Integration of Fourth Industrial Revolution technologies and indigenous knowledge in developing a smart and integrated pollution monitoring system: case of Free State province

PAMELA RAMBA

Dissertation submitted in fulfillment of the requirements for the Degree.

MASTER OF INFORMATION TECHNOLOGY
in the

Department of Information Technology

Faculty of Engineering, Built Environment and Information Technology at the

Central University of Technology, Free State

Supervisor: Ms. Mpho Mbele

Co-supervisor: Prof. Muthoni Masinde

2025

DECLARATION

I, Pamela Ramba, student number , hereby declare that the contents of this dissertation are my original work. This dissertation is submitted for the degree of Master of Information Technology at the Central University of Technology, Free State. This work has not been submitted for any other degree or to any other institution. This work was done under the guidance and close supervision of Ms Mpho Mbele and Prof. Muthoni Masinde.

Pamela Ramba

Date: ____

Signature: ____

In our capacity as supervisors of this dissertation, we certify that the above statements are true to the best of our knowledge.

Ms. Mpho Mbele

Date: ____

Signature: ____

Prof. Muthoni Masinde

Date: ____

Signature: ____

ACKNOWLEDGEMENTS

Firstly, all the praises unto God, who gave me strength throughout this mind- challenging phase of my life. I thank God for the clear vision He embedded in my heart; it was indeed for a time such as this, and it prevailed against all the odds.

My Heartfelt gratitude to my only biological sister, Ms. Fezeka Ramba, who stretched out her faith, declaring light where I grew weary. A special thank you to my mother, Mrs. Ntombomzi Victoria Ramba, who interceded on my behalf. My sisters' kids, Yolanda, Inganathi, and Angellica Ramba, who made sure that I never lost my smile. This would not have been possible without my family's unwavering love and continued support.

A special thank you to my main supervisor, Ms. Mpho Mbele. Her soft voice poured out a glimpse of hope during tense, frustrating, and confusing times. Her sincere and genuine guidance and teachings made it possible for me to face all my academic fears.

A big thank you to my co-supervisor, Professor Muthoni Masinde, for always instilling academic excellence and integrity. She poured out all her distinguished knowledge graciously. I can only say I was groomed by the best in our research field.

A heartfelt gratitude to Mr. Joseph Sandt (IT department manager of Robert Mangaliso Sobukwe Hospital) for all the support and Bible-based words of affirmation. I will forever be grateful for all that you did for me.

A big thank you to my line manager at Siemens-Healthineers, Mr. Bradly Arendse. He believed that my project had potential the first time I spoke with him about it. His support allowed me to learn passionately and step out boldly.

A special thank you to all my friends who supported me.

ABSTRACT

South Africa's National Ambient Air Quality Standards (NAAQS) emphasize clean air as a fundamental pillar of human health and well-being. Yet, in mining-intensive regions of the country, persistent exposure to harmful emissions continues to compromise both occupational safety and community health. Mine workers remain highly susceptible to respiratory illnesses such as silicosis and pulmonary tuberculosis (PTB), at the same time, surrounding populations experience secondary exposure as particulate matter, sulfur dioxide, nitrogen oxides, and other toxic gases are dispersed into residential areas, particularly during windy seasons. Beyond immediate respiratory complications, this exposure contributes to long-term cardiovascular disease, reduced life expectancy, and escalating healthcare costs. Conventional monitoring systems, while present, are often static, resource-intensive, and geographically limited, providing delayed or fragmented data that fail to support real- time decision-making. Moreover, enforcement of NAAQS remains inconsistent, with communities largely excluded from monitoring processes, leaving them without accessible tools to validate or report their lived experiences of pollution. The disconnect between regulatory frameworks, technological capacity, and community participation perpetuates environmental injustice in mining-affected provinces such as the Free State. Air pollution, dominated by particulate matter and toxic gases, therefore remains one of the most significant environmental and public health threats in South Africa, as in many developing countries, with strong links to increased mortality, morbidity, and climate impacts. Addressing this challenge requires innovative monitoring frameworks that transcend the limitations of traditional systems by integrating advanced technological tools with inclusive, culturally resonant approaches rooted in Indigenous Knowledge (IK).

This dissertation explores the integration of Fourth Industrial Revolution (4IR) technologies, specifically Machine Learning (ML) and the Internet of Things (IoT), with an Indigenous Knowledge (IK) approach to design, implement, and evaluate a smart, adaptive, and community-centred air pollution monitoring system in the Free State Province. To achieve this, the study employed a mixed-methods research design. Quantitative techniques included the collection and pre-processing of datasets from the Pelonomi air quality monitoring station in Mangaung, complemented by limited deployments of IoT wireless sensors. Supervised ML algorithms, namely: Random Forest, Gradient Boosting, Support Vector Machine, and Decision Tree Regression, were trained to forecast pollutants such as $PM_{2.5}$, PM_{10}, and SO_2 up to four days in advance. These predictions were benchmarked against NAAQS thresholds. Qualitative techniques, on the other hand, involved structured surveys and interviews with mine workers,

residents, and environmental knowledge holders in the Lejweleputswa district. Indigenous indicators, such as dust storms during windy seasons, odour perception, and respiratory discomfort, were systematically documented and modelled using Fuzzy Cognitive Maps (FCMs), ensuring formal representation, validation, and integration with scientific datasets.

A three-phase, agile systems development approach guided the framework's implementation: (1) Data Acquisition and Prediction — real-time and secondary data collection, cleaning, and ML-based forecasting; (2) Indigenous Knowledge Integration — rigorous modelling and validation of IK indicators through FCMs to complement scientific predictions; and (3) Communication and Dissemination – the development of a mobile Android application that provides accessible forecasts, validates machine outputs against IK, and allows communities to log observations directly. This participatory design ensured inclusivity, especially for semi-literate and illiterate populations often excluded from technology-driven monitoring.

Evaluation confirmed both technical robustness and societal acceptance. Forecasts showed strong alignment with South African Weather Service (SAWS) datasets, demonstrating predictive reliability. Cross-validation revealed a high degree of complementarity between IK observations and scientific outputs, reinforcing the value of hybrid knowledge systems. A community-based evaluation achieved an 89% user approval rating, highlighting the usability, cultural relevance, and potential for wider societal adoption.

The contributions of this research are multi-dimensional. Scientifically, it advances environmental informatics by demonstrating a hybrid framework that systematically integrates qualitative IK indicators with quantitative ML forecasts. Technologically, it delivers one of the first prototypes in South Africa to merge IoT and ML forecasting with IK validation in a mobile platform. Societally, it empowers marginalized communities by transforming them from passive recipients of information to active contributors in environmental monitoring. Policy-wise, it provides empirical evidence for inclusive environmental governance, directly aligning with national frameworks and global agendas, including SDG 3 (Good Health and Well-being), SDG 11 (Sustainable Cities and Communities), and SDG 13 (Climate Action).

Limitations included restricted access to mining shafts, which limited large-scale IoT deployment and necessitated reliance on secondary datasets. Nevertheless, the results underscore the feasibility, scalability, and replicability of the system. Future work should expand sensor coverage, enhance interoperability across regions, and extend the hybrid framework to other domains such as water and soil pollution monitoring.

LIST OF ABBREVIATIONS

Abbreviation	Definition
4IR	Fourth Industrial Revolution
AI	Artificial Intelligence
AMD	Acid Mine Drainage
APK	Android Application Package
CH_4	Methane
CM	Cognitive Maps
CO	Carbon Monoxide
CO_2	Carbon Dioxide
DEA	Department of Environmental Affairs
FCM	Fuzzy Cognitive Maps
FTP	File Transfer Protocol
GPRS	General Packet Radio Service
GPS	Global Positioning System
ICT	Information Communication Technology
IK	Indigenous Knowledge
IKI	Indigenous Knowledge Indicators
IKS	Indigenous Knowledge Systems
IoT	Internet of Things
LK	Local Knowledge
ML	Machine Learning
NAAQS	National Ambient Air Quality Standards
O_3	Ozone
$PM_{2.5}$	Particulate $matter_{2.5}$
PM_{10}	Particulate $matter_{10}$
SAAQIS	South African Air Quality and Information Systems
SAWS	South African Weather Service
SO_2	Sulfur Dioxide
TB	Tuberculosis
ToT	Transfer of Technology

PUBLICATIONS

The following publication directly emanated from some of the work presented in this dissertation:

Ramba, P., Mbele, M. and Masinde, M., 2023, May. Integration of Fourth Industrial Revolution Technologies and Indigenous Knowledge in Developing a Smart and Integrated Pollution Monitoring System. In *2023 IST-Africa Conference (IST-Africa)* (pp. 1-9). IEEE.

TABLE OF CONTENTS

List of Figures

LIST OF TABLES

CHAPTER ONE: INTRODUCTION AND BACKGROUND INFORMATION

1.1. Introduction

South Africa, a country with the deepest mining operations and home to half of the world's gold reserves, faces highly complex environmental problems due to the exploitation of mining ores (Stewart and Malatji, 2018). Gold mining generates significant waste, which is the major contributor to pollution in the country (Gold, 2014). According to Duisebekova et al., (2019), pollution is defined as an unfavourable alteration to the physical, chemical, or biological properties of the air, soil, or water, and this alteration has the potential to negatively impact life or provide a health risk to any living thing. Aguirre-Ayerbe et al., (2020), also emphasise that the environmental consequences of pollution are equally alarming. Air pollution contributes significantly to climate change through the emission of greenhouse gases such as carbon dioxide and methane (Abbasi, 2018).

