UNIVERSITY OF RWANDA

COLLEGE OF SCIENCE AND TECHNOLOGY

AFRICAN CENTER OF EXCELLENCE IN INTERNET OF THINGS (ACEIoT)

A DECENTRALISED BLOCKCHAIN-BASED AIR POLLUTION SPIKES MONITORING FRAMEWORK OVER INTELLIGENT IoT EDGE NETWORKS

PhD. Thesis submitted in the fulfillment of requirements of the award of PhD. Degree in the Internet of Things – Embedded Computing Systems.

Eric NIZEYIMANA

SEPTEMBER 2024

Eric Nizeyimana, September 2024,

UNIVERSITY OF RWANDA

COLLEGE OF SCIENCE AND TECHNOLOGY

AFRICAN CENTER OF EXCELLENCE IN INTERNET OF THINGS (ACEIoT)

A DECENTRALISED BLOCKCHAIN-BASED AIR POLLUTION SPIKES MONITORING FRAMEWORK OVER INTELLIGENT IoT EDGE NETWORKS

Ph.D Thesis submitted in the fulfillment of requirements of the award of PhD. Degree in the Internet of Things – Embedded Computing Systems.

Eric NIZEYIMANA
Reference number: 220014393

Thesis Supervisor: Prof. Damien Hanyurwimfura, PhD

Thesis Co-Supervisors: Dr. Jimmy Nsenga, PhD

Thesis Co-Supervisor: Prof. Junseok Hwang, PhD

SEPTEMBER 2024

Eric Nizeyimana, September 2024,

DECLARATION

I hereby declare that the dissertation entitled "**A Decentralized Blockchain-based Air Pollution Spikes Monitoring Framework over Intelligent IoT Edge Networks**" to be submitted for the Degree of Doctor of Philosophy on the Internet of Things: Embedded Computing Systems at the University of Rwanda, College of Science and Technology, African Center of Excellence in Internet of Things is my original work and has not formed the basis for the award of any degree, diploma, associateship, or fellowship of similar other titles. I also declare that all source of information used was acknowledged by a complete list of references and were cited.

Done at Kigali, on 30 September 2024

Eric Nizeyimana

Eric Nizeyimana, September 2024,

Eric Nizeyimana, a Ph.D student of UR-ACEIoT registration ID 220014393, successfully defended the thesis/dissertation entitled "**A Decentralized Blockchain-Based Air Pollution Spikes Monitoring Framework over Intelligent IoT Edge Networks**", which he prepared after fulfilling the requirements specified in the associated legislation, before the thesis examination members whose signatures are below.

Thesis Supervisor:	**Prof. Damien Hanyurwimfura** University of Rwanda	------------------------
Co-Supervisor:	**Dr. Jimmy Nsenga** University of Rwanda	------------------------
Viva Voce Members:	**Prof. Wali U. Garba (Chair)** University of Rwanda	------------------------
	Dr. Hung Tran National Economics University, Vietnam	------------------------
	Dr. Abdi T. Abdalla University of Dar es Salaam	------------------------

Date of Submission: **10/09/2024**

Date of Defense: **30/09/2024**

Eric Nizeyimana, September 2024,

To my spouse, children, family, and friends,

Eric Nizeyimana, September 2024,

FOREWORD

This thesis is the final deliverable of my Ph.D. research conducted at the African Centre of Excellence in Internet of Things (ACEIoT), with financial support from the Partnership for Skills in Applied Sciences, Engineering, and Technology (PASET) through its Regional Scholarship and Innovation Fund (RSIF).

The motivation for my research stemmed from the alarming number of people worldwide suffering from diseases such as heart disease, lung disease, and pneumonia, often without even knowing the cause, which is linked to poor air quality. With the intent to mitigate the impact of poor air quality on public health, I directed my research toward finding ways to reduce the mortality rate associated with air pollution using advanced knowledge and technology.

Consequently, I submitted a proposal to PASET RSIF for a Ph.D. scholarship. After a rigorous and competitive selection process, I was honored to be among the few selected in the second cohort of the scholarship program. I then joined the University of Rwanda at the African Centre of Excellence in the Internet of Things (ACEIoT). I submitted my application in mid-2019, received my offer in February 2020, and commenced my studies shortly after that with the title "A DECENTRALISED BLOCKCHAIN-BASED AIR POLLUTION SPIKES MONITORING FRAMEWORK OVER INTELLIGENT IoT EDGE NETWORKS".

First and foremost, I give all thanks to Almighty God for His protection and strength. I am nothing without the Lord who strengthens me every day. I want to extend my gratitude to the Government of Rwanda and the University of Rwanda for their initiatives to empower and strengthen human capacity, promoting a knowledge-based economy by delivering high-quality, market-relevant postgraduate education in Rwanda.

My sincere gratitude goes to the ACEIoT administration team for their consistent support and guidance throughout this journey of over four years. Special thanks to Prof. Hanyurwimfura Damien and Dr. Gatera Omar for their inspiration, administrative support, and facilitation in accomplishing various research activities.

I am deeply grateful to my main supervisors, Dr. Jimmy Nsenga and Prof. Damien Hanyurwimfura, for their unwavering support, guidance, and encouragement. Your insights and assistance were crucial to the success of this research. I also extend my thanks to my co-supervisor, Prof. Juneseok Hwang, for his dedicated support, especially during my stay at Seoul National University (SNU) in South Korea. Your guidance was pivotal in accomplishing significant tasks

within a short time. I also appreciate the support from the SNU GRC administration and my co-author, Dr. Dereje Regassa, whose motivation and guidance were invaluable. I also thank the other co-authors who contributed to the journal publications of this research. I extend my gratitude to the doctoral committee for their reviews and comments, which greatly contributed to the success of this thesis.

My heartfelt gratitude goes to my family—my mother, grandmother, aunts, cousins, brothers, sisters, inlaws, and uncles—for their moral support from the beginning of my educational journey to this day. Mum and Grandma, you have been the key to who I am today. I appreciate so much for the love and support from childhood. I appreciated the moral support from various friends, I recognize every motivation from your side. My thanks also go to my classmates, especially Mr. Fidele Maniraguha, Mrs.Barbara Asingwire Kabwiga, and Mr. Prudence M., for your daily motivation, particularly during our time at SNU, which was highly appreciated. Your inspiring words helped me persevere through moments of exhaustion.

Finally, my deepest thanks go to my spouse, Tuyishime M. Ernestine, for your patience and prayers throughout this long journey. To my daughters, Iriho I. Gania, and Ituze N. Galia, and our son Iti Kimptu Gana, as well as my other children, I am their "daddy". I apologize for the times I was preoccupied with writing and could not give you the attention you deserved. Your constant encouragement kept me going, and without it, I may not have been able to complete my studies. You are all a part of the accomplishments that have brought this long journey to a close.

Table of Contents

List of Figures

Eric Nizeyimana, September 2024,

List of Tables

Eric Nizeyimana, September 2024,

ABSTRACT

Air pollution spikes pose significant challenges to public health, environmental stability, and economic development, demanding robust and innovative solutions for effective monitoring and management. This research introduces a decentralized blockchain-based framework for air pollution spike monitoring, leveraging intelligent IoT edge networks to enhance data accuracy, reliability, and timeliness. By integrating blockchain technology with IoT sensors and edge computing, the proposed framework aims to overcome the limitations of traditional air quality monitoring systems, which often struggle with centralized data processing, security vulnerabilities, and delayed response times.

The framework employs IoT devices with multi-pollutant sensors to capture real-time air pollution spike data. These edge devices preprocess and analyze the collected data locally, reducing the need for constant communication with centralized servers and thus enhancing energy efficiency and responsiveness. The edge network architecture ensures that critical information is processed and acted upon promptly, enabling immediate detection and mitigation of pollution spikes.

Blockchain technology is integrated into the framework to ensure the collected data's integrity, transparency, and security. Utilizing a decentralized ledger, the system records all air quality data and transactions immutably, preventing data tampering and fostering trust among stakeholders, including regulatory authorities, industries, and the general public. Smart contracts are deployed on the Ethereum platform to automate the enforcement of air quality regulations, issuing fines to polluters who exceed predefined emission thresholds. This automation streamlines regulatory compliance and enhances accountability and enforcement efficiency.

Machine learning algorithms are applied to the gathered data to predict future pollution trends and detect anomalies in real time. The research compares various time series models, including exponential smoothing, ARIMA, and Recurrent Neural Networks (RNNs), to determine the most effective approach for forecasting air pollution spikes. The exponential smoothing model demonstrated superior performance in terms of accuracy (with Mean Absolute Error (MAE) and Mean Square Error (MSE) of 3.66 and 84.86 respectively for PM2.5) and computational efficiency (due to its simple recursive calculations, making them faster to compute and requiring less memory), enabling reliable prediction and proactive management of pollution events.

This research signifies a paradigm shift in air quality management by combining advanced technologies to create a comprehensive, decentralized monitoring system. The findings underscore the potential of blockchain and intelligent IoT edge networks to revolutionize the way air pollution is monitored and managed, ultimately contributing to improved public health outcomes and environmental sustainability. The results of this research demonstrated the successful design of a prototype framework capable of collecting air pollution spikes while conserving energy for more than 40% by activating only when peaks exceed a predefined threshold. The designed framework demonstrated also that by waking up using the threshold measurement, the possibility of losing some spikes which was 7.1% is no longer there as shown using Monte Carlo. The collected data were analyzed using blockchain technology to design smart contracts for air pollution spikes, ensuring efficient and cost-effective implementation. Additionally, the data were analyzed using machine learning models to predict future spikes. Among the three models evaluated (exponential smoothing, RNNs, and ARIMA), exponential smoothing outperformed the others, proving to be the most effective for this application for all pollutants.

Keywords: Air pollution spikes, IoT edge networks, Machine learning algorithms, Smart contracts

Chapter 1. Introduction

This chapter introduces the study comprehensively, offering a deeper understanding of the problem addressed by this research. It begins by elaborating on the significance and urgency of the issue, detailing the various dimensions and implications of the problem at hand. This includes an exploration of the context in which the problem exists, the challenges it presents, and its impact on relevant stakeholders. The chapter also clearly defines the research questions that guide the investigation, ensuring that the study is focused and directed toward specific, measurable outcomes. Furthermore, it outlines the objectives of the research, specifying what to achieve and how it intends to contribute to the existing body of knowledge. These objectives are framed in a way that highlights their relevance and importance to both the academic community and practical applications. In addition to identifying the problem and objectives, the chapter discusses the anticipated contributions of the research, explaining how the findings could provide solutions or advancements in the field. It also sets the stage for the structure of the remaining chapters, giving a brief overview of what each subsequent chapter will cover, thus providing a roadmap for the reader. This ensures a clear and logical progression of ideas throughout the study, facilitating a better understanding of how the research unfolds and leading to its conclusions.

1.1. The Air Pollution Challenge

Air pollution has posed a major challenge since the beginning of the industrial revolution start of the industrialization revolution on the planet [1] [2]. Automobile emissions, chemical odors, and factory smoke are on the front end of the big pollutants in the air [3]. In every area of activity of Human beings, the intervention of technology plays a clear role in the quick development of the world [4].

In previous years, air pollution has silently become a prolific and invisible killer [1], [5]. While conventional air pollution monitoring systems primarily track long-term peaks, this research underscores the critical danger posed by short-term spikes [6] that can cause various diseases including eye and adnexa [7], [8], brain volume, cognitive decline, and an increased risk of developing dementia [9], heart disease, chronic obstructive pulmonary disease (COPD), lung cancer, migraines, acute lower respiratory infections, and stroke [10]. Air pollution negatively affects the health of nine out of ten children, impeding brain development and compromising their overall well-being [11]. The issue extends to mental health implications, particularly affecting children, with documented instances of brain cell inflammation. Short-term spikes in air pollution are correlated with heightened hospital admissions for childhood psychiatric conditions. The research underscores that children from low-income backgrounds bear a disproportionate impact, evidenced by a 44% increase in hospital visits due to suicidal ideation linked to air pollution spikes [12].

Air pollution is a significant challenge that needs to be addressed in the new industrial development as one of the real-life problems [13]. There are a lot of factors caused by the increase in air pollution such as urbanization, population growth, industrialization, and increased vehicle use. Air pollution results in a low quality of life and dangerous effects on human health by directly affecting the well-being of the population [13]. More than seven million premature deaths are dying due to the problems caused by air pollution every year[5]. The impact of air pollution generates a threat to the quality of life on our planet where children are much affected due to their lungs which are still developing and that can lead to the low capacity of their lungs in adulthood [5]. Air pollution affects our daily activities, health, and wealth as well. Besides other various types of pollutants that come from thermal, soil, noise, and water; air pollution is on the front line for being very dangerous and causing life-threatening diseases and climate change[13]. Air pollution may cause respiratory problems, allergies, diseases, and death in humans and the ecosystem. The report from the World Health Organization (WHO) [14] shows that 94 percent of the world's population is now exposed to air pollution. It can damage the built environment or nature and can cause harm to other living organisms such as food crops and animals[15]. Both natural processes and human activities can create air pollution [16].

The existing systems are inadequate for facilitating compliance with established standards due to their inability to generate useful or meaningful data and their reliance on ineffective techniques for monitoring air pollution spikes[17]. Real-time monitoring using cloud-centric architectures exacerbates this problem, as it rapidly drains the batteries of edge sensors [18]. This, in turn, significantly increases the maintenance costs associated with frequent battery replacements, further highlighting the inefficiency of current monitoring approaches.

1.2. Overview of Air Pollution Legislation

Air pollution is a big challenge for almost all countries of the world. Different governments have taken some measures but there are still problems (security, reliability, fraud of data) with monitoring air pollution from a specific source [19]. The monitoring also is based on looking at the second pollutant called ozone instead of taking factors to the problem of the primary pollutant [20]. This is very challenging because they are trying to solve the problem, but they are still needed by the authorities. Therefore, there is a need to monitor primary pollutants because they are the ones that create secondary pollutants. Being able to measure primary pollutants by having sensors closer to the source, it possible to know exactly who is creating which pollutants; this is almost difficult if analyzing aggregated pollutants. Authorities use global satellites to measure air pollution and most of the time these data take a long time and have some limitations. The report presented by the World Bank in 2015 has shown that PM2.5 experiences high levels of air pollution in some countries. This means having some of their cities with very high pollution levels but comparatively low overall average for the whole country. Some countries have installed additional stations for monitoring air pollution, but they measure only the particle matter of a specific area [21].

A cutting-edge solution for integrating remote sensors and ground measurements was proposed by the World Bank [21]. Even though many countries have advanced in trying to tackle the problem of air pollution, there is still a lot to do to achieve the targeted work in the future [21] like they know which pollutant is most important because they measure only particle matter (PM).

This research proposes the way authorities will use to manage the air pollution sent into the atmosphere. Every source of pollutants will be able to know the quantity exposed and also, if necessary, authorities may charge them accordingly. The objective of this research is to build a smart fraud-free high-resolution air pollution monitoring framework that collects data that is completely open for both source of pollutants and monitoring authorities, thus relying on the integration of emerging technologies such as blockchain, edge computing, and machine learning.

1.3. Motivation and Problem Statement

Air pollution is an important societal challenge threatening the world [22]. Many researchers have adopted technical solutions centered around the cloud. Nowadays, cloud infrastructures are provided by many competitors which has enabled them to push down the cost [23]. However, the issue of the cloud is that a centralized server results in the incidence of communication between user devices [24]. This makes a limitation for applications that entail instantaneous reaction (low latency), operate under limited bandwidth, or are subjected to high privacy requirements. Thus, finding something further than the clouds (such as edge computing) is necessary and air pollution will be showcased in this research.

This PhD research thesis entitled "A Decentralized Blockchain-Based Air Pollution Spikes Monitoring Framework Over Intelligent IoT Edge Networks" revolves around the urgent need for effective monitoring and management of air pollution spikes, particularly in urban environments. Despite various efforts to address air quality issues, traditional monitoring methods often lack real-time capabilities and struggle with data accuracy and trustworthiness. Furthermore, the increasing complexity of urban environments and the dynamic nature of air pollution necessitate innovative solutions that can provide timely and reliable insights into pollution spikes.

Existing air quality monitoring systems face several challenges, including data latency, data integrity concerns, and limited predictive capabilities. These limitations hinder the ability to detect pollution spikes promptly and take proactive measures to mitigate their adverse effects on public health and the environment. Moreover, the lack of a decentralized and transparent framework for monitoring air pollution limits accountability and enforcement mechanisms, leading to challenges in identifying and penalizing offenders.

Therefore, there is a critical need for a decentralized blockchain-based framework that leverages intelligent IoT edge networks to monitor air pollution spikes effectively. Such a framework would address the shortcomings of existing systems by providing real-time monitoring, ensuring data integrity and trustworthiness through blockchain technology, and enabling predictive analytics to anticipate future pollution events. This framework is designed to enhance transparency, accountability, and collaboration among stakeholders involved in air quality management by decentralizing the monitoring process and incorporating blockchain technology.

This research underscores the importance of developing a decentralized blockchain-based framework for monitoring air pollution spikes over intelligent IoT edge networks to address the pressing challenges associated with urban air quality management.

1.4. Objectives of the research

The main purpose of this study was to develop an integrated framework leveraging edge IoT, blockchain, and AI for real-time monitoring and analysis of air pollution spikes, addressing the critical need for low latency, data trust, and predictive capabilities.

Specific Objectives:

i. Design the concept of art for air pollution monitoring based on the weaknesses and limitations of existing systems.
ii. Develop a real-time multi-pollutant (RTMP) sensor integrated and application blockchain identity for tokenizing pollution offenses on the data.
iii. Optimize IoT energy usage in a hardware-software smart IoT device for monitoring short-term exposure to pollution peaks.
iv. Evaluate the performance of air quality monitoring systems by leveraging machine learning to analyze collected spike data and propose fines for peak emitters.

1.5. Research Questions

i. How can edge IoT, blockchain, and AI be integrated to develop a real-time multi-pollutant sensor for monitoring air pollution spikes?
ii. What are the key features and functionalities of short-term prediction AI for alerting emitters of potential risks exceeding exposure standards due to pollution spikes?
iii. How can IoT energy usage be optimized in a hardware-software smart IoT device for monitoring short-term exposure to pollution peaks?
iv. What machine learning techniques can be effectively utilized to analyze collected spike data and propose fines for peak emitters in air quality monitoring systems?

1.6. Research Methodology

This research focused on developing an IoT device capable of capturing various air pollutants. The study began by identifying the significant gap in monitoring air pollution spikes, which can be harmful to human health, particularly affecting the intelligence quotient (IQ) of children. The design of the concept of the art of the framework was generated and enumerated all requirements needed for the development. Recognizing the dangers posed by these pollution spikes, the research team designed a framework to collect comprehensive air pollution data. A Real-Time Multipollutant Sensor (RTMS) was developed using IoT equipment to gather air pollution data directly from the source. The energy is saved by operating the system only for the appearance of the peak that exceeds the threshold of the defined level. This data collection was then integrated with blockchain technology to create smart contracts on the Ethereum platform, ensuring data integrity and transparency.

Subsequent analysis of the collected data was conducted using machine learning techniques, specifically focusing on time series models. The study compared three models to determine the most effective approach for analyzing the air pollution data. The exponential smoothing model emerged as the best performer among the models tested, demonstrating superior accuracy in predicting pollution spikes. This comprehensive approach not only addresses the initial gap in monitoring air pollution spikes but also leverages advanced technologies like IoT, blockchain, and machine learning to provide a robust solution for real-time air quality monitoring and analysis.

1.7. Overview of Major Contributions

The major contributions of this thesis are as follows:

i. **Comprehensive Literature Survey:** Conducted an in-depth literature survey focused on monitoring air pollution spikes.

ii. **Embedded Device Development:** Developed and deployed an embedded device for real-time monitoring of air pollution spikes.

iii. **Machine Learning Prediction Analysis:** Performed predictive analysis using various machine learning algorithms on data gathered from the Real-Time Multipollutant Sensor (RTMS).

iv. **Smart Contract Development:** Created a smart contract to enforce fines on emitters exceeding the World Health Organization (WHO) defined air pollution thresholds.

These contributions address the research questions and align with the objectives of the study. The outcomes have been consolidated into four research papers.

1.8. Thesis Outline

- Chapter 1 introduces an overview of the research topic. Introduces the problem statement, research questions, and objectives; it also highlights the significance and potential impact of the study.
- Chapter 2 is for literature review and presents a comprehensive survey of existing research related to air pollution monitoring by identifying gaps in current knowledge and technological advancements, and also Establishes the context for the research and justifies the need for the study.
- Chapter 3 describes the concept of the art based on the gaps found in Chapter 2 and then proposes the framework to be developed.
- Chapter 4 designs the proposed framework and explains the application of blockchain technology for developing smart contracts on the Ethereum platform.
- Chapter 5 describes the development and deployment of the embedded IoT device for real-time air pollution monitoring. It gives also the details of the data collection process using the Real-Time Multipollutant Sensor (RTMS).
- Chapter 6 then discusses the predictive analysis of air pollution data using various machine learning algorithms. It also compares the performance of different time series models, with a focus on exponential smoothing. Details the creation and functionality of the smart contract for enforcing fines on air pollution emitters. Describes the deployment process on the Ethereum blockchain. Illustrates how the smart contract ensures compliance with WHO-defined air pollution thresholds.
- Chapter 7 concludes the thesis. Below is the reference structure of chapters with their corresponding objectives for this thesis.

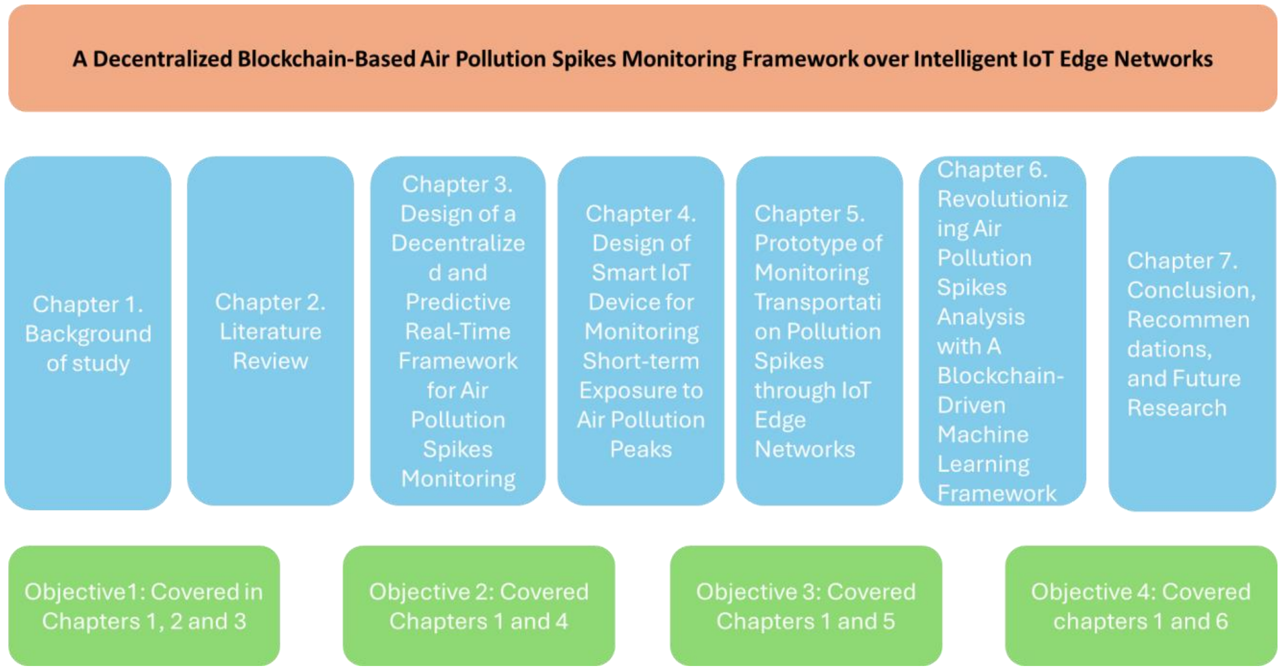


Figure 1. 1. Thesis structure

This Figure 1. 1 provides a detailed correlation between the chapters of this research and their respective objectives. Each chapter is aligned with specific research goals, illustrating how the study progresses logically and cohesively. It highlights the structured approach taken to address the research questions, starting from the initial literature review and methodological framework to the development and implementation of the proposed solutions. Mapping out the chapters alongside their objectives, this Figure 1. 1 offers a clear roadmap of the research process, emphasizing the systematic efforts undertaken to achieve the desired outcomes and contribute to the field of air pollution monitoring and management.

Chapter 2. Literature Review

2.1. Overview and Background

This chapter elaborates on the current strategies employed to address the issue of air pollution, with a predominant reliance on cloud computing solutions. With the global rise in air pollution levels, there is a growing imperative to accommodate the escalating demands of industries, which are primary contributors to atmospheric degradation. Embedded systems emerge as a prominent solution, offering versatile applications across various sectors, including agriculture [25], health [26], Transport [27].

This chapter analyzes some air pollution contests, and the approaches used to solve the problem. There is a subsection for showing the impact of chemical products on the air quality, and the proposed solution that has been recognized for solving the spread of air smog in the atmosphere. The next subsection makes a great comparison between both cloud and edge computing solutions. And lastly, this section gives the impact of AI on air pollution based on the literature from other previous results.

2.2. Air Pollution Challenge

Air pollution has been identified as a big challenge in the industrial revolution [11]. Air pollution is primarily generated from human beings such as combusting fossil fuels. Most air pollution comes from the combustion of coal, and petroleum and this generates sulfur dioxide. The lives of human beings rely on fossil fuels to do their day-to-day activities even though they lead to air pollution which damages the Atmosphere. Many other chemical pollutants generated by manufacturing industries also play a crucial role in damaging the air quality. The world is encountering many manufacturing industries, and the big problem is that they are all releasing very harmful gases in the Atmosphere [28].

The air pollution has demonstrated effects on both humans and the environment. Most problems affect the lives of human beings, and these are death, respiratory diseases, heart disease, eyes, Chronic fatigue, blood illnesses and cancer, genetic disorders, and transmitting the impacts to the following generation, and have all arise [25] , [26]. As for the environment, there is an increase in day-by-day problems like climate change, harm to the soil, vegetation loss, reduced biodiversity, erosion, acid rain, and increasing greenhouse consequences [27].

In urban areas, the problem of air pollution is a big matter to be controlled due to its significant impact on health, the environment, and the economy, especially in developing countries [7]. Different anthropogenic processes in big cities are causing the emission of air pollutants and these can only be categorized into source groups like power plants, motor traffic, industry, trade, and domestic fuel [28]. Air quality is a life-threatening manner that straightforwardly impacts human wellbeing [29] .

The problem of air pollution has challenged many countries and some of them have tried to find solutions [9]. These solutions include an increase in using public transportation, especially in cities reducing private transport(cars), increasing the bike paths network in cities, developing and increasing the urban green spaces in the cities, etc. In the report [27] , the researchers raised some management challenges that did not allow these solutions to be performed. All these management challenges are not allowing the proposed solution to be implemented.

Besides the management challenges also technical ones were presented in the report [27]. These challenges are old technology vehicles and industries; poor fuel quality; and lack of knowledge of using control technologies for pollution.

2.3. Air Pollution Monitoring Technologies

Measures that could be taken can be oriented towards the new technology of IoT. Nowadays, embedded systems are becoming more and more powerful (more processing capacity, better power....) in solving different problems in all areas and can also be implemented for air pollution [2]. Today, most existing air monitoring solutions rely on cloud processing architecture, where "dump" sensors keep sending data to be processed in the cloud. For air pollution, many systems have been designed, and most of them are based on IoT technology [3]. Different researchers have identified proposed approaches. Most of the solutions proposed for the problem are using cloud computing technology which may have some challenges in analyzing air pollution data.

These challenges of cloud solutions can be latency (time to transmit data from sensor to cloud), security (data are transmitted via the Internet and may be intercepted and used in a wrong way), massive infrastructure, fraud of data, and reliability of data [4]. In the paper [5], the authors showed that the latency might take 24 hours to process data generated from edge devices to the cloud. That time is too long and can generate a high risk to the population due to the causes associated with the measured data. There is a need to use edge computing to find the solution to air pollution and generate real-time and reliable data for a given place to be analyzed. Most of all pollutants that are exposed in the Atmosphere are in the form of chemical elements [6].

This research will build on top of the previous analysis to focus on monitoring air contamination caused by chemical products in the context of an industrial zone. The proposed solution analyzes which elements are exposed in the Atmosphere and at which level they are affecting the air quality.

The Figure 2. 1 shows the architecture of edge computing. In general, data are taken at the edge devices and then sent to fog/edge nodes to process data [7]. The figure shows three layers that are needed for edge technology.

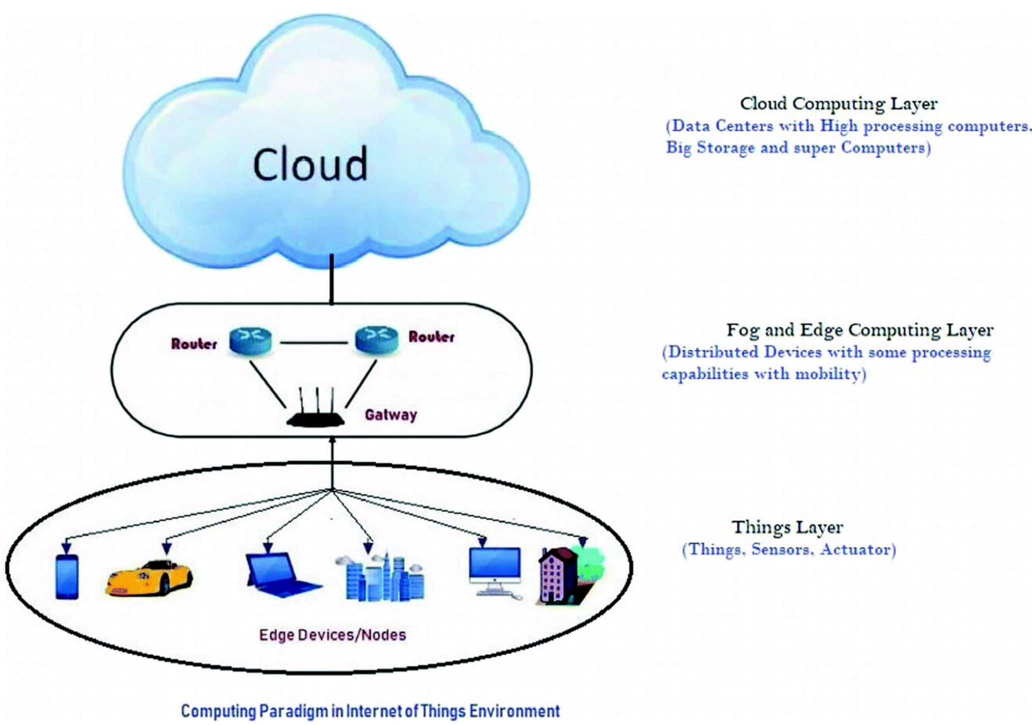


Figure 2. 1 Edge Computing Architecture for Air Quality Monitoring

This research will design a system that will help monitor air pollution in various locations based on the high volume of pollutants in these places. The system will measure the level of pollution in the air based on the chemical elements and compare them to the initial level of pollutants in the Atmosphere at the edge device. The system will notify the supervisors once the level of air pollution exceeds a predefined threshold and will generate the report of each site for future analysis. The system will alert by alarm and message to the administrator, and data will be analyzed accordingly.

2.4. Cloud Architecture for Air Quality

After identifying the problem of air pollution and its effects on both human beings and the environment, researchers have started finding ways to solve this problem. Many proposed solutions have been identified.

These solutions were correcting data from IoT devices and then analyzing them using different algorithms. All these analyses were performed on the cloud and then given predictions. This subsection is going to look at previous solutions that have been identified for air pollution by different research.

In the paper [34], the authors reviewed the published research results obtained relating to air quality using big data and machine learning and they proposed future challenges to work on. These challenges and observations can be summarized as the lack of a strong need for big data quality assurance research in modeling the data quality, the lack of making real-time validation and monitoring air quality automatic, and the lack of tools for increasing the accuracy of air quality evaluation. Other issues like monitoring and analyzing air quality based on different levels and the supervision of air resources. Based on these findings, we are proposing a system that will be a solution for these issues and challenges presented in the paper [34].

Paper [31] estimates the level of pollutants based on the ratio of chemical elements on ozone based on different seasons but without knowing the quantity of emitted. In this research, our smart embedded system will be able to measure the emission of each pollutant in the air. The paper [35] made an accurate assessment of novel approaches for air pollution that can generate environmental health based on clinical results. In that paper, authors based on vital signs and glucose levels. This research will consider the data corrected for analysis and the projection of air pollution based on the results obtained.

In the papers [15] [36], authors demonstrate the use of sensors in monitoring the level of pollution from anywhere with a mobile and computer. Papers do not prove the measurement of air quality to save the lives of premature deaths. The paper [37] explains the way Air pollution is one of the main environmental jeopardies and damages public health throughout the world. As the results of the research found in the paper [37], this mainly affects human health through diverse diseases like lung diseases, hair fall, throat infections, etc. The paper [37] identifies air pollution and provides a suitable recommendation system to reduce or avoid air pollution.

The paper [38] explains the use of high-quality sensors in monitoring air pollution at need high cost. They propose a low-cost sensor system to monitor air quality with real-time systems for predictions using edge computing. This research will provide a system that will be lower cost and easy to use in monitoring air pollution and will know which chemical element is emitted higher than others and the source for that pollutant and the data will be secured.

Different models have been demonstrated in predicting the monitoring of air pollution using machine learning. In the paper [29], researchers proposed the IoT system and presented the modeling based on that system for findings from the predictive result given by machine learning models. However, the paper has not mentioned the automation of the system which can be used in predicting the future of air pollution and measuring the pollutants that are exposed in the Atmosphere.

In papers [39] [40] , authors demonstrated the development solution of air quality based on cloud management and analysis for prediction. However, there is still the need for edge solutions that will allow a quick analysis of data with strong security which solves the problem of latency.

2.5. Spikes in Air Pollution

This section gives an overview of the framework needed for monitoring spikes. As explained above, most of the existing systems are not true real-time due to sensors that sleep in duty cycles to save energy. The shortest defined standard of latency is approximately one hour, and they are only relying on the cloud. This implies existing systems to deal with long-term exposure which led to the availability of long-term data for air pollution. For the existing systems to operate in true real-time, they must reduce the sleeping time from one hour to one minute. This approach seems impossible since the sensor needs time to rest to save the battery with a reduction of energy consumption and reduce the wireless transmission of data. The proposed Framework relies on the system that wakes up sensors when there is a spike/peak. One of the challenges is that there is almost none of dataset of monitoring spikes of air pollution. Therefore, if today's research does not design it, the challenge of the data set will still unsolved.

