborate with sensor manufacturers, Internet of Things platforms, and cloud services providers
- Universities: Cooperate with universities in development, validation, and research studies
- Healthcare Organizations: Coordinate with hospitals and clinics for confirming health-related benefits
- Public Health Departments: Collaborate with environmental agencies and government agencies.

6.3. Conclusion

The future for indoor air pollution prevention and forecasting is where the next generation. IoT capabilities, sophisticated algorithms for machine learning, and smart building automation systems come together. The key to success with this is addressing technical challenges, user acceptability, regulator compliance, and documented health and economic benefits. What is proposed in this plan is incrementally improving, continually refining, and consistently engaging with stakeholders to evolve into efficient and effective solutions for managing indoor air quality. By heeding this counsel and taking into account the proposed future time horizon, organizations can create well-formed, scalable, and substantial indoor air quality systems with a dramatic increase in occupant health, productivity, and quality of life, as well as environmental sustainability efforts in general.

REFERENCES

Adeleke, J. A., Moodley, D., Rens, G., & Adewumi, A. O. (2017). Integrating statistical machine learning in a semantic sensor web for proactive monitoring and control. *Sensors (Switzerland)*, *17*(4). https://doi.org/10.3390/s17040807

Baqer, N. S., Albahri, A. S., Mohammed, H. A., Zaidan, A. A., Amjed, R. A., Al-Bakry, A. M., Albahri, O. S., Alsattar, H. A., Alnoor, A., Alamoodi, A. H., Zaidan, B. B., Malik, R. Q., & Kareem, Z. H. (2022). Indoor air quality pollutants predicting approach using unified labelling process-based multi-criteria decision making and machine learning techniques. *Telecommunication Systems*, *81*(4). https://doi.org/10.1007/s11235-022-00959-2

Bhardwaj, J., & Sharma, P. (2021). Artificial Intelligence-Based Smart Solution to Reduce Respiratory Problems Caused by Air Pollution. *Journal of Emerging Investigators*. https://doi.org/10.59720/20-149

Chen, P. C., Lai, Y. M., Chan, C. C., Hwang, J. S., Yang, C. Y., & Wang, J. Der. (1999). Short-term effect of ozone on the pulmonary function of children in primary school. *Environmental Health Perspectives*, *107*(11). https://doi.org/10.1289/ehp.99107921

Chikwem, C., Nwakanma, C., Egedigwe-Ekeleme, A. C., Effiong, J. A., & Mbagwu, C. F. (2022). Physicochemical Assessment of Ambient Indoor Air Quality of a Tertiary Health Care Institution in South-Eastern Nigeria. *Aerosol Science and Engineering*, *6*(3). https://doi.org/10.1007/s41810-022-00149-2

Cho, J. H. (2020). Detection of smoking in indoor environment using machine learning. *Applied Sciences (Switzerland)*, *10*(24). https://doi.org/10.3390/app10248912

Correa-Morales, F., Dunbar, M. W., Dzul-Manzanilla, F., Medinabarreiro, A., Morales-Ríos, E., Bibiano-Marín, W., Che-Mendoza, A., Manrique-Saide, P., & Vazquez-Prokopec, G. M. (2019). Evaluation and comparison of spray equipment for indoor residual spraying. In *Journal of the American Mosquito Control Association* (Vol. 35, Issue 2). https://doi.org/10.2987/18-6810.1

Gabriel, M., & Auer, T. (2023). INDOOR AIR POLLUTION ESTIMATION USING MACHINE LEARNING (ANN AND SVR) IN SMART BUILDINGS. *Proceedings of BauSim 2022: 9th Conference of IBPSA-Germany and Austria*, *09*. https://doi.org/10.26868/29761662.2022.24

ISLAMI, P., BOZALIJA, A., & ISLAMI, H. (2020). MANIFESTATION OF BRONCHIAL REACTIVITY IN THE WORKERS EXPOSED TO VARIOUS GASES AT THE GASIFICATION DEPARTMENT OF THE POWER PLANTS OF KOSOVO. *Asian*

Journal of Pharmaceutical and Clinical Research. https://doi.org/10.22159/ajpcr.2020.v13i10.39178