Notably, the mining industry is a major contributor to environmental degradation, releasing heavy metals, generating mine dumps, and causing acid mine drainage (Bravo et al., 2012). Moreover, mine workers in South Africa remain susceptible to occupational diseases such as pulmonary tuberculosis (PTB) due to ongoing conditions of daily mining operations, this is despite the efforts by the Chamber of Mines of South Africa to significantly improve the occupational health and safety performance (Stewart and Malatji, 2018). The investments and efforts in technology transfer such as the deployment of technology equipment, machines, and training in the workplace have failed mainly due to underutilization, resulting in stagnant safety performance and dissatisfaction as the technology was not effectively utilized (Lööw, 2022).

Emerging solutions have therefore emphasized the adoption and use of indigenous knowledge systems in environmental monitoring and pollution (Fernández- Llamazares et al., 2020). There is also a need to explore the effectiveness of 4IR technologies such as Artificial Intelligence (AI) and Internet of Things (IoT) alongside non-technological approaches like indigenous knowledge. Integrating these methods could enable the development of a smart and adaptive pollution monitoring system. IoT enables the deployment of interconnected sensors and devices across different and remote areas to collect air quality data continuously and simultaneously (Chen et al., 2018). Among other things, these sensors collect dust, including both particulate matter ($PM_{2.5}$) and particulate matter (PM_{10}), as well as sulfur dioxide (SO_2), nitrogen dioxide (NO_2), carbon monoxide (CO), and ozone (O_3) (Delabrida et al., 2016). In addition, Atzori et al., (2017) indicate that the data from the IoT devices are transmitted in real-time through a centralised database to allow monitoring on a permanent basis, and instant data is available to all stakeholders if relevant. Bzai et al., (2022), allude that Machine Learning (ML) algorithms play a crucial role in analysing an extensive amount of data generated by IoT devices. ML uses algorithms that can identify patterns, predict pollution levels, and provide executable recommendations (Ibrahim and Abdulazeez, 2021).

Predictive models can forecast pollution trends based on a variety of attributes such as historical data, weather conditions, industrial activities, traffic patterns, and other factors, depending on project requirements (Paula et al., 2019). In addition, indigenous systems and community-based monitoring programmes foster collective responsibility and environmental stewardship. Indigenous Knowledge (IK) inculcates in the people the concept of collective responsibility to take care of the environment in a sustainable manner (Dahl and Hansen, 2019). Hiwasaki et al., (2014), mention that collaboration between IK and modern scientists or technologies can lead to innovative and effective solutions for air pollution. An example by Fernández-Llamazares et al., (2020), shows that integrating IK with advanced monitoring technologies can enhance the accuracy and relevance of air quality assessments. The insights brought by communities are also helping to interpret findings and identify sources of pollution that may otherwise be overlooked. Such collaborations can foster mutual learning and respect, paving the way for more inclusive environmental policies and practices (Lusilao-Makiese et al., 2013).

The main aim of this research was to integrate Fourth Industrial Revolution (4IR) technologies with Indigenous Knowledge (IK) in the development of a smart, adaptive, and community-centred air pollution monitoring system tailored to the Free State Province, South Africa. This aim was achieved through the successful design and implementation of a functional prototype that combined multiple knowledge and data sources. The prototype was developed and tested using real datasets from the South African Weather Service (SAWS) ground station in the Free State. On the technological front, machine learning algorithms were applied to forecast air quality up to four days in advance, demonstrating the potential of data-driven early warning systems. In parallel, Indigenous Knowledge indicators were systematically collected through rigorous surveys and modelled using Fuzzy Cognitive Maps (FCMs), which enabled formal representation and validation of community-based environmental knowledge.

To enhance accessibility and community engagement, the research also produced a mobile

Android application that disseminates forecasted pollution levels and allows communities to log IK-based observations. Limited pilot testing of the system demonstrated that community-reported indicators corresponded closely with both ground station and satellite measurements, confirming the feasibility and added value of integrating IK with scientific approaches. Beyond technical validation, the system achieved 89% user approval during evaluation, underscoring its usability, relevance, and potential societal impact. Collectively, these achievements demonstrate that a hybrid monitoring system, bridging advanced 4IR technologies with Indigenous Knowledge, offers a more comprehensive, accurate, and context-sensitive framework for environmental monitoring and management in South Africa.

1.2. Problem Statement

Pollution remains one of the most pressing environmental and public health challenges worldwide, with air pollution being particularly severe due to its strong association with respiratory and cardiovascular diseases, reduced life expectancy, and escalating healthcare costs (Abbasi, 2018). In South Africa, the challenge is compounded by extensive mining activities that release harmful particulate matter, greenhouse gases, and heavy metals into the environment. The Free State Province, characterized by deep-level gold mining and high dust emissions, faces disproportionate exposure to these risks, affecting both mine workers and surrounding communities.

Although South Africa has established regulatory frameworks such as the National Ambient Air Quality Standards (NAAQS), their enforcement and effectiveness are limited by several persistent challenges. Conventional pollution monitoring systems are often **static, resource-intensive, and sparsely distributed**, restricting the availability of timely, localized data. Communities in mining regions remain largely excluded from monitoring and reporting processes, which undermines collective responsibility and weakens adaptive responses to environmental threats.

Investments in advanced equipment and workplace training have not yielded proportional improvements in environmental safety. Studies show that many technological interventions in the mining sector are **underutilized**, primarily due to limited capacity for local adoption, poor integration into decision-making processes, and a lack of cultural or contextual relevance (Lööw, 2022). As a result, environmental management practices remain fragmented, with little meaningful participation from the very communities most affected by pollution.

On the other hand, Indigenous Knowledge (IK) provides valuable, community- embedded insights into environmental change through traditional indicators such as shifts in wind patterns, vegetation conditions, and respiratory symptoms observed in local populations. However, IK has not been systematically integrated with modern technologies in South Africa’s air quality monitoring frameworks. The absence of mechanisms to validate, digitize, and scale such knowledge limits its contribution to evidence-based decision-making.

Furthermore, opportunities offered by Fourth Industrial Revolution (4IR) technologies, including the Internet of Things (IoT), Artificial Intelligence (AI), and machine learning, have not been fully leveraged to address these gaps. While IoT devices and machine learning models are widely used globally for air quality forecasting, their deployment in South Africa has been slow and uneven, constrained by costs, infrastructure limitations, and inadequate local customization. These gaps point to an urgent need for a hybrid, context-sensitive system that integrates 4IR technologies with Indigenous Knowledge to provide adaptive, real-time, and community-validated air pollution monitoring. Such a system would not only generate accurate forecasts based on ground station and satellite data but also incorporate local knowledge and participatory inputs, thereby fostering inclusivity, improving data relevance, and strengthening collective environmental stewardship.

1.3. Research Objectives and Questions

The overall aim of this research was to design, develop, and evaluate an **integrated, smart, and adaptive air pollution monitoring system** for the Free State Province, South Africa, by combining **Fourth Industrial Revolution (4IR) technologies,** specifically Artificial Intelligence (AI) and the Internet of Things (IoT), with **Indigenous Knowledge (IK)**. The goal was to demonstrate that integrating scientific and traditional approaches can yield a more accurate, inclusive, and context-sensitive framework for monitoring and predicting air pollution. The following research questions guided the study:

1.3.1 Research Questions

a) To what extent have artificial intelligence and the Internet of Things been incorporated into pollution monitoring systems in South Africa and beyond?

b) How can indigenous knowledge systems be applied to monitor pollution in South Africa’s underground mines?

c) To what extent can the Fourth Industrial Revolution technologies be used to complement indigenous knowledge systems-based air pollution monitoring for the Free State province in South

Africa and beyond?

To address the questions, the following objectives were generated:

1.3.2 Specific Objectives

a) To investigate the extent to which AI and IoT have been incorporated into pollution monitoring systems in South Africa and beyond.

Achievement: A comprehensive literature and bibliometric analysis was conducted, mapping trends in IoT and AI applications for environmental monitoring globally and nationally. This identified gaps in South African contexts, where adoption remains limited, fragmented, and underutilized compared to global practices.

b) To investigate the indigenous knowledge systems that are applicable in monitoring pollution in South Africa's underground mines.

Achievement: IK indicators were systematically collected through surveys and interviews with community experts in the Free State mining region. These indicators were then rigorously analyzed and modelled using **Fuzzy Cognitive Maps (FCMs)**, enabling formal representation and validation against scientific datasets.

c) To develop and evaluate a generic framework that integrates AI and IoT with indigenous knowledge systems to achieve a smart and interoperable pollution monitoring system.

Achievement: A multi-component framework was designed and implemented, comprising (i.) data acquisition from Arduino wireless sensors and Pelonomi monitoring station in Mangaung. (ii.) Machine Learning–based forecasting of air quality up to four days in advance, and (iii.) Integration of validated IK indicators for enhanced contextual interpretation.

d) To evaluate and assess the effectiveness of the model developed in objective (3) above, using the case of the Free State.

Achievement: The system prototype was evaluated using data from the **Pelonomi monitoring station in Mangaung**. Additionally, a mobile Android application was developed to disseminate forecasts and facilitate community reporting. Pilot tests demonstrated **89% user approval**, confirming the usability, relevance, and societal value of the system.

1.4 The Solution Approach

The problem outlined in Section 1.2 was addressed through the design and implementation of a conceptual and practical framework that integrates Fourth Industrial Revolution (4IR) technologies with Indigenous Knowledge (IK) to create a smart, adaptive, and community-centered air pollution monitoring system for the Free State Province. The framework was structured into three interconnected components:

a) **Data acquisition and prediction** – the collection, preparation, and analysis of air quality datasets, followed by the application of machine learning algorithms to forecast pollution levels.

b) **Indigenous Knowledge gathering and modelling** – the systematic collection of local indicators of pollution through community engagement, and their integration into predictive models using formal techniques.

c) **Communication and dissemination** – the delivery of integrated forecasts and community reports to stakeholders through digital platforms, with emphasis on accessibility and inclusivity.