Existing air pollution monitors designed for long-term constant pollutants are based on cloud-centric architecture, with energy-constrained air pollution sensors operating in duty cycle mode (long period of sleep to save battery energy) and reporting the collected data to the remote cloud at a relatively low frequency (every 1h or even once a day) [47]. This reporting period is in this case acceptable since low latency or quick reactivity is not a requirement for such types of air pollutants. In the case of air pollution spikes, on the one hand, such a long period of sleep would result in missing the detection of several peaks given their short duration. On the other hand, if real-time monitoring is implemented with cloud-centric architecture, it would involve continuous wireless transmission from sensors to the cloud which would quickly drain the battery of those sensors and increase battery replacement maintenance costs [48]. Besides low latency, monitoring air pollution spikes still requires attention to long-term average pollution limits imposed by standards. Finally, given their fast and random occurrence nature, monitoring of pollution spikes requires transparency and integrity of collected data to avoid a lack of trust from both air pollutant emitters and regulatory control agencies [49].

Interestingly, the above requirements of low latency, short-term predictive, and trust in collected air pollution data can be fulfilled respectively by emerging technologies namely (1) edge-centric Internet of Things (IoT), (2) Artificial Intelligence (AI), and (3) blockchain. These technologies have been usually considered separately but recently different research initiatives have explored their convergences to create new application-driven added values [50]. In [51], authors analyzed this convergence in healthcare with a focus on the sensitive data of patients. The proposed system can run a real-time computation protecting user privacy within health data, especially in the case of mission-critical healthcare that requires reactivity to peak values of critical conditions. In [52], the convergence is analyzed in the case of edge-centric IoT nodes to dynamically handle the recovery of data from dead edge nodes within a distributed edge-based immutable public ledger. Regarding air pollution monitoring, the paper [53] has analyzed the integration of edge-centric IoT and AI to solve the problem of computational burden over battery-powered sensing nodes and reported a 70% energy consumption gain thanks to this integration. On the other hand, edge-centric IoT and blockchain have been combined to help air pollution devices self-organize in reaching consensus and storing air pollution data [54]. Our extensive state-of-the-art analysis highlighted that the convergence of the three technologies simultaneously for air pollution monitoring is still limited, to not say not existing.

This research presents the core framework integrating edge-centric IoT, blockchain, and short-term predictive AI applied to monitor particularly air pollution spikes in real time. As components of the proposed framework, our research aims at developing (1) an edge autonomous and low-cost real-time multi-pollutant (RTMP) sensor featuring a blockchain-based identity/wallet and embedding basic intelligence to track and detect spikes, (2) a blockchain smart contract running at fog layer to handle automatic billing as result of air pollution offense and (3) a short term predictive AI distributed between fog and cloud to alert air pollution emitters on risk to exceed long-term average limits defined by the standards.

2.6. Impact of Artificial Intelligence on Air Pollution Spikes

Artificial intelligence has been used in proposing the solution for air pollutants which are serious problems in many regions around the world. Different models have been proposed for the overwhelming effects of air pollution to predict the level of future concentrations. In the paper [45] , authors used deep learning to model architecture on real air pollution.

In the paper [46], machine learning has been proposed as a solution to air pollution for predicting the hourly concentration of some pollutants. The authors used machine learning because they demonstrated it as a popular technique that can proficiently train a model using large-scale optimization algorithms. Authors of paper [46] propose a model that can predict the concentration of air pollution hourly based on refined models to predict the hourly air pollution concentration based on meteorological data of previous days by formulating the prediction over 24 hours as a multi-task learning (MTL) problem.

All technologies used to solve the problem of air pollution have been successful but there are still challenges such as latency, security of data, and privacy. This research proposed the combination of edge computing, blockchain, and artificial intelligence for solving the problem of air pollution. In the next section, there is the proposed model and architecture.

2.7. Short-term Monitoring

In the paper [29] authors described the measurement of short-term exposures to particulate matter (PM) at 1-minute intervals using a small personal nephelometer (pDR; MIE, Inc). This device was worn by ten volunteers over a week, providing detailed data on PM concentrations. The focus on such short intervals allowed for precise capture of PM exposure fluctuations, highlighting the minimal yet significant exposure periods that might otherwise be overlooked in longer monitoring intervals [29]. The study emphasized the critical nature of these short-term measurements in understanding the true impact of PM on health [30]. But for other pollutants, if a substance has a short lifetime, ranging from a few seconds to a few hours, it cannot travel far from its emission source. As a result, short-lived gases are typically observed near their emission origins, such as industrial plants or forest fires. This characteristic makes it easier to detect emission hotspots, as the concentration of these gases will be significantly higher near the source. By monitoring these areas, they can quickly identify and address pollution sources, facilitating more effective environmental management and mitigation strategies [30].

The researchers demonstrated that the short-term duration of air pollution can be as brief as 15 minutes. However, depending on the study's parameters, they also showed that this duration can extend to within a week, particularly when considering the frequency and intensity of pollution peaks. This variability highlights the importance of context-specific monitoring and analysis, as the impact of air pollution can differ significantly based on temporal patterns and pollutant behavior. Understanding these dynamics is crucial for developing targeted strategies to mitigate the adverse effects of air pollution on health and the environment [31], [32].

2.8. Edge Solution for Air Pollution Using Blockchain

In their paper [33], the authors proposed a novel approach for designing a cost-effective and real-time air pollution monitoring system by leveraging edge computing and Internet of Things (IoT) technologies. Current air quality monitoring systems often lack the necessary spatial and temporal resolutions, posing challenges to accurately assessing air pollution. The proposed system employs sensors to collect real-time air quality data to address these issues, which is then transmitted to edge computing devices for processing and analysis [34]. The paper overviews existing monitoring systems, their limitations, challenges, and the proposed edge computing-based IoT architecture for air quality monitoring [34]. From an analysis of these two papers, this research demonstrates the feasibility of applying edge IoT with real-time systems to address air pollution. The approach proposed in these papers has been incorporated into this research to be applied within the realm of transportation, particularly for managing spikes in air pollution.

The paper [35] introduced a modular IoT sensing platform with hybrid learning capabilities for air quality prediction. A collaborative research effort in India focused on developing an IoT-based platform for accurate, affordable air quality monitoring, targeting challenges prevalent in underdeveloped regions. Their system, integrating multiple sensors for various pollutants and utilizing GSM/WiFi for real-time data transmission, aims to bolster environmental intelligence and combat ecological issues exacerbated by urbanization and fossil fuel consumption. However, the current research can capture spikes from vehicles in urban areas and subsequently use policies to act. Furthermore, the prototype proposed in this research encompasses a broader range of pollutants than the previous one.

Another paper [36] introduced a framework for air pollution monitoring in smart cities using IoT and smart sensors. Arshad Ali's study proposes an IoT-based solution to monitor environmental parameters in urban areas, aiding in combating pollution and enhancing city environments. The current prototype solved the issues presented in this paper, which come from the population increase in cities.

The paper [37] introduces a cloud-centric architecture leveraging IoT and smart sensors to monitor urban pollution, addressing challenges like sensor defects and data management for efficient air quality monitoring. The current research is now applied to vehicle pollutants and capturing spikes. En Xin Neo et al., [38] discussed using artificial intelligence (AI) in air quality monitoring for smart city management. The authors, En Xin Neo et al. [38] present a comprehensive study that utilizes machine learning and deep learning techniques to predict air quality. They propose an end-to-end predictive model incorporating various pollution markers and meteorological data for four different urban cities in Selangor, Malaysia. The study highlights the importance of feature optimization to enhance the accuracy of air quality predictions, particularly for PM2.5 concentration. While this paper exclusively focused on a single pollutant from a general perspective, the current research specifically addresses vehicles' spikes.

Moreover, the blockchain-based trusted edge platform enables rapid detection and response to air pollution spikes, facilitating timely interventions to mitigate the impacts on public health and the environment [39]. By fostering greater trust and transparency in pollution monitoring systems, this integrated approach holds immense potential to revolutionize how we monitor, analyze, and address air pollution spikes in a decentralized and collaborative manner [40], [41].

An edge-centric architecture applied to air pollution peak monitoring represents a transformative approach to data collection, analysis, and decision-making in environmental monitoring systems. Unlike traditional centralized architectures where data processing occurs predominantly in remote cloud servers, edge-centric architectures prioritize processing data closer to its source—at the "edge" of the network, where sensors and devices are deployed.

In the context of air pollution peak monitoring, an edge-centric architecture involves deploying sensors and data processing capabilities directly at the locations where pollution data is collected—such as urban areas, industrial sites, or transportation hubs. This allows for real-time analysis of sensor data at the point of collection, enabling immediate detection and response to pollution peaks as they occur.

2.9. Conclusion

The existing air pollution monitoring systems are thoroughly examined in this section, revealing several weaknesses that hinder their effectiveness. Primarily, energy consumption emerges as a significant concern, especially in conventional systems reliant on continuous data transmission and processing. This constant drain on energy resources not only leads to shorter battery life for sensors but also escalates maintenance costs and overall system inefficiencies. Moreover, the issue of periodic data transmission in cloud-centric monitoring architectures is highlighted, underscoring its inadequacy in capturing short-duration pollution spikes. While suitable for long-term pollutant monitoring, this approach falls short in addressing the immediate and transient nature of air pollution spikes. Additionally, cloud-centric face security limitations, with data vulnerability and privacy concerns posing significant challenges. Furthermore, the section underscores the absence of accountability mechanisms for penalizing emitters of air pollution spikes, reflecting gaps in existing regulatory frameworks. The inadequacy of current regulations to address short-term pollution events contributes to challenges in holding responsible parties accountable for their emissions. To tackle these weaknesses, the subsequent chapter delves into a detailed analysis of methodologies employed to develop innovative solutions. This involves a comprehensive examination of existing related work and research initiatives aimed at overcoming energy consumption issues, enhancing data transmission efficiency, strengthening security measures, and establishing accountability frameworks for air pollution spikes. By critically evaluating these challenges and methodologies, the research aims to identify the most effective strategies and approaches. Leveraging emerging technologies such as edge computing, blockchain, and machine learning, the research endeavors to develop a robust and comprehensive framework for air pollution monitoring and management. Through a combination of theoretical analysis and practical experimentation, the research seeks to address the shortcomings of existing systems and pave the way for a more sustainable and resilient approach to air quality monitoring.

Eric Nizeyimana, September 2024,

Chapter 3. Design of a Decentralized and Predictive Real-Time Framework for Air Pollution Spikes Monitoring

3.1. Overview

Exposure to air pollution spikes causes health problems to regularly exposed organisms, raising the need to monitor them in real time. Existing air pollution monitors mainly use a cloud-centric design considering relatively constant pollution, therefore duty-cycling sensors with long sleep periods to save their batteries. Such design is however inefficient for monitoring pollution spikes. Furthermore, since spikes vanish rapidly, the integrity of monitoring data is very important. This paper presents a framework integrating edge-centric design and blockchain in monitoring air pollution spikes while using short-term prediction artificial intelligence to timely alert pollution emitters about exceeding long-term average pollution limits defined by standards.

3.2. Introduction

The growth of the industrial revolution, road transportation, solid waste disposal, and fuel combustion have increased air pollution problems in different regions of the world, causing deaths, diseases, climate change, and other severe environmental degradations [1]. While air pollution characterized by long-term relatively constant levels has been the main consideration in air quality standards [42] and consequently in air pollution monitoring as well, several studies have shown that repetitive exposure to air pollution peaks/spikes may have important immediate or/and long-term health impacts for neighborhood beings [8], [10], [9]. Monitoring relatively constant air pollutants does not require real-time, meaning low latency, since the pollution level stays constant for a long period to allow sensing at low frequency. On the contrary, real-time monitoring is a strong requirement for air pollution spikes, the latter being characterized by a short lifetime, a period over which its pollution level is considerably higher than the one of a long-term period within which it occurs [11]. The importance of monitoring such types of pollutants has been demonstrated in [12] with a focus on blocking carbon peak exposure caused by transportation.

Existing air pollution monitors designed for long-term constant pollutants are based on cloud-centric architecture, with energy-constrained air pollution sensors operating in duty cycle mode (long period of sleep to save battery energy) and reporting the collected data to the remote cloud at a relatively low frequency (every 1h or even once a day) [43]. This reporting period is in this case acceptable since low latency or quick reactivity is not a requirement for such types of air pollutants. In the case of air pollution spikes, on the one hand, such a long period of sleep would result in missing the detection of several peaks given their short duration. On the other hand, if real-time monitoring is implemented with cloud-centric architecture, it would involve continuous wireless transmission from sensors to the cloud which would quickly drain the battery of those sensors and increase battery replacement maintenance costs [44]. Besides low latency, monitoring air pollution spikes still requires attention to long-term average pollution limits imposed by standards. Finally, given their fast and random occurrence nature, monitoring of pollution spikes requires transparency and integrity of collected data to avoid a lack of trust from both air pollutant emitters and regulatory control agencies [45].

Interestingly, the above requirements of low latency, short-term predictive, and trust in collected air pollution data can be fulfilled respectively by emerging technologies namely (1) edge-centric Internet of Things (IoT), (2) Artificial Intelligence (AI), and (3) blockchain. These technologies have been usually considered separately but recently different research initiatives have explored their convergences to create new application-driven added values [41], [46], [47], [48]. In [49], authors analyzed this convergence in healthcare with a focus on the sensitive data of patients. The proposed system can run a real-time computation protecting user privacy within health data, especially in the case of mission-critical healthcare that requires reactivity to peak values of critical conditions. In [41], the convergence is analyzed in the case of edge-centric IoT nodes to dynamically handle the recovery of data from dead edge nodes within a distributed edge-based immutable public ledger. Regarding air pollution monitoring, on the one hand [48] has analyzed the integration of edge-centric IoT and AI to solve the problem of computational burden over battery-powered sensing nodes and reported a 70% energy consumption gain thanks to this integration. On the other hand, edge-centric IoT and blockchain have been combined [50] to help air pollution devices self-organize in reaching consensus and storing air pollution data. Our extensive state-of-the-art analysis highlighted that the convergence of the three technologies simultaneously for air pollution monitoring is still limited, to not say not existing.

In this work, we presented our core framework integrating edge-centric IoT, blockchain, and short-term predictive AI applied to monitor particularly air pollution spikes in real time. As components of the proposed framework, our research aims at developing (1) an edge autonomous and low-cost real-time multi-pollutant (RTMP) sensor featuring a blockchain-based identity/wallet and embedding basic intelligence to track and detect spikes, (2) a blockchain smart contract running at fog layer to handle automatic billing as result of air pollution offense and (3) a short term predictive AI distributed between fog and cloud to alert air pollution emitters on risk to exceed long-term average limits defined by the standards.

3.3. Requirement for Efficient Monitoring of Air Pollution Spikes

As stated in the introduction, air pollution spikes exhibit special characteristics that constrain their monitoring to be real-time, predictive in a short time, and produce trusted data. This section discusses each constraint in more detail to derive qualitative and where applicable measurable design specs.

3.3.1. Characterization of Air Pollution Spikes

Figure 3. 1 shows both constant-level and peak-like pollution respectively associated with long-term and short-exposure. Long-term exposure refers to the actual contact with a toxic substance for a long time. Short-time exposure refers to the contact of chemical particles within a short time, like 15 minutes. For instance, a peak or spike of pollution can happen during a temporary intensive process within an industry production. A spike is characterized by a relatively short-duration period over which its pollution level is considerably higher than the one of a long-term period within which it occurs. Exposure to repetitive peaks can yield drastic and potentially deadly impacts on exposed living beings.

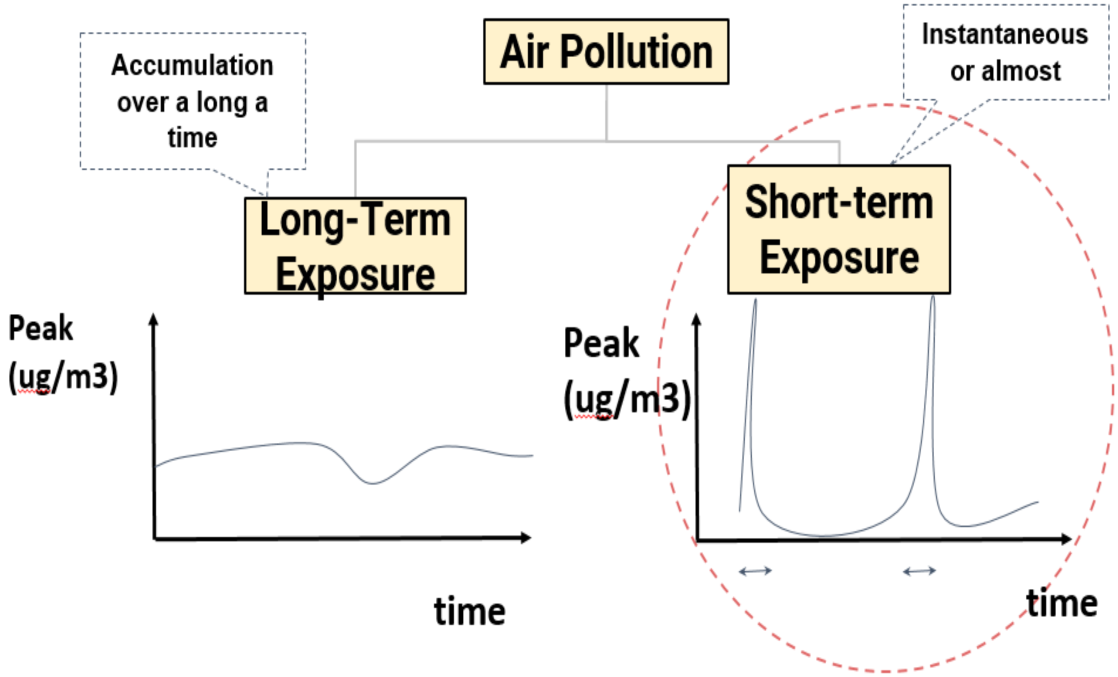


Figure 3. 1 Air pollutant classification: Constant-level peak versus repetitive peak

3.3.2. Main Design Requirements for Efficient Monitoring of Spike Air Pollutants

3.3.2.1. Low latency, Quick reactivity

In traditional cloud-centric architecture, edge sensors must send all collected data to the cloud for processing and analysis. However, given the huge energy consumption of continuous wireless transmission, edge sensors are constrained to operate in duty cycle mode to increase their battery lifetime. Concerning the average time defined in the World Health Organization (WHO) standard for pollution limitation, the smallest sleep mode is 1h. Unfortunately, within sleep time, many short-term pollution peaks may occur and therefore be missed by cloud-centric monitoring. Edge-centric architecture offers an interesting alternative to their cloud counterpart since collected data are first preprocessed and analyzed on the edge sensor, thus removing the unnecessary wireless transmission, like for instance when the air pollution is still relatively constant concerning the previous one. In our scenario, wireless transmission will only occur when a spike has been recorded by the sensor. If not, the cloud analytics will assume that the air pollution stays relatively equal to the end value of the last peak. With this technique, sensing the spike can take place continuously and be reported as soon as it is recorded.

3.3.2.2. Short-term AI prediction

The standards for air pollution limitations define the allowed exposure by considering a minimum average period of 1h in the case of Nitrogen dioxide and up to 24h for other pollutants like particle matter PM2.5. Even though we are concerned with monitoring short-term air pollution peaks, it is still important to comply with long-term limitations defined by the standards as well. Therefore, there is a need to forecast the average pollution over the current hour based on the pattern of a given pollution emitter in terms of producing air pollution peaks. This prediction could for instance warn him/her that the average pollution including detected short-term peaks is likely to go over the allowed average limits within the hour, thus allowing for instance to postpone or slow down the process that is currently creating the air pollution peaks. Short-term AI prediction is currently mainly used in forecasting the electrical energy consumption load to avoid blackouts, especially in the case of time-varying renewable energy production [28].

3.3.2.3. Data transparency and trust

Short-term air pollution peaks are seen as fast-occurring and vanishing events requiring observation at the right moment to not miss them. With long-term average pollution, the level stays stable for a relatively long term allowing anyone to notice and measure it indisputably. Air pollution spikes, on the other hand, are likely to yield disputes between air pollution emitters and air pollution legislation authorities since it is hard to prove posterior. Therefore, the monitoring method must rely on technologies that ensure trust, transparency, and integrity of recorded air pollution data. Blockchain technologies with their decentralized immutable data storage offer the right solution for such types of requirements.

3.4. Convergence of Edge-centric IoT, Blockchain, and AI for Real-Time Air Pollution Monitoring

As explained in the previous section, the inherent characteristics of short-term air pollution peaks require effective monitoring to be low latency, guarantee data integrity, and support short-term prediction. The 3 requirements are respectively fulfilled by edge-centric IoT, Blockchain, and AI. This section presents our IoT layered design for converging the three technologies into a single framework with a particular focus on its application in monitoring air pollution spikes. Figure 3. 2

shows the proposed IoT architecture that distributes the 3 fundamental technologies of our framework in edge, fog, and cloud layers.

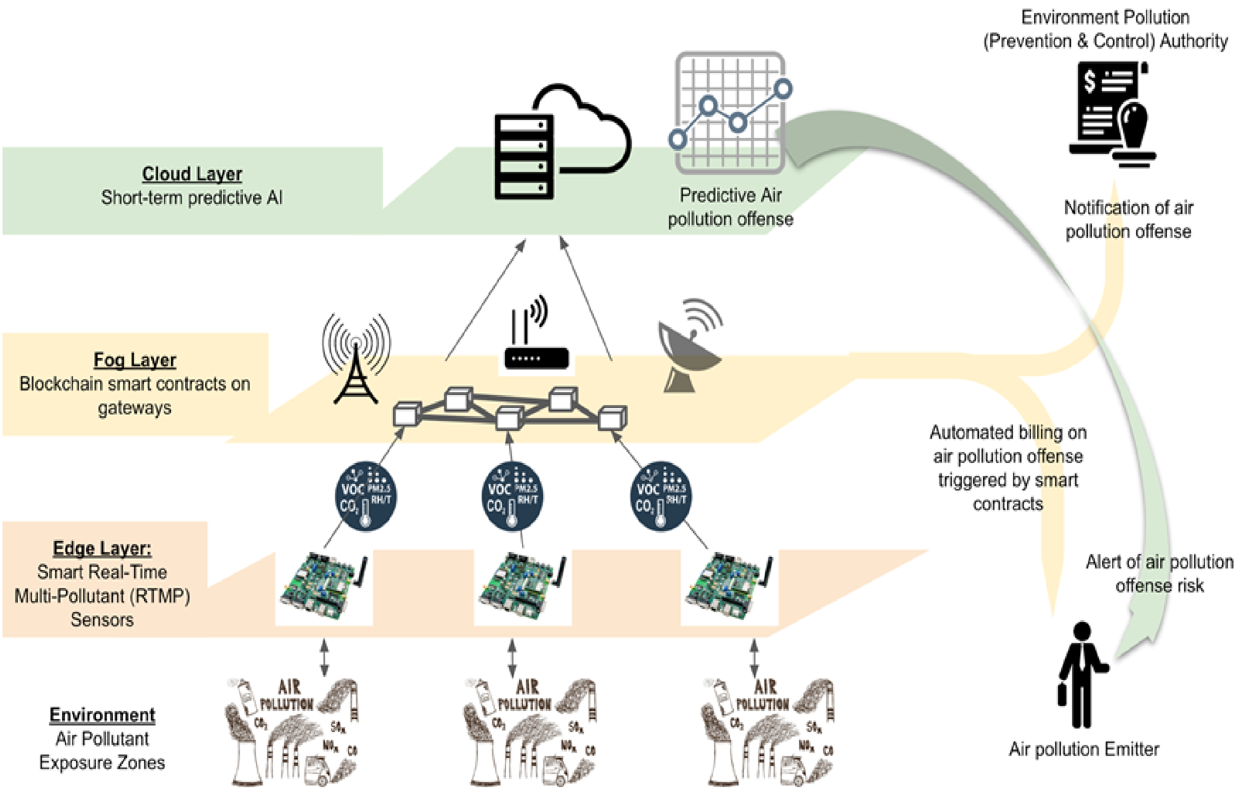


Figure 3. 2 Architecture of htree technologies in design

3.4.1. Edge Layer

The edge layer contains a distributed network of smart multi-pollutant sensors located as close as possible to the origin of air pollution. Those sensors rely on hardware and software co-design to combine the interrupt feature of embedded hardware with a spike tracking algorithm to efficiently detect the beginning of an air pollution peak and start continuous sensing until the peak vanishes. The full signature of the air pollution peak constitutes the content of the blockchain transaction to be sent from the edge sensor to the fog layer distributed ledger. The tokenization of this multi-pollutant sensor asset could create a new type of investment since investors are directly rewarded with a share of the return generated by the tokenization asset, meaning money collected from air pollution offenses.

3.4.2. Fog Layer

The fog layer, which comprises internet gateways such as cellular base stations, and wireless and satellite gateways, provides the infrastructure to support the decentralized blockchain. This blockchain features a smart contract for automating billing for air pollution offenses based on transactions whose contents are air pollution peak signatures collected by multi-pollutant sensors. A smart contract is a set of promises that automatically execute the terms of the contracts. In our case, the smart contract may execute either when the peak exceeds a given threshold, either instantaneously or during a minimum consecutive duration of time. Furthermore, long-term average pollution offenses can also trigger the execution of a smart contract. It may be also interesting to envision a reward mechanism, for instance when a given pollutant emitter controls its pollution below a defined threshold during a period of busy pollution.

3.4.3. Cloud Layer

Using the blockchain as its data store, the cloud layer of our framework will implement analytics for short-term forecasting of average air pollution within the current hour. If there is a high probability that the aggregated pollutants exceed the maximum limits allowed by the standard, an alert will be sent to the air pollution emitters so that he/she can take action to control the air pollution.

3.5. Conclusion

In conclusion of this chapter, the integration of edge IoT, blockchain, and AI enables the development of new systems and services that jointly require low latency, data trust, and prediction of future events. Our research particularly analyzed this integration in the context of monitoring air pollution spikes. This thesis presented the framework of a real-time multi-pollutant (RTMP) sensor featuring a blockchain identity to tokenize pollution offenses. Thanks to short-term prediction AI, the proposed sensor will alert air pollution emitters of potential risks to exceed the average exposure allowed by standards during a given period.

Eric Nizeyimana, September 2024,

Chapter 4. Design of Smart IoT Device for Monitoring Short-Term Exposure to Air Pollution Peaks

4.1. Overview

Air pollution spikes have been causing harm to human beings and the environment. Most exposure to Air pollution spikes has demonstrated a significant impact on mental health, especially children at an early age. That leads to suicide or depression. Previous research concentrated on air pollution in general. Existing monitoring systems do not consider Short-term air pollution peaks. This section presents the co-design of the hardware and software for IoT to monitor air pollution spikes for a short duration in real-time monitoring. The system comprises two technologies edge computing to capture short-term exposure and a mathematical model for distribution in analyzing the captured data. This system ensures the presence of the spikes start and end for each pollutant. Monte Carlo simulation has been used in this research to predict the next spike of each pollutant. Artificial intelligence was used to analyze immutable data for a short-term prediction. After the analysis, legislators based on intelligent contracts created using blockchain to reduce pollution based on its source.

4.2. Introduction and Background

Air pollution is the silent, prolific, and invisible killer for many years ago [1]. Most existing systems for monitoring air pollution are measuring the long-term peaks. Instead, the research shows that the short-term peaks are perilous [2] and can lead to different diseases such as eye and adnexa [3], brain volume, cognitive decrements, dementia development [4], heart, chronic obstructive pulmonary disease(COPD), lung cancer, migraine, acute lower respiratory infections, and stroke [5]. Air pollution is exposed to more than nine to ten children and is stunting their brains, affecting their health [6]. That leads to the problem of mental health, especially in children (brain cell inflammation). Short-term spikes in air pollution are the source of increased hospital visits for childhood psychiatric. The research shows that children from low-income families are more affected, leading to an increase of 44% of those who visited the hospital with suicidal thoughts due to the spikes in air pollution [7] [8].

The spikes of air pollution have more severe effects on the brains of children. The research shows that air pollution spikes can cause mental health, depression, and anxiety. It can lead to the children having a lower intelligence quotient (IQ), and poorer memory, delaying their development, and leaving women infertile earlier. Spikes affect brain chemistry differently; for example, industries and traffic may carry toxins using tiny passageways and then enter directly into the brain [7] . In 2020, spikes increased higher than five years [9]. But until now, minor effects of short-term exposure to air pollution are known.

Spikes of air pollution are damaging the future generation of humans, and emitters do not condemn the creation of that harm. Governments have tried to prevent air pollution in general, but the research shows that none of them has been viewing the spikes as a dangerous and long-lasting killer of children. The research suggested that legislators should protect children's exposure to air pollution to advance the initiative for their public health [10].

This chapter is composed of the following sections: Section 4.2 of Background covers all literature reviews related to the monitoring system of Air Pollution. Section 4.3. is about the co-design of hardware and software for the prototype monitoring of the spikes. Section 4.4. is about performance analysis and simulation, where the analysis is made for some data and tools used to do simulation, and the results are shown in that section. The last section 4.5. is the conclusion, which summarizes the section and the proposal for future research.

The section deeply explains the previous research and enumerates some challenges that are still in this area that can be solved using this research.

WHO has put the Global Air quality guidelines in different years to prevent air pollution in general, but there are no measures taken for spikes [11]. Many people are victims of air pollution, and emitters of pollutants are not charged for anything because no one knows the air pollution they produce. Some measures have been taken [11], but they do not regularize the correspondence between emitters and victims of spikes. When these spikes continue to be repetitive, they cause more problems in health [12].

Many people live in big and small cities. Developing countries used to have high populations exposed to air pollution. It is also where most sources of pollutants are found [13]. Once the spikes appear in the cities, they affect a large population [14]. Spikes can appear anytime, so this may come from different sources. If the spikes are not monitored, they can affect the lives of human beings, as explained above. Spikes occur in a short time, and most existing monitoring systems for air pollution cannot recognize their appearance. Children are the most affected, primarily their mental, leading to their future loss [7]. Emitters of spikes may not even know how they affect human beings' health [15].

Authorities are oriented toward monitoring air pollution in general [16] [17]. Instead, spikes are affecting the future generation and the population as well. The source of spikes allows the identification of emitters, and then authorities may take advanced majors accordingly. Victims of these spikes are more in danger once they are repetitive. That may lead to many unexpected severe health problems.

The author of [18], proposed a system that can monitor the spikes in air pollution and predict the next spike using road management data.

Air pollution spikes have a short lifetime, requiring monitoring in a smaller time resolution [19]. These spikes need to be monitored at each appearance so as not to affect the living. Spikes need particular ways of monitoring them that differ from the existing methods. They appear in a short time, and then they disappear. If they do not monitor their appearance, they mix with other collected results of pollution and then may result in the average instead of the over-level for pollutants. Spike monitoring can enable counting all spikes passed, predicting the following occurring spikes in the system. Once they occur from their sources, these unexpected pollutants may damage many things because they were not prepared before. They do not last for an extended period, leading to the big mess of not monitoring them. Spikes generated due to some occurred events planned before, but no analysis of the effect may cause. Also, spikes may occur due to unexpected events from the environment or any other source without any prior planning of the event. Spikes need a real-time and a low-time resolution to react to the effects that may occur due to their presence [20] [21], sometimes to the loss of life [22].

Most existing systems for monitoring air pollution are based on cloud-centric architecture [23] [24]. These systems measure air pollution with long-term exposure. The average of peaks for air pollution in a specific time is considered the result of a monitored place. That is because of allows sensors to capture information during a specific time and wake up to send the data to the cloud, known as duty cycle mode (taking a long sleep period to save the battery energy). That is for saving battery life during the wireless communication mode of sending data to the cloud. In the design of the sensor node of IoT applications, battery life is one of the critical parameters to consider. The reporting of collected data to the remote centric-cloud architecture of air pollution has a low frequency for at least 1 hour to extend the battery lifetime.

The centric-cloud architecture uses wireless communication for transferring data from sensors to the remote. That leads to high energy consumption, and the sensors sleep within a certain period of collecting data and storing them locally. That makes sensors monitor long-term average peaks instead of capturing all peaks [25]. That leads to the miss of monitoring short-duration peaks. These spikes may appear periodically or not. The air pollution peak average threshold may exceed for specific pollutants, and the system may not be aware of that unexpected change. No system can capture spikes for a short duration from the existing cloud-based systems.

Cloud-centric has failed to monitor spikes because of transferring data by waking up the sensor. The cloud-centric architecture collects data on air pollution using sensors at data gathering. It uses wireless communication to send the data to the cloud [26, 27]. Then, the system does the data management for the given application in the cloud. In the first phase of data gathering, sensors collect information related to air pollution. This information is transferred to the cloud through the wireless communication channel. This communication channel consumes high energy [28]. The last phase is data management, which analyses, processes, and stores data in the cloud. These data can be used to predict air quality [29].

The cloud-centric architecture also has a latency problem due to transferring the data after a specific time. These systems also take time to react to the processed data [30]. Predicting the possible air pollution event may take longer as data processing is based on the cloud, not on edge. There is a need for data transparency and trust, and this may be difficult for the data passing in the network without additional security measures.

Therefore, there is a need for an edge-centric system to monitor short-term peaks in air pollution. This thesis helps to understand the design of the edge-centric smart sensor to monitor air pollution by waking up the sensor once the air pollution variation attained the given threshold.

There are not many several systems developed to monitor and predict air pollution spikes. Artificial Intelligence technology is the predictor developed by a new wireless company, and that system can predict the following levels of air pollution within an hour. This system uses AI for analyzing weather measurements, images of CCTV cameras, air pollution sensing devices, Bluetooth, and history readings. The system links the existing real-time data to predict the next coming hour for traffic jams and air pollution. This predictor is accurate at 97%. It has been tested for implementation in some cities like Wolverhampton [18]. It is excellent and friendly to the existing technologies, but it doesn't take pollutant data on the roads or nearby since it is linked to the load management system. This leads to the lack of identifying the source of each pollutant and the quantity. The predictor predicts air pollution in general but doesn't identify spikes that come and may arrive.

Following the previous works that researchers have done, the existing systems only measure a few pollutants. Most monitor air pollution at the cloud-centric, leading to latency, security, cost, and control problems. The existing system woke up sensors periodically, leading to the danger of an unexpected increase in pollutants. Existing systems haven't mentioned the identification of spikes, and the time stay based on their appearance level.

This research takes all six primary pollutants as explained by WHO in [11], and it can be installed and not based on historical readings.

4.4. Hardware and Software Co-design

4.4.1. Improve IoT Energy Management

In designing the embedded system for performing a real-time environment, there is the issue of the increase of power dissipation. IoT hardware design increases power dissipation from the real-time application and the device for the best performance. The problem was created during the deployment of the number of transistors comparable with the power consumption. There are two causes of power dissipation in designing lower-power IoT systems. The first one is when the power dissipation for each transistor increases with impact to the increase of density gate, which implies the increase of power density for the whole system. The second is the increase in the frequency of IoT systems for better performance.