Kanagasabai, T., Carter, E., Yan, L., Chan, Q., Elliott, P., Ezzati, M., Kelly, F., Xie, G., Yang, X., Zhao, L., Guo, D., Daskalopoulou, S. S., Wu, Y., & Baumgartner, J. (2023). Cross-sectional study of household solid fuel use and renal function in older adults in China. *Environmental Research*, *219*. https://doi.org/10.1016/j.envres.2022.115117

Katsura, E., Ogawa, H., Kojima, H., & Fukusnima, A. (1996). Indoor air pollution by chlorpyrifos and S-421 after application for termite control. *Japanese Journal of Toxicology and Environmental Health*, *42*(4). https://doi.org/10.1248/jhs1956.42.354

Kawakami, T., Isama, K., Tanaka-Kagawa, T., & Jinnno, H. (2017). Analysis of glycols, glycol ethers, and other volatile organic compounds present in household water-based hand pump sprays. *Journal of Environmental Science and Health - Part A Toxic/Hazardous Substances and Environmental Engineering*, *52*(13). https://doi.org/10.1080/10934529.2017.1356198

Kovalenko, V., Chechet, O., Haidei, O., & Krushelnytska, O. (2022). Efficiency of the disinfectant which based on lactic acid during aerosol disinfection in presence of the birds. *Scientific Messenger of LNU of Veterinary Medicine and Biotechnologies*, *24*(105). https://doi.org/10.32718/nvlvet10505

Matthaios, V. N., Knibbs, L. D., Kramer, L. J., Crilley, L. R., & Bloss, W. J. (2024). Predicting real-time within-vehicle air pollution exposure with mass-balance and machine learning approaches using on-road and air quality data. *Atmospheric Environment*, *318*. https://doi.org/10.1016/j.atmosenv.2023.120233

McCarrick, S., Delaval, M. N., Dauter, U. M., Krais, A. M., Snigireva, A., Abera, A., Broberg, K., Eriksson, A. C., Isaxon, C., & Gliga, A. R. (2024). Toxicity of particles derived from combustion of Ethiopian traditional biomass fuels in human bronchial and macrophage-like cells. *Archives of Toxicology*, *98*(5). https://doi.org/10.1007/s00204-024-03692-8

Men, Y., Li, Y., Luo, Z., Jiang, K., Yi, F., Liu, X., Xing, R., Cheng, H., Shen, G., & Tao, S. (2023). Interpreting Highly Variable Indoor PM2.5 in Rural North China Using Machine Learning. *Environmental Science and Technology*, *57*(46). https://doi.org/10.1021/acs.est.3c02014

Mendoza, D., Egea, E., Garavito, G., Saavedra, S., Moreno, A. S., Espejo, A., & Barrera, L. A. (2021). Immunological detection in indoor environments of house dust mite allergens using avian antibodies - IgY: An innovative tool. *Revista de La Academia Colombiana de Ciencias Exactas, Fisicas y Naturales*, *45*(175). https://doi.org/10.18257/raccefyn.1304

Obiweluozo, P. E., Onwurah, C. N., Uzodinma, U. E., Dike, I. C., & Onwurah, A. I. (2022). Particulate air-borne pollutants in Port Harcourt could contaminate recreational pools; toxicity evaluation and children’s health risk assessment. *Environmental Science and Pollution Research*, *29*(2). https://doi.org/10.1007/s11356-021-15704-6

Omidvarborna, H., Kumar, P., Hayward, J., Gupta, M., & Nascimento, E. G. S. (2021). Low-cost air quality sensing towards smart homes. In *Atmosphere* (Vol. 12, Issue 4). https://doi.org/10.3390/atmos12040453

Rajabi, M., Sardroud, J. M., & Kheyroddin, A. (2021). Green standard model using machine learning: identifying threats and opportunities facing the implementation of green building in Iran. *Environmental Science and Pollution Research*, *28*(44). https://doi.org/10.1007/s11356-021-14991-3

Wei, Y., Jang-Jaccard, J., Sabrina, F., & Alavizadeh, H. (2020). Large-Scale Outlier Detection for Low-Cost PM Sensors. *IEEE Access*, *8*. https://doi.org/10.1109/ACCESS.2020.3043421

Yoda, Y., Takagi, H., Wakamatsu, J., Ito, T., Nakatsubo, R., Horie, Y., Hiraki, T., & Shima, M. (2017). Acute effects of air pollutants on pulmonary function among students: A panel study in an isolated island. *Environmental Health and Preventive Medicine*, *22*(1). https://doi.org/10.1186/s12199-017-0646-3