To ensure adaptability and iterative improvement, the framework was developed using an agile systems development life cycle. The implementation followed three phases:

1.4.1 Phase 1: Data Collection and Forecasting

Air quality data was obtained from the Pelonomi air pollution monitoring station in Mangaung, which served as the main secondary dataset for training and testing predictive models. In addition, a limited number of IoT sensors were installed and tested in selected locations to assess the feasibility of real-time monitoring in mining-affected areas. Although large- scale deployment in mines was restricted due to access limitations, the sensor testing confirmed the potential for future integration of real-time data streams.

The datasets were processed using supervised machine learning algorithms, including Random Forest, Gradient Boosting, Support Vector Machine (SVM), and Decision Tree Regression, to forecast air pollution levels ($PM_{2.5}$, PM_{10}, SO_2) up to four days in advance. Forecasts were benchmarked against the National Ambient Air Quality Standards (NAAQS), ensuring that predictions were meaningful for public health and policy purposes.

1.4.2 Phase 2: Indigenous Knowledge Integration

Indigenous Knowledge indicators of air pollution were systematically collected through structured questionnaires administered to mine workers, community members, and IK experts in the Lejweleputswa district. The responses provided valuable insights into local perceptions of air quality, including observations such as increased dust during windy seasons, respiratory discomfort, and visible atmospheric changes. These indicators were analyzed and represented using Fuzzy Cognitive Maps (FCMs), which allowed the qualitative IK responses to be formally modelled and linked to scientific datasets. This step ensured that the system incorporated both measurable environmental data and lived community experiences, enhancing its contextual relevance.

1.4.3 Phase 3: Dissemination through Mobile Application

A mobile Android application was developed to serve as the communication interface between the system and its end-users. The app provided air quality forecasts, delivered early warnings, and allowed communities to report IK-based pollution observations directly into the system. Limited pilot testing confirmed that community-reported indicators closely corresponded with scientific predictions. The application was highly rated for its ease of use and relevance, particularly among semi-literate and illiterate users who had previously been excluded from scientific monitoring processes.

1.4.4 Case Study and Evaluation

The complete framework was tested in Free State province. Evaluation results demonstrated that the system was both scientifically robust and socially relevant, achieving an 89% user approval rating. This confirms that hybrid approaches combining 4IR technologies with Indigenous Knowledge not only improve predictive accuracy but also enhance community acceptance, usability, and ownership of environmental monitoring systems.

1.4.5 The Methodological approach

The solution approach was implemented through a mixed-methods research design, which combined both quantitative and qualitative techniques to strengthen the validity, reliability, and contextual relevance of the study. The choice of mixed methods was guided by the recognition that air pollution is not only a measurable physical phenomenon but also a lived social experience, particularly for communities affected by mining- related emissions in the Free State.

1.4.6 Quantitative Techniques

Quantitative methods included the application of supervised machine learning algorithms, Support Vector Machine (SVM), Decision Tree Regression, Random Forest, and Gradient Boosting, trained on air quality datasets from the Pelonomi monitoring station. A statistical correlation analysis was conducted to determine the relationship between pollution levels and proximity to mining shafts, and a comparative analysis was undertaken against the National Ambient Air

Quality Standards (NAAQS) thresholds to establish compliance. These techniques enabled the development of predictive models capable of forecasting pollution levels up to four days in advance.

Quantitative techniques were also employed to capture the Indigenous Knowledge (IK) of local communities and mine workers. This was primarily achieved through online questionnaires administered using the Survicate platform in two phases:

- **Phase 1 (July–September 2022):** Aimed at uncovering the broader spectrum of challenges communities face due to mining-related pollution and identifying local IK indicators.
- **Phase 2 (June–October 2024):** Focused on evaluating the developed system, confirming whether it met its objectives, and assessing community acceptance.

The questionnaires, which included both open- and closed-ended questions, targeted mine workers, residents, medical professionals, and guest contributors. Responses provided insights into the lived experiences of pollution, common indicators recognized by the community (e.g., dust clouds, odours, respiratory discomfort), and feedback on the usability of the mobile application. The responses were analyzed and modelled using Fuzzy Cognitive Maps (FCMs) and the Mental Modeler tool, which enabled formal representation of qualitative IK data alongside scientific datasets.

1.4.7 Value of the Mixed Methods Approach

Adopting a mixed methods design brought several key advantages to the study:

- **Collaboration:** Quantitative methods provided objective forecasts and statistical rigor, while qualitative approaches contextualized these findings by explaining community experiences. This synergy offered a more holistic understanding than either method could achieve alone.
- **Complementarity:** Results from one method informed and validated the other. For example, machine learning forecasts were cross-checked against community questionnaire responses, which revealed that local residents consistently associated high pollution with windy seasons when dust was blown into residential areas. Similarly, miners reported continuous daily exposure, which aligns with model outputs showing elevated concentrations closer to the shafts.
- **Expansion:** Employing multiple methods broadened the scope of the research. By combining ground station datasets, machine learning predictions, and IK indicators, the study expanded the range of knowledge sources integrated into one monitoring framework.

- **Participatory and Appreciative of Local Knowledge:** The questionnaire ensured that the voices of semi-literate and illiterate community members were systematically captured. Their input helped define IK indicators, which were then embedded into the system design, ensuring that the monitoring framework reflected lived experiences and cultural realities.
- **Acceptance and Usability:** By integrating community knowledge, the system was not perceived as a top-down technological imposition but as a collaborative solution. This inclusivity increased end-user confidence and contributed to the **89% approval rating** recorded during system evaluation.

1.5 Significance and Contribution of the Study

This study makes significant contributions to scientific knowledge, technological innovation, and community-centered environmental management by demonstrating how Fourth Industrial Revolution (4IR) technologies can be integrated with Indigenous Knowledge (IK) to address air pollution challenges in the Free State Province, South Africa.

1.5.1 Scientific Contribution

The study advances environmental informatics by developing a **hybrid framework** that integrates quantitative datasets (ground station measurements, limited IoT sensor readings, and machine learning forecasts) with qualitative indicators derived from Indigenous Knowledge.

It demonstrates the application of **Fuzzy Cognitive Maps (FCMs)** to systematically model and validate IK indicators, ensuring that community knowledge is represented in a scientifically robust manner.

The research further contributes to predictive modelling by showing that supervised machine learning algorithms can reliably forecast air quality up to **four days in advance**, offering early warning capacity in contexts with limited monitoring infrastructure.

1.5.2 Technological Contribution

A **functional prototype system** was developed, consisting of three integrated components: (i) machine learning-based forecasting of air quality using SAWS datasets and tested IoT sensors, (ii) formal modelling of IK indicators, and (iii) an Android mobile application for dissemination and community engagement.

The system provides **real-time access** to forecasts while enabling communities to log IK-based observations, thereby establishing one of the first **participatory mobile platforms for air**

pollution monitoring in South Africa.

1.5.3 Societal Contribution

By incorporating Indigenous Knowledge, the system ensured that **local communities were not passive data recipients but active contributors** to environmental monitoring. This enhanced acceptance and usability of the system, as evidenced by the **89% user approval rating** obtained during evaluation in the Free State.

The early detection and four-day forecasting capability of the system empower **mine workers, households, and decision-makers** to take proactive measures, thereby reducing health risks and improving resilience against pollution exposure.

The approach is particularly valuable for semi-literate or illiterate populations, who rely heavily on lived experience and local indicators. Integrating these perspectives into the system increases accessibility, relevance, and community ownership.

1.5.4 Policy and Research Contribution

The study provides empirical evidence for the **integration of IK into formal pollution monitoring systems**, contributing to national discussions on inclusive and sustainable environmental management.

The framework supports compliance with **National Ambient Air Quality Standards (NAAQS)**. It aligns with global agendas such as the **Sustainable Development Goals (SDGs)**, particularly SDG 3 (Good Health and Well-Being), SDG 11 (Sustainable Cities and Communities), and SDG 13 (Climate Action).

The system offers a replicable model for future research, where hybrid approaches combining 4IR technologies and IK can be extended to other provinces and to different domains such as water or soil pollution.

1.6 System Evaluation Criteria

The effectiveness of the developed air pollution monitoring system was assessed against the research objectives using a **three-tier evaluation strategy**: scientific validation, Indigenous Knowledge (IK) validation, and user- centered evaluation.

1.6.1 Scientific Validation

The predictive capacity of the system was evaluated by comparing the four-day-ahead forecasts generated by machine learning algorithms with the actual values recorded at the Pelonomi Ground

Station in Mangaung. Performance was measured using standard metrics such as Mean Absolute Error (MAE) and R^2 values to assess accuracy and reliability. The results confirmed that the supervised learning models (Random Forest, Gradient Boosting, Support Vector Machine, and Decision Tree Regression) provided accurate forecasts of $PM_{2.5}$, PM_{10}, and SO_2, aligned with the thresholds of the National Ambient Air Quality Standards (NAAQS).

1.6.2 Indigenous Knowledge Validation

To evaluate the role of Indigenous Knowledge in the framework, the IK indicators gathered through questionnaires and logged via the Android application were systematically compared against the same-day readings from the Pelonomi Ground Station. Screenshots from the mobile app, which captured user-reported conditions (e.g., dust, visible smog, odours, or respiratory discomfort), were cross-referenced with scientific measurements on the corresponding timescale. This direct comparison demonstrated a strong correlation between IK observations and measured pollution levels, confirming the validity of local knowledge as a complementary data source.

Furthermore, the IK indicators were formally modelled using Fuzzy Cognitive Maps (FCMs), which helped to distinguish between polluted, not polluted, and moderate states. This provided a structured way of validating community perceptions against quantitative datasets, showing that IK not only reflected actual pollution dynamics but also enriched the interpretation of scientific results.

1.6.3 User-Centered Evaluation

System usability and societal relevance were tested through questionnaires administered during pilot evaluations in the Free State. Participants included miners, local residents, and IK experts. Evaluation criteria included the ease of use of the mobile app, clarity of the displayed information, cultural relevance, and perceived usefulness of the forecasts. Results showed that 89% of participants agreed that the system effectively addressed real pollution challenges and was easy to use, while 11% expressed reservations, mainly due to limited access to smartphones.