The power dissipation problem is improving IoT energy management by waking up the device using analog interrupts.

Energy management is still a crucial problem in today's sensors [31]. There is a need to continuously allow the sensor to stay in energy-saving deep sleep mode to solve that issue. The system needs to wake up on measurement appearance with low energy consumption. Since the energy consumption implies a decrease in battery life, the system should monitor sensors connected to use little energy.

The system stays asleep most of the time and only wakes up for the threshold's quick and effective measurement. The CPU of the system uses much energy in comparison to the rest of the parts. That means that reducing the CPU system's busy time is the best way to reduce the consumption of energy.

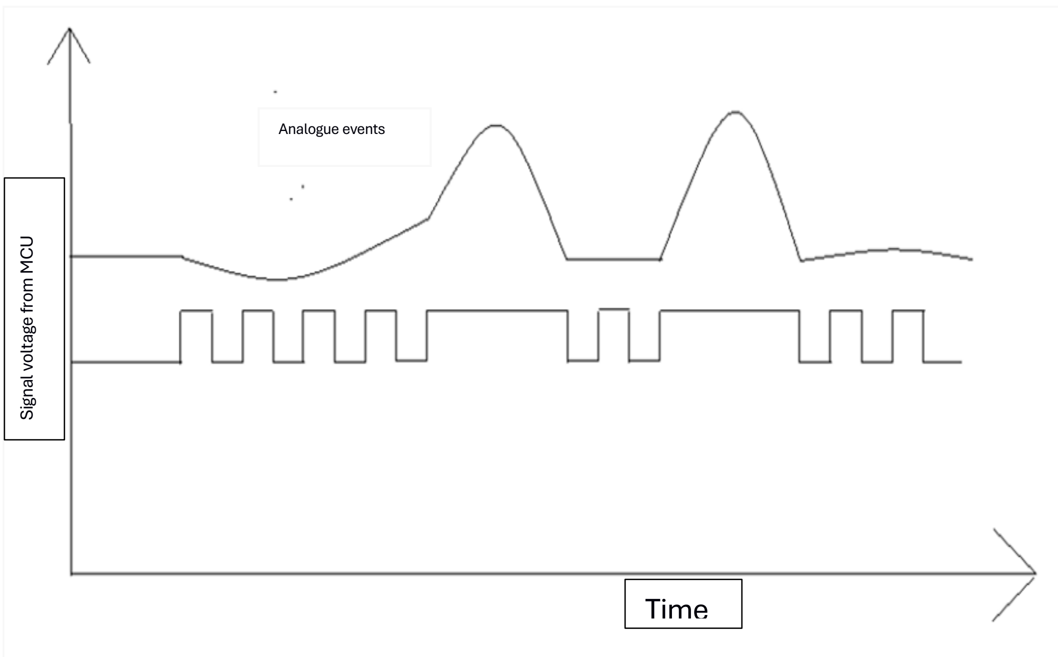


Figure 4. 1 The system Wakes up Periodically to Detect Events.

Figure 4. 1, the system wakes up periodically to detect events, making the CPU continuously active. The sensors capture data from the environment and send these data for processing. The analog event creates a signal which is transformed directly into a digital signal.

Figure 4. 2 gives the intention to look at the signal after using analog interrupts. The sensors can record all analog events, and once the threshold passes, the system wakes up for recording and processing.

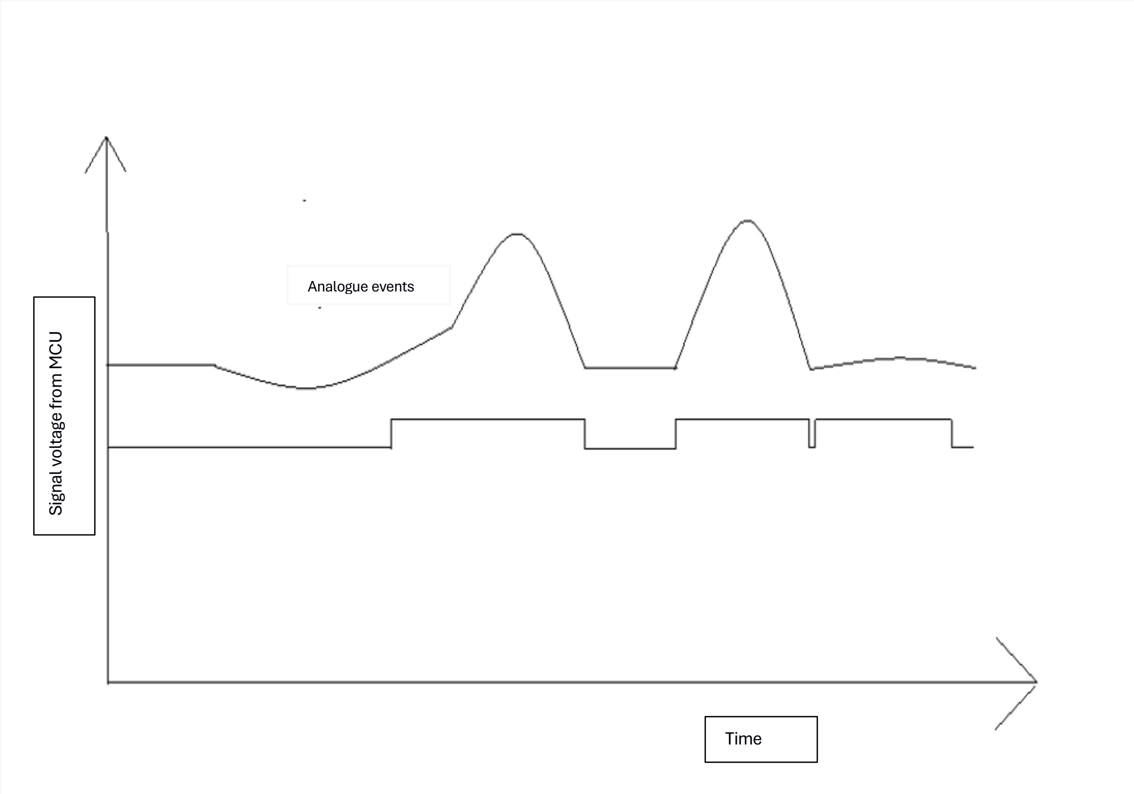


Figure 4. 2 The System wakes up on the threshold

There are two options for using analog interrupts: The first is ADC wake-up and the second is an external op-amp-based voltage comparator.

4.4.1.1. ADC Wake up

Sensors convert analog measurements into electronic signals. The ADC (Analog to Digital Converter) converts the produced analog signal to a digital signal using the frequency sampling mode based on the Nyquist theorem. Interrupts that are alerting electronic signals are sent to the Microcontroller Unit (MCU) processor, which may come as an external part of the internal peripheral or external one.

The ADC (Analog-to-Digital Converter) component designed using the Proteus simulator is intended to monitor air pollution by detecting spikes in analog signals and converting them into digital signals that trigger specific actions, such as switching LEDs on or off. The design criteria for this ADC system include the following:

- **Components selection**: one ADC0804 Integrated Circuit, eight LEDs, one resistor of 1k, one variable resistor or potentiometer, one push button, one wero board, one nonpolar capacitor with 150pf, and some jumper wires.
- **Threshold-Based Switching Mechanism**: a system is designed that switches ON and OFF based on the voltage once the input exceeds the threshold. Then, it helps to monitor all events that come and exceed the threshold. That is useful in air pollution monitoring based on spikes only.
- **User Interaction**: The button is activated by the measurement of the node sensor for air pollution. The input signal is an analog signal generating the output that can switch on LEDs.
- **Visual Feedback**: The LEDs serve as a straightforward visual indication of whether the input signal has surpassed the threshold, making it easy to see at a glance whether a pollution spike has been detected.

Figure 4. 3 shows that the data are generated from the physical environment, and then the sensor accepts these analog measurements in the form of analog signals. The analog signal transforms into a digital signal and is then sent to the processor. At the input, the environment creates a signal using physical quantity. The sensor takes that signal and makes it in the presence of an electrical signal. The signal is in analog form and needs to transform into digital form, and then using the ADC tool; it generates the digital signal used to monitor air pollution using the given threshold.

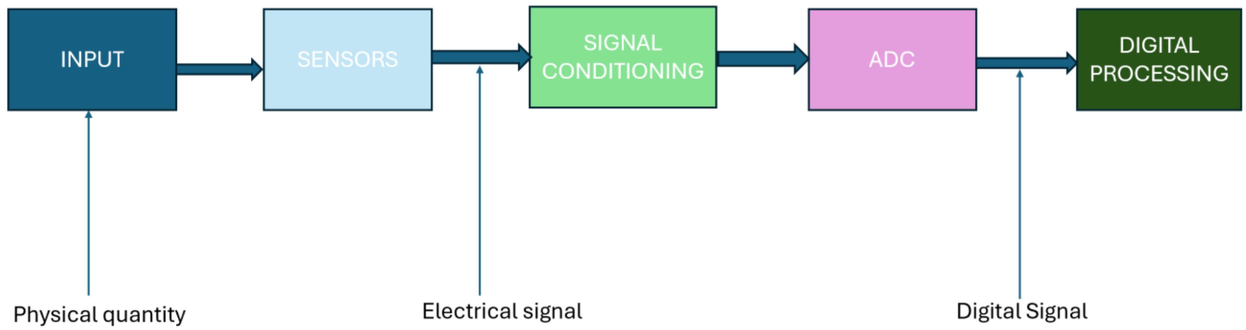


Figure 4. 3 Measurement to signal processing

This chapter applies the sampling of analog signals, and the system acts based on the threshold. That should be done using ADC to trigger timers precisely. That uses many MCU resources, which leads to high power consumption since timers must be active to perform ADC. Another methodology is not to use timers and allow signals to be monitored continuously by the ADC, which consumes a high-power consumption.

The solution to be adapted that may not consume huge amounts of power is to integrate ADC in the MCU without dependency on the CPU (Central Processing Unit). That allows the CPU to disable all clocks except the one of ADC. Then the ADC wakes up the CPU and other parts of the MCU by using the logical conditions. The ADC creates interrupts to wake up the rest of the system by referencing the configured threshold.

The ADC wake-up uses the voltage comparator to activate the CPU and the system. The reference voltage V_{ref} is compared to the input voltage Vin for deciding whether to wake up the system or not. The V_{out} is in digital mode, and from there, the decision to wake up the system is taken. Since that is on the sensor by detecting the measurement of the event to wake up the system, there is the optimality of this strategy because no loss of data appears and the optimization of the sleeping time.

4.4.1.2. External op-amp-based voltage comparator

The other way to wake up the system is to use an external operational amplifier based on the voltage comparator. This way requires extra resources to add to the system. Adding this wake-up circuit to the sensor node decreases the average power consumption, but also it may create a loss of information since the original signal was amplified.

The external op-amp compares one analog voltage level to another and generates output based on the comparison. It detects the voltage from measurements and then switches from the sleeping mode of the system to an active mode. The switching time of the op-amp voltage comparator slows the system even though it operates on analog voltage.

The op-amp voltage comparator uses input, amplification, and output terminals. It uses negative feedback voltage, which leads to compensation capacitance to prevent oscillation in that integrated circuit. That creates an inside power dissipation, which may increase the temperature of the chip and the self-heating.

The operational amplifier voltage comparator may present an error voltage called the input offset voltage caused by the characteristics of transistors of each terminal or by the input bias current.

Figure 4. 4 shows of the op-amp voltage comparator with its five terminals.

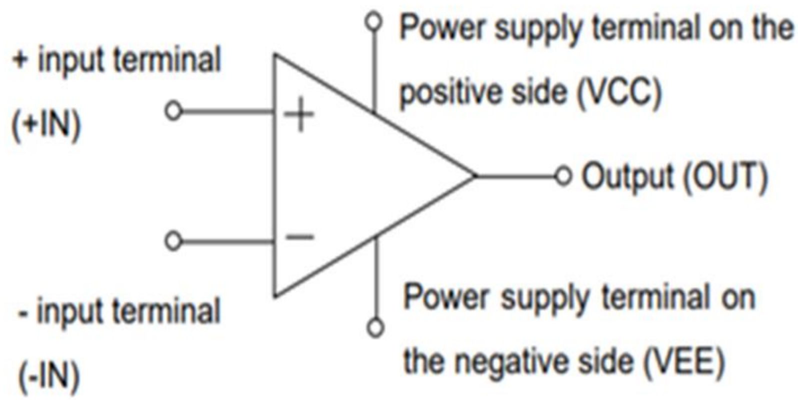


Figure 4. 4 Op-amp voltage comparator

Therefore, based on the above comparison of ADC wake-up and external op-amp voltage comparator wake-up, this research suggests using ADC wake-up to activate the system from sleeping mode to active mode. The measurements are taken from the environment and create analog input to the sensor node, and that analog input changes to an analog signal with a certain amount of voltage. Then the analog signal needs to transform into a digital signal using an ADC converter, and during that conversation, the ADC decides if it wakes up the whole system based on the comparison of the input voltage and the reference voltage. The system needs to identify the starting voltage above the threshold voltage and record these values. The following subsection explains how the system used to pick these signals that attained the threshold.

4.4.2. Peak Digital Signal Processing

Measurements captured from the environment need to be processed and analyzed. Once the ADC wakes up, it gets an analog signal and converts it to a digital signal, quickly processing it scientifically. The system is woken up based on the data that exceeds the predefined threshold, allowing the system to record that event.

This research is working on air pollution spikes. These spikes are identified based on the minimum predetermined value of pollutants. WHO has defined each pollutant's threshold as shown in Table 4. 1. These values are measured in micrograms per cubic meter.

Table 4. 1 Air pollutants

Pollutant name	Minimum Concentration (μg/m^3)
Particulate Matter $PM_{2.5}$	35
Particulate Matter PM_{10}	70
Carbone monoxide/ Carbone dioxide (CO/CO_2)	1000
Nitrate Oxide (NO_x)	80
Sulfur of Oxide (SO_x)	50
Ozone (O_3)	120

This research detects a peak in a signal and measures its position, height, width, and/or area. When the sensor node identifies the signal that exceeds a threshold value of any type of pollutant, the system starts to record the event, and when it attains the peak, it starts to decrease, going to the value that should always be less than the threshold. The first derivative of the peak is applied to downward-going zero-crossing (threshold) at the maximum of the peak. Since the signal may have noise from measurement due to the environment, this can lead to false zero-crossing. Therefore, the smooth technique can detect only the desired peaks and ignore peaks that are too small, too wide, or too narrow.

Once they become high frequency, air pollution spikes (peaks) cause mental health problems that can lead to hypertension, suicide, and heart diseases. These peaks imply the concentration of pollutants concerning the given time. If not reduced, that concentration causes health problems compared to normal pollutants that do not pass the threshold.

The input signal is taken in the window size measured based on the length of the signal above the threshold. The height of that signal is also identified.

The digital signal processing from ADC is set low or high. We examine all signals with high since they are above the threshold. Signals which are less than the threshold are identified as low. Then finding the peaks in the given signal that we can call X describes all points above the threshold. Each peak has its amplitude or the height of the signal.

Figure 4. 5 is for peak detection of digital signals with a height of 1 and a length of the period of 500 microseconds.

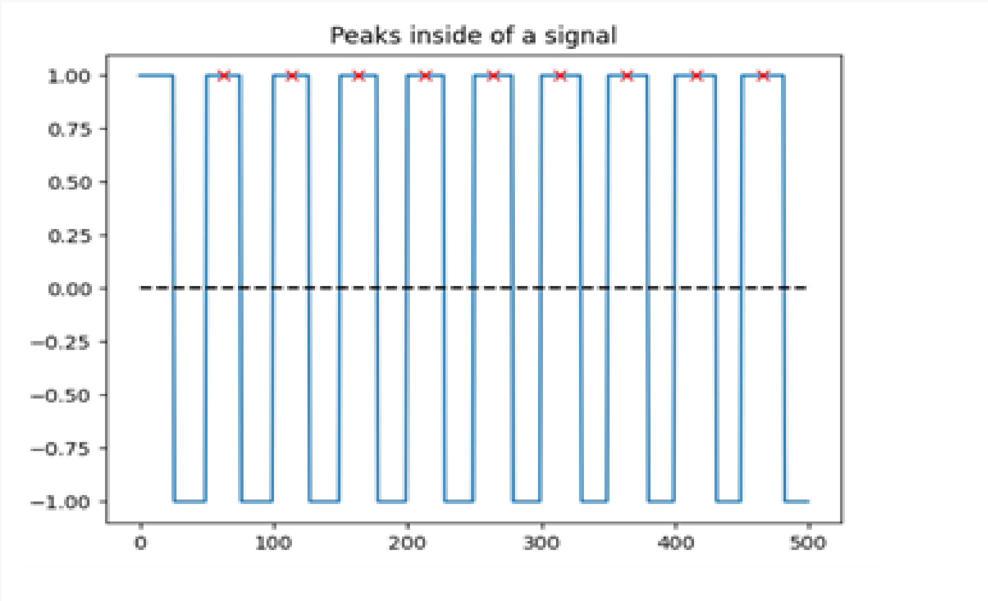


Figure 4. 5 Peaks detection

From the input signal of the sensor node, ADC wakes up and activates the system once the peak has appeared. The probability of obtaining a peak at a particular point t_i of the input signal from the environment depends on its incoming voltage and the given standard threshold voltage t_v.

$$P(\text{Spike at } t_i | \text{ Spike at } t_v) = \int_{v \in Ci} \mathbf{G}[\mathbf{V}(\mathbf{t})] \qquad (4.\ 1)$$

With C_i, the set of possible voltages can rise in the interval i. V presents the voltage signal. And G[V(t)] is the Gaussian distribution.

Once peaks are detected from the given interval time and their location, the system is applied to collect these peaks. These peaks are stored in the array of integers for future analysis using

Machine Learning. The system can analyze the collected peaks for notifying the appearance of pollutants and the expectation of the next peak.

4.5. Performance Analysis/Simulation

4.5.2. Distribution Patterns of Air Pollution Spikes

Air pollution spikes are coming from the increase of unexpected pollution generated by emitters. All these spikes are from the environment and can be distributed in the atmosphere. The system for monitoring spikes can capture distributed pollutants either directly or indirectly from the source.

Direct spikes are captured by the system and then analyzed without adding the environment to it and for example, having the sensor node connected to the place generates pollutants. The indirect spikes are those that pass into the environment and meet with other pollutants before being measured by the system.

The system accepts measurement in any of three patterns or their combination. Spikes can be presented to the system either: uniform random and/or clumped. Uniform spikes are these spikes that come within a given period. They occur periodically in the system. These spikes make it easy to predict the next peak. These peaks can appear in different sizes and densities.

Random spikes are these peaks that are entered into the system randomly. These peaks can be very harmful since they are not easily predictable. This research suggests using a certain period to analyze all peaks appearing, and then using Machine learning, can predict the next peak. Clumped spikes are predicted or unpredicted peaks with a heavy density. These peaks are perilous, and they need profound observations to analyze their prediction. The nature of the air pollution environment can have all these above patterns of distribution of spikes. All these spikes are based on the period to predict the next appearance of the peak.

The distribution model has been used to predict the next within a specific period. Since the generated signal from the ADC wakes-up converter is a digital signal, it has discrete values. The Poisson distribution is used in this Research to model the arrival rate of spikes in a specific fixed interval of time. The performance parameters are based on the mean of signals in a period λ and the number of spikes k.

Let λ be the parameter greater than 0 and let distribution k = 1,2,3, … n be the appearance of spikes in the input signal to the sensor node; in other words, k is presenting a discrete random variable counted. The probability density function (*pdf*) is used to specify the random variable's probability of being in the range of the values.

Then the *pdf* that a Poisson random variable X with the mean λ is equal to a given by the formula.

$$pdf = P(X = a) = \frac{\lambda k e^{-\lambda}}{k!} \qquad (4.2)$$

Where e is a constant of approximately 2.71828.

The *pdf* gives the probability of getting spikes each time by using the meaning of the spikes and the number of spikes counted.

The system can identify spikes and predict finding peaks in a given time. Most existing systems for monitoring air pollution are using cloud-centric duty cycle mode. The sensor collects data related to air pollution while it is in sleeping mode to save the battery's lifetime or the harvested power for energy conservation. The sensor only wakes up after a specific period to transmit collected data to the cloud-centric for analysis.

The duty cycle D can be defined as a ratio of pulse width (PW), a busy time, and the total period of T of the signal and then expressed in percentage.

$$D = \frac{PW}{T} \times 100 \qquad (4.3)$$

Therefore the 60% duty cycle means that the signal is on 60% of the time but off 40%. That implies the power consumption in recording data. There is a need for much energy during the transmission of the data to the cloud-centric server. This energy consumption reduces the sensor battery lifetime or the harvest power storage of energy.

The solution for optimizing the use of sensor energy or harvest power is edge centric HW/SW Codesign smart sensor. That allows the analog interrupt to wake up the sensor once the coming signal voltage is higher than the threshold voltage. Only the system is active in collecting spikes for quick analysis of data. Once the spikes are over, the intelligent sensor goes back to sleep mode.

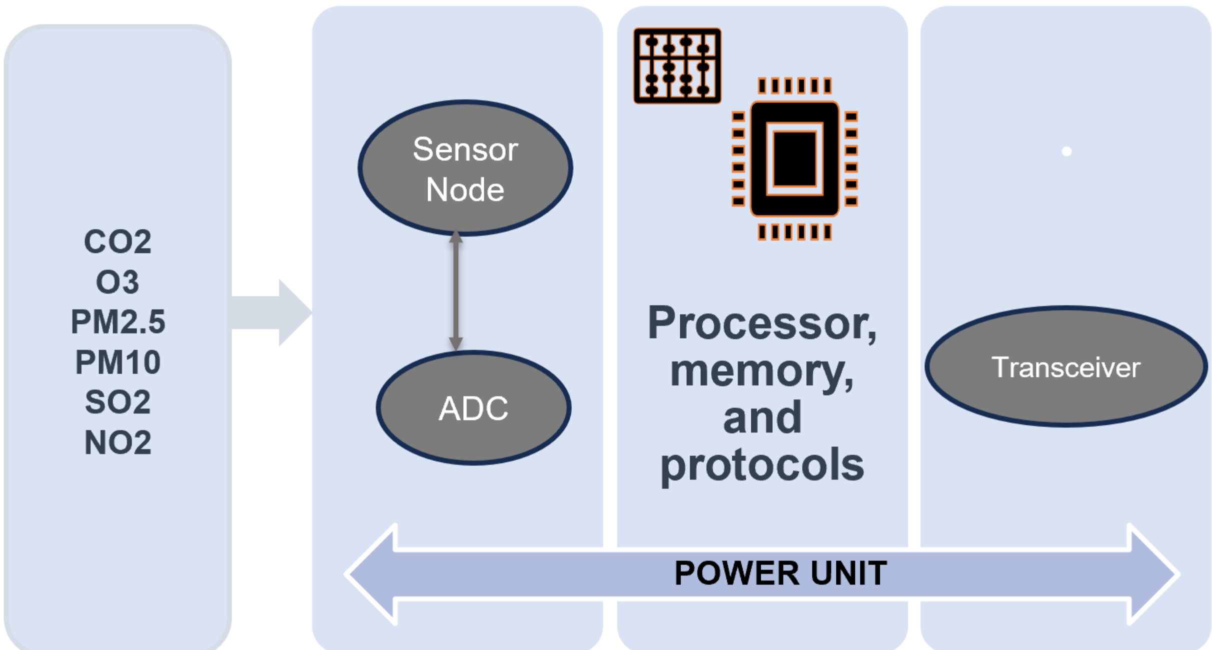


Figure 4. 6 Edge-centric HW/SW codesign system

The Figure 4. 6 describes the edge-centric HW/SW co-design system. The measurement from the environment is those pollutants CO, CO_2, $PM_{2.5}$, PM_{10}, SO_X, and NO_X. Once one of these pollutants sends a value more significant than the threshold defined in the table [1], the sensor node will send all signals to ADC, which wakes up the processor, memory, and protocol for analyzing the coming signal since it is a spike. The power unit is there to empower each device. That reduces the power consumption since it is waking up for recording spikes.

In edge-centric HW/SW co-design, data analysis is performed locally, and there is no transmission energy of a short periodical time. Only transmission can be done once in a while for further analysis. The edge-centric smart sensor is a real-time monitoring system for air pollution spikes and can react to its appearance. From those spikes, it does analysis locally, and the decision is taken quickly.

The cloud-centric is still needed to analyze big data collected by sensors, while edge-centric can be considered an operator of instant data. At the edge, centric analytical tools and AI tools are nearest the system, implying operational efficiency. Security and privacy are strong at the edge-centric smart sensor. This system is reliable. Since one node can go down and is unreachable, the other system parts continue to operate. The speed of data at edge computing implies analytical, and computational resources to the end-users, bringing quick responses and applications.

4.5.2. Performance Metrics

The energy consumption at the edge-centric HW/SW co-design smart sensor and cloud-centric can be distinguished in the below metrics:

- ***Throughput:*** output at the edge is generated in real-time while data is transmitted in the cloud, which consumes much energy.
- ***Collecting data:*** at this stage, edge computing wakes up only during the collection of spikes while the cloud uses a duty cycle which implies power consumption.
- ***Processing:*** at computing only, the system wakes up on the threshold, while the CPU and other parts of the system operate periodically for cloud-centric—the probability of identifying the subsequent spikes at edge computing within the length of the interval of spikes.

Table 4. 2 Comparison of the existing system and our proposed system

	Throughput	**Data collection**	**Processing / Performance**
Designed system	Save energy consumption over 40%	Spikes to wake up the system	Real-time processing and prediction of data beyond the threshold depends on the wake-up frequencies.
Existing systems [32, 33]	Energy consumption	Periodically wake up	It takes periodic time to predict the next spike.

From Table 4. 2, the designed system performs better in all performing metrics. The designed system can save energy consumption at 40% on the throughput metric since the existing systems (mostly cloud computing systems) are using a duty cycle. On the second metric of data collection, the designed system woke up on the appearance of spikes, and that can lead to quickly identifying the spike while the existing systems wake up periodically and can miss some spikes, which may be dangerous. Lastly, the processing metric is so quick at the designed system while existing systems take enough time to process and predict the next appearance of the spike in air pollution.

Peaks can be harmful to human beings, and there is a need to monitor them. For cloud computing, some peaks may be lost during the sleep mode of the sensor at the sleeping mode of the sensor. At the sleeping mode of the sensor, all peaks arrive and can be combined with the whole signal to present the mean of the whole period. On edge-centric, the system is woken up by spikes in the input signal.

4.5.3. Monte - Carlo Simulation

This section uses Monte-Carlo simulation as a mathematical technique used to estimate the probability of possible outcomes in a process that cannot be predicted due to its uncertain appearance.

It is based on making a computational algorithm to find the numerical results of repeated random sampling. These uncertainty events can be predicted and forecasted using the Monte Carlo technique to model them.

This research uses Monte-Carlo simulation to predict the subsequent spikes using their probabilities of occurring. As our data are stored discretely after the ADC converter, we estimate the probability of occurring in a specific period.

Let's use the same example of PM2.5 for its Poisson distribution, which was 7.1%. Then that means in the interval of a period there is a 7.1%. Probability to find the spikes of PM2.5. The figure below is for particulate matter data, and it uses the synthetic data generated mostly.generate. The mean λ of the data is 29, and the appearance of spikes is 27. Then the probability of getting the spikes is 0.071.

The performance of the proposed system has good accuracy since it can identify each appearance of the spike of each pollutant. Most other existing systems measure air pollution in general and are not specific on the spike prediction of each pollutant.

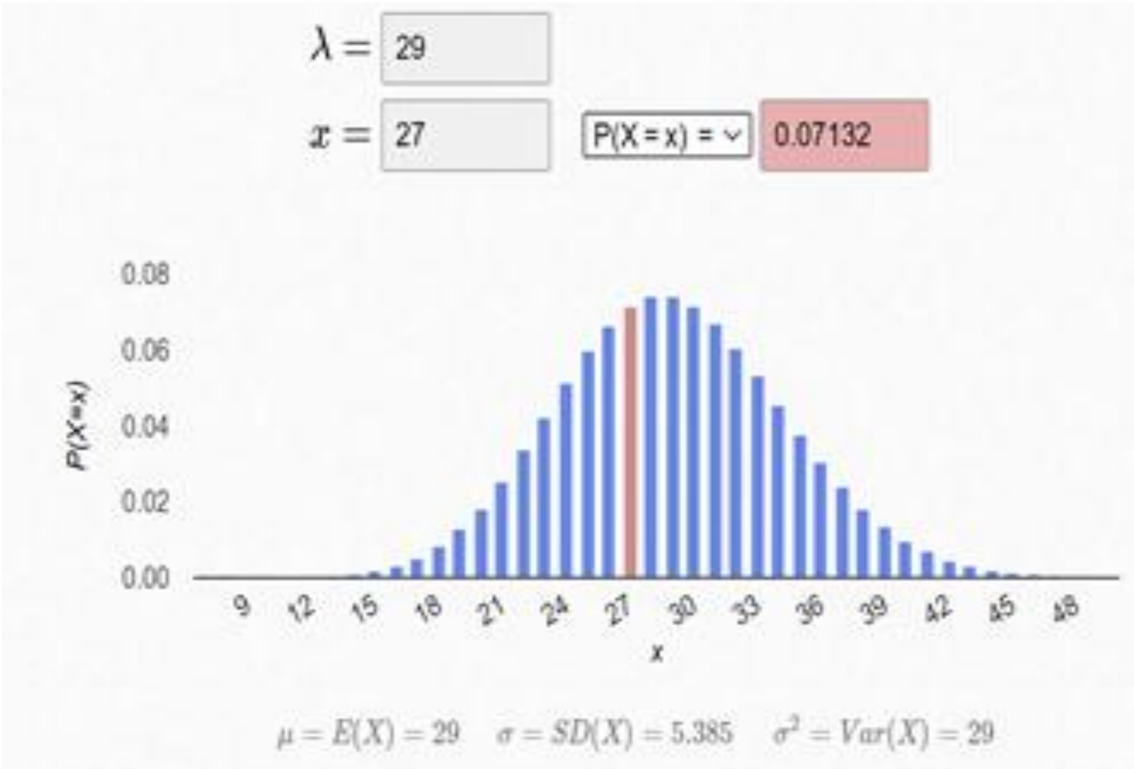


Figure 4. 7 Probability of the following peaks for PM2.5

The signal length is sampled at 30 samples for the whole period, and the probability of finding the next peak of PM2.5 is 7.1%, as shown in Figure 4. 7

4.6. Conclusion

This chapter designed a hardware-software smart IoT device to monitor short-term exposure to air pollution peaks. Spikes are causing significant health problems, especially in children, leading to mental problems, suicide, stroke, heart disease, and lung. This research improves the IoT energy by waking up the system through the appearance of digital signals using ADC wake-up. The designed system performs better than the existing system through the performance metrics, as explained in Table 7. 1. The chapter explained the finding of peaks that are stored in the array. The mathematical model was generated using Poisson distribution to find the appearance of peaks. Monte Carlo has been introduced to predict the next coming peak. The prediction showed that PM2.5 could be predicted at a 7.1% probability of spikes appearing. This probability is high since it showed that in the appearance of spikes, there should be a 7.1% of PM2.5.

The system and authorities analyze the collected peaks to compensate for the peak emitters. As peaks are dangerous to health, there should be a proposal of fining people accordingly if they exceed the threshold. The hardware-software co-design generates a dataset of spike signatures that will be used by machine learning for the next chapters.

Eric Nizeyimana, September 2024,

Chapter 5. Prototype of Monitoring Transportation Pollution Spikes Through IoT Edge Networks

5.1. Overview

Air pollution is a critical problem in densely populated urban areas, with traffic significantly contributing. To mitigate the adverse effects of air pollution on public health and the environment, there is a growing need for real-time monitoring and detection of pollution spikes in transportation. This Research presents a novel approach to using Internet of Things (IoT) edge networks for real-time detection of air pollution peaks in transportation, specifically designed for innovative city applications. The proposed system uses IoT sensors in buses, cabs, and private cars. These sensors are equipped with air quality monitoring capabilities, including the measurement of pollutants such as particulate matter (PM2.5 and PM10), nitrogen dioxide (NO_2), ozone (O_3), sulfur dioxide (SO_2), and carbon dioxide (CO_2). The sensors continuously collect air quality data and transmit it to edge devices within the transportation infrastructure. The data collected by these sensors are analyzed, and alerts are generated when pollution levels exceed predefined thresholds. By deploying this system within IoT edge networks, transportation authorities can promptly respond to pollution spikes, improving air quality, public health, and environmental sustainability. This Research details the sensor technology, data analysis methods, and the practical implementation of this innovative system, shedding light on its potential for addressing the pressing issue of transportation-related pollution. The proposed IoT edge network for real-time air pollution spike detection in transportation offers significant advantages, including low-latency data processing, scalability, and cost-effectiveness. By leveraging the power of edge computing and IoT technologies, smart cities can proactively monitor and manage air pollution, leading to healthier and more sustainable urban environments.

5.2. Introduction and Background

Air pollution is one of cities' biggest environmental and public health challenges worldwide [1, 2]. The adverse effects of air pollution on human health, ecosystems, and the climate are well-documented, making it a critical issue that requires immediate attention [15]. In urban areas, transportation is a significant contributor to air pollution, releasing various pollutants such as particulate matter (PM2.5 and PM10), nitrogen dioxide (NO2), ozone (O3), Sulfur dioxide (SO2), carbon dioxide (CO2), and carbon monoxide (CO [52]). This research deals with air pollution in transportation, especially in urban areas.

Air pollution is characterized as a silent but deadly menace, leading to an annual death toll of seven million individuals [53]. Alarmingly, approximately 90% of the global population is consistently exposed to its harmful effects[53]. Both long and short exposure are dangerous to our health[54]. Many associated diseases with air pollution include Respiratory and cardiovascular [55] Air pollution exerts significant adverse effects on human health, manifesting in various forms, such as respiratory diseases, including asthma and lung cancer, as well as cardiovascular dysfunctions and malignancies[56]. Research has identified a strong correlation between male infertility and exposure to air pollution, and it has established a link between air contamination and an elevated risk of immune dysfunction, neuroinflammation, neurobehavioral hyperactivity, criminal behaviors, premature aging, Alzheimer's disease, and Parkinson's disease[56]. Notably, traffic-related air pollutants have been implicated in skin aging and the development of pigmented facial spots.

Furthermore, air pollution is associated with eye irritation, dry eye syndrome, retinopathy risk, and adverse ocular outcomes [56]. Chronic exposure to air pollutants during pregnancy is linked to negative effects on fetal development, including low birth weight and stillbirth.

Additionally, air pollution is recognized as a significant contributor to the increased prevalence of allergic diseases in children [56]. In implementing this project, the expectation is to decrease the deaths from air pollution. The research elaborates on policies that can be adopted to enhance the well-being of individual car owners and the broader community, ultimately contributing to an increase in the overall life expectancy of the global population.