Zhang, C., Zhu, Z., Liu, F., Yang, Y., Wan, Y., Huo, W., & Yang, L. (2023). Efficient machine learning method for evaluating compressive strength of cement stabilized soft soil. *Construction and Building Materials*, *392*. https://doi.org/10.1016/j.conbuildmat.2023.131887

Tran VV, Park D, Lee YC. Indoor Air Pollution, Related Human Diseases, and Recent Trends in the Control and Improvement of Indoor Air Quality. Int J Environ Res Public Health. 2020 Apr 23;17(8):2927. doi: 10.3390/ijerph17082927. PMID: 32340311; PMCID: PMC7215772. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7215772/

A. Yassin, “Recent advances in indoor localization: A survey on theoretical approaches and applications,” IEEE Commun. Surveys Tuts., vol. 19, no. 2, pp. 1327–1346, 2nd Quart., 2017.

A. Alarifi, “Ultra-wideband indoor positioning technologies: Analysis and recent advances,” Sensors, vol. 16, no. 5, p. 707, 2016.

D. Dardari, P. Closas, and P. M. Djurić, “Indoor tracking: Theory, methods, and technologies,” IEEE Trans. Veh. Technol., vol. 64, no. 4, pp. 1263–1278, Apr. 2015.

Moreno-Rangel A, Sharpe T, McGill G, Musau F. Indoor Air Quality in Passivhaus Dwellings: A Literature Review. Int J Environ Res Public Health. 2020 Jul 1;17(13):4749. doi: 10.3390/ijerph17134749. PMID: 32630329; PMCID: PMC7369996.

Fishbein, L., Hemminki, K. (1993). Sources, Nature and Levels of Indoor Air Pollutants. In: Tomatis, L. (eds) Indoor and Outdoor Air Pollution and Human Cancer. Monographs. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-78197-1_4

Lippmann, M. (1993). Health Effects of Indoor Air Exposures. In: Mohr, U., Bates, D.V., Fabel, H., Utell, M.J. (eds) Advances in Controlled Clinical Inhalation Studies. ILSI Monographs. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-77176-7_7

Chen, R.C.; Guo, H.Y.; Lin, M.P.; Lin, H.S. The carbon dioxide concentration detection using mobile phones combine Bluetooth and QR code. In Proceedings of the 6th IEEE International Conference on Awareness Science and Technology (iCAST), Paris, France, 29–31 October 2014.

Kim, J.Y.; Chu, C.H.; Shin, S.M. ISSAQ: An integrated sensing systems for real-time indoor air quality monitoring. IEEE Sens. J. **2014**, 14, 4230–4244. [CrossRef]

Albalak R. Cultural practices and exposure to particles pollution from indoor biomass cooking: effects on respiratory health and nutritional status among the Aymara Indians of the Bolivian Highlands [Doctoral dissertation]. University of Michigan, 1997.

-, R. S. W., -, B. U., -, P. G., & -, Prof. D. M. B. (2023). Indoor Air Pollution Monitoring System. *International Journal For Multidisciplinary Research*, *5*(3). https://doi.org/10.36948/ijfmr.2023.v05i03.3281

Adeleke, J. A., Moodley, D., Rens, G., & Adewumi, A. O. (2017). Integrating statistical machine learning in a semantic sensor web for proactive monitoring and control. *Sensors (Switzerland), 17*(4). https://doi.org/10.3390/s17040807

Arano, K. A. G., Sun, S., Ordieres-Mere, J., & Gong, B. (2019). The use of the internet of things for estimating personal pollution exposure. *International Journal of Environmental Research and Public Health*, *16*(17). https://doi.org/10.3390/ijerph16173130

Baqer, N. S., Albahri, A. S., Mohammed, H. A., Zaidan, A. A., Amjed, R. A., Al-Bakry, A. M., Albahri, O. S., Alsattar, H. A., Alnoor, A., Alamoodi, A. H., Zaidan, B. B., Malik, R. Q., & Kareem, Z. H. (2022). Indoor air quality pollutants predicting approach using unified labelling process-based multi-criteria decision making and machine learning techniques. *Telecommunication Systems*, *81*(4). https://doi.org/10.1007/s11235-022-00959-2

Bhardwaj, J., & Sharma, P. (2021). Artificial Intelligence-Based Smart Solution to Reduce Respiratory Problems Caused by Air Pollution. *Journal of Emerging Investigators*. https://doi.org/10.59720/20-149