The integration of IK into the mobile app increased acceptance by ensuring that the system reflected community knowledge and lived experiences. This participatory approach made the system more relatable, increasing trust and engagement among semi-literate or illiterate populations who had previously been excluded from technology-driven monitoring efforts.

1.7 Scope and Limitation

This study focused on air quality in the Free State Province, drawing primarily on historical datasets from the Pelonomi monitoring station, supported by limited testing of IoT sensors and the integration of Indigenous Knowledge (IK) indicators from communities in the Lejweleputswa district. These combined inputs were sufficient to design, build, and evaluate a functional prototype system that was both scientifically robust and socially validated.

The main limitation was the **restricted access to mining sites**, which prevented large-scale deployment of configured wireless sensors and resulted in reliance on secondary data. While this reliance may have reduced the spatial precision of forecasts for locations further from Mangaung, the hybrid framework nevertheless achieved reliable results and strong user approval. Future studies could strengthen the system by deploying primary sensor networks within mining areas for continuous real-time data collection, expanding coverage to additional monitoring stations across the province, and broadening the pool of IK contributors to ensure even greater representativeness of community experiences.

1.8 Dissertation Structure

The rest of the dissertation is organized along the chapters highlighted below.

1.8.1 Chapter 2: Literature Review

This chapter presents a comprehensive analysis of the existing body of literature, which serves as the conceptual framework and basis for this investigation. It highlights the theoretical foundations and technological advancements that underpin the primary achievements of this research. In addition, a bibliometric analysis is presented in this chapter to give a broad overview of the current trends in air pollution prediction and monitoring systems.

1.8.2 Chapter 3: Research Methodology

This chapter discusses the research design and methods employed in the study, as well as the steps taken to achieve the research goals. The methodology that the research took is discussed in depth in the chapter. The chapter also addresses the research questions of the study. It also shows how an integrated system that combines IK, IoT, and machine learning was implemented with a detailed explanation of each component.

1.8.3 Chapter 4: Framework Architecture, Design and Implementation

This chapter presents the framework architecture, analysis, and design of the system as well as the implementation of the system. This chapter also outlines the thorough framework that was followed in creating the air quality monitoring system.

1.8.4 Chapter 5: Results and Discussion

This chapter discusses the prototype results, the performance of the learning algorithm, and the predictions of air pollution. It also discusses the analysis and integration processes that were conducted throughout the research.

1.8.5 Chapter 6: Evaluation, Discussion and Conclusion

This chapter discusses, evaluates and concludes the entire study. It highlights a developed, smart, and adaptive air pollution monitoring framework. The Chapter also highlights the limitations and problems that were encountered throughout the research process. It also highlights probable future work.

CHAPTER TWO: LITERATURE REVIEW

2.1. Introduction

This chapter offers an in-depth exploration of existing literature that forms the foundation of this study. The chapter begins with a bibliometric analysis in section 2.2, which provides a comprehensive overview of the research landscape in air pollution monitoring and prediction research. The findings of the bibliometric analysis guided the identification of research gaps and informed the choice of key descriptors for the subsequent sections of the literature review. In section 2.3, an overview of environmental pollution and the guidelines that must be followed are discussed. Additionally, sections 2.4 to 2.7 highlight the integration of 4IR technologies, and IK, highlighting the transformative potential of these integrative approaches in reshaping air pollution monitoring in underground mining environments, and within mining communities. The last section 2.8 presents a summary of the chapter.

2.2. Bibliometric Analysis on Air Pollution Monitoring and Prediction

To begin this research study, a bibliometric analysis was conducted to gain a comprehensive understanding of the existing literature in this field. Bibliometric analysis is a rigorous and effective technique for examining large volumes of scientific data, enabling researchers to explore the complexities of an evolving field while identifying its frontiers (Donthu et al., 2021). It is used to analyze research patterns within a specific field or by individual researchers, assess the impact of research contributions, identify emerging research areas, find potential research partners, and select appropriate publication outlets (Dilanjani et al., 2025). The goal of the analysis was to evaluate the progress of research on leveraging AI and the IoT for monitoring and predicting air pollution using a bibliometric program called VOSviewer (Fitria et al., 2021). In addition to identifying research trends, key focus areas, and existing gaps, the analysis examined the relationship between the frequency of air pollution monitoring and the development of publications.

The Web of Science (WoS) was used as a primary resource to identify relevant papers on pollution monitoring and prediction. The database was chosen because it is one of the largest databases of abstracts and citations for peer-reviewed scientific publications, books, and conference proceedings, covering a wide range of scientific disciplines (Adisa et al., 2020). Diverse topics

linked to monitoring air pollution using 4IR technologies were combined using the advanced search option. The only restriction was on the review period, which was from 2018 to 2022. The reason behind
the selection dates were to examine notable trends within the five years preceding the conceptualization of this study.

A three-stage approach was followed. These are:(i) data compilation; (ii) data arrangement and data cleaning; (iii) analysis, interpretation, and visualisation (Matandirotya, 2021). The study developed a search string that included the following terms (TS= ("air pollution") AND TS= ("monitor*" or "predict") AND TS= ("machine learning" or "artificial intelligence" or "internet of things" or "wireless sensor network*") for selecting documents. The query was conducted on topic fields, including titles, keywords, indexing fields, and abstract.
In total, 852 documents were found after the search. Eventually, 420 papers were selected after duplicates, studies unrelated to the prediction of air pollution, and studies primarily concerned with impacts, as well as review papers, were filtered out. Most of the retrieved documents are scientific articles, followed by conference papers.
To evaluate the state of air pollution monitoring and prediction, various subgroups were analyzed. These subfields include countries that are most engaged in the field's research, as well as keywords co-occurrence analysis. Section 2.2.1 assesses publication trends and identifies emerging themes related to pollution monitoring and prediction, and Section 2.2.2 focuses on the network analysis.

2.2.1. Predicting Air Pollution Research Trends

Figure 2.1 shows the trend of publication numbers and years from 2018 to 2022. The data show that the highest number of publications was in 2022, with 158 publications, and the lowest was in 2018, with 21 publications.

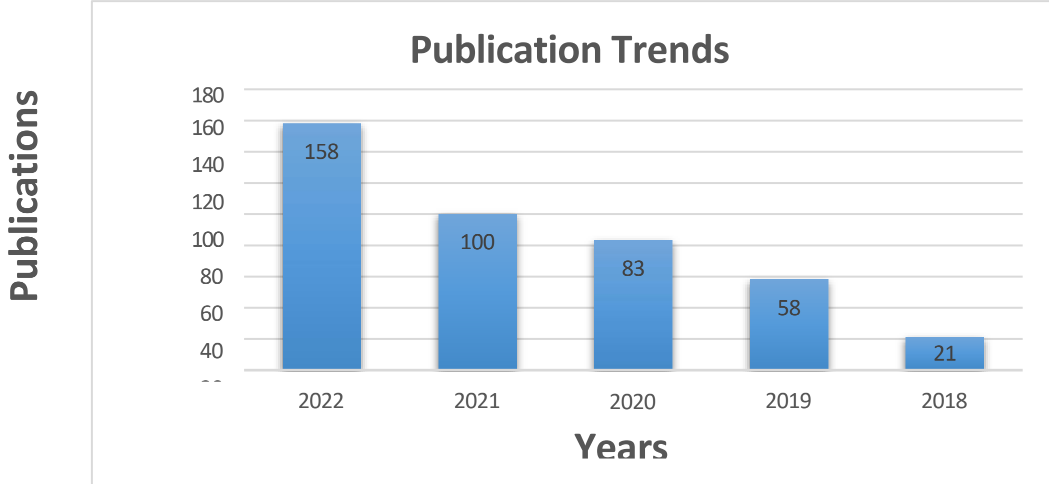


Figure 2.1: Distribution of annual publications air pollution prediction research from 2018 to 2022.

Diverse countries have contributed to the publications on air pollution prediction. Based on the first author's country affiliation, the countries were given ranks, and only countries ranking in the top ten were considered. Figure 2.2 illustrates the above information. The leading countries are China, with 111 publications from the total of 420 selected publications. The second, third, and fourth are the United States of America (USA), India, and England with 84, 70, and 26 publications, respectively. It is important to note that no African country appears in the top 10, as shown above and illustrated in Figure 2.2 below.

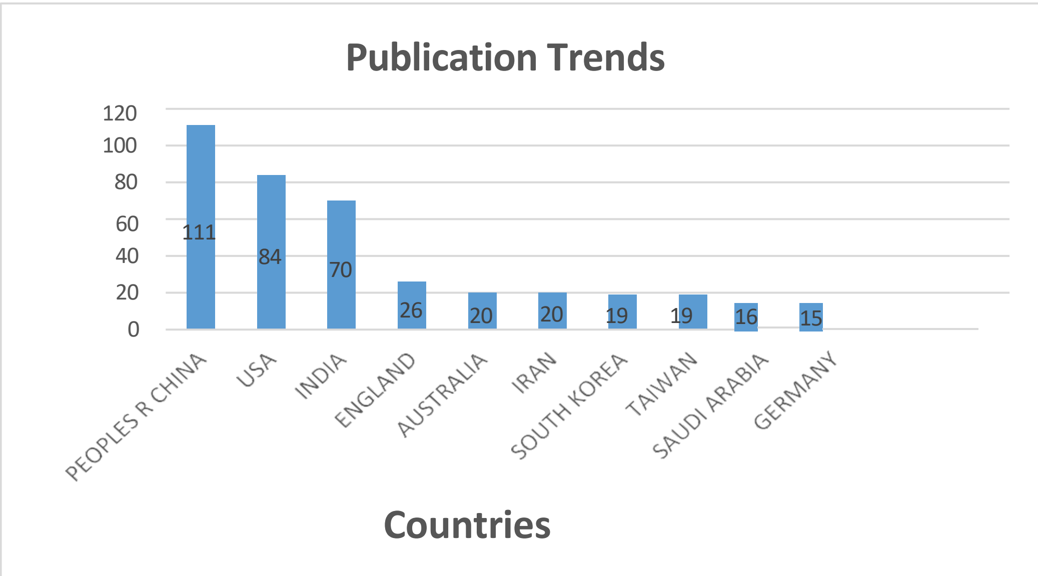


Figure 2.2: Countries in the top ten ranks in publications on air pollution predictions from 2018 to 2022.