The adverse impacts of air pollution affect pregnant women, children, elderly adults, and individuals residing in rural areas [55]. Numerous studies have linked exposure to air pollutants with impaired brain development, learning disabilities, and children's children's lower intelligence quotient (IQ) [57]. Children's developing brains are more vulnerable to the harmful effects of air pollution than adults [57].

Here are some ways in which air pollution can impact the IQ of children:

- Neurodevelopmental Effects: Exposure to air pollutants, especially fine particulate matter (PM2.5) and nitrogen dioxide (NO2), has been associated with neuroinflammation and oxidative stress in the brain [57], [58], [59], [60]. These inflammatory responses can disrupt normal brain development, affecting cognitive functions and potentially leading to lower IQ scores [60].

- Cognitive Decline: Long-term exposure to air pollution during childhood has been linked to cognitive decline later in life[61][62]. This decline may manifest as memory deficits, reduced attention span, and impaired decision-making abilities, all of which can influence IQ scores.

- Impact on Academic Performance: High levels of air pollution have been associated with reduced academic performance among children. Students exposed to chronic air pollution may experience difficulties learning and achieving academic milestones, affecting their IQ development.

- Structural Brain Changes: Studies have shown that exposure to air pollution can lead to structural changes in the brain, including alterations in the hippocampus and frontal cortex, areas crucial for learning, memory, and IQ development [63].

- Prenatal Exposure: Prenatal exposure to air pollution can also significantly impact children's cognitive development. Studies suggest that exposure to air pollutants during pregnancy can lead to lower IQ scores and increased risk of cognitive impairments in children [57].

- Socioeconomic Disparities: Children from low-income communities often face higher levels of air pollution due to proximity to industrial facilities and traffic congestion. As a result, they may experience more significant cognitive impairments, exacerbating existing socioeconomic disparities [64].

It's important to note that while there is a growing body of evidence linking air pollution to cognitive deficits in children, the specific mechanisms and causal relationships are still subjects of ongoing research. Furthermore, individual susceptibility can vary based on genetics, socioeconomic status, and preexisting health conditions [64], [65]. Efforts to address this issue include advocating for stricter air quality standards, promoting cleaner transportation options, and raising awareness about the potential risks of air pollution, especially for vulnerable populations like children [66], [67]. Due to the adverse effects of air pollution, vulnerable individuals are disproportionately impacted. Children, the future of our society, are particularly susceptible to the adverse effects of air pollution, which can affect their cognitive development. As a result, it is imperative to implement measures to safeguard their health and well-being, and this research will do that.

In urban environments, the escalating issue of air pollution poses substantial risks to public health, with transportation systems being significant contributors. Air pollution spikes can harm children's cognitive development and intelligence quotient (IQ) [59]. Rapid and accurate detection of air pollution spikes caused by vehicular emissions is imperative to protect citizens' well-being [52]. However, current monitoring methods often lack real-time capabilities and struggle to capture dynamic pollution fluctuations [68]. The emergence of Internet of Things (IoT) edge networks offers a promising avenue for establishing efficient, responsive, localized air quality monitoring systems. This research aims to harness IoT edge networks to detect real-time air pollution spikes in transportation, enabling timely interventions and safeguarding public health [69]. This research proposes a system capable of capturing spikes in pollution from transportation, assessing their impact on health, and suggesting methods to mitigate these effects.

Measuring spikes in air pollution from transportation is paramount due to its significant implications for public health and the environment. Air pollution spikes have a health impact on vulnerable populations such as Children, the elderly, pregnant women, and individuals with preexisting health conditions who are particularly susceptible to the health effects of air pollution. Short-term spikes in air pollution can trigger respiratory and cardiovascular issues, exacerbating conditions like asthma, bronchitis, and heart disease [70]. Increased pollution levels often lead to a surge in hospital admissions and emergency room visits, placing a strain on healthcare systems [71]. With this prototype system, the levels of air pollution that can impact vulnerable individuals will be continuously monitored. When these levels exceed a predetermined threshold, notifications will be sent to the relevant authorities.

Pollution spikes from transportation sources can occur suddenly due to traffic congestion or meteorological conditions, resulting in immediate health risks [72], [73]. Even brief exposure to elevated pollution levels can lead to adverse health effects. Repeated exposure to pollution spikes can contribute to chronic health problems, including reduced lung function and long-term cardiovascular issues [74]. Ecosystem Damage is in danger due to air pollutants, which can harm plants, water bodies, and soil, impacting biodiversity and ecosystem health. Some pollutants can lead to the formation of acid rain, which harms aquatic life and damages buildings and infrastructure [60]. This research introduces a developed prototype system for monitoring air pollution resulting from transportation, specifically from cars. When any spikes in pollution occur, the system will automatically trigger an analysis by relevant authorities to assess its impact. Consequently, this proposed system holds the potential to enhance both public health and environmental quality by effectively reducing air pollution originating from transportation, particularly in urban areas.

Black carbon, a component of particulate matter emitted from transportation, contributes to climate change by absorbing sunlight and accelerating ice melt in polar regions [75]. Air pollution spikes can lead to violations of air quality standards set by regulatory bodies, necessitating corrective measures to ensure compliance. Real-time monitoring of spikes informs the development and implementation of effective pollution control policies and traffic management strategies [76]. Providing real-time information about pollution spikes empowers individuals to take protective actions, such as reducing outdoor activities or adjusting travel plans [72], [73]. Accessible air quality data encourages public advocacy for cleaner transportation options and policies [77]. This research aims to inform the development of policies that can provide real-time solutions for deviations in transportation based on the specific locations of pollutants.

Some IoT edge network systems were proposed by previous researchers to monitor air pollution [39], [78], [79]. The paper [28] proposed a framework for designing air pollution using IoT edge networks, blockchains, and Artificial Intelligence. Most of these proposed systems are only for monitoring air pollution in general; they are not specific for transportation spikes pollutants. In the current research, this Research focuses on vehicle air pollution.

Monitoring spikes help authorities identify pollution hotspots and target interventions where they are most needed [80]. Precise data allows for efficient allocation of resources for pollution mitigation efforts. Lastly, studying pollution spikes contributes to a better understanding of air quality dynamics, pollutant behavior, and their impacts. Data on pollution spikes can be used in educational campaigns to inform people about the importance of reducing pollution sources [81]. Monitoring spikes in air pollution from transportation sources is critical for safeguarding public health, protecting the environment, and fostering informed decision-making at individual and policy levels [69]. Timely action to mitigate these spikes can lead to improved air quality, enhanced public well-being, and a more sustainable future [66].

Based on the analysis of existing systems, it becomes evident that this paper introduces a novel system poised to play a pivotal role in combatting transportation-related pollution. There are dynamic Pollution Patterns where Air pollution levels can vary widely across time and locations due to traffic density, weather conditions, and industrial activities. Existing monitoring approaches may miss rapid pollution spikes, delaying necessary actions. The need for real-time Data can prompt the identification of pollution spikes demands real-time data collection and processing. Traditional centralized monitoring systems face latency issues, hindering timely responses. There should be spatial variability as the air quality can significantly differ even within short distances due to local sources of pollution. Establishing localized monitoring to capture these variations is a challenge. Edge network reliability is still challenging in IoT edge networks since they rely on distributed sensors deployed on vehicles or infrastructure. Ensuring sensor accuracy, reliability, and data synchronization in these dynamic networks is crucial. Integrating data from various sensors, vehicle fleets, and stationary sources necessitates robust data fusion and analytics techniques to identify pollution spikes accurately. The last challenge we examined in this paper was privacy and security concerns for collecting real-time data from transportation sources. Designing systems that respect individuals' privacy while providing valuable insights is essential.

5.3. Related Works

This section serves as an introduction to existing systems relevant to the development of the system in this research. It aims to provide an overview of these systems, highlighting their respective strengths and weaknesses compared to the developed system. Furthermore, it will compare various research aspects between the related systems and the one set in this study.

In their paper [48], the authors proposed a novel approach for designing a cost-effective and real-time air pollution monitoring system by leveraging edge computing and Internet of Things (IoT) technologies. Current air quality monitoring systems often lack the necessary spatial and temporal resolutions, posing challenges in accurately assessing air pollution. The proposed system employs sensors to collect real-time air quality data to address these issues, which is then transmitted to edge computing devices for processing and analysis [16]. This approach reduces the computational load on battery-powered sensing nodes while maintaining accuracy. The paper overviews existing monitoring systems, their limitations, challenges, and the proposed edge computing-based IoT architecture for air quality monitoring [16]. From an analysis of these two papers, this research demonstrates the feasibility of applying edge IoT with real-time systems to address air pollution. The approach proposed in these papers has been incorporated into this research to be applied within the realm of transportation, particularly for managing spikes in air pollution.

The study on developing an Artificial intelligence (AI) model to measure traffic-related air pollution using multisensory and weather data investigates the impact of various input variables on training air quality indexes through fuzzy logic combined with simulated annealing (SA) and particle swarm optimization (PSO) [82]. The model predicts concentrations of NO2 and CO based on resistivity values from multisensory devices and weather variables. PSO outperforms SA in the optimization process, and the study highlights the sensitivity of input resistivities for predicting NO2 and CO concentrations, which is crucial for addressing air pollution challenges. The paper [82] focused on only two specific pollutants. However, this research developed a prototype to address six significant pollutants that significantly impact air quality worldwide.

The paper [83] introduced portable IoT air-quality monitoring devices for vehicle installation in Ibarra City, Ecuador. These devices use Message Queuing Telemetry Transport (MQTT) to send data to a time series database in edge computing and cloud computing for visualization. The research employs outlier detection and supervised classification to analyze data, determining air pollution levels categorized as low, normal, and high. Results show over 90% performance for IoT nodes inferring air quality, with memory consumption of 14 Kbytes in flash and 3 Kbytes in Random Access Memory (RAM). This approach presents an efficient, cost-effective solution for air-quality measurement through IoT technology. However, it is important to note that the system, in its current form, does not measure air pollution directly from the vehicle itself. In the proposed prototype developed within this research, the system will be integrated with individual vehicles, allowing for the measurement of pollutants emitted by each car over time to ensure their continued compliance with air quality standards. If a vehicle exceeds established pollution thresholds, specific policies and measures will be implemented accordingly.

The paper [84] discussed the integration of IoT, edge intelligence, 5G, and blockchain technologies into autonomous vehicles (AVs) to enhance their efficiency and sustainability. AVs are becoming a vital part of intelligent transportation systems. The study reviews the impact and implementation of these technologies, addressing challenges and insights into seamless integration. The integration aims to provide self-verifying, self-executing, and secure AV systems. This current research is based on some challenges defined in the paper [84] and addresses solutions accordingly.

The study from the paper [85] investigated the potential application areas of low-cost PM (particulate matter) sensors for air quality monitoring. The focus is on evaluating the performance of two PM sensor models, PMS5003 and SPS30, based on the US Environmental Protection Agency (EPA) guidelines. The sensors are assessed in terms of indicators like coefficient of determination (R2), root mean squared error (RMSE), mean absolute error (MAE), mean normalized bias (MNB), and coefficient of variation (CV). The aim is to determine if these sensors could be used as supplemental monitoring tools to enhance existing air quality measurement networks. The study contributes to understanding the effectiveness of low-cost sensors for air quality assessment. The previous system only focused on monitoring a single pollutant, PM. However, in the current research, a low-cost monitoring system has been developed to track six different pollutants simultaneously.

The paper [40] study focused on improving indoor air quality (IAQ) monitoring systems. It addresses the need to balance well-being and energy efficiency in building design, especially given the impact of indoor air quality on occupants' health and comfort. The authors propose an innovative approach that considers pollutant concentrations and exposure time, developing new indices. They integrate sensors and IoT technology to create a monitoring system that can autonomously or remotely control ventilation based on IAQ data and predefined thresholds. The study introduces decision-making algorithms to account for measurement uncertainty and validate the approach using simulated data. The current research leverages IoT edge networks and an algorithm to monitor outdoor air pollution from vehicles while considering spikes.

The authors of the paper [86] introduced an adaptive hardware-software platform for predictive air quality monitoring, explicitly focusing on carbon dioxide (CO_2) concentrations. The aim is to forecast CO_2 trends accurately to prevent sudden increases and enable Heating, Ventilation, and Air Conditioning (HVAC) systems to control indoor conditions effectively while avoiding energy waste. The system employs deep learning algorithms to analyze a limited window of recent data, making it adaptable to changing habits and environmental conditions. The platform achieves high accuracy, mainly using the Long Short-Term Memory network. This innovative approach addresses IAQ concerns, especially in increased indoor time due to factors like COVID-19. The current research employs a similar approach to test the developed prototype within a laboratory setting, but with a notable difference: it assesses the system's performance for six different pollutants.

The paper [87] discussed the significance of air quality monitoring and control in the context of smart cities. It presents a detailed framework for air pollution monitoring and prediction using the Internet of Things (IoT) architecture. The paper [88]examined low-cost air quality monitoring devices to assess road transport-related emissions, aiming to supplement high-precision monitoring stations. A case study conducted in Munich focuses on PM10 measurements. Ten low-cost devices were deployed to gather hourly PM10 values. Statistical analysis of historical data revealed correlations between PM10 concentrations, weather conditions, and traffic volumes. The study found that some expected relationships were evident in the data from low-cost devices, but others were inconclusive. The research highlights the importance of extended measurement periods, frequent calibration, and the challenges of interpreting point measurements. It suggests that dispersion models can aid in understanding complex factors affecting air quality. The current research delves deeper into the proposed framework, particularly in calibrating all six pollutants for application in transportation.

The paper [89] introduced a modular IoT sensing platform with hybrid learning capabilities for air quality prediction. A team of researchers from various institutions in India conducted the study. The focus is addressing the challenges of accurate and low-cost air quality measurements in underdeveloped countries. The proposed platform includes an IoT node with multiple sensors to monitor pollutants like Ammonia (NH_3), NO_2, CO, and PM2.5, along with air humidity and ambient temperature. The system uses Global System for Mobile Communication (GSM)/WiFi technology to transmit real-time air quality data, generate alerts, and facilitate data analysis for environmental intelligence applications. The goal is to enhance the field of distributed, low-cost sensing devices for environmental monitoring. The study emphasizes the need for precise measurements and predictions of pollutants to combat the increasing ecological issues caused by urbanization and fossil fuel usage. However, the current research can capture spikes from vehicles in urban areas and subsequently use policies to take action. Furthermore, the prototype proposed in this research encompasses a broader range of pollutants than the previous one.

The paper [90] introduced a framework for air pollution monitoring in smart cities using IoT and smart sensors. The study addresses the challenges of urbanization and the increasing city population, leading to heightened pollution levels. The proposed solution involves IoT and intelligent sensors continuously monitoring various environmental parameters such as humidity, carbon emissions, temperature, smoke, and hazardous particulates. The collected data is transmitted to a central office for analysis and action, contributing to improving the city's environment. The current prototype solved the issues presented in this paper, which come from the population increase in cities.

The authors of the paper [91] addressed the increasing challenges of urbanization and pollution by leveraging the IoT paradigm and smart sensors. The framework aims to monitor various pollutants using a network of static and mobile sensors integrated through gateways. Data is transmitted wirelessly to the cloud, undergoing processing and analysis. The architecture accounts for sensor defects, power consumption, and data management. The study presents an overview of current trends and introduces a comprehensive cloud-centric architecture for efficient air quality monitoring in urban environments. The current research is now focusing on vehicle pollutants and capturing the spikes.

The article [92] discussed using artificial intelligence (AI) in air quality monitoring for smart city management. The authors, En Xin Neo et al., present a comprehensive study that utilizes machine learning and deep learning techniques to predict air quality. They propose an end-to-end predictive model incorporating various pollution markers and meteorological data for four different urban cities in Selangor, Malaysia. The study highlights the importance of feature optimization to enhance the accuracy of air quality predictions, particularly for PM2.5

concentration. While this paper exclusively focused on a single pollutant from a general perspective, the current research specifically addresses vehicles' spikes.

The study from the paper [93] introduced a vehicle sensor network (VSN) approach for monitoring air quality in urban areas using low-cost IoT devices mounted on vehicles. The devices collect geolocated particulate matter (PM) measurements transmitted to an IT infrastructure. Real-time spatial and temporal pollutant distribution maps are generated and made accessible online. The VSN system was deployed in Trieste, Italy, with support from volunteers and local transportation authorities. Results reveal areas with poor air quality linked to increased vehicular traffic. The VSN approach offers valuable insights for urban planning and encourages reduced private car usage in favor of public transportation, enhancing air quality.

This research has the potential objective of harnessing IoT edge networks for real-time air pollution spike detection in transportation to safeguard public health. This research develops a distributed sensor network that establishes a network of IoT sensors deployed on vehicles and at strategic locations along transportation routes to capture real-time air quality data. This research implements edge computing capabilities to process sensor data locally, enabling rapid detection of air pollution spikes without significant latency. In this paper, the algorithms to analyze data patterns and identify sudden increases in pollutant levels that characterize pollution spikes were designed. This research proposed policies to collaborate with governmental agencies responsible for environmental protection and public health to ensure data integration into decision-making processes and generate evidence-based recommendations for policy changes and regulatory measures to reduce transportation-related air pollution. This research uses machine learning techniques to develop predictive models for anticipating pollution spikes based on historical data and environmental conditions. This research contributes to the scientific community by sharing insights, data, and methodologies that can advance the understanding of air pollution's impact on health and the effectiveness of IoT-based solutions.

The table 5.1 provides a summary of existing systems and the current prototype developed. It highlights all the unresolved challenges in the previous design and outlines the solutions implemented by the current research.

Table 5. 1 Comparison of the existing systems and the current system

Existing work	**Current research**

Introduces a novel approach for designing a cost-effective and real-time air pollution monitoring system using edge computing and IoT technologies [48]	Implements the novel by applying it in urban areas with vehicles
Using multisensory and weather data, developing an AI model to measure traffic-related air pollution[48].	Only a few pollutants were considered; the current research accesses six pollutants. In addition, air pollution was measured in general, yet this research will consider the pollutants from each vehicle.
Development of portable IoT air-quality monitoring devices for vehicles for intelligent transportation systems [48].	These systems have been developed and installed on the car. But they are installed on the top of the vehicle and can measure air pollution for the environment in general without considering the pollutants from the vehicles' exhaust. Also, it captures pollution without considering spikes of them to be reported as remarkable as is done in the current research.
Development of systems that are only specific to some pollutants [48].	The current research developed a system that comprises six dangerous pollutants.
Development of a framework for the smart city for air pollution monitoring [48]	The current research has designed a prototype based on the framework proposed for smart city air pollution monitoring, with the addition of the capability to record and monitor the exhaust of air pollution from vehicles.

In Table 5. 1, we aim to compare the previous systems and the innovations introduced by the current research, addressing previously unmet challenges.

Therefore, the global challenge of urban air pollution's impact on health and the environment, with a focus on transportation, is a major contributor to harmful pollutants. Children's cognitive development, often neglected, is highlighted as vulnerable to air pollution's detrimental effects, with studies showing links to lowered IQ scores and learning disabilities. Rapid pollution spikes from transportation pose urgent risks, particularly for vulnerable groups, emphasizing the need for real-time detection and intervention. Leveraging IoT edge networks emerges as a solution to revolutionize air quality monitoring, necessitating collaboration, accurate sensors, and privacy safeguards. In the next section, the paper highlights the materials and methodology used to achieve the results required.

5.4. Materials and Methods

This section explains the materials and methodology used in this research. The presented study aims to detect air pollution peaks in transportation using real-time IoT edge networks. The prototype was developed in the laboratory of Seoul National University and includes a comprehensive setup of materials and a well-defined methodology. Sensors such as high-precision air quality sensors were used as materials to measure various pollutants such as particulate matter (PM2.5 and PM10), carbon dioxide (CO), ozone (O3), sulfur dioxide (SO2), and nitrogen dioxide (NO2). For prototyping, ioT-enabled microcontrollers, such as Arduino UNO, were used to interface with the sensors for local data processing and communication with edge networks. ESP8266 WiFi wireless communication modules enabled data transfer between edge devices and central systems. Local processing functions allow the preprocessing of data and rapid identification of pollution spikes before relevant information is transmitted using a spike detection algorithm. A central repository was set up to collect and store the transmitted data for further analysis and visualization. The real-time air pollution spike detection method involves a multi-step process that takes advantage of edge computing and IoT technologies, as explained in this section[94]. The paper proposes to generate and share real-time alerts and reports with relevant stakeholders, including local authorities, health departments, and the public.

5.4.2. Hardware Components

The prototype system comprises a Microcontroller, sensors, and communication and transmission devices. These hardware devices are connected to capture the air pollution from different sensors on different cars that are in transportation.

Figure 5. 1 Arduino UNO board

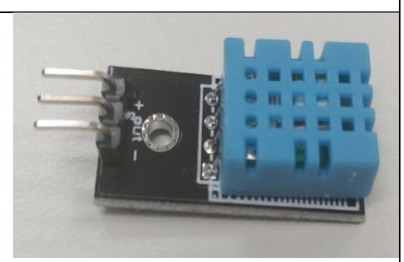
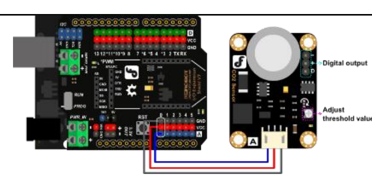
From Figure 5. 1, the ATmega328P is the core of Arduino Uno, with 32 KB flash memory for code storage, 2KB SRAM for data, and 1KB EEPROM for non-volatile storage. It offers digital I/O pins for interfacing with sensors and actuators, six analog input pins, timers for precise timing, and PWM for tasks like motor control. Communication is possible through UART, I2C, and SPI interfaces, supporting various clock frequencies. This microcontroller is the brain behind Arduino projects, enabling code control of connected components.

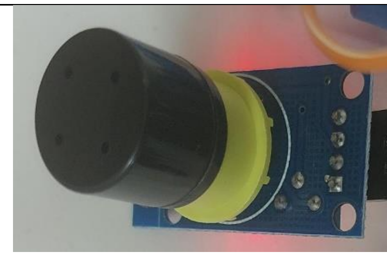
The system comprises different types of sensors from other manufacturers with high precision. All these sensors have threshold values for good air quality as defined by the World Health Organization (WHO), as shown in Table 4. 1.

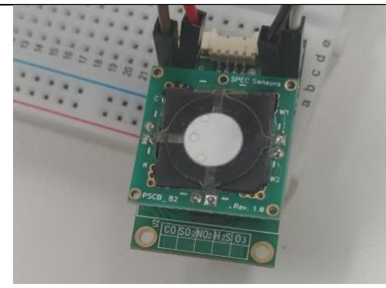
Table 5. 2 provides a comprehensive overview of all sensors utilized in this research, offering detailed descriptions of each sensor's functionality and purpose. The table also includes visual representations of these sensors, which can measure specific environmental pollutants.

Table 5. 2 Sensors description

Serial Number	**Sensor name**	**Description**	**Image**

1	Particulate Matter sensor (PM2.5 and PM10):	PM sensors measure fine airborne particles like PM10 and PM2.5, impacting air quality and health. They employ optical scattering and absorption principles, including a light source, sample chamber, detector, and signal processing, which are crucial for monitoring respiratory and cardiovascular risks. The research incorporates the Plantower PMS5003 sensor, offering real-time, precise particle concentration data via a digital interface.	
2	Temperature and Humidity Sensor	A temperature and humidity sensor, or hygrothermometer, measures ambient temperature and relative humidity. These sensors are vital for weather monitoring and indoor climate control, utilizing capacitance or electrical resistance changes to gauge humidity and providing temperature and humidity readings in Celsius or Fahrenheit and percentage relative humidity (% RH).	
3	CO2 sensor	DFRobot's CO2 sensor utilizes voltage output to detect CO2 levels, with a customizable threshold for digital signal activation. It employs an MG-811 gas sensor, offering	

		CO_2 sensitivity and stable performance, while its industrial-grade design ensures reliability. Calibration is necessary, operating at 5V, with compatibility for the Gravity interface, and featuring a compact 32x42mm size.	
4	Ozone Sensor	In this research, the MQ131 O_3 sensor is employed to measure ozone (O_3) gas concentration, essential for Earth's stratosphere but harmful at ground level. Figure 5 depicts the MQ131 sensor used in the project, which operates through chemo-resistive gas sensing, detecting ozone-induced resistance changes in its sensitive material. The sensor includes a heater for temperature control and signal conditioning circuitry, facilitating ozone concentration determination through output signal processing.	
5	SO2 Sensor	SO_2 sensors are pivotal in monitoring environmental air quality and health by detecting sulfur dioxide (SO_2) gas concentrations. They utilize diverse technologies such as electrochemical, optical, and chemical absorbance methods, each offering distinct detection	

		mechanisms while outputting visual changes that correlate with SO2 concentration.	
6	NO2 Sensor	Nitrogen dioxide (NO2) sensors are designed to measure the concentration of this harmful gas generated from combustion processes in vehicles, power plants, and industry. They employ surface adsorption, detecting changes in properties like electrical conductivity or mass to quantify NO2 levels. These sensors find applications in vehicles for emissions monitoring and engine optimization due to their critical role in health and environmental assessments.	
7	ESP8266 Wifi module	The ESP8266 has been a driving force behind the proliferation of IoT projects and applications due to its affordability, capabilities, and ease of integration. The WiFi module supports data transmission. It only transmits once the value exceeds the threshold.	

5.4.3. Hardware Architecture

Figure 5. 1 presents the designed architecture that combines all hardware components explained above.

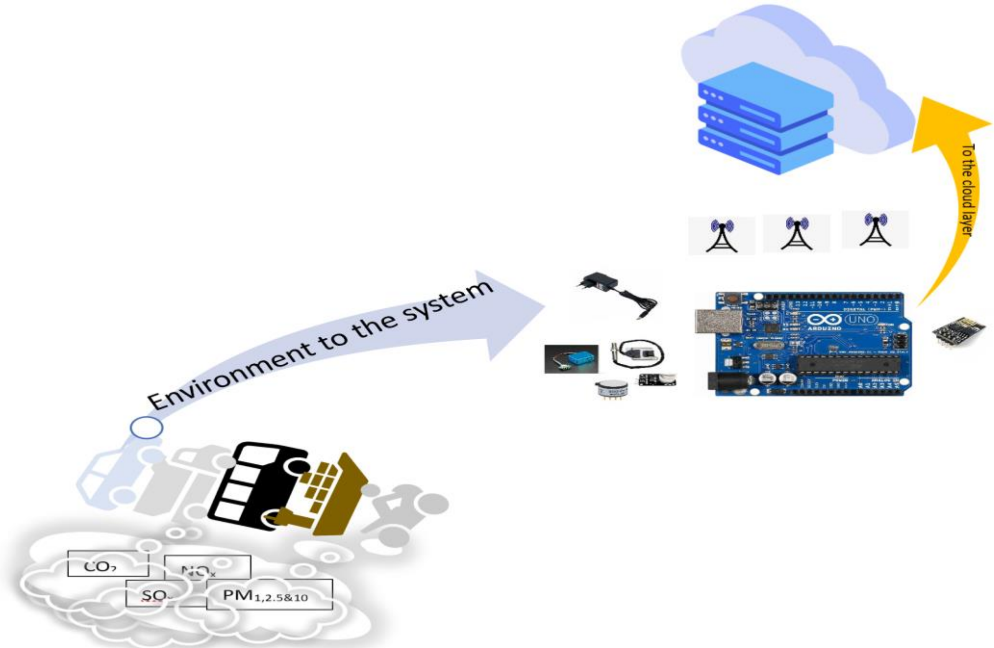


Figure 5. 2 The prototype architecture of the developed system

Figure 5. 2 the architecture shows how all hardware components are connected to capture the air pollution from the environment on cars and where the cars are passing.

Figure 5. 3 is displaying the prototype meticulously designed with all sensors seamlessly integrated onto an Arduino Uno board. This innovative prototype can detect all pollutants commonly associated with vehicles, as elucidated in this paper. Moreover, the system is seamlessly integrated into the car's power distribution system, enabling it to initiate automatically when the vehicle is powered on. As depicted in Figure 5. 3, the system is securely enclosed within a protective box, safeguarding it from potential damage while preserving the necessary air capture space from the car's exhaust to which it is affixed.

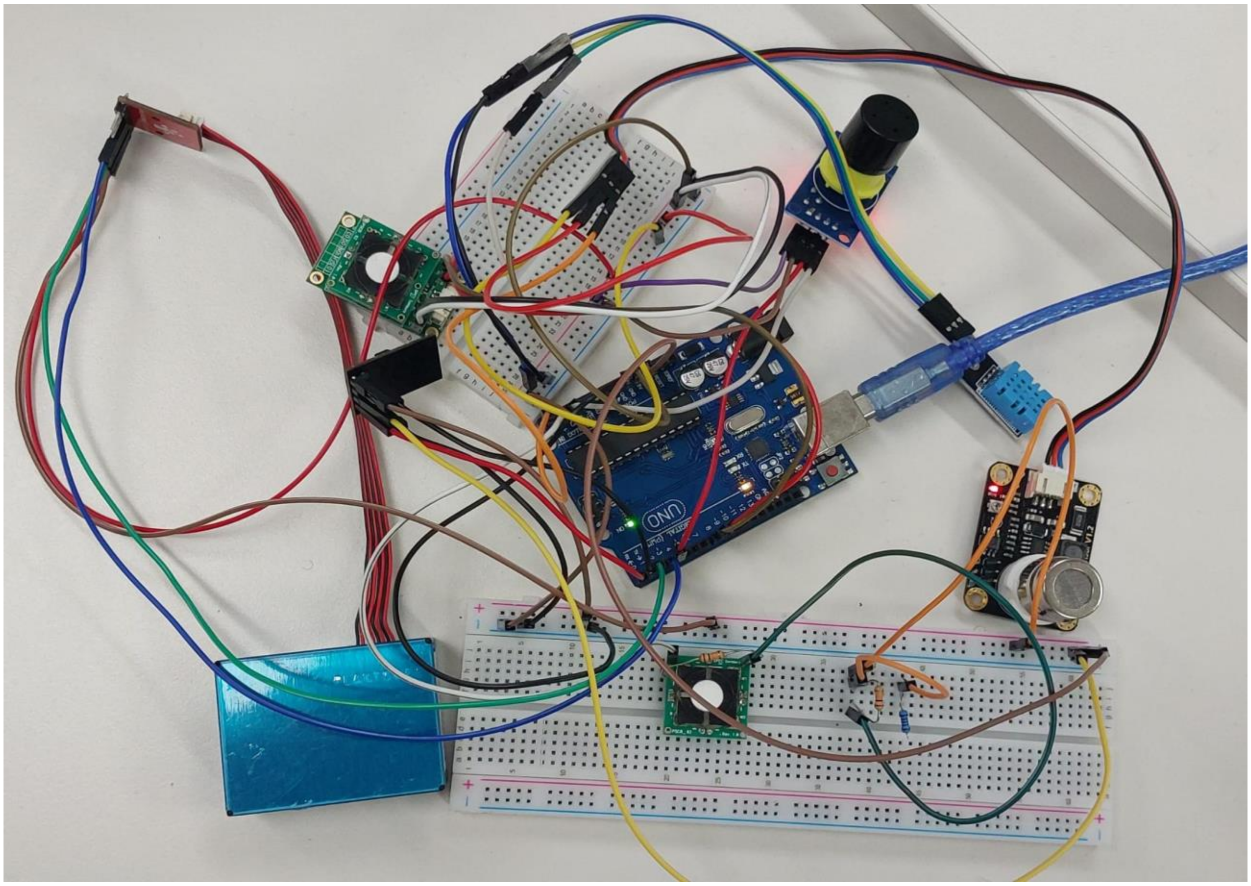


Figure 5. 3 Prototype designed

5.4.3. Mathematical Tools

This research uses mathematical concepts to introduce a novel approach to capturing spikes in air pollution data. It identifies spikes in the sensor data by employing the Heaviside step function and Dirac delta function on data corrected by a specially designed prototype on the figure. This process involves applying these mathematical functions to the sensor data and calculating the spike function.

This research applies both the Heaviside step function and the Dirac delta function from each sensor's dataset. The application of the Heaviside step function serves to emphasize sudden changes or abrupt increases in pollution levels. Simultaneously, the Dirac delta function helps pinpoint specific time points where pollution levels experience instantaneous spikes.

Two essential mathematical functions are pivotal in this analysis: the Heaviside step function and the Dirac delta function. The Heaviside step function, denoted as H(t), is used to model the sudden onset of a phenomenon. The Dirac delta function, denoted as δ(t), represents an infinitely narrow impulse at a specific point.

Heaviside step function is:

$$V(x) = H(x - ti) \qquad x \geq \theta \qquad (5.1)$$

The $V(x)$ is the input signal of the sensor data at x time. Then $H(x - ti)$ is the Heaviside step function at a time x is great or equal to the threshold θ. That means the function gives 0 when the signal from the sensor is less than the threshold θ. Then, once the value at time x is greater than the threshold θ, the value becomes 1. The ti denotes the peak at a specific point.

The Dirac Delta function denoted as $\delta(x)$, is a distribution or generalized function used to represent point sources or impulses.

$$\delta(x) = \begin{cases} 0, & for\ x \neq 0 \\ \infty, & for\ x = 0 \end{cases} \qquad (5.2)$$

The $\delta(x)$ has an unusual property where its integral over the entire real line equals 1.

$$\int_{-\infty}^{+\infty} \delta(x)dx = 1 \qquad (5.3)$$

Then, by applying the equation (5.3 on the Heaviside function (5.1, which will give the peak value at a time $(x - ti)$. The Dirac delta function ensures that the spike is localized at ti. Then the peak function will be the equation (5.1 multiply by equation (5.3:

$$Peak(x - ti) = H(x - ti) \times \delta(x - ti) \qquad (5.4)$$

We calculate the spike function for each sensor dataset based on applying the Heaviside step function and the Dirac delta function. The spike function quantitatively represents the detected spikes, providing insight into the intensity and duration of pollution events.

The spike function will be measured based on the height of the peak H at the specific point ti for the value greater than or equal to the threshold θ. Then the spike function from the equation (5. 4 can be denoted as:

$$Spike(x) = H \times \ Peak(x - ti) \ \ for\ x \geq \theta \qquad (5.5)$$

Then, from the equation (5. 5, let's analyze spike behaviors when data comes from the sensors by applying Sigmoid functions. Therefore, that function will give the continuous doted function from different points calculated in the equation (5. 5. The sigmoid function is a function that creates a gradual rise in the sensor data when it approaches the threshold. Let's assume a sensor measures a physical quantity (e.g., temperature, NO2), then makes a model of spikes in the data whenever it goes above a certain threshold θ. The combination of the sigmoid function and a scaling factor can be used to achieve this.