Boesgaard, C., Hansen, B. V., Kejser, U. B., Mollerup, S. H., Ryhl-Svendsen, M., & Torp-Smith, N. (2022). Prediction of the indoor climate in cultural heritage buildings through machine learning: first results from two field tests. *Heritage Science*, *10*(1). https://doi.org/10.1186/s40494-022-00805-3

Cho, J. H. (2020a). Detection of smoking in indoor environment using machine learning. *Applied Sciences (Switzerland)*, *10*(24). https://doi.org/10.3390/app10248912

Cho, J. H. (2020b). Detection of smoking in indoor environment using machine learning. *Applied Sciences (Switzerland)*, *10*(24). https://doi.org/10.3390/app10248912

Chu, K. U., & Ho, Y. H. (2022). Max Fast Fourier Transform (maxFFT) Clustering Approach for Classifying Indoor Air Quality. *Atmosphere*, *13*(9). https://doi.org/10.3390/atmos13091375

Dorst, T., Schneider, T., Eichstädt, S., & Schütze, A. (2023). Influence of measurement uncertainty on machine learning results demonstrated for a smart gas sensor. *Journal of Sensors and Sensor Systems*, *12*(1). https://doi.org/10.5194/jsss-12-45-2023

Gabriel, M., & Auer, T. (2023). INDOOR AIR POLLUTION ESTIMATION USING MACHINE LEARNING (ANN AND SVR) IN SMART BUILDINGS. *Proceedings of BauSim 2022: 9th Conference of IBPSA-Germany and Austria*, *09*. https://doi.org/10.26868/29761662.2022.24

Guo, Q., Ren, M., Wu, S., Sun, Y., Wang, J., Wang, Q., Ma, Y., Song, X., & Chen, Y. (2022). Applications of artificial intelligence in the field of air pollution: A bibliometric analysis. *Frontiers in Public Health*, *10*. https://doi.org/10.3389/fpubh.2022.933665

Kapoor, N. R., Kumar, A., Kumar, A., Kumar, A., Mohammed, M. A., Kumar, K., Kadry, S., & Lim, S. (2022). Machine Learning-Based CO2Prediction for Office Room: A Pilot Study. *Wireless Communications and Mobile Computing*, *2022*. https://doi.org/10.1155/2022/9404807

Kim, D., Cho, S., Tamil, L., Song, D. J., & Seo, S. (2020). Predicting asthma attacks: Effects of indoor PM concentrations on peak expiratory flow rates of asthmatic children. *IEEE Access*, *8*. https://doi.org/10.1109/ACCESS.2019.2960551

Kim, J., Bang, J. il, Choi, A., Moon, H. J., & Sung, M. (2023). Estimation of Occupancy Using IoT Sensors and a Carbon Dioxide-Based Machine Learning Model with Ventilation System and Differential Pressure Data. *Sensors*, *23*(2). https://doi.org/10.3390/s23020585

Li, C., Xia, Y., & Wang, L. (2024). Household unclean fuel use, indoor pollution and self-rated health: risk assessment of environmental pollution caused by energy poverty from a public health perspective. *Environmental Science and Pollution Research*, *31*(12). https://doi.org/10.1007/s11356-023-27676-w

Lin, Z., Lin, S., Neamtiu, I. A., Ye, B., Csobod, E., Fazakas, E., & Gurzau, E. (2021). Predicting environmental risk factors in relation to health outcomes among school children from Romania using random forest model - An analysis of data from the SINPHONIE project. *Science of the Total Environment*, *784*. https://doi.org/10.1016/j.scitotenv.2021.147145

Long, H., Luo, J., Zhang, Y., Li, S., Xie, S., Ma, H., & Zhang, H. (2023). Revealing Long-Term Indoor Air Quality Prediction: An Intelligent Informer-Based Approach. *Sensors*, *23*(18). https://doi.org/10.3390/s23188003

Matthaios, V. N., Knibbs, L. D., Kramer, L. J., Crilley, L. R., & Bloss, W. J. (2024). Predicting real-time within-vehicle air pollution exposure with mass-balance and machine learning approaches using on-road and air quality data. *Atmospheric Environment*, *318*. https://doi.org/10.1016/j.atmosenv.2023.120233