2.2.2. Network Analysis

Further VOSViewer software was used to conduct a bibliometric analysis for presentation and visualization of the trends. This was achieved by exporting the Web of Science (WoS) literature data in a text (.txt) format into VOSVierwer. This software is useful for tracking the evolution of scholarship in a variety of fields (Matandirotya, 2021).

2.2.3. Keyword Co-occurrence Analysis

Keyword co-occurrence focuses on understanding the knowledge components and knowledge structure of a scientific/technical field by examining the links between keywords in the literature (Radhakrishnan et al., 2017). It is used to explore the hotspots or subject evolution of a certain field or subject (Tan and Hao, 2022). In this study, keywords that occurred six times or more in the WoS database were selected for the final analysis, and 118 out of 1912 met the threshold. It is indicated in Figure 2.3 how the keywords' co-occurrences are grouped into five network clusters based on their distinct colours. The insights gained from the keyword co-occurrence and clustering analysis highlight the prominence of air pollution prediction, machine learning, and health impacts as research hotspots, thereby providing a strong basis for turning to the substantive discussion of environmental pollution and its monitoring guidelines in the next section.

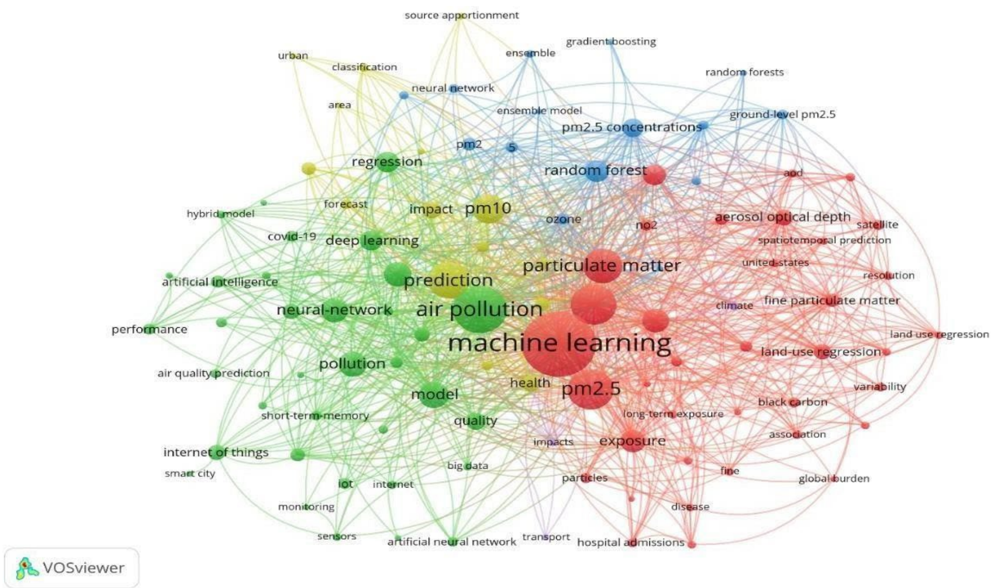


Figure 2.3: Keywords co-occurrence network

Clustering is an activity to divide several data points into groups based on similarities that have been determined beforehand. Clusters are groups of data objects that are alike in the same cluster,

and like different cluster objects (Nainggolan and Eviyanti Purba, 2020). The red cluster as seen in Figure 2.3 has the most elements, meaning that the research on air pollution prediction has several papers with keywords in that cluster. The dominating keywords in the red cluster include machine learning, exposure, black-carbon, and disease to mention a few. This is followed by the green cluster, which contains keywords such as prediction, AI, PM_{10}, and IoT, amongst others. The third cluster is blue, and it has keywords such as neural network, ozone, random forest, and particulate matter ($PM_{2.5}$) concentrations, amongst others. The fourth cluster is yellow with keywords such as urban, forecast, prediction, and health to mention a few. The final cluster, which is purple, includes keywords such as artificial intelligence, climate, impacts, and transport.
Furthermore, prominent keywords represented with bigger circles reflect research direction and active hot topics in researching air pollution prediction. These topics include ML, AI, air pollution, particulate matter, and prediction. They appear in different clusters, thus suggesting that these keywords have gained widespread attention from researchers. Particulate matter ($PM_{2.5}$) is the keyword that appears the most. This indicates that this pollutant is the most extensively researched in terms of air pollution. There are also keywords such as health and hospital admissions indicating that air pollution is harmful to humans, the environment, and other living organisms (Manisalidis et al., 2020).

Building on the patterns and gaps revealed by the bibliometric mapping, the following section shifts to a thematic exploration of environmental pollution, beginning with its definitions, sources, and monitoring guidelines as the foundation for contextualizing air pollution in underground mining environments.

2.3. Environmental Pollution

Underground mining is used to collect ore from under the earth's surface securely, economically, and with the least amount of waste feasible (Ousman et al., 2012). Underground mining, being the primary economic sector in the Lejweleputswa district, results in the formation of diverse types of pollution that affect the surrounding communities (Fashola et al., 2016). The major forms of pollution in this area are air pollution, water pollution, mine dumps, acid mine drainage, heavy metal contamination, and leaching.
Pollution is the presence of or introduction of a substance that has poisonous or harmful effects into the environment (Muduli et al., 2018b). Pollution occurs mostly when there are changes in the biological, chemical, or physical constituents of the environment (air masses, temperature,

climate, etc) (Manisalidis et al., 2020).

Additionally, air pollution consists of chemicals in the air that are present at concentrations above levels deemed safe for human health (Almetwally et al., 2020). It primarily has two environmental effects. Firstly, acid rain can acidify soils and threaten food security (Hou et al., 2021). Secondly, ground-level ozone, which can destroy crops and commercial forests, ultimately impacts the quality of life (Hou et al., 2021). According to the World Health Organization, as described by (Sharifi-Rad et al., 2020), air pollution is one of the biggest environmental risks to human health, contributing to increased mortality and morbidity. The primary pollutants that cause health issues include particulate matter, sulfur dioxide, nitrogen dioxide, and ozone (Orellano et al., 2020). Pollutants like particulate matter can also be caused by dust deposition from the underground mines.

2.3.1 Underground Mine Pollution

The underground mining industry plays a significant role in economic growth and development (Mohsin et al., 2021). However, pollution from underground mines impacts mine workers through direct inhalation of harmful substances, and it also affects residents and local farmers through contaminated air and noise pollution (Muimba-Kankolongo et al., 2022). Mining-related air quality management concerns are therefore mostly focused on particulate matter impacts, such as dust deposition, health effects related to PM_{10} and $PM_{2.5}$, and the mineralogy and chemical makeup of the particles (Aluru et al., 2023).
Specifically, mine workers are mostly affected by pollution, as they directly inhale the polluted air while being underground and while in the mine shaft during their working hours (Wu et al., 2019). Research has also established that many South African mine workers are susceptible to occupational diseases, such as silicosis and pulmonary tuberculosis (PTB), due to daily mining operations. The surrounding communities are also affected by the contaminated air that is widely blown from the mines to homes, especially during the windy season (Mpanza et al., 2020).
Table 2.1 summarises some of the main toxic pollutants that mine workers and residents near the mines are exposed to, and the impact it has on their health.

Table 2.1:Underground mining operations, pollutants, and their health impacts

Pollutants	Health Effects
Carbon monoxide (CO)	This pollutant can reduce oxygen capacity of the blood cells and leads to reduction in oxygen delivery to the body's organs and tissues. Extremely prominent levels can cause fatality (Bull et al., 2020).
Nitrogen dioxide (NO2)	Substantial risk factors of emphysema, asthma, and bronchitis are associated with NO2. This pollutant aggravates existing heart disease in most of the affected people, and it increases premature death (Alonge et al., 2019).
Ozone (O3)	Triggers chest pain, coughing, throat irritation, and congestion. This pollutant worsens bronchitis, emphysema, and asthma (Gopichandran et al., 2016).
Sulfur dioxide (SO2)	This is linked to a substantial risk factor of bronchoconstriction and increased asthma symptoms (Bull et al., 2020).
Particulate matter (PM2.5 & PM10)	They can cause premature death to people who already have heart and lung diseases. They may also aggravate asthma, decrease lung functionality, and increase respiratory symptoms such as coughing and difficulty in breathing (Alonge et al., 2019).
Lead (Pb)	It can accumulate in bones and affect the nervous system, kidney functionality, immune system, reproductive system, developmental system, and cardiovascular system. This pollutant also affects the oxygen capacity of blood (Bull et al., 2020).

In Table 2.1, it is vivid that air pollution has a negative impact on miners' health. Gold miners are highly prone to developing silicosis with attendant TB (latent or active) because of their long-term exposure to silica dust deep underground.

Meanwhile, a study reported in (Tibane and Mamba, 2022) outlined some of the risks associated with underground mining operations and the pollution they cause. The authors looked specifically at coal mining and the specific threat of gangue pollution. Aguirre-Ayerbe et al. (2020) outline the pollution risks that come with mining; they outline the pollution risks that come with mining, the study laid out usable variables suitable for testing environmental impact after mining processes. In that study, the data indicated that the gold mines were more polluting than others due to the multiple processes involved in mining. The finding underscores the critical need to investigate pollution monitoring guidelines that are meant to manage and mitigate air pollution.