The sigmoid function is often defined as:

$$S(x) = \frac{1}{1 + e^{-x}} \qquad (5.6)$$

To control the steepness of the rise near the threshold, you can multiply the input by a scaling factor 'k':

$$ScaledS(x) = \frac{1}{1 + e^{(-k \times (x - \theta))}} \qquad (5.7)$$

Then, applying equation (5. 7 with the height H, the result can be a scaled obtained for scaled spikes in the below equation:

$$SpikesSensorData(x) = \begin{cases} 0, & for\ x < 0 \\ H \times Scaled(x), & for\ x \geq \Theta \end{cases} \tag{5. 8}$$

In this function, the sensor data remains low for values below the threshold θ. As the input x crosses θ, the sigmoid function causes the data to rise gradually, and then the spike is introduced by scaling the sigmoid function by the magnitude H. The parameters θ, H, and k can be adjusted to control the function's behavior.

This paper applies these mathematical functions as tools to analyze behaviors of spikes during the data capturing from the air pollution sensors. Some software components have been used to compute the algorithm from the mathematical model and the hardware components.

5.4.4. Software Components

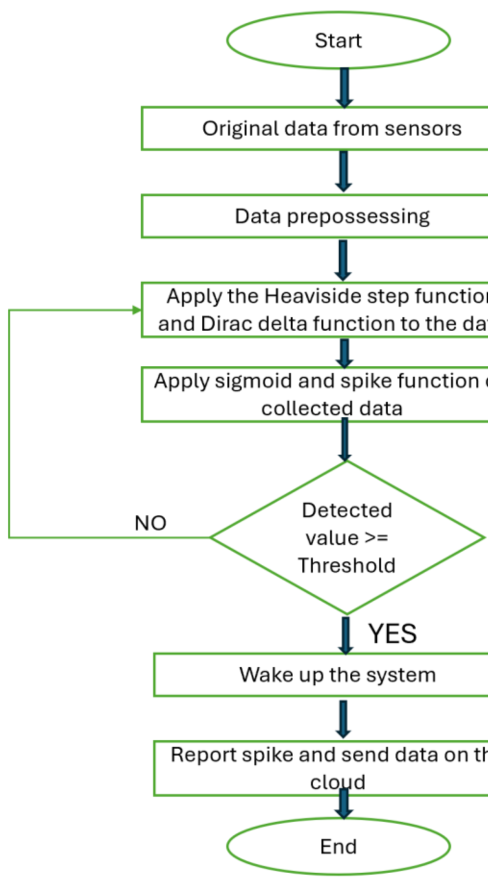
This research has used different types of software to analyze data captured by the developed prototype in Figure 5. 3 and to implement the algorithm generated from the mathematical models explained in the previous subsection.

This research harnesses the Arduino IDE to facilitate the programming and control of an Arduino Uno microcontroller. The program uploaded to the Arduino Uno is designed to oversee the operation of multiple sensors, ensuring they operate by their respective calibrations and measurement parameters. Subsequently, the acquired sensor data is transmitted to cloud servers utilizing the server's designated IP address for data transfer.

In addition to the Arduino IDE, this research also leverages the Python programming language. Python is employed to conduct an in-depth analysis of the spike behaviors generated by each sensor integrated into the system. By combining these software tools into the research framework, a comprehensive approach is taken to collect, process, and analyze data, thereby facilitating a thorough investigation of the research objectives.

5.4.5. Algorithm for Spikes Detection

This research has developed an algorithm specifically designed to address the critical issue of spike detection within air pollution monitoring systems deployed in transportation networks. The overarching objective of this algorithm is to contribute to improving public health in densely populated urban areas by identifying and characterizing elevated air pollution levels, often caused by factors like vehicular emissions and industrial activities. As depicted in the accompanying flowchart, the algorithm represents a systematic approach to data analysis and pattern recognition, enabling timely and accurate identification of pollution spikes. This research endeavors to detect spikes and provide valuable insights for developing proactive pollution control strategies and mitigation measures, ultimately fostering healthier urban environments and enhancing the well-being of the communities residing therein.



Algorithm 1. Flowchart for spikes algorithm

Algorithm 1 is for the flowchart of the algorithm. The process commences with the sensors generating data from the surrounding environment. This data is meticulously collected and subsequently stored, preparing it for the preprocessing stage. The system employs the Heaviside step function and the Dirac delta function to ensure the accurate capture of spikes within the data. Following this step, a sigmoid function is applied to the discrete data points, transforming them into a continuous signal. Finally, the system proceeds to generate the spike function. The system will analyze the spike and determine if it is greater than the defined threshold of the sensor, as described in Table 4. 1.

Algorithm for Air Pollution Spikes Detection in Transportation for Better Health

Input: Sensor data (Particulate matter levels, Temperature, Humidity, O3, NO2, SO2, CO2)

Output: Pollution level, alerts for spikes or dangerous levels

Step 1: Data Collection

- Set up air quality monitoring sensors at predetermined locations.
- Continuously collect sensor data, including readings for various pollutants such as Temperature, Humidity, PM2.5, PM10, CO2, NO2, O3, and SO2.
- Record the timestamp for each data point.

Step 2: Data Preprocessing

- Check for sensor anomalies and calibration issues.
- Handle missing or erroneous data through interpolation or data imputation techniques.
- Smooth the data to reduce noise using filters or moving averages.

Step 3: Spike Detection

- Define a threshold for each pollutant.
- Compare the current data by applying the Heaviside step, Direc delta, Sigmoid, and Spike functions with the threshold values.
- If any parameter exceeds the threshold, generate an alert indicating a pollution spike.

Step 4: Data Reporting and Visualization

- Display real-time or periodic pollution data on a dashboard.
- Provide historical pollution data for analysis and comparison.

Step 5: Mitigation and Response

- Implement strategies to reduce pollution sources if persistent high pollution levels are detected.

Step 6: Data Storage

- Store collected data in a secure and accessible database for future analysis and research.

Step 7: Continuous Monitoring

- Continuously run the algorithm, repeating steps 1 to 5, to ensure ongoing air quality monitoring.

This algorithm provides a high-level overview of the process involved in air pollution spike detection.

The research section explains the materials and methodology used to detect air pollution peaks in transport using real-time IoT edge networks. It describes high-precision air quality sensors to measure various pollutants and IoT-enabled microcontrollers for data processing and communication with edge networks. Mathematical tools such as the Heaviside step function and the Dirac delta function are used to identify pollution peaks, and a software framework using Arduino IDE and Python is used for data analysis. Furthermore, the study introduces a comprehensive algorithm for detecting air pollution spikes, which can significantly contribute to enhancing public health in urban regions through the identification and characterization of heightened pollution levels. This algorithm includes data acquisition, preprocessing, spike detection, reporting, and continuous monitoring.

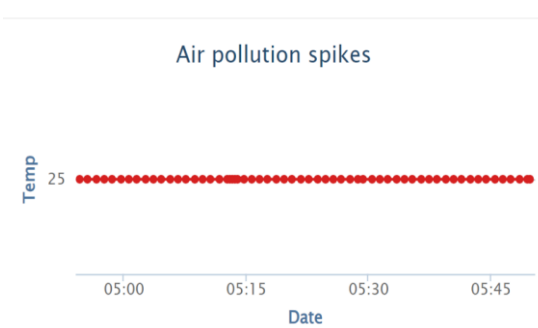
5.5. Experimental Evaluation and Results

This section deals with a comprehensive examination of the research findings. It contains a detailed account of the results obtained from using each of the materials and methods explained in the previous section. Currently, the prototype is being tested in a controlled laboratory environment where the temperature and humidity values remain relatively stable, mainly due to the active operation of an air conditioning system during daytime hours. Nevertheless, spikes in sensor readings have been observed due to external environmental factors. In this section, the authors present the results of the newly developed prototype and explain the application of an algorithm to identify and categorize these spikes across the different sensors integrated into the system.

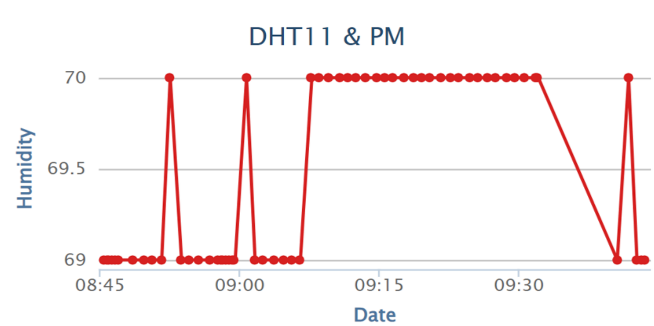
5.5.1. Sensor Caption Data for all Sensor

The sensors collect data from the physical environment. If the data falls below the predefined threshold, the system will not transmit the data to the cloud. Conversely, when the data exceeds the predefined threshold, the system will send the data to the cloud. Figure 5. 4 to Figure 5. 6 display information from all sensors from the server.

All the below figures (Figure 5. 4, Figure 5. 5, Figure 5. 6,) illustrate data collected from various sensors. These sensors are connected to a single Arduino board, enabling the system to transmit relevant data based on pollutant levels exceeding predefined thresholds from the vehicle's exhaust port, where the system is installed. As a result, the system can now conserve energy, aligning with the proposal presented in the paper. Furthermore, it facilitates a broad spectrum of monitoring capabilities by situating a measurement sensor at crucial locations where vehicles operate, thereby providing real-time pollutant concentration data. Moreover, by individually placing measurement sensors within cars, users can promptly assess pollutant concentrations generated within the vehicle environment. This feature allows for swift actions to minimize pollutant emissions. Additionally, individually locating measurement sensors in vehicles can be a policy tool for users to reduce soot pollutant emissions during vehicle operation actively. Furthermore, this research has the potential to contribute to effective decision-making and proactive pollution control strategies by integrating with existing transportation infrastructure and pollution monitoring systems.

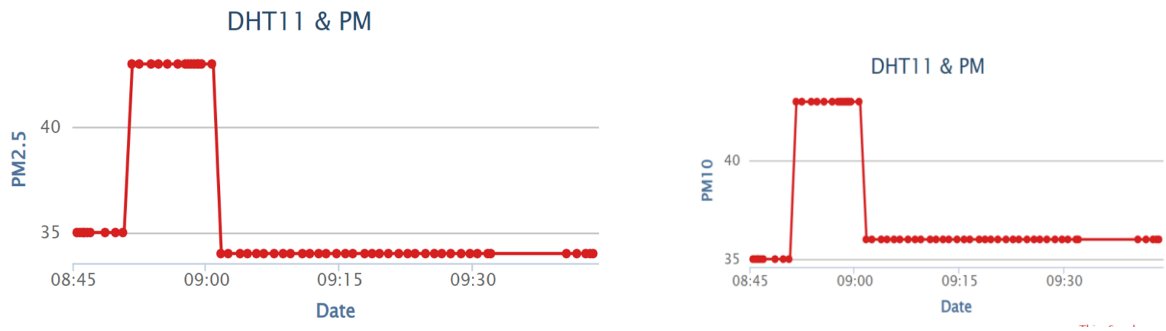


(a)



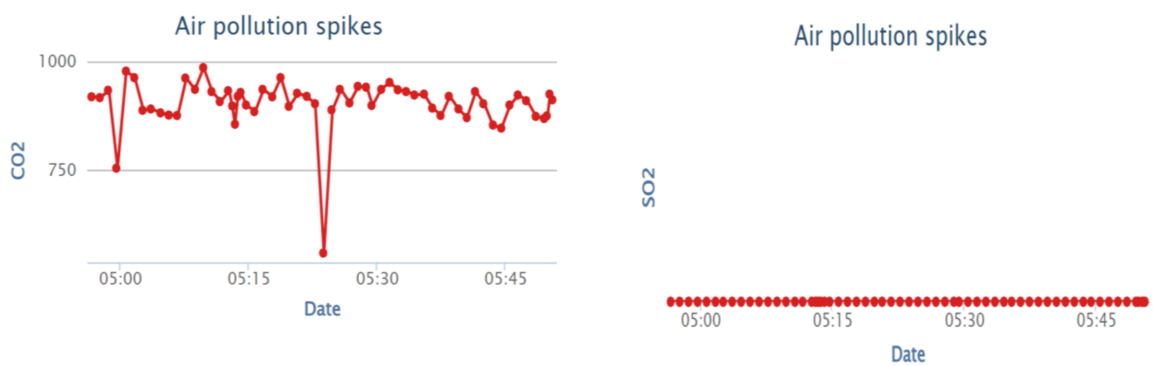
(b)

Figure 5. 4 (a) Data on the server for temperature. (b) data from the server for humidity

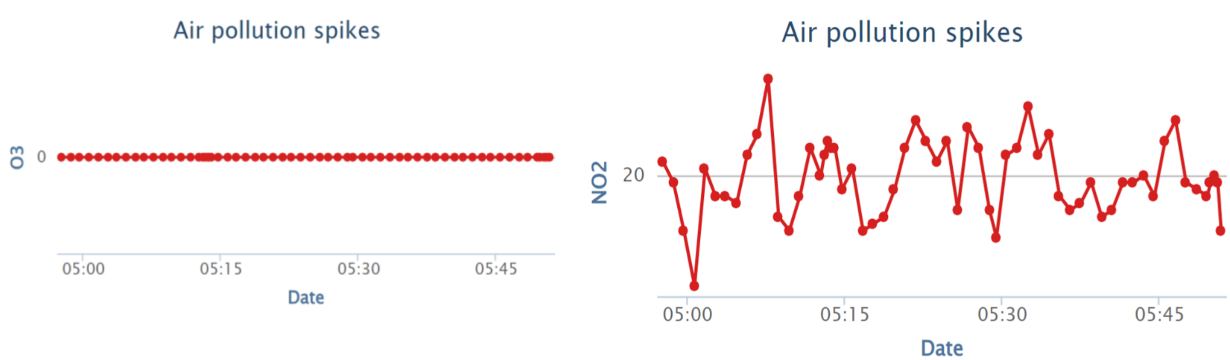


(a) **(b)**

Figure 5. 5 (a) PM2.5 data from the server. (b) PM10 data from the server



(a) **(b)**



(c) (d)

Figure 5. 6 (a) CO2 data on the server. (b) SO2 data on the server. (c) Ozone data from the server. (d) NO2 data from the server.

Figure 5. 4(a), The data emanating from the temperature sensors presents a remarkably consistent profile, marked by minimal fluctuations. This constancy can be attributed to the controlled laboratory environment, meticulously maintained at a steady temperature of 25 degrees Celsius, courtesy of an active air conditioning system. Under these controlled conditions, the temperature readings remain virtually constant, reflecting the precision and stability of the laboratory setting. However, it's essential to note that occasional variations do manifest, typically on days when the air conditioner (AC) is temporarily deactivated. During these interludes, temperature levels may exhibit minor deviations, which can have a cascading effect on the other sensors in the system. Such variations are particularly significant in the context of air pollution monitoring, as they can potentially catalyze an increase in pollution levels. The interconnectedness of these sensor readings highlights the importance of a stable temperature environment to ensure the accuracy and reliability of the pollution data collected.

Figure 5. 4 (b) shows that the data derived from the humidity sensor portrays a dynamic profile characterized by noticeable fluctuations in response to the operational status of the air conditioning (AC) system. When the AC is inactive or temporarily turned off, humidity levels show a discernible uptick. Conversely, when the AC is engaged, humidity levels register a decrease. These fluctuations in humidity readings provide valuable insights into the interplay between environmental conditions and the operation of the air conditioning system. The data underscores the direct impact of AC systems on local humidity levels, a factor that must be considered when interpreting air pollution data. Importantly, it's crucial to acknowledge that the system is slated for deployment in an outdoor environment, where temperature and humidity are not subject to the controlled conditions imposed by an AC system. Consequently, ensuring suitable and consistent environmental conditions will be of paramount importance. To ensure the accuracy and reliability of the data, it is imperative to implement measures aimed at mitigating the potential influence of uncontrolled humidity fluctuations that may manifest in outdoor environments.

Moving to Figure 5. 5 (a), The dataset derived from the PM2.5 sensor reveals a crucial aspect of air quality monitoring. It encapsulates a range of particulate matter concentrations, specifically within the spectrum of 35 to 70 micrograms per cubic meter. This range is significant because it aligns with the air quality standards established by the World Health Organization (WHO), which defines these threshold values as indicative of acceptable air quality conditions. An intricate system operation comes into play as the sensor data is processed in real-time. Data points falling below the lower threshold of 35 micrograms per cubic meter or exceeding the upper limit of 70 micrograms per cubic meter prompt an intelligent response from the WiFi module, as illustrated in Figure 8. This response entails the module withholding data transmission, an energy-conservation strategy designed to optimize the system's power usage. In contrast, data points within the defined range of 35 to 70 micrograms per cubic meter indicate satisfactory air quality. These data points are then promptly and automatically relayed to the cloud servers. The selective transmission of this data, in alignment with established air quality standards, ensures that only the most pertinent information is communicated, reducing unnecessary data transfer while preserving energy resources. This distinctive approach underscores the system's commitment to precision and efficiency, making it adept at conserving power and delivering real-time, actionable air quality information to the cloud. By adhering to the designated PM2.5 thresholds, this intelligent system strikes an optimal balance between energy preservation and accurate air quality reporting, contributing to a more sustainable and eco-friendly deployment.

Figure 5. 5 (b) visualizes PM10 data to provide a compelling insight into the air quality monitoring system's responsiveness to varying pollution levels. The dataset, monitored by the PM10 sensor, is especially critical in identifying particulate matter levels that exceed the WHO-defined threshold of 70 micrograms per cubic meter. These elevated levels can indicate deteriorating air quality and potentially harmful environmental conditions. One of the remarkable features of this visualization is the display of an "empty" PM10 window, symbolizing instances when the particulate matter levels fall below the critical threshold of 70 micrograms per cubic meter. This deliberate absence serves as a powerful visual cue, indicating that, during such periods, air quality is deemed acceptable and within the defined safety standards. Conversely, when the PM10 data surpasses the 70 microgram per cubic meter threshold, it instantaneously populates the PM10 window with real-time readings. This instantaneous display signifies a critical event when air quality is of concern, warranting immediate attention and action. By presenting PM10 data in this manner, the visualization intuitively guides users' attention to critical variations in air quality, ensuring that instances of elevated particulate matter levels above the defined threshold are instantly recognized. Such clear and visual reporting empowers individuals and authorities to take

prompt measures to address air quality issues, reinforcing the system's commitment to enhancing environmental safety and public health.

Figure 5. 6 (a) provides valuable insights into the system's capacity to monitor carbon dioxide levels, a key indicator of indoor air quality. With a defined threshold set at 1000 micrograms per cubic meter ($\mu g/m^3$), this visualization is crucial for assessing environmental conditions and air safety. The laboratory data showcased in this visualization notably reveals that CO_2 levels consistently remain below the defined threshold of 1000 $\mu g/m^3$. This can be attributed to the laboratory's-controlled environment, which includes air conditioning (AC) and air purification systems. These controlled conditions effectively maintain clean and healthy air quality, safeguarding against harmful concentrations of CO_2. However, it's essential to recognize that during real-world deployment, the monitoring system will operate in environments that may lack the luxuries of air conditioning and air purification. The CO_2 sensor's data will become even more critical in such settings. The data will be actively transmitted only when the CO_2 levels exceed the critical threshold of 1000 $\mu g/m^3$. This transmission strategy optimizes energy usage and network resources, ensuring that information is shared when it matters most – in the presence of elevated CO_2 concentrations that could potentially impact human health and comfort. This visualization thus underscores the adaptability of the air quality monitoring system, demonstrating its ability to differentiate between safe, well-controlled environments and those where the threat of elevated CO_2 levels looms. It empowers users with the timely information necessary to take appropriate actions, safeguarding against potential health risks and optimizing indoor air quality.

In Figure 5. 6 (b), SO_2 sensor data visualization is a testament to the precision and reliability of the monitoring system. In the controlled environment of the laboratory, the SO_2 sensor consistently registers a reading of 7.3 micrograms per cubic meter ($\mu g/m^3$). This level is significantly below the threshold defined in Table 4. 1, which underscores the absence of this particular pollutant in the laboratory setting. The consistency in the SO_2 sensor data is not surprising, given that the laboratory is a clean and controlled space, free from the emissions and pollutants typically associated with transportation-related activities. In such an environment, the absence of SO_2 is expected, and the sensor readings reaffirm this expectation. However, the significance of this visualization lies in its implications for transportation-related pollution analysis. While the laboratory environment remains unpolluted by SO_2, the monitoring system's capability to measure even trace amounts of this pollutant is crucial for real-world applications. In urban and industrial areas where transportation activities are prevalent, SO_2 emissions from sources like vehicle exhaust can pose significant health and environmental risks. Although well below the defined threshold, the stable reading of 7.3 μg/m³ serves as a baseline for SO_2 levels in clean environments and highlights the system's sensitivity to even minimal concentrations of this pollutant. When deployed in transportation networks or areas with potential SO_2 emissions, the system can detect and report any increases above this baseline, providing essential data for pollution analysis and enabling timely responses to maintain air quality and public health. In essence, this visualization showcases the system's accuracy and consistency and emphasizes its critical role in monitoring pollutants like SO_2 in real-world scenarios. It is a testament to the system's adaptability and the importance of collecting comprehensive air quality data to benefit public health and environmental well-being.

Figure 5. 6(c) gives the data representation from the MQ131 sensor, tasked with capturing ozone (O_3) levels, and offers a unique perspective on the capabilities and potential of the monitoring system. The data in the controlled laboratory experiment shows a constant value of 0. This result is entirely predictable, given that the study was conducted indoors, and indoor settings are typically devoid of ozone exposure, resulting in a stable and unchanging reading. The MQ131 sensor's constant reading of 0 in the laboratory environment serves as a reminder of the controlled conditions of the experiment. Its future role in outdoor deployments promises to provide critical data for ozone concentration, which will contribute significantly to our understanding of air quality and the impact of ozone on public health and the environment. This shift from static indoor reading to real-world outdoor measurements underscores the system's adaptability and potential to contribute to a healthier, more informed society.

Lastly, Figure 5. 6(d) offers a valuable glimpse into the dynamic nature of the NO_2 sensor's measurements. As observed in the laboratory experiment, the data fluctuates, exhibiting variations in response to the environmental conditions. These fluctuations are completely anticipated since nitrogen dioxide (NO_2) levels in the atmosphere can be affected by various factors, including traffic emissions, industrial operations, and meteorological conditions. In the laboratory environment, the NO_2 sensor's data remains consistently below the threshold defined in Table 4. 1, which is 80 micrograms per cubic meter. This outcome is mainly due to the controlled indoor setting, which is not subjected to the various pollution sources and atmospheric dynamics encountered in outdoor environments. However, it is essential to note that the NO_2 sensor is specifically designed for real-world outdoor deployments, where it will fulfill its vital role in monitoring nitrogen dioxide levels. The sensor will transmit data only when the measured NO2 levels exceed the predefined threshold of 80 micrograms per cubic meter. This threshold is under the World Health Organization's (WHO) guidelines for NO2 concentrations in outdoor air. Nitrogen dioxide is a significant air pollutant, primarily from combustion processes in vehicles, power plants, and industrial facilities. Elevated levels of NO2 in the atmosphere can adversely affect human health, particularly the respiratory system. Monitoring and regulating NO2 levels are essential for assessing air quality, protecting public health, and implementing effective pollution control measures. The NO2 sensor's capability to transmit data selectively, focusing on instances when pollution levels surpass the established threshold, is a practical approach to data management. It ensures that the system efficiently utilizes resources and provides the most relevant information for environmental and health assessments. As a result, the NO2 sensor's data will serve as a critical tool in understanding and addressing the impact of nitrogen dioxide on air quality, offering valuable insights and contributing to the overall mission of creating healthier and more sustainable urban environments.

5.5.2. Analysis of Spikes

This research scrutinizes data from each sensor employing the designated algorithm. Presented below are the resultant figures of this analysis. Figure 5. 7 vividly illustrates all sensors, each represented by distinct colors in the dataset.

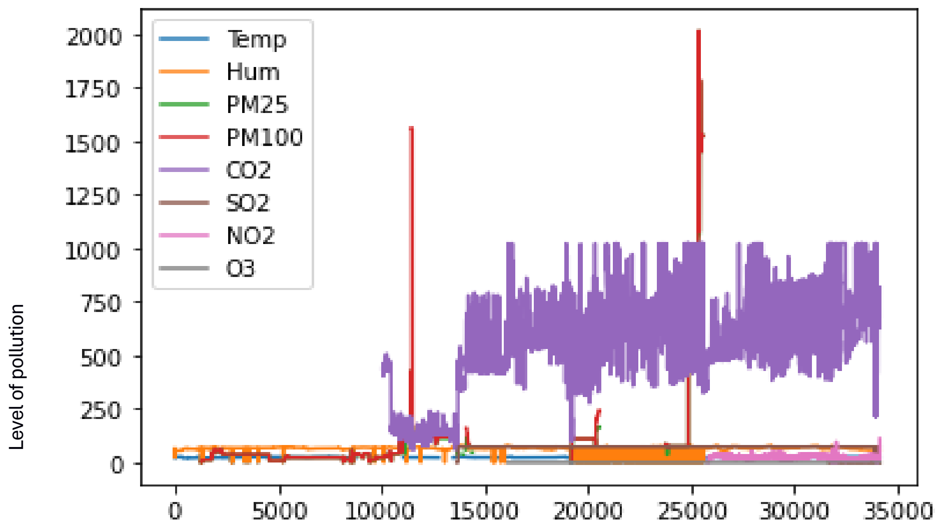


Figure 5. 7 Data from all sensors

Figure 5. 7 offers a comprehensive view of the extensive dataset acquired from various sensors, all meticulously corrected by the prototype developed within the confines of the laboratory at Seoul National University. This rich dataset encompasses a staggering total of approximately 35,000 rows of meticulously collected data. An essential point to consider is that the data points within this dataset do not synchronize with the same starting time due to the unique initiation process of the system. This initiation process commenced precisely when the very first sensor, the temperature and humidity, initiated its data collection. To visually represent these datasets, temperature readings are labeled as 'Temp' and distinguished by a unique shade of blue, while humidity values are denoted as 'Hum' and vividly depicted in a striking orange hue. This figure serves as a window into the intricate world of data collected and is instrumental in understanding the dynamic environmental conditions observed throughout the data acquisition process.

Moving on to the pollution-related sensors, Figure 5. 7 showcases PM2.5 and PM10 data in green and red, respectively. Notably, their values align closely, and further details regarding these pollutants will be presented in subsequent figures, which will delve into each sensor's performance in detail. Notably, the PM2.5 sensor was added shortly after the system's inception, so its data does not initiate at index 0. The presence of peaks in both PM2.5 and PM10 data suggests the occurrence of spikes, emphasizing the necessity for the system to detect these spikes using Algorithm 1.

The purple-colored data points represent readings from the CO_2 sensor, with the graph revealing that these measurements were initiated after the installation of the previously explained sensors. Meanwhile, the SO_2 sensor's data, depicted in brown, seems relatively constant. The gray-colored data signifies ozone levels, which consistently register at zero. Finally, the pink data points correspond to NO_2 measurements, representing the most recent addition to the sensor array. The subsequent section will expound on the discussion of each sensor's results.

This section is an illuminating showcase of the outcomes achieved through the meticulously designed prototype, complemented by the ingenious algorithm at the heart of the research. The primary focus of this section has been detecting pollution spikes within the extensive dataset amassed through the sensor network. As a result, we have gained invaluable insights into air quality dynamics within transportation networks. Furthermore, this section has unveiled the results for each sensor, showcasing their respective performance as observed on the cloud server. These results offer a glimpse into the real-world behavior of these sensors as they capture data from the environment, providing a rich source of information for subsequent analysis and decision-making. In essence, this section represents the culmination of a multifaceted research endeavor, combining cutting-edge technology, mathematical rigor, and practical implementation to achieve a holistic understanding of air pollution in transportation networks. The dataset presented here is a testament to the comprehensive nature of this study, encompassing a wide array of environmental parameters crucial for assessing air quality and its impact on public health. As we embark on the forthcoming section, we will delve even deeper into the interpretation and implications of these results. This analytical journey will unlock the hidden meanings within the data, extract actionable insights, and explore the ramifications of these findings for urban planning, public health, and environmental sustainability.

5.6. Discussions

This section delves into a comprehensive discussion of the results previously presented. It entails thoroughly examining and explaining the outcomes obtained from each sensor. The significance of the data hinges on the algorithm's design and the specific thresholds delineated in Table 4. 1 for each sensor. These pivotal results are graphically depicted in

Figure 5. 8.

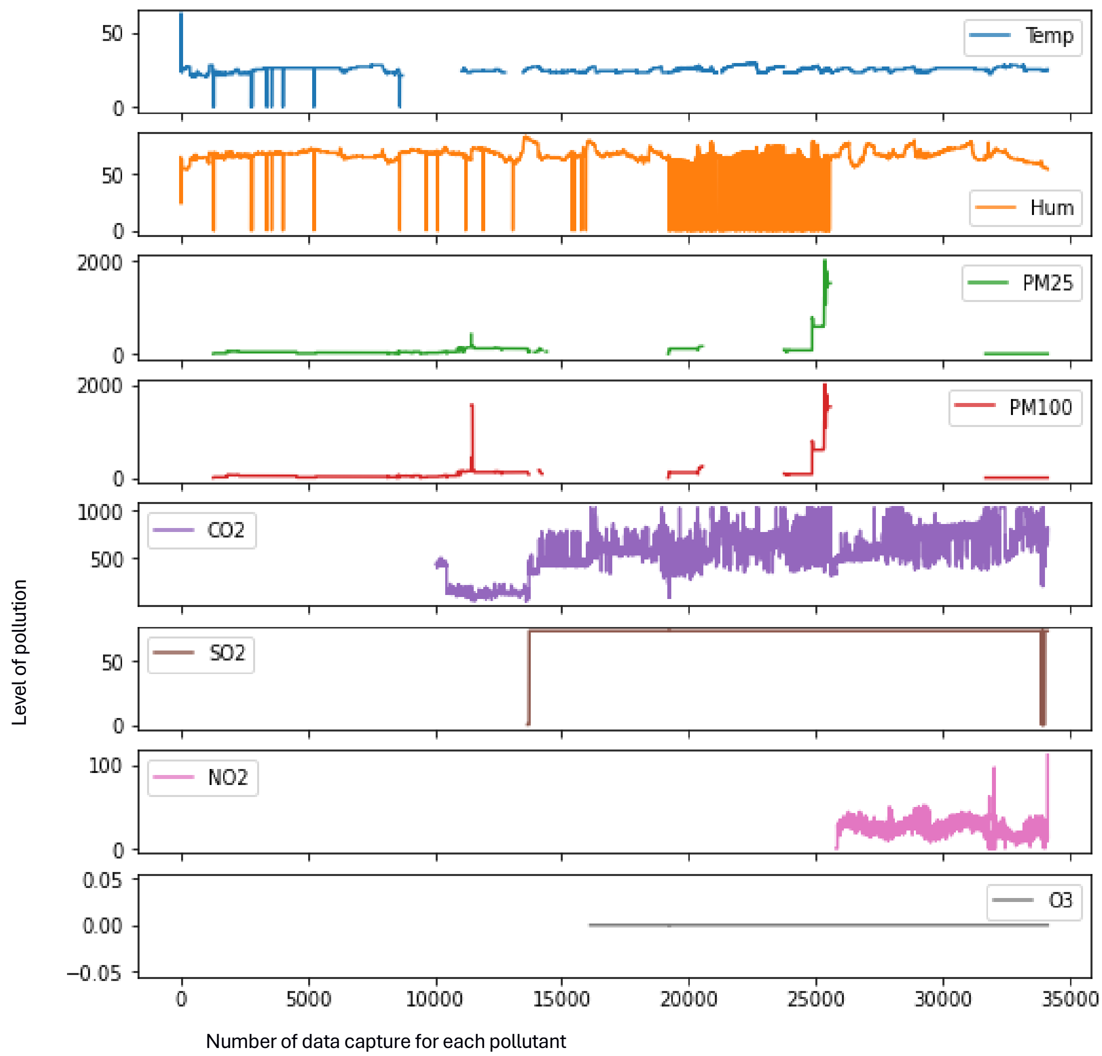


Figure 5. 8 Data manipulation from each sensor.

In

Figure 5. 8, the analysis of the Temperature sensor data reveals intriguing insights. During this experiment, the direct temperature values, prominently indicated within the temperature portion of the dataset, were recorded exclusively during intervals when the sensor refrained from transmitting data to the cloud. This intriguing behavior stemmed from the ambient temperature being consistently maintained below the sensor's predefined threshold of 22 degrees Celsius. Consequently, the temperature remained relatively stable throughout the experiment. On the contrary, the Humidity data exhibited a stark contrast, maintaining a remarkable consistency and

showing negligible fluctuations. However, it is important to note that discrete data points might emerge by marginally lowering the system's defined threshold, shedding light on the subtle nuances of environmental conditions.

The data stemming from the PM2.5 sensor, capable of capturing particles with diameters of 2.5 micrometers, and the PM10 sensor, which accommodates larger 10-micrometer particles, exhibit a discrete pattern. When these particulate matter levels fall below their predefined thresholds, they are incorporated into the dataset, offering a distinct visualization of air quality. With its stable air conditioning (AC) system, the laboratory environment facilitated the successful capture and transmission of data from the CO2 sensor. However, it is imperative to define appropriate thresholds for this sensor since it sporadically detects signal spikes, a topic of forthcoming discussion in the subsequent section.

The performance of the SO2 sensor is equally fascinating. The data originating from this sensor remains constant throughout the experiment, consistently measured well below the threshold of 50 micrograms per cubic meter, equivalent to 0.05 parts per million (ppm). To visualize these minuscule values more effectively on the graph, they have been ingeniously scaled by a factor of 1000. In the controlled laboratory setting, the SO2 sensor consistently registers a low 0.0073 ppm, highlighting its precision and reliability in monitoring this pollutant, which, while largely absent within the laboratory, plays a pivotal role in transportation-related pollution analysis. These intriguing findings set the stage for deeper discussions and explorations in the subsequent sections of our research.

The data obtained from the NO2 sensor indicates that the expected values are typically below 80 micrograms per cubic meter, which aligns with the threshold defined in Table 4. 1. However, it's important to note that the NO2 sensor installed in the laboratory setting demonstrates some variability in its readings over time. This variability is evident in the fluctuations observed in the data. Upon closer examination, it becomes apparent that the NO2 sensor readings occasionally exhibit spikes, as illustrated in

Figure 5. *8*.

These spikes represent instances where the NO2 levels temporarily rise above the anticipated range, reaching values around 100 micrograms per cubic meter. This phenomenon underscores the sensor's capability to detect short-lived pollution events or localized sources of NO2 emissions.

These occasional spikes provide valuable insight into the dynamic nature of air quality, even in a controlled laboratory environment. It serves as a reminder that many factors, including environmental conditions, human activities, and industrial processes, influence pollution levels. The NO2 sensor's capacity to record these variations and spikes establishes it as a vital component in our air quality monitoring system, mainly when utilized in real-world environments characterized by a broader range of pollution sources that can be highly unpredictable. The subsequent sections of our research will delve deeper into interpreting such spikes and their implications for monitoring and managing air quality in transportation networks.