Men, Y., Li, Y., Luo, Z., Jiang, K., Yi, F., Liu, X., Xing, R., Cheng, H., Shen, G., & Tao, S. (2023). Interpreting Highly Variable Indoor PM2.5 in Rural North China Using Machine Learning. *Environmental Science and Technology*, *57*(46). https://doi.org/10.1021/acs.est.3c02014

Omidvarborna, H., Kumar, P., Hayward, J., Gupta, M., & Nascimento, E. G. S. (2021). Low-cost air quality sensing towards smart homes. In *Atmosphere* (Vol. 12, Issue 4). https://doi.org/10.3390/atmos12040453

Rajabi, M., Sardroud, J. M., & Kheyroddin, A. (2021). Green standard model using machine learning: identifying threats and opportunities facing the implementation of green building in Iran. *Environmental Science and Pollution Research*, *28*(44). https://doi.org/10.1007/s11356-021-14991-3

Ramirez-Alcocer, U. M., Tello-Leal, E., Macías-Hernández, B. A., & Hernandez-Resendiz, J. D. (2022). Data-Driven Prediction of COVID-19 Daily New Cases through a Hybrid Approach of Machine Learning Unsupervised and Deep Learning. *Atmosphere*, *13*(8). https://doi.org/10.3390/atmos13081205

Saminathan, S., & Malathy, C. (2023). Ensemble-based classification approach for PM2.5 concentration forecasting using meteorological data. *Frontiers in Big Data*, *6*. https://doi.org/10.3389/fdata.2023.1175259

Shi, T., Yang, W., Qi, A., Li, P., & Qiao, J. (2023). LASSO and attention-TCN: a concurrent method for indoor particulate matter prediction. *Applied Intelligence*, *53*(17). https://doi.org/10.1007/s10489-023-04507-6

Sokhi, R. S., Moussiopoulos, N., Baklanov, A., Bartzis, J., Coll, I., Finardi, S., Friedrich, R., Geels, C., Grönholm, T., Halenka, T., Ketzel, M., Maragkidou, A., Matthias, V., Moldanova, J., Ntziachristos, L., Schäfer, K., Suppan, P., Tsegas, G., Carmichael, G., … Kukkonen, J. (2022). Advances in air quality research - current and emerging challenges. In *Atmospheric Chemistry and Physics* (Vol. 22, Issue 7). https://doi.org/10.5194/acp-22-4615-2022

Sonawani, S., & Patil, K. (2024). Air quality measurement, prediction and warning using transfer learning based IOT system for ambient assisted living. *International Journal of Pervasive Computing and Communications*, *20*(1). https://doi.org/10.1108/IJPCC-07-2022-0271

Tong, X., Ho, J. M. W., Li, Z., Lui, K. H., Kwok, T. C. Y., Tsoi, K. K. F., & Ho, K. F. (2020). Prediction model for air particulate matter levels in the households of elderly individuals in Hong Kong. *Science of the Total Environment*, *717*. https://doi.org/10.1016/j.scitotenv.2019.135323

Wang, J., Du, W., Lei, Y., Chen, Y., Wang, Z., Mao, K., Tao, S., & Pan, B. (2023). Quantifying the dynamic characteristics of indoor air pollution using real-time sensors: Current status and future implications. In *Environment International* (Vol. 175). https://doi.org/10.1016/j.envint.2023.107934

Wei, Y., Jang-Jaccard, J., Sabrina, F., & Alavizadeh, H. (2020). Large-Scale Outlier Detection for Low-Cost PM Sensors. *IEEE Access*, *8*. https://doi.org/10.1109/ACCESS.2020.3043421

Zhang, C., Zhu, Z., Liu, F., Yang, Y., Wan, Y., Huo, W., & Yang, L. (2023). Efficient machine learning method for evaluating the compressive strength of cement-stabilized soft soil. *Construction and Building Materials*, *392*. https://doi.org/10.1016/j.conbuildmat.2023.131887

Zhou, S., Guo, Y., Su, T., Chen, G., Liu, H., Li, Q., Bao, H., Ji, Y., Luo, S., Liu, Z., Wang, H., Liu, J., Han, N., & Wang, H. J. (2023). Individual and joint effect of indoor air pollution index and ambient particulate matter on fetal growth: A prospective cohort study. *International Journal of Epidemiology*, *52*(3). https://doi.org/10.1093/ije/dyad021.