2.3.2 Pollution Monitoring and Guidelines

Standards for ambient air quality are crucial in managing air quality. In South Africa, the implementation of the National Ambient Air Quality Standards (NAAQS) had highlighted the transition in the management of air quality from source to receptor (Manisalidis et al., 2020). The criteria for setting limits in the NAAQS are illustrated in Table 2.2 below. Once the pollutants in the air reach these threshold limits (concentration in µg/m³), the air is considered gaseous, polluted and toxic to humans. Notably, high quantities of these pollutants can affect the human respiratory system and increase their susceptibility to respiratory ailments and breathing issues, which may be severe if not treated earlier. The criteria for pollution limits are set by the NAAQS. Pollutants covered in the NAAQS are ozone(O_3), particulate matter (PM), lead (Pb), carbon monoxide (CO), sulfur oxide (SOx), nitrogen oxide (NOx), and lead as illustrated in Table 2.2. High concentrations of these pollutants can affect the respiratory system and increase susceptibility to severe respiratory ailments and breathing difficulties (Delavar et al., 2019).

Table 2.2: NAAQS, source: National Gazette and Notice 2012

Pollutant	Averaging Time	Concentration in µg/m³
SO_2	10 minutes	500
	1 hour	350
	24 hours	125
	1 year	50
NO_2	1 hour	200
	1 year	40
PM_{10}	24 hours	75
	1 year	40
$PM_{2.5}$	24 hours	40
	1 year	20
Ozone	8-hour running average	120
Benzene	1 year	5
Lead	1 year	0.5
CO	1 hour	30 mg/m³
	8-hour running average	10 mg/m³

Proper monitoring and predicting of these pollutants for miners could reduce their impact on well-being. This can be achieved through 4IR technologies such as AI and IoT, integrated with the indigenous knowledge miners use to cope with pollution effects.

2.4. Fourth Industrial Revolution

The Fourth Industrial Revolution (4IR) differs from the previous three revolutions in two significant ways. First, it combines cyber-physical systems, the IoT, and the Internet of Systems. The results include concepts such as smart monitoring and control systems, in which machines are enhanced with web connectivity and connected to a system that can visualize the entire system. The result of this are systems that can 'think' for themselves and make judgments (Ghobakhloo et al., 2021). Second, the technologies that support the 4IR are no longer limited to mechanical tasks (Power et al., 2023) as they are capable of cognitive processing in the same way that people do. Therefore, a variety of possibilities for developing autonomous systems in every sector have become a reality. The 4IR is the ongoing automation of old manufacturing and industrial operations using modern, innovative technology (Shilenge and Telukdarie, 2021). 4IR combines both the IoT and large- scale machine-to-machine (M2M) communication for enhanced automation, self- monitoring, and the production of smart machines that can evaluate and diagnose issues without the need for human intervention (Fanoro et al., 2021).

Across all industries, the 4IR is projected to bring the highest levels of digitisation, automation, virtualization, and decentralisation (Peters and Trunschke, 2021). Importantly, AI, robotics, cloud computing, and the IoT are all necessary components of the 4IR. Most significantly, none of these technologies are considered in isolation by the 4IR, instead, it refers to a fusion in which these high-tech instruments coexist (Shilenge and Telukdarie, 2021).

2.4.1. Internet of Things (IoT)

According to Chen et al., (2022), IoT is often easier to define than to describe. One of the defining components of the IoT is that several devices, smartphones, and sensors have gone online, and these devices have a way to communicate with each other (Chen et al., 2022). This means that besides their core functionality like controlling smart lights at home or regulating temperature in an apartment, these devices also generate substantial amounts of data on their own, which can be made available to other devices and used to expand the usefulness of these sensors. As more devices come online and communicate with each other, there is a vast opportunity to harvest the data from these devices and channel it towards a common and defined goal, like measuring the air quality in each locale (Shinde and Siddiqui, 2018).

According to Atzori et al., (2017), IoT connects physical devices to the Internet, offering new insights and opportunities to enhance operations, expedite automation, and cut costs. As more machines and devices incorporate network capacity, the number of smart things that cloud- based applications can manage, and monitor is increasing. Clark et al., (2016) further support this by stating that, in a nutshell, IoT is the concept of connecting any device (as long as it has an on/off switch) to the Internet and to other connected devices. The IoT is a giant network of connected things and people, all of which collect and share data about the way they are used and about the environment around them. Wireless Sensor Networks (WSN) are an essential part of this giant network and have a crucial role in overseeing and managing different environmental factors.

2.4.2. Wireless Sensor Networks

Wireless sensor networks have the potential to enable a wide range of innovative monitoring and control applications. Examples include target tracking, intrusion detection, animal habitat monitoring, temperature control, and disaster management (Wu et al., 2019). The implementation of wireless sensor networks for monitoring the complex, dynamic, and harsh environment of underground workstations has been explored worldwide to enhance workers safety (Honghui et al., 2017).

Additionally, wireless sensor network (WSN) systems have broad applications in various environmental monitoring frameworks (Chen et al., 2018). To establish a wireless sensor-based network, ZigBee technology, based on the Institute of Electrical and Electronics Engineers (IEEE) 802.15.4 wireless networking standard, is recommended as it is well-suited for operation in harsh environments (Honghui et al., 2017). WSN systems have the potential to collect reliable and accurate information, enabling early warnings and rapid, coordinated responses to potential hazards. This capability is particularly valuable in environmental monitoring, where it can help mitigate natural disasters and save lives (Sun et al., 2019). Most WSN frameworks follow a multi-level structure, as shown below for clarity.

a) **Sensor Deployment**—Wireless sensors are randomly distributed across the study area for data collection.

b) **Base Station Implementation** – A centralized base station is set up to receive and process data.

c) **Network Establishment** – Wireless connections between sensor nodes and the base station are configured.

d) **Data Computation and Analysis** — The aggregated data is processed and transformed into meaningful insights.

e) **Data Storage** – The analyzed data is stored in a database for future reference.

f) **Data Dissemination** – The processed information is distributed to authorized end users for decision-making.

By integrating these levels, WSN systems enhance environmental monitoring efforts, ensuring timely responses to potential hazards while improving data accuracy and accessibility. The most important feature of the WSN is its wireless communication capability, which enables fast networking, high bandwidth, and support for Mobile Ad hoc Network (MANET) (Zhao and Yang, 2018). Unlike traditional networks, WSNs do not require any physical infrastructure (Zhao and Yang, 2018). Once successfully configured, adjacent nodes within the network can automatically connect without the need for any human intervention.

2.4.3. Wireless Connection

The connection between nodes is one of the most crucial aspects of the network. A stable connection is necessary for data to flow from one point to another. For an underground mining environment, it is ideal to utilize a wireless communication standard. Sigfox is a global network operator that builds wireless networks to connect low-power objects, which need to be continuously on and emit small amounts of data (Carandell et al., 2018). The signal generated by Sigfox technology can also be used to effortlessly cover wide areas and penetrate underground objects for enhanced connectivity (Fujdiak et al., 2018).

One of the most attractive characteristics of Sigfox is that it can cover a large area with the least number of base stations and consumes a low amount of power (Gomez et al., 2019).

Figure 2.4 illustrates an example of a SigFox network architecture, the generic components of which usually consist of.

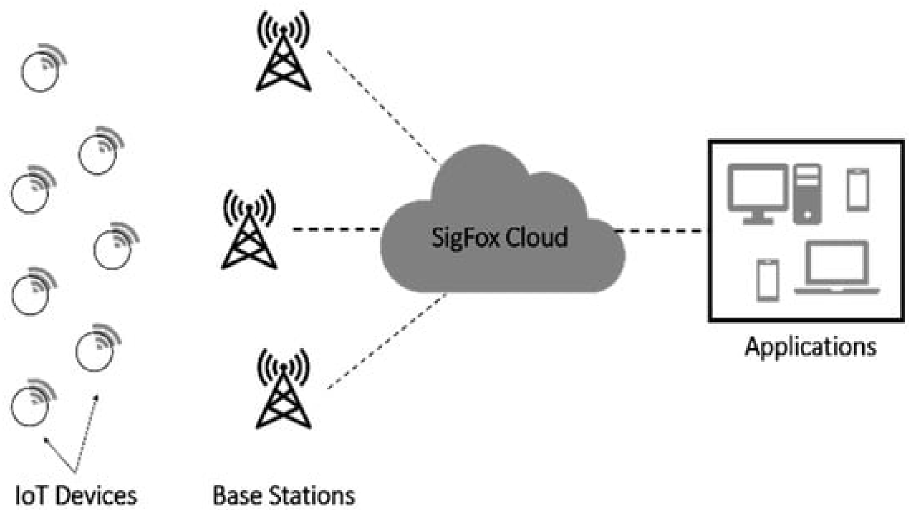


Figure 2.4: Sigfox network architecture, source: (Gomez et al., 2019).

2.4.4. Artificial Intelligence

As Artificial Intelligence (AI) becomes ubiquitous, it is tempting to think of it as a modern construct that only came through after the era of the 4IR. Some of the concepts considered as the building blocks of AI include intelligent data retrieval systems, automatic theorem proving, machine vision, robotics, automatic programming, natural language processing, and intelligent and autonomous decisions based on real- time data.

Poola, (2017), Outlines the diverse ways in which AI has permeated society and made itself indispensable in many facets of life. For example, many smartphones on the market today can anticipate user actions through the use of AI. Mitchell, (2019) highlights advancements made by Google in AI, including the ability to make automated calls and respond to questions in a human-like manner.

Machine Learning is another component of AI. It can be defined as a process whereby machines learn from collected data, and over time, they perceive patterns, trends, and subtle changes. Because the machines are constantly learning, this eventually makes them smart, which means they can be relied upon to give 'intelligent' responses when they are asked questions based on the questions they are being asked.