The data derived from the ozone sensor within the laboratory environment consistently registers at a value of zero. This outcome is attributed to the specific conditions within the room, characterized by continuous air conditioning (AC) and limited exposure to outdoor pollutants. The indoor atmosphere, as a controlled environment, starkly contrasts the conditions in which this system is intended to operate when deployed outdoors. In outdoor settings, environmental conditions tend to be more dynamic and pollutant-laden. The ozone levels in these settings are anticipated to be substantially higher than the constant zero values observed in the laboratory. Outdoor environments inherently contain many pollutants stemming from various sources, including vehicular emissions, industrial activities, and natural factors. The exhaust emitted by vehicles, where this system is primarily designed to be deployed, is particularly significant in contributing to elevated ozone levels and other harmful pollutants. As a result, the ability to capture and analyze real ozone measurements in these dynamic outdoor conditions is of paramount importance. The system is poised to provide invaluable insights into the complex interplay of pollutants and environmental factors, furthering our understanding and enhancing our ability to address air quality challenges effectively. The forthcoming sections of our research will explore the implications of these findings in greater detail.

In Figure 5. 9, the data from various sensors are visually represented in their respective distributions based on the median values. The graph clearly illustrates that the data about different pollutants are entirely independent.

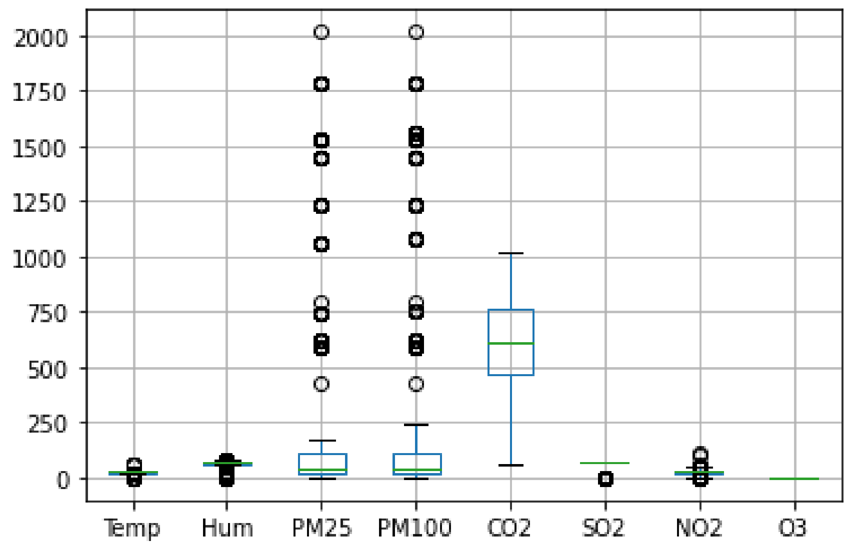


Figure 5. 9 Distribution of data from all sensors

In Figure 5. 9, the data extracted from the PM2.5 sensor paints a compelling picture, revealing a substantial presence of spikes within the dataset. The dataset contains over 2000 instances of elevated values for PM2.5, surpassing the established threshold of 35 micrograms per cubic meter, as indicated in Figure 5. 9. These thresholds serve as essential benchmarks, setting the upper limits for acceptable air quality. To put this into perspective, the World Health Organization (WHO) and other environmental regulatory bodies have defined these thresholds to safeguard public health. In ideal conditions, air quality should not exceed these limits to prevent adverse health effects, particularly on vulnerable populations. Hence, the abundance of spikes, each indicative of elevated PM2.5 levels, is a cause for concern.

Moreover, it's essential to recognize that these spikes in PM2.5 levels are intertwined with factors that can directly impact air quality. Transportation activities, industrial processes, and environmental conditions can all contribute to fluctuations in PM2.5 concentrations. Monitoring and addressing these spikes is paramount to mitigate the potential health risks posed by exposure to delicate particulate matter, which can infiltrate the respiratory system and lead to various health issues, including respiratory diseases, cardiovascular problems, and even premature mortality. The revelation of over 2000 elevated PM2.5 values underscores the urgent need for proactive air quality management and stringent measures to curb pollution sources. Whether through targeted policies, emissions controls, or public awareness campaigns, addressing the root causes of these spikes is essential to ensure clean and breathable air for all. These findings serve as a potent reminder of the interconnectedness of environmental conditions and human health, emphasizing the critical role that real-time monitoring and data analysis play in safeguarding our well-being and the environment. As the paper proves, these values can harm children [25].

Based on the results obtained, it becomes evident that this paper has successfully developed the proposed system, which finds direct applicability in the vehicles discussed in Paper [85]. Furthermore, the findings of this research shed light on the considerable impact of memory consumption when utilizing this system, as demonstrated in Paper [83]. Papers [41-45] presented various suggestions for addressing air pollution in smart cities, and this research offers a concrete solution to these proposals, marking a significant contribution to the field.

The "Prototype for Monitoring Transportation Pollution Spikes through IoT Edge Networks" study offers valuable insights into real-time air pollution detection, particularly concerning the dynamic context of transportation. In this section, we delve into a comprehensive analysis of the study's findings, shedding light on the behavior of various sensors employed to measure critical environmental parameters and pollutant levels within a meticulously controlled laboratory setting.

The study's results cast a spotlight on several key observations. Firstly, the data acquired from the temperature sensor unveils a clear correlation with data transmission. Specifically, the temperature data was recorded when the sensor was not transmitting to the cloud, coinciding with temperatures below 22 degrees Celsius. In contrast, the humidity data remained consistent throughout the experiment and did not exhibit notable fluctuations below the predefined threshold. However, the study raises the possibility that the system's entry should be slightly adjusted; discrete data points could emerge in the humidity dataset.

Moreover, the data from the PM2.5 sensor, which measures particulate matter levels with 2.5mm and 10mm diameters, exhibited discrete patterns. These discreet data points were primarily observed when the sensor's readings fell below the predefined threshold values. This underscores the sensor's ability to discern and transmit data effectively when pollution levels surpass the specified thresholds.

Considering that the laboratory environment in which the system was situated maintained a stable air conditioning (AC) system, the data captured by the CO2 sensor remained consistent and aligned with expectations. However, it's worth noting that, on occasion, this sensor detected spikes in the signals. These occasional spikes raise intriguing questions about the dynamics of indoor air quality, hinting at the need for further exploration to comprehend these phenomena better.

The study also examined the data from the SO_2 sensor, which exhibited a remarkable degree of stability. The data consistently registered at a low level, well below the defined threshold of 50 micrograms per cubic meter, equivalent to 0.05 parts per million (ppm). To enhance the visualization of these small values on the graph, they were scaled by a factor of 1000. The laboratory environment was notably free of this pollutant, crucial for transportation-related pollution analysis.

In contrast, the ozone (O_3) sensor readings remained zero within the indoor laboratory setting. However, it's imperative to acknowledge that these values are expected to significantly deviate from this consistent zero level when deployed in outdoor environments. Outdoor settings inherently contain a broader spectrum of pollutants, making the accurate measurement and monitoring of ozone levels a pivotal aspect of the study's real-world applicability, particularly in transportation.

Last, the study scrutinized data obtained from the NO_2 sensor. It was determined that the expected values should generally fall below 80 micrograms per cubic meter, adhering to the threshold outlined in Table 4. 1. However, as the sensor was also situated within the laboratory, it exhibited variation in readings over time. Notably, the data displayed occasional spikes, as shown in Figure 15, with readings occasionally surging to around 100 micrograms per cubic meter. These intermittent spikes underscore the sensor's potential to detect and respond to short-lived pollution events or localized sources of NO_2 emissions, an invaluable attribute for monitoring air quality in dynamic transportation scenarios.

In conclusion, the "Prototype for Monitoring Transportation Pollution Spikes through IoT Edge Networks" study offers a wealth of insights into real-time air pollution detection. These findings collectively advance our understanding of this critical field, particularly within the context of transportation. The results provide valuable implications for health and environmental management, underscoring the significance of proactive and precise air quality monitoring to ensure the well-being of communities and the sustainable management of our shared environment.

5.7 Conclusion

In summary, the chapter titled "Prototype for Monitoring Transportation Pollution Spikes through IoT Edge Networks" stands as a significant milestone in air pollution monitoring and management, with a particular focus on the transportation sector. The research showcased an innovative approach that leverages the power of IoT edge networks and a diverse array of sensors to enable real-time detection of air pollution spikes. Sensors for PM2.5, PM10, CO2, NO2, SO2, and O3 were used to capture pollutants from vehicles during the movement for vehicles. The study's results speak volumes about the system's efficiency in capturing and analyzing pollutant spike data, offering valuable insights into the intricacies of environmental dynamics within the controlled confines of a laboratory setting. Results from the laboratory show that O3 is at zero and SO2 is not modifying but cannot cause any harm. PM 2.5, PM10, CO2, and NO2 all these sensor values are modified accordingly with some spikes to the environment from the laboratory. These findings transcend the confines of the laboratory and hold immense significance for public health and urban environmental management. The ability to promptly identify pollution spikes in real time equips us with the means to initiate swift interventions and preventive measures, ultimately safeguarding the well-being of urban populations. This research lays a strong foundation for advancing air quality monitoring in the context of transportation, promising to contribute to creating healthier and more sustainable urban environments. The research is poised to take a significant step forward by shifting its focus to encompass a broader analysis of data collected from various locations where the system is deployed. This expansion will usher in new challenges, particularly regarding data management and security. In this regard, integrating blockchain technology plays a pivotal role in addressing these challenges, ensuring the integrity and traceability of the data collected. Additionally, adopting machine learning models is set to enhance the system's predictive capabilities, allowing for a more proactive approach to pollution management. As the research embarks on this next phase, it holds the potential to revolutionize how we perceive and address air quality issues in transportation, fostering a cleaner and healthier future for urban dwellers.

Eric Nizeyimana, September 2024,

Chapter 6. Revolutionizing Air Pollution Spikes Analysis with a Blockchain-Driven Machine Learning Framework

6.1. Overview

Air pollution spikes pose significant health risks and environmental challenges that demand innovative solutions for effective analysis and mitigation. This paper introduces a groundbreaking approach to revolutionize air pollution spikes analysis using a blockchain-driven machine learning framework. Leveraging the transparency and immutability of blockchain technology, coupled with the predictive power of machine learning algorithms, our framework offers real-time monitoring, accurate prediction, and proactive management of air pollution spikes. By integrating data from diverse sources, including IoT sensors, our framework provides comprehensive insights into air quality dynamics. Furthermore, the decentralized nature of blockchain ensures data integrity and enhances trust among stakeholders, including regulatory authorities, industries, and communities. Through case studies and simulations, we demonstrated the efficacy and scalability of our framework in addressing air pollution spikes across diverse geographical regions. The Machine learning techniques for the Timeseries model (RNNs, ARIMA, and Exponential Smoothing) were analyzed and compared using statistical metrics (R-squared (R^2), Mean Absolute Error (MAE), and Mean Squared Error (MSE)). The exponential Smoothing model performed well compared to the other two models for all parameters. This research signifies a paradigm shift in air quality management, empowering stakeholders to make informed decisions and mitigate the adverse impacts of air pollution spikes on public health and the environment. This research demonstrated that machine learning and blockchain can be put together to analyze data on air pollution spikes and make predictions by ensuring the emitter of pollutants. This solution will ensure the prevention of exposure which can harm human beings and the environment.

6.2. Introduction

Air pollution stands as one of the most urgent and far-reaching challenges of our time, impacting public health, environmental stability, and economic prosperity on a global scale [51], [95], [96], [97], [98]. Despite concerted efforts to curb pollution levels, the occurrence of air pollution spikes remains a persistent threat, marked by sudden and drastic surges in pollutant concentrations [97]. These spikes pose significant challenges for policymakers, researchers, and citizens worldwide, demanding innovative solutions to effectively monitor, analyze, and mitigate their impacts [99], [100].

The consequences of air pollution spikes are profound and multifaceted [101]. From exacerbating respiratory illnesses to damaging ecosystems and jeopardizing economic growth, the ramifications are far-reaching and often long-lasting [9], [102], [103]. Moreover, vulnerable populations, including children, the elderly, and those with pre-existing health conditions, are disproportionately affected, exacerbating existing health disparities and placing additional strain on healthcare systems [61], [98][61], [98].

Traditional methods for analyzing and addressing air pollution spikes are often inadequate in addressing these complex challenges [104]. Conventional monitoring systems, reliant on static sensors and manual data collection, struggle to provide timely and accurate insights into evolving pollution dynamics [105]. As a result, decision-makers are often left with incomplete or outdated information, hindering their ability to implement targeted interventions and mitigate the impacts of pollution spikes effectively.

Furthermore, the lack of transparency and real-time responsiveness inherent in traditional approaches exacerbates the challenge of addressing air pollution spikes[106]. Without timely and reliable data, policymakers, researchers, and citizens face significant barriers to understanding the root causes of pollution spikes and devising evidence-based strategies to address them.

In response to these challenges, there is a growing recognition of the need for innovative approaches that leverage emerging technologies such as IoT (Internet of Things) and AI (Artificial Intelligence) to revolutionize the monitoring and analysis of air pollution spikes [107], [108]. By harnessing the power of IoT-enabled sensor networks, researchers can collect real-time data on pollutant levels, meteorological conditions, and other relevant factors, providing a more comprehensive understanding of pollution dynamics [107], [108].

6.3. Related Works

This section delves into the related work concerning the integration of machine learning and blockchain technologies, particularly in the context of air pollution spike detection and prediction with an emphasis on security applications. It explores various studies that have harnessed machine learning algorithms to analyze and predict air quality trends, effectively identifying pollution spikes. Additionally, it examines the role of blockchain technology in ensuring data integrity, transparency, and security within these systems. By integrating the predictive capabilities of machine learning with the decentralized and immutable nature of blockchain, researchers have developed robust solutions to enhance the reliability and trustworthiness of air pollution monitoring. This section provides a comprehensive review of methodologies, models, and frameworks employed in previous research, offering valuable insights into the current advancements and identifying potential areas for further exploration in this field.

AI-driven analytics holds the promise of unlocking valuable insights from vast and complex datasets, enabling predictive modeling, trend analysis, and early warning systems for pollution spikes [108]. By combining advanced analytics with real-time monitoring capabilities, stakeholders can gain actionable insights into pollution trends, identify high-risk areas, and implement targeted interventions to mitigate the impacts of pollution spikes effectively [109].

In response to the persistent challenges posed by air pollution spikes, a groundbreaking approach is emerging at the intersection of blockchain technology and machine learning (ML) frameworks [110]. This innovative fusion of technologies holds the promise of revolutionizing the analysis of air pollution spikes, offering a transformative solution to address these complex and pressing environmental issues [111].

At its core, this innovative framework integrates the inherent strengths of blockchain technology and ML algorithms to create a robust and transparent system for monitoring and analyzing air pollution spikes [112]. Blockchain, renowned for its immutable ledger capabilities, serves as the foundation for securely recording and storing pollution data, ensuring data integrity, transparency, and traceability throughout the analysis process [113]. By leveraging blockchain's tamper-proof ledger, stakeholders can have confidence in the accuracy and reliability of the data used for pollution analysis, mitigating concerns regarding data manipulation or tampering.

Complementing blockchain, ML algorithms bring powerful analytical capabilities to the table, enabling the extraction of valuable insights from the vast and complex datasets generated by air pollution monitoring systems [112]. Through advanced data analytics techniques such as pattern recognition, anomaly detection, and predictive modeling, ML algorithms can uncover hidden patterns, trends, and correlations within the pollution data, providing deeper insights into the dynamics of air pollution spikes [110]. By harnessing the analytical power of ML, stakeholders can gain a comprehensive understanding of air pollution trends, identify contributing factors to pollution spikes, and even predict future occurrences with greater accuracy [114].

The paper [115] utilizes reinforcement learning (RL) to develop an optimal Bitcoin-like blockchain mining strategy that adapts to changing network conditions without requiring detailed knowledge of the blockchain model. The introduced multidimensional RL algorithm achieves near-optimal performance by addressing the non-linear objectives of the mining Markov Decision Process (MDP), demonstrating effectiveness in dynamic blockchain networks.

The research [116], [117], [118] introduces a novel reference architecture for blockchain-enabled federated learning (BCFL), integrating smart contracts, IPFS, and decentralized identifiers to enhance the flexibility and extensibility of federated learning processes, and verifies the architecture in a real-world Ethereum environment.

The integration of blockchain technology and ML frameworks offers a host of potential benefits for environmental management and public health [119]. By providing a secure and transparent platform for data sharing and collaboration, blockchain facilitates greater cooperation among stakeholders, including government agencies, research institutions, and citizen scientists, in addressing air pollution challenges [41], [45], [46], [120]. Furthermore, the analytical capabilities of ML enable more effective decision-making and intervention strategies, allowing policymakers to implement targeted measures to mitigate the impacts of pollution spikes and protect public health [77].

Blockchain technology, initially conceived as the backbone of cryptocurrencies, has emerged as a transformative force across diverse sectors due to its capacity to augment transparency, security, and decentralization [121]. In the realm of air pollution analysis, blockchain's inherent features present a unique opportunity to address longstanding challenges related to data integrity and accountability.

The capability of blockchain to establish immutable records of data sources, measurements, and transactions is particularly relevant in the context of air pollution analysis [110]. By leveraging blockchain technology to store pollution data on a decentralized ledger, stakeholders can effectively safeguard against data manipulation and ensure the veracity of information throughout the analysis process [113]. This tamper-proof nature of blockchain records instills confidence in the reliability and authenticity of pollution data, thereby enhancing trust among stakeholders.

Furthermore, the decentralized nature of blockchain ensures that pollution data is not controlled by a single entity, mitigating the risk of data manipulation or bias [113]. This decentralization fosters greater transparency and accountability, as all transactions recorded on the blockchain are visible to participants in the network. As a result, stakeholders can trace the origin of pollution data, verify its accuracy, and hold accountable those responsible for data collection and analysis.

Complementing blockchain technology, machine learning (ML) algorithms present powerful tools for extracting valuable insights from vast datasets, including those related to air pollution analysis [112]. By leveraging historical pollution data, meteorological variables, and other relevant factors, ML models can uncover hidden patterns, trends, and anomalies, offering a deeper understanding of the dynamics of air pollution spikes.

Historical pollution data serves as a rich source of information for ML algorithms, enabling them to identify correlations and dependencies between different variables. By analyzing historical trends, ML models can discern patterns indicative of pollution spikes and understand the underlying factors contributing to their occurrence [119]. This retrospective analysis lays the groundwork for more accurate predictions and proactive interventions in the future.

Moreover, ML-driven analysis facilitated the prediction of future pollution occurrences by extrapolating from historical data and incorporating real-time inputs from monitoring systems [122]. By learning from past events and adapting to changing environmental conditions, ML models can forecast the likelihood and severity of pollution spikes with increasing accuracy. This predictive capability empowers stakeholders to take proactive measures and implement targeted interventions to mitigate the impacts of pollution spikes on public health and the environment.

This study introduced an innovative approach to revolutionizing the analysis of air pollution spikes by harnessing the power of a groundbreaking blockchain-driven machine-learning framework. Air pollution stands as a significant threat to public health and environmental well-being, demanding novel solutions for effective monitoring and analysis. Traditional methods often fall short due to their lack of real-time capabilities and challenges in interpreting complex datasets.

In this study, the paper addresses these shortcomings by developing a machine-learning model capable of predicting pollution data transmitted to the cloud in real-time. We evaluated various types of machine learning models and found that exponential smoothing time series outperformed alternative approaches, demonstrating superior performance in detecting pollution spikes.

The utilization of machine learning models enables us to analyze pollution data with unmatched precision and efficiency, facilitating the timely detection of spikes in pollutant concentrations. By leveraging historical data and adaptive algorithms, our framework can forecast future pollution events, empowering stakeholders to take proactive interventions and make well-informed policy decisions.

6.4. Material and Methods

In this section, the paper elaborates on all materials and methods used to combine blockchain techniques with machine learning models for analyzing air pollution spikes data. The paper used the dataset for air pollution collected at Seoul National University (SNU) for a month for blockchain. In contrast, for Machine learning the dataset [123] for air pollution in Seoul city for three years was used to train and test models.

6.4.1. IoT Data Collection and Integration

Data used in this research work are of two types. Firstly, there is some data collected using the developed IoT devices presented in the paper [109]. These data were collected in the SNU and have been used in this paper for the blockchain technology part. Sensor data was collected based on the threshold stored in a local database and then hashed on the blockchain to ensure immutability and transparency. Then the smart contract deployed on the Ethereum blockchain is designed to manage air quality data and enforce thresholds for various pollutants. This contract enables the creation of individual air quality contracts, each tailored to specific pollutant thresholds and associated fines.

Nodes which are individual computers maintain a copy of the entire blockchain and participate in the network by validating and propagating transactions and blocks. Smart contracts are self-executing contracts with the terms of the agreement directly written into code, which are deployed on the blockchain. These contracts automatically enforce and execute the terms of the agreement when predefined conditions are met, ensuring trust, transparency, and immutability in transactions without the need for intermediaries.

6.4.2. Blockchain Smart Contract Design

This research used the Ethereum blockchain-based platform for its robustness and support for smart contracts. Air pollution spikes data are sent across different geographical nodes to ensure decentralization. These data are sent in the way of blocks and hash and encryption keys are generated and applied to them. Then smart contracts were developed to ensure the individuals contract based on the threshold of air pollution spikes.

The smart contracts are defined by the contract name and in this paper is AirQualityContract. It has three functionalities which are the creation of individual air quality contracts, setting and retrieving air quality data, and enforcement of pollutant thresholds and fines.

Figure 6. 1 shows the structure of the contract.

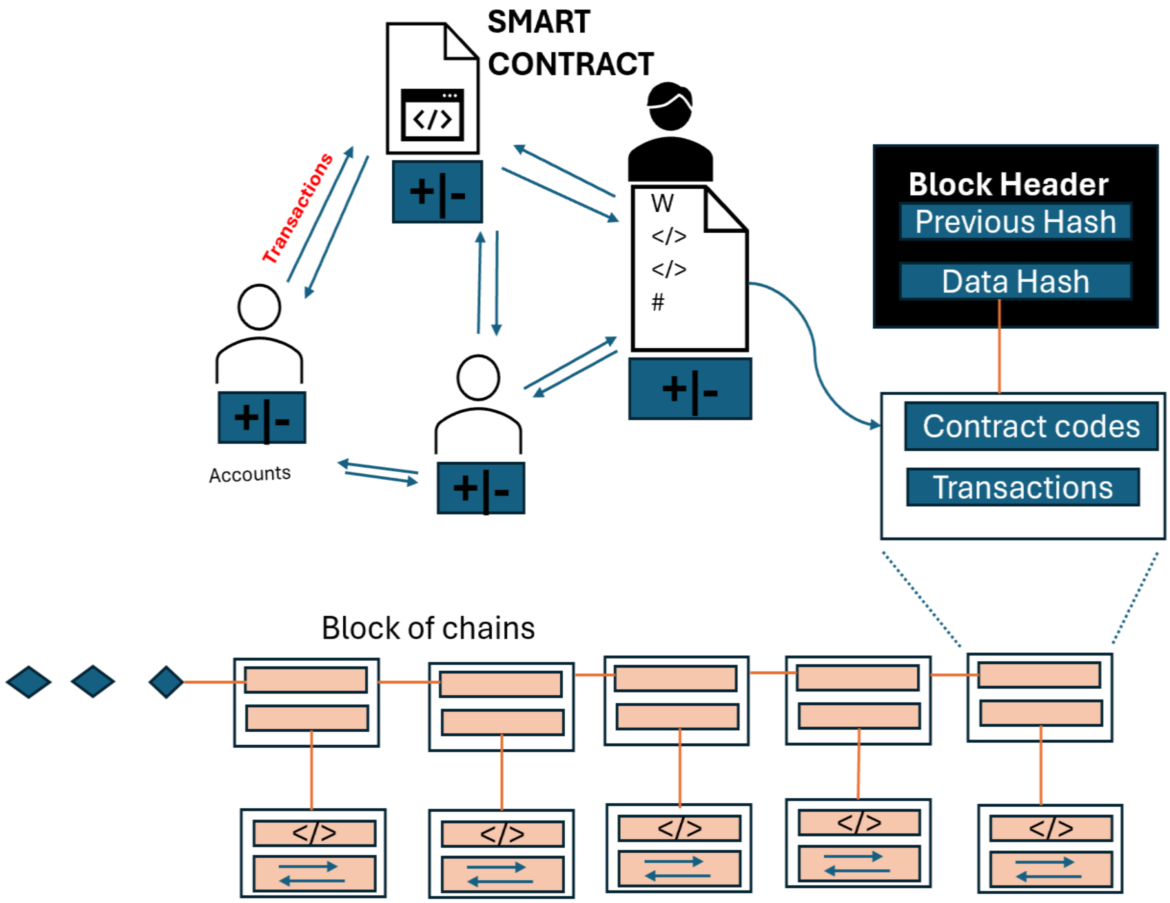


Figure 6. 1 Blockchain ledger system with smart contract build-in

In the Figure 6. 1, the design of our blockchain system is composed of different parties including the smart contract, the transaction, and the data security to be sent to different nodes.

The smart contract consists of the following main components in this thesis:

- **Constructor:** Initializes the contract with pollutant thresholds and fine amounts.
- **setAirQualityData**: Allows for the setting of air quality data for PM2.5, PM10, O3, SO2, NO2, and CO2.
- **getAirQualityData:** Retrieves the stored air quality data.
- **enforceThresholds:** Checks if the current air quality data breaches any thresholds and applies fines accordingly.

The smart contract interacts with external data sources, such as CSV files containing air quality data. The process involves:

1. **Data Import:** The user uploads a CSV file containing air quality data to an external platform.
2. **Data Processing:** A Python script reads the CSV file, extracts the air quality data, and formats it for interaction with the smart contract.
3. **Smart Contract Interaction**: The formatted air quality data is passed to the smart contract using Web3.js, where it is stored and processed.
4. **Threshold Enforcement**: The smart contract compares the received air quality data against predefined thresholds and applies fines if necessary.

The smart contract presented in this chapter offers a robust framework for managing air quality data and enforcing pollutant thresholds. By incorporating key components such as the constructor for initialization, setAirQualityData for data input, getAirQualityData for data retrieval, and enforceThresholds for threshold enforcement, the contract ensures comprehensive monitoring and regulation of air quality parameters. The integration with external data sources through a seamless process involving CSV file uploads, Python-based data extraction, and Web3.js interaction highlights the contract's versatility and practicality. This approach not only enhances the accuracy and reliability of air quality management but also automates the enforcement of environmental standards, thereby contributing to improved public health and environmental protection.

6.4.3. Modeling Air Pollution Spikes Using Machine Learning

This research employs machine learning to predict future air pollution spikes by evaluating multiple models. The analysis identified the exponential smoothing time series model as the best performer among the top three models tested. All models, which include exponential smoothing time series, Recurrent Neural Networks (RNNs), and ARIMA, rely on historical time series data for predictions. The performance of these models was rigorously assessed using Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2) metrics.

This research uses a time series model since it is designing an ML algorithm and analyzing it based on the time-ordered data points to predict future values based on historical patterns in a dataset.

The appropriate model for time series is chosen based on the data characteristics and forecasting goals. The historical data were used to train the model by adjusting its parameters (SO_2, NO_2, PM2.5, PM10, O_3, and CO) to minimize the error between predicted and actual values. To validate the model, the dataset of each parameter was split into training (80% of data) and validation (20% of the data) sets to tune hyperparameters and prevent overfitting. Once trained and validated, the model is used to make future predictions. The model generates 10 forecasts based on past data and learned patterns.

In the comparison of these three models, the paper is based on the way they are built mathematically. Each of these models relies on various mathematical concepts to analyze and predict future data points. Let's look at their concept below:

6.4.3.1. Exponential Smoothing

Exponential smoothing is a widely used technique for time series forecasting that assigns exponentially decreasing weights to past observations [124]. This means that more recent data points are given higher importance in the forecast than older ones. Here's a brief note on exponential smoothing applied to historical data: Exponential smoothing models work by recursively updating a forecast based on the previous forecast and the most recent observed value. The basic idea is to calculate a smoothed forecast by combining the previous forecast with a fraction of the latest observation, adjusted by a smoothing parameter. Below are the mathematical components used to build it:

$$y'_t = \alpha y_{t-1} + (1-\alpha)\, y'_{t-1} \qquad (6.1)$$

Where y'_t is the forecasted value at time t, y_{t-1} is the observed value at time t-1, α is the smoothing parameter ($0 < \alpha < 1$).

6.4.3.2. Recurrent Neural Networks (RNNs)

Recurrent Neural Network (RNN) time series models leverage historical data to forecast future values or perform other tasks such as anomaly detection or classification[125]. These models are particularly well-suited for analyzing time series data because they capture temporal dependencies and patterns. RNNs are designed to learn from sequential data, making them effective at capturing temporal dependencies within a time series [125]. By analyzing historical data points in sequence, RNNs can identify patterns and trends that influence future values.

Once trained, RNN time series models can be used to forecast future values based on historical data. These predictions can provide valuable insights for decision-making in various domains, including finance, energy, healthcare, and climate science. Below are the mathematical components used to build it:

$$h_t = \sigma(W_h h_{t-1} + W_x x_t + b) \qquad (6.2)$$

$$y_t = \phi(W_y h_t + c) \qquad (6.3)$$

Where h_t is the hidden state at time t, x_t is the input, y_t is the output, W_y , W_x , W_h are weight matrices, b, and c are biases, and σ, ϕ are activation functions.

6.4.3.3. Autoregressive Integrated Moving Average (ARIMA)

ARIMA (Autoregressive Integrated Moving Average) modeling is a widely used technique for time series analysis and forecasting. It combines three key components: Autoregressive (AR) terms: These represent the relationship between a variable and its own lagged values. AR terms capture the linear dependence between the variable and its past values [124]. Integrated (I) term: This component accounts for the differencing of the time series data to make it stationary. Stationarity is crucial for many time series models, including ARIMA. Moving Average (MA) terms: These represent the relationship between a variable and the residual errors from a moving average model applied to lagged values of the variable. MA terms capture the short-term changes or fluctuations in the time series [124]. ARIMA models are particularly useful when the time series data exhibit non-stationarity, trends, and seasonal patterns. By fitting an ARIMA model to historical data, analysts can make forecasts and predictions about future values of the time series. Below are the mathematical components used to build it:

AR(p) Model:

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \cdots + \Phi_p y_{t-p} + \epsilon_t \qquad (6.4)$$

Where Φ_i are the autoregressive coefficients, $\in_t$ is white noise.

MA(q) Model:

$$y_t = \in_t + \theta_1 \in_{t-1} + \theta_2 \in_{t-2} + \cdots + \theta_q \in_{t-q} \qquad (6.5)$$

Where θ_i are the moving average coefficients.

ARIMA(p, d, q) Model:

$$\Delta^d y_t = \Phi_1 \Delta^d y_{t-1} + \cdots + \Phi_p \Delta^d y_{t-p} + \in_t + \theta_1 \in_{t-1} + \cdots + \theta_q \in_{t-q} \qquad (6.6)$$

where Δ represents differencing to achieve stationarity, and d is the order of differencing.

This section outlines the materials and methods used in this study. Blockchain contracts were designed and developed to handle data collected by the devices described earlier in the paper. These contracts were deployed to enforce fines on air pollution emitters. Additionally, machine learning models were designed and analyzed to determine the most suitable model for the dataset from [123]. The study identified the use of time series analysis and compared three models: ARIMA, RNNs, and Exponential Smoothing. The following section presents the results and discussion.

6.5. Results and Discussions

This section presents the results based on the materials and methods described previously. The primary outcome is the implementation of a smart contract on the Ethereum platform. These contracts were successfully deployed, enabling the enforcement of fines on air pollution emitters that exceed set thresholds. Furthermore, various machine learning models were compared to identify the most suitable model for air pollution data. Time series models, including ARIMA, RNNs, and Exponential Smoothing, were analyzed and compared, and the best-performing model was selected based on its performance metrics.

6.5.1. Smart Contract Implementation

In implementing the smart contract, data were uploaded and read as on the below graph Figure 6. 2:

```
Choose Files  Eric - All data new.csv
• Eric - All data new.csv(text/csv) - 457452 bytes, last modified: 5/13/2024 - 100% done
Saving Eric - All data new.csv to Eric - All data new (4).csv
                        Time  Temp  Hum  PM25  PM100  CO2    SO2  NO2  O3
0  2023-08-23T08:06:50+00:00    25   66    22     23  474  73.14   21   0
1  2023-08-23T08:07:16+00:00    24   66    22     23  500  73.14   24   0
2  2023-08-23T08:07:36+00:00    24   66    22     23  528  73.14   26   0
3  2023-08-23T08:07:56+00:00    25   66    22     23  502  73.14   28   0
4  2023-08-23T08:08:16+00:00    24   66    22     23  481  73.14   22   0
```

Figure 6. 2 The output of the loaded data for preparing their smart contract

These data represent the air pollution data collected from the developed device presented in the paper [109]. These are the data used to develop and deploy the smart contracts to fine all emitters exceeding the threshold.

The development of the contract was carried out on the Solidity platform, a widely used programming language for writing smart contracts on the Ethereum blockchain. The contract was designed specifically to impose fines on air pollution emitters who exceed predefined thresholds. This design ensures that the contract can autonomously monitor emissions data and enforce penalties, promoting compliance with environmental regulations.

Once the contracts were developed, they were deployed on the Ethereum blockchain. Deployment involves uploading the contract code to the blockchain network, making it immutable and accessible to all network participants. This step is crucial as it ensures that the contract is publicly verifiable and tamper-proof, thereby enhancing transparency and trust among stakeholders.

After deployment, the contracts became active and ready to perform transactions. When an emitter's pollution levels surpass the threshold, the contract automatically triggers a fine. This transaction is recorded on the blockchain, providing a transparent and permanent record of the enforcement action. The use of blockchain technology thus ensures that the process of monitoring and penalizing pollution is both efficient and reliable, leveraging the inherent benefits of decentralization and immutability.

Figure 6. 43 and Figure 6. 4 illustrates the deployment process within the system, which fines all emitters exceeding the threshold based on the developed contract. When an emitter surpasses the set pollution limits, the system calculates the equivalent gas fee and initiates a transaction to impose the fine. Once the fine is applied, the system records the transaction on the blockchain. This ensures a transparent and permanent log of all enforcement actions, maintaining accountability and allowing for easy verification by all stakeholders.