ML applications in agriculture offer an innovative way to monitor and forecast agricultural inputs and outputs across diverse commodities (Lee et al., 2017). The case for digitization in agriculture is further improved by the fact that there are now widely accessible commercial platforms for developing ML models, as well as several open-source alternatives.

Shilenge and Telukdarie, (2021) asserts that much of the 4IR is built atop ML. This is because artificial intelligence (AI) is embedded in most modern technologies, enabling them to anticipate user needs before input is made. However, for these systems to gain their intelligence, they must first learn (Michalski et al., 2024). This is where ML comes in. While collecting a vast amount of data, either from the applications and sensors built into smartphones or from shopping habits, the system will continually learn some patterns. It is these predictable patterns that make the machines smart and capable of making some predictions based on previous observed behaviour (Thomas et al., 2018), For example, a shopping behavior that shows beer and bread bought together every Saturday morning will trigger the 'system' to send a notification to the shoppers if bread is on a special sale towards the weekend.

The use of ML in controlling air pollution and monitoring systems, including the relative efficacy of ML systems. The study reported in (Ghazali et al., 2012) tested empirically a machine learning air pollution monitoring system with predictive models. Using an array of gaseous and meteorological sensors with wireless transmitting capability, and three machine learning models, the researchers sought to determine the capacity of ML in the prediction of concentrations of sulfur dioxide, nitrogen dioxide, and ground-level ozone. The models also sought to predict 1, 8, 12, and 24 hours ahead of concentration values. The findings, while varying across the various gases being predicted and the time periods for which the predictions were made, indicate that ML is a capable technology for monitoring and predicting air quality in multiple locations.

Meanwhile, ML is a specific subset of AI, based on the idea that the systems can learn from previous computations to produce reliable results, identify patterns, and make decisions with minimal human intervention. Ding et al., (2021) describe ML as the method of data analysis that automates analytical model building. When it comes to network data analysis, ML is regarded as one of the most promising methodological approaches (Praveen Kumar et al., 2019), enabling automatic network self-configuration and fault management. ML is being promoted due to the 4IR's highest levels of automation, digitalization, decentralization, and virtualization (Peters and Trunschke, 2021). Moreover, ML is an effective method for managing, organizing, maintaining, and optimizing networking systems, given the rapid growth of contemporary internet and mobile communication technologies (Yazici et al., 2018).

2.4.5. Machine Learning Techniques for Data Analysis

There are three different types of ML algorithms, namely supervised, unsupervised, and reinforcement learning. Each of these algorithms is employable in every sector. Nonetheless, it is the desired output that determines which machine learning algorithm will produce the best results, based on the size, quality, and diversity of the datasets. Peng et al., (2022) claims that, even the most experienced data scientists cannot tell you which algorithm will perform the best before experimenting with other algorithms. There are three algorithm categories under the umbrella of ML, as indicated in Table 2.3 below.

Table 2.3: Machine learning categories

Supervised machine learning	Supervised ML algorithm works well for pollution prediction because the output is already known. Using labelled datasets from sensors, the readings are compared against the standard threshold to determine whether the air is polluted or not. The desired outcome is a classification of "true" (polluted) or "false" (not polluted**),** which is then further categorized accordingly (Uddin et al., 2019).
Unsupervised Machine Learning	In an unsupervised machine learning algorithm, the training dataset is not labelled, and there is no guide (Osunmakinde, 2013). The machine must analyse the given dataset and identify hidden patterns to make predictions about the output. The result of this algorithm is based on similarities and differences. One commonly used unsupervised learning is K-Means Clustering algorithm.
Reinforcement Machine Learning	Reinforcement learning focuses on structured learning procedures, the learning algorithm then attempts to explore several options and possibilities after creating the rules, monitoring, and analysing each output to decide which is the best. Trial and error are taught to the machine through reinforcement learning (Baraneetharan, 2020). It draws lessons from previous mistakes and starts to modify its strategy in reaction to the circumstance to get the optimal outcome.

This research employed supervised learning algorithms. In supervised learning, algorithms are trained on labelled datasets to learn and return the desired labelled data explicitly. Uddin et al. (2019) mention that: supervised algorithm directs the learning method to produce the desired results, and the most commonly used supervised ML algorithms for pollution monitoring are the Naïve Bayes Classifier Algorithm and Decision Trees.

Table 2.4 illustrates the three supervised ML techniques.

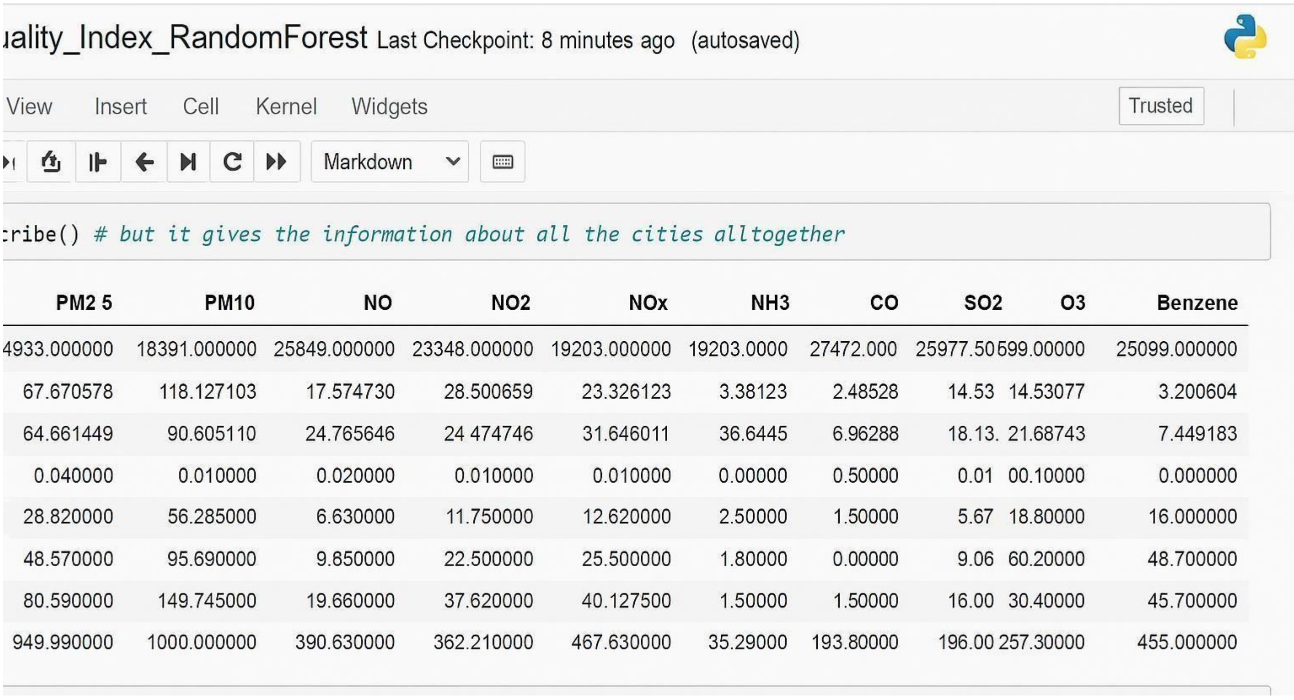
Table 2.4: Supervised machine learning technique

Classification	The ML program must draw a decision from the input datasets, determine the output, and classify new observations (Osunmakinde, 2013). For example, filtering to determine if an email is spam or not spam, based on the existing experimental data or enforced rules.

Regression	The learning algorithm must estimate and understand the relationship between the datasets. Regression analysis focuses on one dependent variable (Jayasundara et al., 2025) and a series of other changing variables. Regression is mainly used for predictions and forecasting.
Forecasting	This is frequently used to analyze patterns and make future predictions based on previous and present data.

Most IoT-related monitoring systems utilize supervised machine learning to validate and preprocess data on programming platforms such as JupyterHub and GitHub, depending on the programmer's preferences or project requirements. These ML platforms utilize the Python programming language to process the collected datasets/pollutants.

Figure 2.5 illustrates JupyterHub, where the Random Forest ML algorithm was employed to train the datasets and calculate the minimum, maximum, and average values of pollutants.



PM2 5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3	Benzene
4933.000000	18391.000000	25849.000000	23348.000000	19203.000000	19203.0000	27472.000	25977.50	599.00000	25099.000000
67.670578	118.127103	17.574730	28.500659	23.326123	3.38123	2.48528	14.53	14.53077	3.200604
64.661449	90.605110	24.765646	24 474746	31.646011	36.6445	6.96288	18.13.	21.68743	7.449183
0.040000	0.010000	0.020000	0.010000	0.010000	0.00000	0.50000	0.01	00.10000	0.000000
28.820000	56.285000	6.630000	11.750000	12.620000	2.50000	1.50000	5.67	18.80000	16.000000
48.570000	95.690000	9.850000	22.500000	25.500000	1.80000	0.00000	9.06	60.20000	48.700000
80.590000	149.745000	19.660000	37.620000	40.127500	1.50000	1.50000	16.00	30.40000	45.700000
949.990000	1000.000000	390.630000	362.210000	467.630000	35.29000	193.80000	196.00	257.30000	455.000000

Figure 2.5: Random Forest machine learning algorithm training datasets, source: (Paksi, 2022)

2.5. Integrating Indigenous Knowledge in monitoring systems

Carm (2014) defines Indigenous Knowledge (IK) as local knowledge that is unique to a culture or society and a general understanding of the surroundings. It is the basis for local-level decision-making in areas such as healthcare, agriculture, education, natural resource management, and a host of other activities in the community (Paula et al., 2019). Combining IK with other kinds of knowledge or information results in hybrid or integrated knowledge (Mbele et al., 2016). Furthermore, it is crucial to integrate various knowledge to promote a sustainable environment and create an adaptive system that is both acceptable and well understood by local communities' Indigenous knowledge systems.