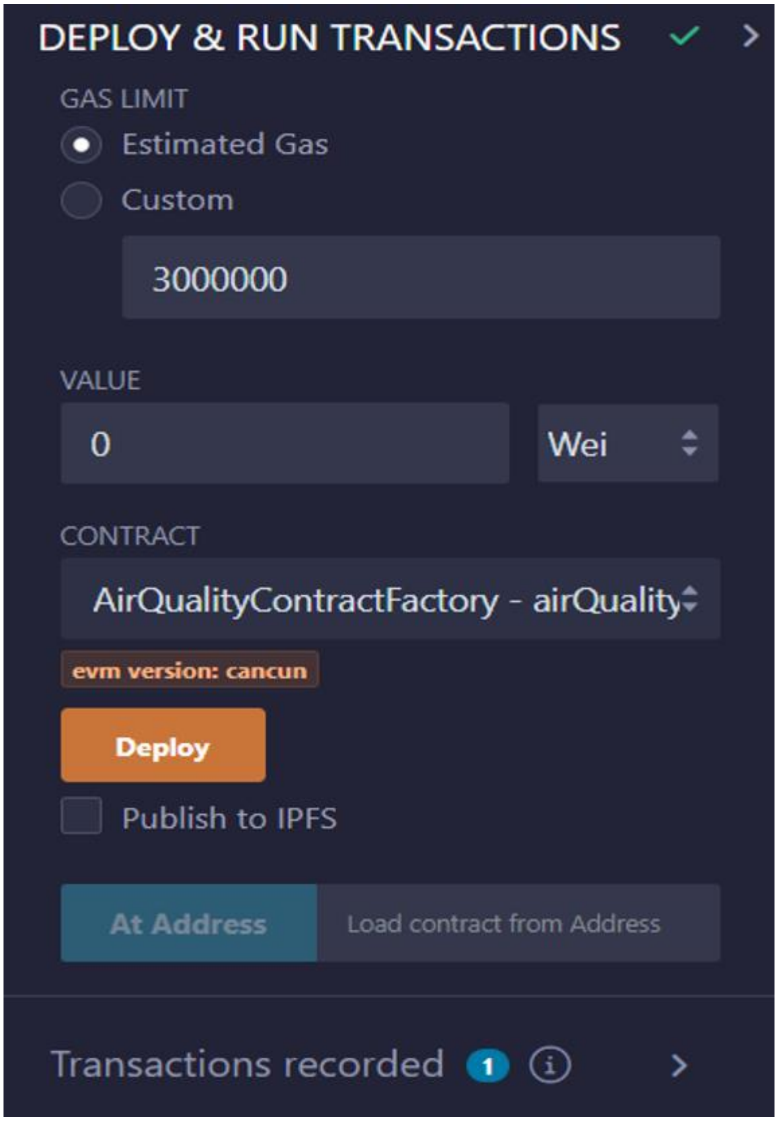


Figure 6. 3 Deployment of the smart contract

Once the deployment of the smart contracts is completed, the transaction process follows based on the criteria defined within the contracts. These criteria include specific thresholds for various pollutants, which, when exceeded, trigger the system to initiate a fine. The process begins with the system continuously monitoring emission levels against the set thresholds. The smart contract is programmed to automatically assess these levels and determine when an emitter has breached the allowable limits. This automated assessment ensures that the enforcement of fines is both prompt and accurate, minimizing the potential for human error and ensuring consistent application of environmental regulations.

The Figure 6. 4 below provides a detailed overview of the transaction process. It illustrates how the system evaluates the necessary conditions before applying fines. Upon detecting that the emission levels have exceeded the thresholds, the system calculates the appropriate fine amount. This calculation is based on predefined parameters set within the contract, ensuring that fines are proportionate to the severity of the violation. The system then prepares to execute the transaction, which includes recording the breach, calculating the equivalent gas fee required for the transaction on the Ethereum network, and initiating the transfer of the fine amount from the emitter to the designated authority.

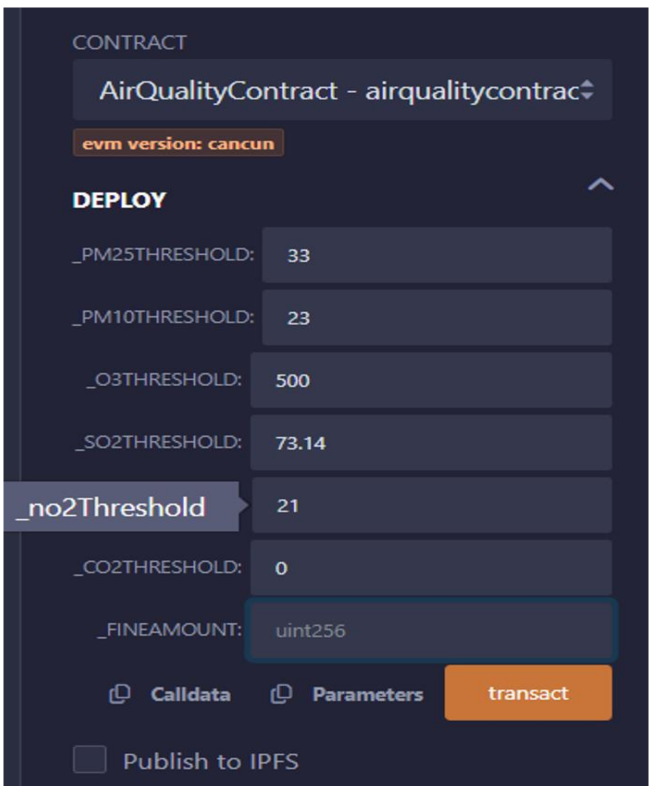


Figure 6. 4 Transaction process from Ethereum

Once the fine amount is determined, the system applies the transaction, enforcing the penalty on the offending emitter. This transaction is recorded on the blockchain, creating an immutable and transparent log of the enforcement action. By recording every transaction, the system ensures accountability and provides a verifiable history of all fines imposed. This transparency is crucial for maintaining trust among stakeholders, including regulatory bodies, industries, and the public. The blockchain's decentralized nature further enhances this trust by preventing any single entity from altering the records. Consequently, the system not only enforces environmental regulations effectively but also upholds the principles of transparency and accountability in environmental governance.

6.5.2. Model Selection and Design

This chapter identifies and analyzes three time series models to determine the best-fitting algorithm for the dataset. The models, developed using data collected over three consecutive years, were chosen to address the challenge of predicting air pollution levels. The analysis focused on three key metrics: R-squared (R^2), Mean Absolute Error (MAE), and Mean Squared Error (MSE). These metrics provided a comprehensive evaluation of each model's accuracy and reliability in predicting air pollution levels. By comparing the performance of the models across these metrics, the study aimed to identify the most effective approach for forecasting and managing air quality data.

Each pollutant's data was rigorously evaluated using the specified metrics. The R-squared value assessed how well the model explained the variability of the data, while the MAE and MSE measured the average magnitude of the errors and the square of the errors, respectively. By applying these metrics to the data for each pollutant, the study ensured a thorough assessment of model performance. The model that demonstrated the best balance of high R-squared values and low error metrics was selected as the most suitable for the dataset. This approach ensured that the chosen model not only provided accurate predictions but also maintained robustness across different pollutants, ultimately aiding in more effective air quality management.

This chapter compares these metrics by using the metric table for each model. Let’s look at all models starting with RNNs, ARIMA then Exponential smoothing.

1. RNNs

In this chapter, all parameters defining air pollution were identified and analyzed to develop an effective predictive model. Figure 6. 5 illustrates the performance of the data and the forecasting results for each pollutant after training the RNNs model. These visualizations highlight how well the model captures historical trends and its ability to predict future pollution levels accurately. The analysis encompassed various pollutants, ensuring that the model could handle diverse data inputs and provide reliable forecasts. By thoroughly training the model and evaluating its performance, the study demonstrated its capability to offer precise and actionable insights for air quality management.

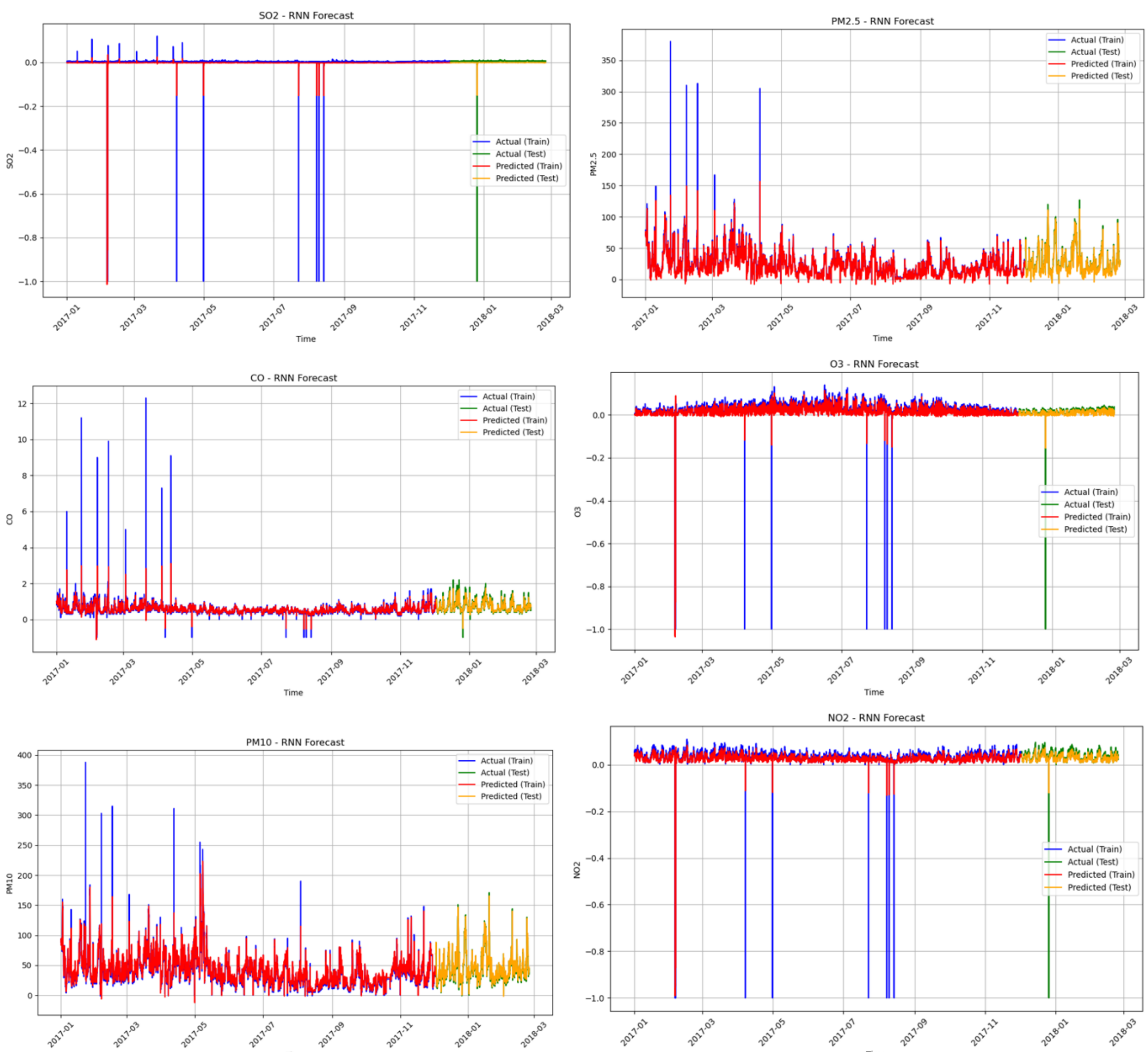


Figure 6. 5 RNNs model for all six pollutants

The above Figure 6. 5 shows all six parameters analyzed using RNN models, illustrating their performance and predictive capabilities. Actual values for trained data are presented in blue color while the actual test data are colored in green. Then the predictive values for the trained set are in red while the predictive values for testing are in yellow. The table below provides the evaluation metrics, including R-squared, Mean Absolute Error (MAE), and Mean Squared Error (MSE), which quantify the accuracy and reliability of the model for each parameter.

The RNN model yielded some negative R-squared values for SO2, indicating poor predictive performance for this specific pollutant. Despite performing well on other parameters, the negative R-squared values for SO2 led to the rejection of the RNN model as a reliable predictor. The table below provides detailed metrics for the RNN model, including R-squared, Mean Absolute Error (MAE), and Mean Squared Error (MSE) for all analyzed pollutants. These metrics highlight the model's performance across different parameters.

Table 6. 1 RNNs metric table

Parameters	**Mean Absolute Error (MAE)**	**Mean Squared Error (MSE)**	**R-Squared (R^2)**
PM2.5	3.39	25.96	0.94
PM10	4.54	46.10	0.92
NO2	0.01	0.00	0.29
CO	0.09	0.02	0.83
SO2	0.01	0.00	-0.12
O3	0.01	0.00	0.10

Based on the data from the above Table 6. 1, while some parameters performed very well, there were notable issues with certain pollutants. Specifically, the R-squared value for O3 was relatively low, indicating a weaker predictive performance. Additionally, the R-squared value for SO2 was negative, further highlighting significant inaccuracies in the model's predictions for this pollutant. These shortcomings suggest that, despite the model's overall effectiveness, it may not be suitable for accurately forecasting all types of air pollutants.

2. ARIMA

The second model analyzed was ARIMA, which was applied to all six pollutants to evaluate its performance. The figure below provides detailed insights into the performance of the ARIMA model. The graph Figure 6. 6 illustrates both the actual and predicted data for the training and testing phases, allowing for a clear comparison of the model's accuracy. This visual representation highlights how well the ARIMA model captures historical data trends and its effectiveness in forecasting future pollution levels.

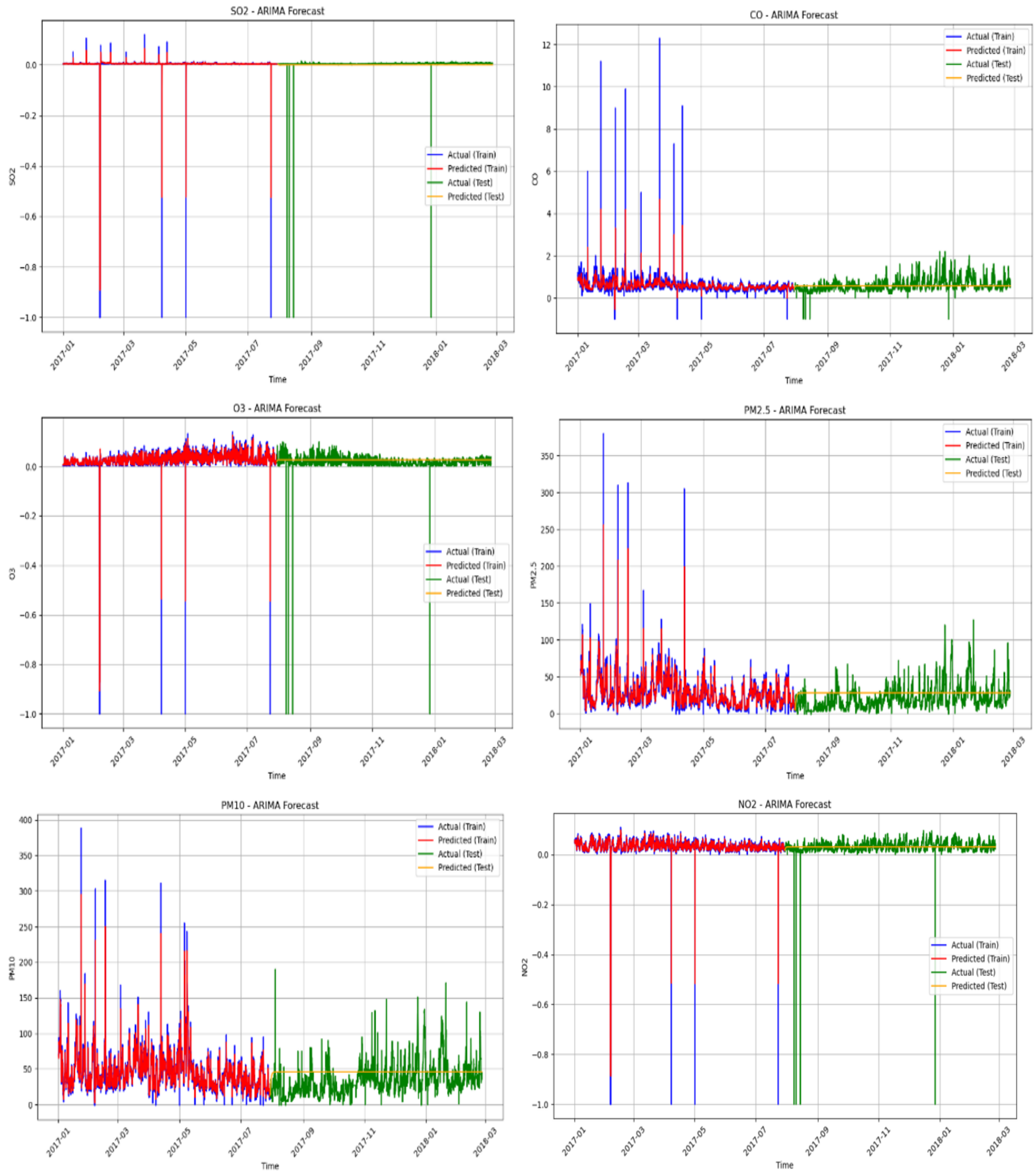


Figure 6. 6 ARIMA Model for all six pollutants

The ARIMA model performed poorly, as evidenced by the negative R-squared values for most pollutants. This indicates that the model failed to capture the underlying patterns in the data and was not effective in predicting future values. Table 6. 2 provides a detailed breakdown of the ARIMA metrics, including R-squared, Mean Absolute Error (MAE), and Mean Squared Error (MSE) for each pollutant, further highlighting the model's inadequacies in delivering accurate predictions.

Table 6. 2 ARIMA metric table

Parameters	**Mean Absolute Error (MAE)**	**Mean Squared Error (MSE)**	**R-Squared (R^2)**
PM2.5	14.37	316.72	-0.15
PM10	20.05	602.82	-0.22
NO2	0.01	0.00	-0.00
CO	0.21	0.08	-0.00
SO2	0.01	0.00	-0.02
O3	0.01	0.00	-0.05

The metric from Table 6. 2 indicates that this model is not suitable for analyzing the data. With most R-squared values being negative, it is evident that the model fails to accurately capture and predict the patterns in the dataset. Consequently, due to these consistently poor performance metrics, the model should be rejected as an effective tool for air pollution analysis.

3. Exponential smoothing

The exponential smoothing model was analyzed in this paper after determining that the other time series models, ARIMA and RNNs, did not perform well. The exponential smoothing model showed a marked improvement across all pollutants compared to the other models. The Figure 6. 7 illustrates the performance of the exponential smoothing model, displaying both the actual and predicted values for the training and testing sets. This comparison highlights the model's ability to forecast air pollution levels more accurately, demonstrating its effectiveness over the previously tested models.

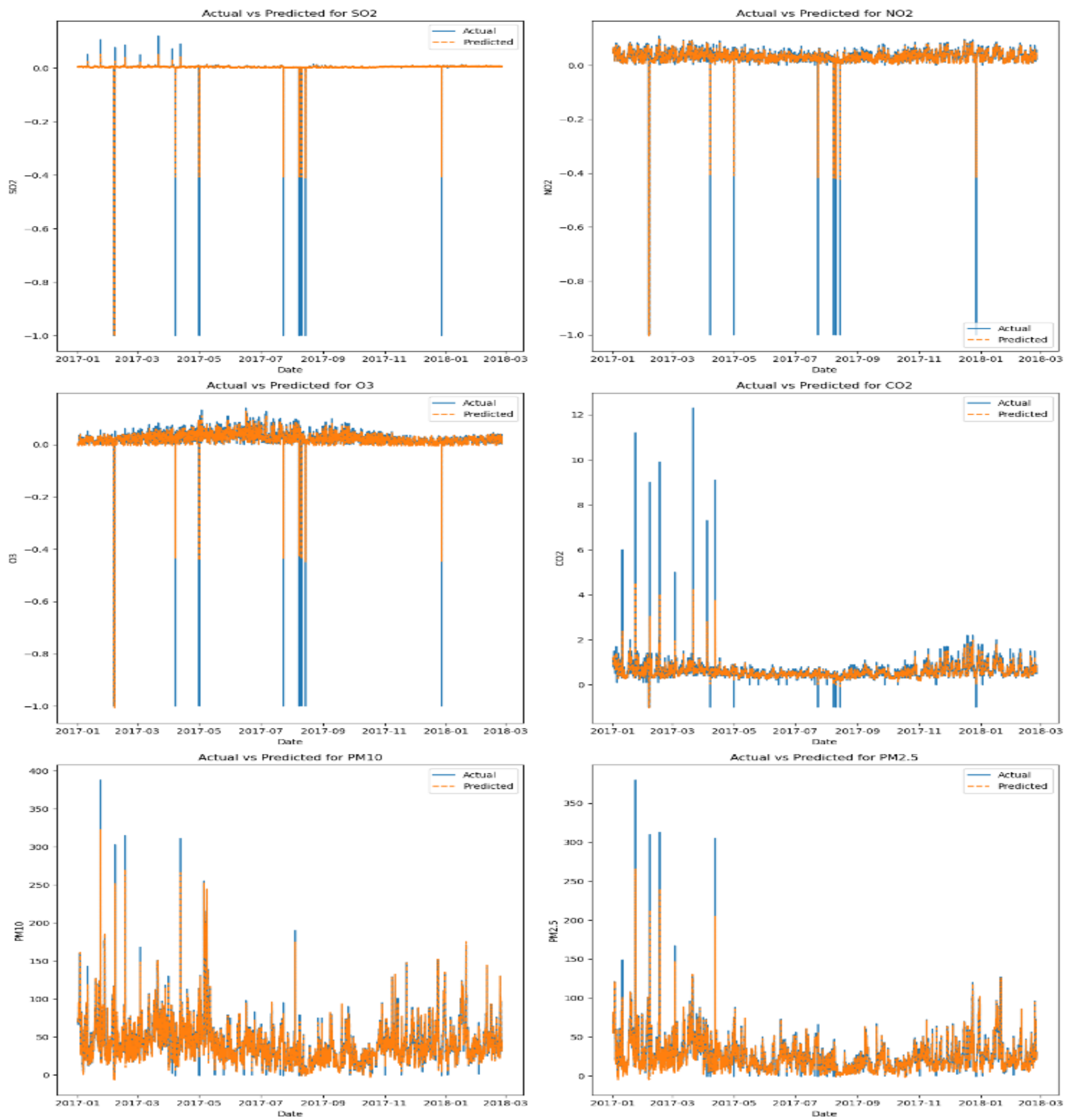


Figure 6. 7 Exponential Smoothing model for all six pollutants

From Figure 6. 7, the exponential smoothing model is performing exceptionally well. The model's ability to align closely with the actual data points for both the training and testing sets demonstrates its robustness and accuracy in predicting air pollution levels. Unlike the RNN and ARIMA models, which produced negative R-squared values for several pollutants, the exponential smoothing model consistently yielded positive values across all metrics. This indicates a reliable and effective performance, making it a superior choice for this dataset.

The metric Table 6. 3 further substantiates these findings, presenting the values for R-squared, Mean Absolute Error (MAE), and Mean Squared Error (MSE) for each pollutant. None of the R-squared values were negative, as seen with the previous models, reaffirming the exponential smoothing model's accuracy. The positive values across all metrics illustrate the model's ability to make precise predictions, ensuring that air quality trends are well-identified and reliable for future forecasting. This comprehensive analysis highlights the exponential smoothing model as the most suitable for predicting air pollution spikes, providing a solid foundation for proactive air quality management.

Table 6. 3 Metric table for exponential smoothing model

Parameters	**Mean Absolute Error (MAE)**	**Mean Squared Error (MSE)**	**R-Squared (R^2)**
PM2.5	3.66	84.86	0.7709
PM10	4.95	116.16	0.8241
NO2	0.01	0.00	0.5578
CO	0.10	0.09	0.3581
SO2	0.00	0.00	0.5239
O3	0.01	0.00	0.5551

The results of this research demonstrate strong performance in terms of MAE and MSE values, particularly for PM2.5, where our framework exhibits significantly lower errors compared to models used in previous studies. The comparison highlights that our framework consistently outperforms existing systems, as shown in the Figure 6. 8, which references four established models [126], [127], [128], [129]. These findings underscore the effectiveness of our approach in delivering more accurate predictions, solidifying its superiority over the referenced systems.

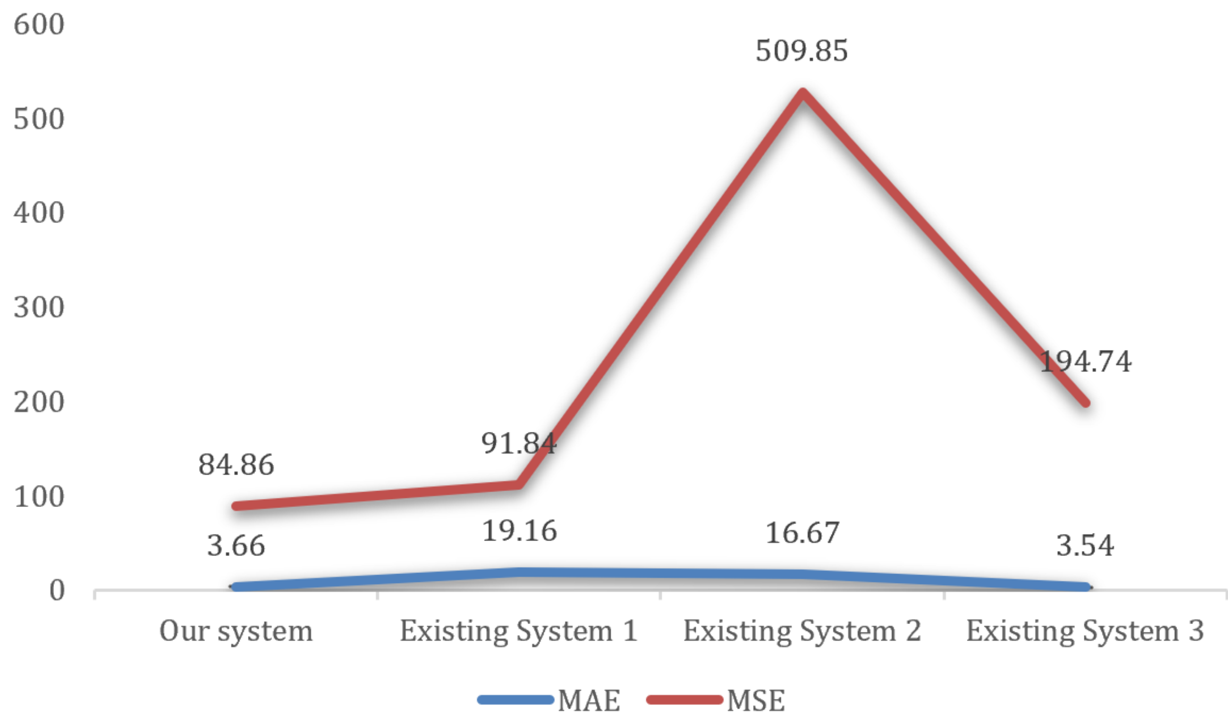


Figure 6. 8 Compare existing systems and developed framework

By analyzing the error bars, it is evident that our designed framework significantly outperforms other existing systems. The narrower error bars indicate lower variability and higher precision in predictions, highlighting the robustness and reliability of our approach. This superior performance, compared to the broader error margins observed in other models, underscores the effectiveness of our framework in providing accurate and consistent results in air pollution monitoring.

6.6. Conclusion

In conclusion, this chapter presents a pioneering approach to revolutionizing the analysis and mitigation of air pollution spikes by integrating blockchain technology and machine learning. The proposed framework effectively combines blockchain's transparency and immutability with the predictive accuracy of machine learning, ensuring reliable real-time monitoring and enforcement. The machine learning models, particularly exponential smoothing, provide robust predictions, enabling proactive management of air pollution spikes. This integration results in a trustworthy and efficient system for maintaining air quality. Through the integration of data from diverse sources, including IoT sensors, comprehensive insights into air quality dynamics are provided, ensuring informed decision-making processes. Furthermore, the decentralized nature of blockchain enhances data integrity and fosters trust among stakeholders. Through case studies and simulations, the efficacy and scalability of the framework in addressing air pollution spikes across diverse geographical regions are demonstrated, signifying a paradigm shift in air quality management.

Additionally, this chapter presents the results of implementing a smart contract on the Ethereum platform, enabling the enforcement of fines on air pollution emitters that surpass set thresholds. Various machine learning models, including ARIMA, RNNs, and Exponential Smoothing, were compared to identify the most suitable model for air pollution data. The analysis focused on key metrics such as R-squared, Mean Absolute Error (MAE), and Mean Squared Error (MSE), comprehensively evaluating each model's accuracy and reliability. By selecting the model that demonstrated the best balance of high R-squared values and low error metrics, the study ensured accurate predictions and effective management of air quality data. The comparison of these metrics across different models, including RNNs, ARIMA, and Exponential Smoothing, underscores the importance of selecting the most appropriate model for ensuring robust air quality management practices.

Chapter 7. Conclusion, Recommendations, and Future Research

The integration of edge IoT, blockchain, and AI in our research presented a pioneering approach to monitoring air pollution spikes, addressing the critical need for low latency, data trust, and predictive capabilities. By analyzing this integration within the context of air quality monitoring, we have designed a real-time multi-pollutant (RTMP) sensor featuring blockchain identity by saving above 40% of the energy consumption compared to the existing system. This sensor, coupled with short-term prediction AI, alerts emitters of potential risks exceeding exposure standards, thus mitigating health risks associated with pollution spikes, especially for vulnerable populations like children. The proposed hardware-software smart IoT device monitors short-term exposure to pollution peaks, optimizing IoT energy usage through innovative signal wake-up mechanisms. The system's performance surpassed existing models, as evidenced by performance metrics and the successful identification and prediction of pollutant spikes using mathematical models like Poisson distribution and Monte Carlo simulations. AI was integrated with the developed system Machine learning models were compared and the exponential smoothing time series performed very well on all pollutant data.

Moving forward, our research focused on utilizing machine learning to analyze the collected spike data, proposing fines for peak emitters. The prototype developed for monitoring transportation pollution spikes through IoT edge networks represents a significant advancement in urban air quality monitoring. By enabling real-time detection and analysis of pollutant spikes caused by vehicular emissions, this research contributes to creating healthier and more sustainable urban environments.

This section provides an overview comparing the developed system to existing systems for monitoring air pollution. The table below outlines the specific problems identified in current systems, such as inadequate real-time data, high energy consumption, and limited predictive capabilities. It then details how the developed system addresses these issues through innovations like enhanced real-time monitoring, efficient energy usage, and advanced predictive algorithms. This comparative analysis highlights the advantages and improvements offered by the newly developed system over traditional approaches. The Table 7. 1 gives a summary of the comparison.

Table 7. 1 Comparison of existing systems and developed framework

Existing systems	Developed system
Centralized cloud servers result in communication delays, making them unsuitable for applications requiring low latency, limited bandwidth, or high privacy.	The developed system is a real-time system developed using Edge centric system.
Transmission of data which implies energy consumption with the sensor’s battery-dried	Transmission once the value passes the defined threshold by applying interrupts. Improve power management by over 40%. And capture all spikes since the probability of losing it is 7.1% for periodic systems.
Unable to identify who is polluting the air and at which level	Identify the source of the pollution from the emitters.
Unknown level of pollution from emitters compared to the defined standards and based on historical emission	Prediction of the emission and taking measures.
Existing system predictions performed with MAE and MSE have bigger compared values 19.16,16.67, 5.10, and 91.81, 509,85, 194,74 for PM2.5	The developed system has only 3.66 value for PM2.5 for MAE and 84.86 for MSE
Lack of trust in the cloud data leads to the lack of taking some measures of reduction	Smart contracts were developed to penalize emitters of air pollution who exceed the threshold.

For future works and recommendations, we will work on further optimizing IoT energy consumption and exploring additional low-power wake-up mechanisms to enhance the sustainability and longevity of the monitoring devices. We will also explore the interoperability of blockchain with other emerging technologies to enhance data sharing, security, and the integration of multiple environmental monitoring systems. We will conduct long-term studies to evaluate the impact of real-time air quality monitoring and AI-driven interventions on public health, particularly in vulnerable communities. This system will be tested in real-world scenarios in the transportation and industrial areas. Finally, we will Work on integrating the system's outputs with policy-making processes, enabling more effective regulatory measures and encouraging widespread adoption by governmental and environmental agencies.

In conclusion, integrating blockchain technology and machine learning in our framework revolutionizes air pollution spike analysis, offering crucial insights for timely interventions and informed decision-making. This innovative approach empowers proactive measures to mitigate pollution's adverse effects, paving the way for cleaner and healthier urban environments globally. Through ongoing refinement and optimization, our research holds the potential to transform air quality monitoring systems and foster environmental sustainability for future generations.

Eric Nizeyimana, September 2024,

References

[1] H.-B. Ly *et al.*, "Development of an AI Model to Measure Traffic Air Pollution from Multisensor and Weather Data," *Sensors*, vol. 19, no. 22, p. 4941, Nov. 2019, doi: 10.3390/s19224941.

[2] C. F. Rider and C. Carlsten, "Air pollution and DNA methylation: Effects of exposure in humans," Sep. 03, 2019, *BioMed Central Ltd.* doi: 10.1186/s13148-019-0713-2.

[3] T. J. Wallington, J. E. Anderson, R. H. Dolan, and S. L. Winkler, "Vehicle Emissions and Urban Air Quality: 60 Years of Progress," May 01, 2022, *MDPI*. doi: 10.3390/atmos13050650.

[4] M. Bishop, R. Burgess, and C. Zipfel, "Technology and Development," in *Introduction to Development Engineering*, Cham: Springer International Publishing, 2023, pp. 17–57. doi: 10.1007/978-3-030-86065-3_2.

[5] WHO, "7 Million Deaths Annually Linked to Air Pollution-WHO Report."

[6] A. Goshua, C. A. Akdis, and K. C. Nadeau, "World Health Organization global air quality guideline recommendations: Executive summary," *Allergy*, vol. 77, no. 7, pp. 1955–1960, Jul. 2022, doi: 10.1111/all.15224.

[7] A. E. Millen, S. Dighe, K. Kordas, B. Z. Aminigo, M. L. Zafron, and L. Mu, "Air Pollution and Chronic Eye Disease in Adults: A Scoping Review," *Ophthalmic Epidemiol*, vol. 31, no. 1, pp. 1–10, Jan. 2024, doi: 10.1080/09286586.2023.2183513.

[8] Y. J. Choi *et al.*, "Short-term effects of air pollution on blood pressure," *Sci Rep*, vol. 9, no. 1, Dec. 2019, doi: 10.1038/s41598-019-56413-y.

[9] S. Zhong, Z. Yu, and W. Zhu, "Study of the Effects of Air Pollutants on Human Health Based on Baidu Indices of Disease Symptoms and Air Quality Monitoring Data in Beijing, China," *Int J Environ Res Public Health*, vol. 16, no. 6, p. 1014, Mar. 2019, doi: 10.3390/ijerph16061014.

[10] W. Zhou *et al.*, "Exposure scenario: Another important factor determining the toxic effects of PM2.5 and possible mechanisms involved," *Environmental Pollution*, vol. 226, pp. 412–425, Jul. 2017, doi: 10.1016/j.envpol.2017.04.010.

[11] G. Miskell, W. Pattinson, L. Weissert, and D. Williams, "Forecasting short-term peak concentrations from a network of air quality instruments measuring PM2.5 using boosted gradient machine models," *J Environ Manage*, vol. 242, pp. 56–64, Jul. 2019, doi: 10.1016/j.jenvman.2019.04.010.

[12] E. Dons *et al.*, "Transport most likely to cause air pollution peak exposures in everyday life: Evidence from over 2000 days of personal monitoring," *Atmos Environ*, vol. 213, pp. 424–432, Sep. 2019, doi: 10.1016/j.atmosenv.2019.06.035.

[13] I. Manisalidis, E. Stavropoulou, A. Stavropoulos, and E. Bezirtzoglou, "Environmental and Health Impacts of Air Pollution: A Review," *Front Public Health*, vol. 8, Feb. 2020, doi: 10.3389/fpubh.2020.00014.

[14] J. Rentschler and N. Leonova, "Global air pollution exposure and poverty," *Nat Commun*, vol. 14, no. 1, Dec. 2023, doi: 10.1038/s41467-023-39797-4.

[15] I. Manisalidis, E. Stavropoulou, A. Stavropoulos, and E. Bezirtzoglou, "Environmental and Health Impacts of Air Pollution: A Review," Feb. 20, 2020, *Frontiers Media S.A.* doi: 10.3389/fpubh.2020.00014.

[16] E. Bainomugisha, J. Ssematimba, and D. Okure, "Design Considerations for a Distributed

Low-Cost Air Quality Sensing System for Urban Environments in Low-Resource Settings," *Atmosphere (Basel)*, vol. 14, no. 2, Feb. 2023, doi: 10.3390/atmos14020354.
[17] U.S. Environmental Protection Agency (EPA), "Air Quality Standards and Implementation.," https://www.epa.gov/air-quality-implementation-plans.
[18] D. Wang, D. Zhong, and A. Souri, "Energy management solutions in the Internet of Things applications: Technical analysis and new research directions," *Cogn Syst Res*, vol. 67, pp. 33–49, Jun. 2021, doi: 10.1016/J.COGSYS.2020.12.009.
[19] J. S. Turiel and R. K. Kaufmann, "Evidence of air quality data misreporting in China: An impulse indicator saturation model comparison of local government-reported and U.S. embassy-reported PM2.5 concentrations (2015–2017)," *PLoS One*, vol. 16, no. 4, p. e0249063, Apr. 2021, doi: 10.1371/journal.pone.0249063.
[20] T. Egorova, E. Rozanov, P. Arsenovic, and T. Sukhodolov, "Ozone Layer Evolution in the Early 20th Century," *Atmosphere (Basel)*, vol. 11, no. 2, p. 169, Feb. 2020, doi: 10.3390/atmos11020169.
[21] The World Bank, "Understanding Air Pollution and the Way It Is Measured," https://www.worldbank.org/en/news/feature/2015/07/14/understanding-air-pollution-and-the-way-it-is-measured.
[22] R. Singh, B. Kandpal, R. Garg, T. Vasudeva, Y. Shrma, and M. Arif Kamal, "Challenges Associated with Air Pollution in Sustainable City Development: Case of New Delhi and New York," *American Journal of Civil Engineering and Architecture*, vol. 10, no. 3, pp. 116–125, Aug. 2022, doi: 10.12691/ajcea-10-3-2.
[23] S. M. S. D. Malleswari and T. K. Mohana, "Air pollution monitoring system using IoT devices: Review," *Mater Today Proc*, vol. 51, pp. 1147–1150, 2022, doi: 10.1016/j.matpr.2021.07.114.
[24] V. K. Tembhurne *et al.*, "IoT-based Air Pollution Monitoring System to Measure Air Quality on Cloud Storage," in *2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing (PCEMS)*, IEEE, Apr. 2023, pp. 1–6. doi: 10.1109/PCEMS58491.2023.10136085.
[25] D. Hanyurwimfura, E. Nizeyimana, F. Ndikumana, D. Mukanyiligira, A. Bakar Diwani, and F. Mukamanzi, "Monitoring system to strive against fall armyworm in crops case study: Maize in Rwanda," in *Proceedings - 2018 IEEE SmartWorld, Ubiquitous Intelligence and Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People and Smart City Innovations, SmartWorld/UIC/ATC/ScalCom/CBDCo*, 2018. doi: 10.1109/SmartWorld.2018.00046.
[26] D. Oladimeji, K. Gupta, N. A. Kose, K. Gundogan, L. Ge, and F. Liang, "Smart Transportation: An Overview of Technologies and Applications," Apr. 01, 2023, *MDPI*. doi: 10.3390/s23083880.
[27] N. Arandia, J. I. Garate, and J. Mabe, "Embedded Sensor Systems in Medical Devices: Requisites and Challenges Ahead," *Sensors*, vol. 22, no. 24, Dec. 2022, doi: 10.3390/s22249917.
[28] World Health Organization, "Ambient (outdoor) air pollution," chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.francedatacenter.com/wp-content/uploads/2020/11/FINALSTUDYEnglishKK-03-20-210-EN-N13072020pdf.pdf.
[29] P. J. E. Quintana, J. R. Valenzia, R. J. Delfino, and L.-J. S. Liu, "Monitoring of 1-Min Personal Particulate Matter Exposures in Relation to Voice-Recorded Time–Activity Data," *Environ Res*, vol. 87, no. 3, pp. 199–213, Nov. 2001, doi: 10.1006/enrs.2001.4304.

[30] Royal Belgian Institute for Space Aeronomy, “Lifespan of gases is relevant to air quality & climate,” https://www.aeronomie.be/en/encyclopedia/lifespan-gases-relevant-air-quality-climate.
[31] Y. Liu *et al.*, “Short-Term Exposure to Ambient Air Pollution and Mortality From Myocardial Infarction,” *J Am Coll Cardiol*, vol. 77, no. 3, pp. 271–281, Jan. 2021, doi: 10.1016/j.jacc.2020.11.033.
[32] P. de Prado-Bert *et al.*, “Short- and medium-term air pollution exposure, plasmatic protein levels and blood pressure in children,” *Environ Res*, vol. 211, Aug. 2022, doi: 10.1016/j.envres.2022.113109.
[33] Z. Idrees, Z. Zou, and L. Zheng, “Edge Computing Based IoT Architecture for Low Cost Air Pollution Monitoring Systems: A Comprehensive System Analysis, Design Considerations & Development,” *Sensors*, vol. 18, no. 9, p. 3021, Sep. 2018, doi: 10.3390/s18093021.
[34] E. Bainomugisha, J. Ssematimba, and D. Okure, “Design Considerations for a Distributed Low-Cost Air Quality Sensing System for Urban Environments in Low-Resource Settings,” *Atmosphere (Basel)*, vol. 14, no. 2, p. 354, Feb. 2023, doi: 10.3390/atmos14020354.
[35] K. Sridhar, P. Radhakrishnan, G. Swapna, R. Kesavamoorthy, L. Pallavi, and R. Thiagarajan, “A modular IOT sensing platform using hybrid learning ability for air quality prediction,” *Measurement: Sensors*, vol. 25, p. 100609, Feb. 2023, doi: 10.1016/j.measen.2022.100609.
[36] A. Ali, “A Framework for Air Pollution Monitoring in Smart Cities by Using IoT and Smart Sensors,” *Informatica*, vol. 46, no. 5, May 2022, doi: 10.31449/inf.v46i5.4003.
[37] J. Kalajdjieski, B. R. Stojkoska, and K. Trivodaliev, “IoT Based Framework for Air Pollution Monitoring in Smart Cities,” in *2020 28th Telecommunications Forum (TELFOR)*, IEEE, Nov. 2020, pp. 1–4. doi: 10.1109/TELFOR51502.2020.9306531.
[38] E. X. Neo *et al.*, “Artificial intelligence-assisted air quality monitoring for smart city management,” *PeerJ Comput Sci*, vol. 9, p. e1306, May 2023, doi: 10.7717/peerj-cs.1306.
[39] Y. Han, B. Park, and J. Jeong, “A novel architecture of air pollution measurement platform using 5G and blockchain for industrial IoT applications,” in *Procedia Computer Science*, Elsevier B.V., 2019, pp. 728–733. doi: 10.1016/j.procs.2019.08.105.
[40] S. Singh, P. K. Sharma, B. Yoon, M. Shojafar, G. H. Cho, and I.-H. Ra, “Convergence of blockchain and artificial intelligence in IoT network for the sustainable smart city,” *Sustain Cities Soc*, vol. 63, p. 102364, Dec. 2020, doi: 10.1016/j.scs.2020.102364.
[41] Y. He, Y. Wang, C. Qiu, Q. Lin, J. Li, and Z. Ming, “Blockchain-Based Edge Computing Resource Allocation in IoT: A Deep Reinforcement Learning Approach,” *IEEE Internet Things J*, vol. 8, no. 4, pp. 2226–2237, Feb. 2021, doi: 10.1109/JIOT.2020.3035437.
[42] World Health Organization, “Air quality guidelines: global update 2005: particulate matter, ozone, nitrogen dioxide and sulfur dioxide,” https://iris.who.int/handle/10665/107823.
[43] J. Grant, M. Eltoukhy, and S. Asfour, “Short-Term Electrical Peak Demand Forecasting in a Large Government Building Using Artificial Neural Networks,” *Energies (Basel)*, vol. 7, no. 4, pp. 1935–1953, Mar. 2014, doi: 10.3390/en7041935.
[44] Q. D. La, M. V. Ngo, T. Q. Dinh, T. Q. S. Quek, and H. Shin, “Enabling intelligence in fog computing to achieve energy and latency reduction,” *Digital Communications and Networks*, vol. 5, no. 1, pp. 3–9, Feb. 2019, doi: 10.1016/j.dcan.2018.10.008.
[45] F. Casino, T. K. Dasaklis, and C. Patsakis, “A systematic literature review of blockchain-based applications: Current status, classification and open issues,” *Telematics and*

Informatics, vol. 36, pp. 55–81, Mar. 2019, doi: 10.1016/j.tele.2018.11.006.
[46] H. F. Atlam, M. A. Azad, A. G. Alzahrani, and G. Wills, “A Review of Blockchain in Internet of Things and AI,” *Big Data and Cognitive Computing*, vol. 4, no. 4, p. 28, Oct. 2020, doi: 10.3390/bdcc4040028.
[47] F. Jameel, U. Javaid, W. U. Khan, M. N. Aman, H. Pervaiz, and R. Jäntti, “Reinforcement Learning in Blockchain-Enabled IIoT Networks: A Survey of Recent Advances and Open Challenges,” *Sustainability*, vol. 12, no. 12, p. 5161, Jun. 2020, doi: 10.3390/su12125161.
[48] Z. Idrees, Z. Zou, and L. Zheng, “Edge computing based IoT architecture for low cost air pollution monitoring systems: A comprehensive system analysis, design considerations & development,” *Sensors (Switzerland)*, vol. 18, no. 9, Sep. 2018, doi: 10.3390/s18093021.
[49] A. N. J. P. Q. H. T. & T. W. Tuan Nguyen Gia, “Artificial Intelligence at the Edge in the Blockchain of Things,” *Springer Link*, pp. 267–280, May 2020.
[50] Lei. Kai, *Proceedings of 2018 1st IEEE International Conference on Hot Information-Centric Networking (HotICN 2018) : Aug 15-17, 2018, Shenzhen, Guangdong, China*. IEEE Press, 2018.
[51] B. Chen and H. Kan, “Air pollution and population health: A global challenge,” in *Environmental Health and Preventive Medicine*, Mar. 2008, pp. 94–101. doi: 10.1007/s12199-007-0018-5.
[52] P. Wang *et al.*, “Aggravated air pollution and health burden due to traffic congestion in urban China,” *Atmos Chem Phys*, vol. 23, no. 5, pp. 2983–2996, Mar. 2023, doi: 10.5194/acp-23-2983-2023.
[53] “Compendium of WHO and other UN guidance on health and environment Chapter 2. Air pollution,” 2022.
[54] E. R. Kulick, J. D. Kaufman, and C. Sack, “Ambient Air Pollution and Stroke: An Updated Review,” Mar. 01, 2023, *Wolters Kluwer Health*. doi: 10.1161/STROKEAHA.122.035498.
[55] N. Institute of Environmental Health Sciences, “Air Pollution and Your Health.” [Online]. Available: https://niehs.nih.gov
[56] G. Singh Sarla and G. Singh Sarla -, “Air pollution: Health effects Contaminación del aire: efectos sobre la salud.”
[57] C. T. Loftus *et al.*, “Prenatal air pollution and childhood IQ: Preliminary evidence of effect modification by folate,” *Environ Res*, vol. 176, Sep. 2019, doi: 10.1016/j.envres.2019.05.036.
[58] F. P. Perera *et al.*, “Prenatal airborne polycyclic aromatic hydrocarbon exposure and child IQ at age 5 years,” *Pediatrics*, vol. 124, no. 2, Aug. 2009, doi: 10.1542/peds.2008-3506.
[59] Z. E. M. Morgan *et al.*, “Prenatal exposure to ambient air pollution is associated with neurodevelopmental outcomes at 2 years of age,” *Environ Health*, vol. 22, no. 1, Dec. 2023, doi: 10.1186/s12940-022-00951-y.
[60] J. T. Lee, “Review of epidemiological studies on air pollution and health effects in children,” Jan. 01, 2021. doi: 10.3345/cep.2019.00843.
[61] A. Milojevic, P. Dutey-Magni, L. Dearden, and P. Wilkinson, “Lifelong exposure to air pollution and cognitive development in young children: The UK Millennium Cohort Study,” *Environmental Research Letters*, vol. 16, no. 5, May 2021, doi: 10.1088/1748-9326/abe90c.
[62] M. J. Zare Sakhvidi *et al.*, “Outdoor air pollution exposure and cognitive performance: findings from the enrolment phase of the CONSTANCES cohort,” *Lancet Planet Health*, vol. 6, no. 3, pp. e219–e229, Mar. 2022, doi: 10.1016/S2542-5196(22)00001-8.

[63] G. P. Bate, S. P. Velos, G. B. Gimena, and M. B. Go, "Influence of IQ and Personality on College Students' Academic Performance In A Philippine State University." [Online]. Available: http://journalppw.com
[64] S. Mathiarasan and A. Hüls, "Impact of environmental injustice on children's health—interaction between air pollution and socioeconomic status," Jan. 02, 2021, *MDPI AG*. doi: 10.3390/ijerph18020795.
[65] G. T. Wodtke, K. Ard, C. Bullock, K. White, and B. Priem, "Concentrated poverty, ambient air pollution, and child cognitive development," 2022. [Online]. Available: https://www.science.org
[66] "Air pollution: how it affects our health."
[67] R. Saadeh, Y. Khader, M. Malkawi, and M. Z. Allouh, "Communicating the Risks of Air Pollution to the Public: A Perspective from Jordan and Lebanon," *Environ Health Insights*, vol. 16, 2022, doi: 10.1177/11786302221127851.
[68] S. Yan *et al.*, "Breathing Green: Maximising Health and Environmental Benefits for Active Transportation Users Leveraging Large Scale Air Quality Data," Jul. 2023, [Online]. Available: http://arxiv.org/abs/2307.15401
[69] H. Wang *et al.*, "Health benefits of on-road transportation pollution control programs in China," 2015, doi: 10.1073/pnas.1921271117/-/DCSupplemental.
[70] P. Rani and A. Dhok, "Effects of Pollution on Pregnancy and Infants," *Cureus*, Jan. 2023, doi: 10.7759/cureus.33906.
[71] N. Brusselaers, C. Macharis, and K. Mommens, "The health impact of freight transport-related air pollution on vulnerable population groups," *Environmental Pollution*, vol. 329, Jul. 2023, doi: 10.1016/j.envpol.2023.121555.
[72] T. Afrin and N. Yodo, "A survey of road traffic congestion measures towards a sustainable and resilient transportation system," Jun. 01, 2020, *MDPI*. doi: 10.3390/su12114660.
[73] S. Allabakash, S. Lim, K.-S. Chong, and T. Yamada, "Particulate Matter Concentrations over South Korea: Impact of Meteorology and Other Pollutants," *Remote Sens (Basel)*, vol. 14, no. 19, p. 4849, Sep. 2022, doi: 10.3390/rs14194849.
[74] E. Nizeyimana, J. Nsenga, R. Shibasaki, D. Hanyurwimfura, and J. Hwang, "Design of Smart IoT Device for Monitoring Short-term Exposure to Air Pollution Peaks," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 1, 2022, doi: 10.14569/IJACSA.2022.0130103.
[75] S. Kang, Y. Zhang, Y. Qian, and H. Wang, "A review of black carbon in snow and ice and its impact on the cryosphere," Nov. 01, 2020, *Elsevier B.V.* doi: 10.1016/j.earscirev.2020.103346.
[76] J. Lu, B. Li, H. Li, and A. Al-Barakani, "Expansion of city scale, traffic modes, traffic congestion, and air pollution," *Cities*, vol. 108, Jan. 2021, doi: 10.1016/j.cities.2020.102974.
[77] E. Nizeyimana, D. Hanyurwimfura, R. Shibasaki, and J. Nsenga, "Design of a Decentralized and Predictive Real-Time Framework for Air Pollution Spikes Monitoring," in *2021 IEEE 6th International Conference on Cloud Computing and Big Data Analytics, ICCCBDA 2021*, 2021. doi: 10.1109/ICCCBDA51879.2021.9442611.
[78] S. Krishna and J. Guptha, "Edge Computing Based Air Pollution Monitoring System," *International Journal of Information Technology*, vol. 7, [Online]. Available: www.ijitjournal.org
[79] K. Biondi *et al.*, "Air Pollution Detection System using Edge Computing."

[80] D. Bousiotis, L. N. S. Alconcel, D. C. S. Beddows, R. M. Harrison, and F. D. Pope, "Monitoring and apportioning sources of indoor air quality using low-cost particulate matter sensors," *Environ Int*, vol. 174, Apr. 2023, doi: 10.1016/j.envint.2023.107907.
[81] A. S. Ramírez, S. Ramondt, K. Van Bogart, and R. Perez-Zuniga, "Public Awareness of Air Pollution and Health Threats: Challenges and Opportunities for Communication Strategies To Improve Environmental Health Literacy," *J Health Commun*, vol. 24, no. 1, pp. 75–83, Jan. 2019, doi: 10.1080/10810730.2019.1574320.
[82] H. B. Ly *et al.*, "Development of an AI model to measure traffic air pollution from multisensor and weather data," *Sensors (Switzerland)*, vol. 19, no. 22, Nov. 2019, doi: 10.3390/s19224941.
[83] V. E. Alvear-Puertas, Y. A. Burbano-Prado, P. D. Rosero-Montalvo, P. Tözün, F. Marcillo, and W. Hernandez, "Smart and Portable Air-Quality Monitoring IoT Low-Cost Devices in Ibarra City, Ecuador," *Sensors*, vol. 22, no. 18, Sep. 2022, doi: 10.3390/s22187015.
[84] A. Biswas and H. C. Wang, "Autonomous Vehicles Enabled by the Integration of IoT, Edge Intelligence, 5G, and Blockchain," Feb. 01, 2023, *MDPI*. doi: 10.3390/s23041963.
[85] D. Suriano and M. Prato, "An Investigation on the Possible Application Areas of Low-Cost PM Sensors for Air Quality Monitoring," *Sensors*, vol. 23, no. 8, Apr. 2023, doi: 10.3390/s23083976.
[86] G. Rescio, A. Manni, A. Caroppo, A. M. Carluccio, P. Siciliano, and A. Leone, "Multi-Sensor Platform for Predictive Air Quality Monitoring," *Sensors*, vol. 23, no. 11, Jun. 2023, doi: 10.3390/s23115139.
[87] J. Kalajdjieski, M. Korunoski, B. R. Stojkoska, and K. Trivodaliev, "Smart City Air Pollution Monitoring and Prediction: A Case Study of Skopje," in *Communications in Computer and Information Science*, Springer Science and Business Media Deutschland GmbH, 2020, pp. 15–27. doi: 10.1007/978-3-030-62098-1_2.
[88] N. Celikkaya, M. Fullerton, and B. Fullerton, "Use of Low-Cost Air Quality Monitoring Devices for Assessment of Road Transport Related Emissions," in *Transportation Research Procedia*, Elsevier B.V., 2019, pp. 762–781. doi: 10.1016/j.trpro.2019.09.125.
[89] K. Sridhar, P. Radhakrishnan, G. Swapna, R. Kesavamoorthy, L. Pallavi, and R. Thiagarajan, "A modular IOT sensing platform using hybrid learning ability for air quality prediction," *Measurement: Sensors*, vol. 25, Feb. 2023, doi: 10.1016/j.measen.2022.100609.
[90] A. Ali, "A Framework for Air Pollution Monitoring in Smart Cities by Using IoT and Smart Sensors," *Informatica (Slovenia)*, vol. 46, no. 5, pp. 129–138, 2022, doi: 10.31449/inf.v46i5.4003.
[91] J. Kalajdjieski, B. R. Stojkoska, and K. Trivodaliev, "IoT based framework for air pollution monitoring in smart cities," in *2020 28th Telecommunications Forum, TELFOR 2020 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., Nov. 2020. doi: 10.1109/TELFOR51502.2020.9306531.
[92] E. X. Neo *et al.*, "Artificial intelligence-assisted air quality monitoring for smart city management," *PeerJ Comput Sci*, vol. 9, 2023, doi: 10.7717/peerj-cs.1306.
[93] P. Diviacco *et al.*, "Monitoring Air Quality in Urban Areas Using a Vehicle Sensor Network (VSN) Crowdsensing Paradigm," *Remote Sens (Basel)*, vol. 14, no. 21, Nov. 2022, doi: 10.3390/rs14215576.
[94] L. Gerevini *et al.*, "An end-to-end real-time pollutants spilling recognition in wastewater based on the IoT-ready SENSIPLUS platform," *Journal of King Saud University* -

Computer and Information Sciences, vol. 35, no. 1, pp. 499–513, Jan. 2023, doi: 10.1016/j.jksuci.2022.12.018.

[95] J. Awewomom *et al.*, "Addressing global environmental pollution using environmental control techniques: a focus on environmental policy and preventive environmental management," *Discover Environment*, vol. 2, no. 1, Feb. 2024, doi: 10.1007/s44274-024-00033-5.

[96] A. Lammers *et al.*, "The Impact of Short-Term Exposure to Air Pollution on the Exhaled Breath of Healthy Adults," *Sensors*, vol. 21, no. 7, p. 2518, Apr. 2021, doi: 10.3390/s21072518.

[97] Sam Nickerson, "Even Small Spikes in Air Pollution Can Threaten Children's Mental Health, Research Suggests," *EcoWatch*, Sep. 27, 2019.

[98] W. A. Suk *et al.*, "Environmental Pollution: An Under-recognized Threat to Children's Health, Especially in Low- and Middle-Income Countries," *Environ Health Perspect*, vol. 124, no. 3, Mar. 2016, doi: 10.1289/ehp.1510517.

[99] N. S. Chipangamate and G. T. Nwaila, "Assessment of challenges and strategies for driving energy transitions in emerging markets: A socio-technological systems perspective," Apr. 01, 2024, *KeAi Communications Co.* doi: 10.1016/j.engeos.2023.100257.

[100] S. Lu and F. Zhou, "Impact of Ship Emission Control Area Policies on Port Air Quality—A Case Study of Ningbo Port, China," *Sustainability (Switzerland)* , vol. 16, no. 9, May 2024, doi: 10.3390/su16093659.

[101] S. Mayor, "Spikes in air pollution raise heart risk as much as sustained exposure, study suggests," *BMJ*, p. k772, Feb. 2018, doi: 10.1136/bmj.k772.

[102] J. Liu *et al.*, "Transition in air pollution, disease burden and health cost in China: A comparative study of long-term and short-term exposure," *Environmental Pollution*, vol. 277, p. 116770, May 2021, doi: 10.1016/j.envpol.2021.116770.

[103] S. Hussain and M. Reza, "Environmental Damage and Global Health: Understanding the Impacts and Proposing Mitigation Strategies."

[104] K. Bhui *et al.*, "Air quality and mental health: evidence, challenges and future directions," *BJPsych Open*, vol. 9, no. 4, Jul. 2023, doi: 10.1192/bjo.2023.507.

[105] B. Janga, G. P. Asamani, Z. Sun, and N. Cristea, "A Review of Practical AI for Remote Sensing in Earth Sciences," Aug. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/rs15164112.

[106] Ika Idris, "Jakarta air pollution: The challenges of turning evidence into policy," https://lens.monash.edu/@politics-society/2023/12/12/1386254/jakarta-air-pollution-the-challenges-of-turning-evidence-into-policy.

[107] M. E. E. Alahi *et al.*, "Integration of IoT-Enabled Technologies and Artificial Intelligence (AI) for Smart City Scenario: Recent Advancements and Future Trends," Jun. 01, 2023, *MDPI*. doi: 10.3390/s23115206.

[108] S. M. Popescu *et al.*, "Artificial intelligence and IoT driven technologies for environmental pollution monitoring and management," 2024, *Frontiers Media SA*. doi: 10.3389/fenvs.2024.1336088.

[109] E. Nizeyimana, D. Hanyurwimfura, J. Hwang, J. Nsenga, and D. Regassa, "Prototype of Monitoring Transportation Pollution Spikes through the Internet of Things Edge Networks," *Sensors (Basel)*, vol. 23, no. 21, Nov. 2023, doi: 10.3390/s23218941.

[110] H. Nazir and J. Fan, "Revolutionizing Retail: Examining the Influence of Blockchain-Enabled IoT Capabilities on Sustainable Firm Performance," *Sustainability (Switzerland)* ,

vol. 16, no. 9, May 2024, doi: 10.3390/su16093534.
[111] M. Ghamari, "Journal of Pollution Emerging Technologies and Strategies to Combat Air Pollution: A Comprehensive Review," 2023, doi: 10.37421/2684-4958.2023.6.298.
[112] A. Valencia-Arias, J. D. González-Ruiz, L. Verde Flores, L. Vega-Mori, P. Rodríguez-Correa, and G. Sánchez Santos, "Machine Learning and Blockchain: A Bibliometric Study on Security and Privacy," *Information (Switzerland)*, vol. 15, no. 1, Jan. 2024, doi: 10.3390/info15010065.
[113] G. Habib, S. Sharma, S. Ibrahim, I. Ahmad, S. Qureshi, and M. Ishfaq, "Blockchain Technology: Benefits, Challenges, Applications, and Integration of Blockchain Technology with Cloud Computing," Nov. 01, 2022, *MDPI*. doi: 10.3390/fi14110341.
[114] C. Shetty *et al.*, "A Machine Learning Approach for Environmental Assessment on Air Quality and Mitigation Strategy," *Journal of Engineering*, vol. 2024, pp. 1–16, Feb. 2024, doi: 10.1155/2024/2893021.
[115] I. Ahmed, Y. Zhang, G. Jeon, W. Lin, M. R. Khosravi, and L. Qi, "A blockchain- and artificial intelligence-enabled smart IoT framework for sustainable city," *International Journal of Intelligent Systems*, vol. 37, no. 9, pp. 6493–6507, Sep. 2022, doi: 10.1002/int.22852.
[116] A. Dahiya, B. B. Gupta, W. Alhalabi, and K. Ulrichd, "A comprehensive analysis of blockchain and its applications in intelligent systems based on IoT, cloud and social media," *International Journal of Intelligent Systems*, vol. 37, no. 12, pp. 11037–11077, Dec. 2022, doi: 10.1002/int.23032.
[117] E. Goh, D.-Y. Kim, K. Lee, S. Oh, J.-E. Chae, and D.-Y. Kim, "Blockchain-Enabled Federated Learning: A Reference Architecture Design, Implementation, and Verification," Jun. 2023, [Online]. Available: http://arxiv.org/abs/2306.10841
[118] E. Goh, D.-Y. Kim, K. Lee, S. Oh, J.-E. Chae, and D.-Y. Kim, "Blockchain-Enabled Federated Learning: A Reference Architecture Design, Implementation, and Verification," Jun. 2023, [Online]. Available: http://arxiv.org/abs/2306.10841
[119] I. Essamlali, H. Nhaila, and M. El Khaili, "Supervised Machine Learning Approaches for Predicting Key Pollutants and for the Sustainable Enhancement of Urban Air Quality: A Systematic Review," Feb. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/su16030976.
[120] H. Xue, D. Chen, N. Zhang, H.-N. Dai, and K. Yu, "Integration of blockchain and edge computing in internet of things: A survey," *Future Generation Computer Systems*, vol. 144, pp. 307–326, Jul. 2023, doi: 10.1016/j.future.2022.10.029.
[121] S. Dong, K. Abbas, M. Li, and J. Kamruzzaman, "Blockchain technology and application: an overview," *PeerJ Comput Sci*, vol. 9, p. e1705, Nov. 2023, doi: 10.7717/peerj-cs.1705.
[122] N. S. Gupta, Y. Mohta, K. Heda, R. Armaan, B. Valarmathi, and G. Arulkumaran, "Prediction of Air Quality Index Using Machine Learning Techniques: A Comparative Analysis," *J Environ Public Health*, vol. 2023, pp. 1–26, Jan. 2023, doi: 10.1155/2023/4916267.
[123] Bappe Kim, "Air Pollution in Seoul," https://www.kaggle.com/datasets/bappekim/air-pollution-in-seoul.
[124] Rob J Hyndman and George Athanasopoulos, *Forecasting: Principles and Practice* , 3rd ed. 2021.
[125] K. Benidis *et al.*, "Deep Learning for Time Series Forecasting: Tutorial and Literature Survey," Apr. 2020, doi: 10.1145/3533382.

[126] Y. Lyu, Q. Ju, F. Lv, J. Feng, X. Pang, and X. Li, “Spatiotemporal variations of air pollutants and ozone prediction using machine learning algorithms in the Beijing-Tianjin-Hebei region from 2014 to 2021,” *Environmental Pollution*, vol. 306, p. 119420, Aug. 2022, doi: 10.1016/J.ENVPOL.2022.119420.
[127] L. X. ZHANG *et al.*, “Enrichment Reporter System of Genome Editing Positive Cells,” *Chinese Journal of Analytical Chemistry*, vol. 48, no. 1, pp. 1–12, Jan. 2020, doi: 10.1016/S1872-2040(19)61206-5.
[128] T. Xayasouk, H. M. Lee, and G. Lee, “Air pollution prediction using long short-term memory (LSTM) and deep autoencoder (DAE) models,” *Sustainability (Switzerland)*, vol. 12, no. 6, Mar. 2020, doi: 10.3390/su12062570.
[129] E. Sharma, R. C. Deo, R. Prasad, A. V. Parisi, and N. Raj, “Deep Air Quality Forecasts: Suspended Particulate Matter Modeling with Convolutional Neural and Long Short-Term Memory Networks,” *IEEE Access*, vol. 8, pp. 209503–209516, 2020, doi: 10.1109/ACCESS.2020.3039002.

List of Publications

1. Eric Nizeyimana, Damien Hanyurwimfura, Ryosuke Shibasaki, Jimmy Nsenga; April 2021, Design of a Decentralized and Predictive Real-Time Framework for Air Pollution Spikes Monitoring; 2021 IEEE 6th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), 10.1109/ICCCBDA51879.2021.9442611 / https://ieeexplore.ieee.org/document/9442611

   Conference metrics

Journal Name	IEEE Xplore
Publisher	IEEE
Indexing	Scopus

2. Eric Nizeyimana, Jimmy Nsenga, Ryosuke Shibasaki, Damien Hanyurwimfura, JunSeok Hwang, 2022. Design of Smart IoT Device for Monitoring Short-term Exposure to Air Pollution Peaks, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 13, No. 1, 2022, DOI: 10.14569/IJACSA.2022.0130103

   Journal Metrics

Journal Name	International Journal of Advanced Computer Science and Applications
Publisher	SCIENCE & INFORMATION SAI ORGANIZATION LTD
Indexing	Scopus

3. Nizeyimana, Eric, Damien Hanyurwimfura, Junseok Hwang, Jimmy Nsenga, and Dereje Regassa. 2023. "Prototype of Monitoring Transportation Pollution Spikes through the Internet of Things Edge Networks" Sensors 23, no. 21: 8941. https://doi.org/10.3390/s23218941

   Journal Metrics

Journal Name	Sensors
Publisher	MDPI
Indexing	SCI

4. Eric Nizeyimana, Damien Hanyurwimfura, Jimmy Nsenga, Jules Zirikana, Bonaventure Karikumutima, Junseok Hwang, Irene Niyonambaza Mihigo; Revolutionizing Air Pollution Spikes Analysis with A Blockchain-Driven Machine Learning Framework; Paper under review

<u>Journal Metrics</u>

Journal Name	International Journal of Intelligent Systems
Publisher	HINDAWI
Indexing	SCI

Appendix

The table below provides an overview of the research questions aligned with their corresponding objectives, summarizing the materials and methods used along with the outcomes. Each research question is mapped to specific objectives, detailing the experimental setup, tools, and techniques employed in the study. This structured presentation ensures clarity in understanding the relationship between the research questions, the methodologies applied, and the results obtained, thereby highlighting the effectiveness and relevance of the study's approach.

SN	Research Question	Research Objectives	Materials and Methods	Outcomes (Paper generated)
1	How can edge IoT, blockchain, and AI be integrated to develop a real-time multi-pollutant sensor for monitoring air pollution spikes?	Review the existing air pollution monitoring for their weakness and limitations	• Existing literature review • Design the framework	Design of a Decentralized and Predictive Real-Time Framework for Air Pollution Spikes Monitoring
2	What are the key features and functionalities of short-term prediction AI for alerting emitters of potential risks exceeding exposure standards due to pollution spikes?	Development and evaluation of a real-time multi-pollutant (RTMP) sensor integrated and application blockchain identity for tokenizing pollution offenses on the data.	• Improve IoT power management • Apply interrupt to wake up sensors	Design of Smart IoT Device for Monitoring Short-term Exposure to Air Pollution Peaks

3	How can IoT energy usage be optimized in a hardware-software smart IoT device for monitoring short-term exposure to pollution peaks?	Optimize IoT energy usage in a hardware-software smart IoT device for monitoring short-term exposure to pollution peaks.	• IoT edge network for prototype design • Mathematical tools for the algorithm design	Prototype of Monitoring Transportation Pollution Spikes through the Internet of Things Edge Networks
4	What machine learning techniques can be effectively utilized to analyze collected spike data and propose fines for peak emitters in air quality monitoring systems?	Enhance the performance of air quality monitoring systems by leveraging machine learning to analyze collected spike data and propose fines for peak emitters.	• Develop smart contracts • Machine learning Models for forecasting	Revolutionizing Air Pollution Spikes Analysis with A Blockchain-Driven Machine Learning Framework