Dahl and Hansen (2019) document some of the Indigenous Knowledge Systems (IKS) that have been employed in mining by different communities over the years. Considering that the communities in the pre-industrial world lacked many sophisticated technologies that have become ubiquitous in contemporary times, a series of unique knowledge systems were passed down from generation to generation. This view is also supported by Barnhardt et al., (2005), who noted that while the information was not specifically about mining, certain natural occurrences served as triggers and cues for various activities. Observations such as the position of the moon, subtle changes in tides, soil composition of a specific location, and the occurrence of certain vegetation acted as key indicators of the habitat on which communities relied. Collectively, these cues could serve as prospecting data, enabling early communities to detect the presence of minerals underground without the need for sophisticated machinery or technology to test samples. Additionally, the absence of mercury, which has become indispensable in contemporary mining, ensured that traditional mining methods were environmentally sustainable.

2.5.1. IK Analysis using Fuzzy Cognitive Maps and Mental Modeler

According to Singh and U-Gauhati (2017), local knowledge is defined as the knowledge that members of a particular community have acquired over time and continue to acquire. It is based on experience, frequently tested through centuries of use, tailored to the local environment and culture, and embedded in relationships, institutions, and practices within the community. Local knowledge is also dynamic and always evolving. Studies involving people are not always linear, therefore, several deciding factors can take priority over each other, such as identifying whose knowledge to represent, how the data will be collected, and the best process for what they are going to represent (Lusilao-Makiese et al., 2013). Due to the nature of the local Knowledge, it requires a sophisticated analysis. Fuzzy Cognitive Maps (FCM) and Mental Modeler were designed to model and analyze such types of data (Gray et al., 2013).

2.5.2. Fuzzy Cognitive Maps and their Origin

FCMs facilitate a holistic view of the mind (Gray et al., 2013). While traditional approaches might examine cognitive modules in isolation, FCMs emphasize the interconnectedness of these modules (Lusilao- Makiese et al., 2013). This holistic perspective can provide deeper insights into how cognitive functions work together to produce complex behaviours (Gray et al., 2012). FCMs are a powerful computational tool that can be used to model complex systems and processes. They combine aspects of neural networks and fuzzy logic, making them particularly suitable for tackling problems where uncertainty and imprecision are significant.

One of the intriguing applications of FCMs is in understanding and mapping mental modularity, which is the concept that the human mind is composed of distinct and semi-independent modules, and that each specializes in different cognitive functions (Gray et al., 2012). These modules are designed to process specific types of information, including language, visual perception, and social interactions. The modular view of the mind offers a compelling way to explain the efficiency and specialization of human cognition (Kadaifci et al., 2024).

Practically, the use of FCMs can extend to various applications, ranging from and not limited to AI and mental health (Sarmiento et al., 2024). In AI, for instance, creating systems that mimic human cognition could benefit from FCM models that incorporate human-like modularity and adaptability (Cleveland et al., 2024). In mental health, FCMs could be used to simulate the impacts of therapeutic interventions on different cognitive functions, offering personalized treatment strategies (Cleveland et al., 2024).

One of the primary strengths of using FCMs to model mental modularity is their adaptability. FCMs can be updated as new data becomes available (Barbrook-Johnson and Penn, 2022) making them a dynamic tool for exploring cognitive processes. This adaptability is critical in cognitive science, where our understanding is constantly evolving. By continually refining the weights and nodes based on new insights, researchers can create ever more accurate models of mental functioning (Solana-Gutiérrez et al., 2017). For example, consider the interaction between the language processing module and the visual perception module. In a simple FCM representation, these modules would be nodes or components connected by edges, with weights indicating the strength and nature of their interaction (Henly-Shepard et al., 2015). A positive weight might indicate that enhanced language processing improves visual perception, whereas a negative weight could indicate interference. The fuzzy nature of these connections allows for gradual degrees of influence, rather than strict binary relationships, better reflecting the complexities of real-world mental processes (Henly-Shepard et al., 2015).

2.5.3. Related Systems and Application of FCM

FCMs are an influential modeling and decision-support tool used in complex systems analysis (Henly-Shepard et al., 2015). They integrate both the aspects of cognitive maps and fuzzy logic to represent and analyze the behavior of complex systems. FCMs have been employed in a wide range of disciplines, including, but not limited to, environmental sciences, engineering, and AI. FCMs are a versatile tool used in modeling and decision-making in complex systems where uncertainty, ambiguity, and dynamic interactions are key elements. By integrating aspects of cognitive mapping and fuzzy logic, Gray et al. (2012) affirm that FCMs enable the representation and analysis of relationships between interconnected variables.

One of the most prominent areas where FCMs are applied is in environmental management, particularly in modeling ecosystems and predicting the impacts of various environmental interventions (Solana- Gutiérrez et al., 2017). FCM was developed and used in coastal ecosystem management to model the impact of different management strategies on a coastal ecosystem. By incorporating expert knowledge from ecologists and local stakeholders in this study, the FCM simulated the effects of various interventions, such as reducing pollution or limiting fishing on the health of the ecosystem (Nor et al., 2022). Another application is in water resource management, where FCMs have been employed to model the interactions between different water users, environmental conditions, and government policies (Rooney et al., 2023). FCM was used to assess the potential impacts of climate change on water resources in the Nile River basin (Sarmiento et al., 2024), The model incorporated variables, such as rainfall, temperature, irrigation demand, and water governance policies (Hou et al., 2021).

FCMs have found significant applications in the healthcare sector, particularly in decision support systems. The inherent uncertainty and complexity of medical diagnosis and treatment planning make FCMs well-suited for modeling patient treatment strategies and outcomes (Kadaifci et al., 2024). In medical diagnosis, FCMs have been used to model the relationships between the symptoms identified in a patient's presentation, the underlying condition, and treatment outcomes (Obiero et al., 2023). To diagnose Alzheimer’s disease, the model incorporated factors such as cognitive decline, genetic predisposition, brain imaging results, and patient history (Salinas Salmeron et al., 2019). Moreover, the interaction of factors provided physicians with tools to assess Alzheimer's disease as well as explore treatment options.

Sarmiento et al. (2024) said that FCMs are used to model cancer treatment strategies, particularly in cases where multiple treatment options such as chemotherapy, radiation or surgery need to be considered. Sarmiento et al. (2024) modelled a decision support system for breast cancer treatment to simulate how various treatment combinations might influence patient outcomes. The system enabled physicians to test different treatment plans, considering factors such as tumor type, patient age, and comorbidities, and make more informed decisions about personalized treatment strategies.

Yun et al. (2022) modelled public health policies, and FCMs were employed to simulate the spread of infectious diseases and the effects of various interventions. Recently, Cleveland et al., (2024) developed FCMs to simulate the spread of COVID-19 and evaluate the effectiveness of various policy measures, such as lockdowns, mask mandates, and vaccination. It incorporated variables such as infection rates, healthcare capacity, economic impacts, and public compliance to help decision makers identify strategies that balance economic sustainability and public health.

FCMs have been applied in energy management, particularly in the control and optimization of energy grids. Kumar and Pande (2023) developed an FCM system to optimize the operation of a smart grid where renewable energy sources, storage systems, and energy consumption interact dynamically. The system simulated how several factors, such as energy demand, weather conditions, and storage capacity, influenced grid stability, allowing for real-time optimization of energy flow and ensuring efficient energy use while maintaining grid reliability.

Gu et al. (2023) propose an innovative spatiotemporal hybrid early warning system based on adaptive feature extraction and improved FCMs. A spatial spill-over analysis model, based on the Moran index and local gravitational clustering, was proposed to capture the diffusion and concentration characteristics of air pollution between regions. A hesitant fuzzy information-optimized fuzzy cognitive maps model was proposed to forecast the air quality of urban agglomerations. The experimental results demonstrate that the air quality forecasting accuracy of urban agglomerations can be significantly enhanced when geographical conditions and interactions among cities are comprehensively considered (Gu et al., 2023).

Furthermore, a fuzzy-based analysis of air particle pollution data is presented, along with a fuzzy-based analysis of magnetic biomonitoring data. The fuzzy inference system membership functions were built by standardizing the data to make them independent of the values. The magnetic index of contamination was used as a driver component for summarising this information about the sample's fundamental magnetic properties in only one fuzzy value. The methodology developed for building this magnetic index of contamination was also utilised for magnetic monitoring data from other environmental matrices, such as soils and sediments (Chaparro et al., 2024).

In the field of AI and ML, FCMs have been applied to develop models that can learn and adapt based on new data. FCMs are particularly useful in systems where knowledge representation and reasoning under uncertainty are critical. Nápoles et al. (2018) developed a hybrid FCM- neural network model capable of both learning from data and reasoning with expert knowledge. An FCM-neural network hybrid was used for fault diagnosis in electrical systems Ibrahim and Abdulazeez, (2021) and the model learned from historical data about system failures while also incorporating expert knowledge about potential causes of faults.
Table 2.5 represents a summary of other practical work that has been carried out using the FCM mental modeler.

Table 2.5: FCM systems

RESEACH AREA	**RESEACH PURPOSE**	**DATA COLLECTION METHOD**	**FINDINGS/RESULTS**	**SOURCE**
Social ecological decision making.	Modelling the integration of stakeholder knowledge in social–ecological decision-making, Benefits, and limitations to knowledge diversity.	Collected Fuzzy-Logic Cognitive Maps from several stakeholder groups such as managers, scientists, harvesters, pre- and post-harvest sectors, and environmental Nonprofit organizations.	The study suggests that, incorporating stakeholder information into natural resource governance will increase the flexibility of social-ecological systems by reducing rigidity, representing different viewpoints, and encouraging adaptability in decision-making.	Gray *et al.*, (2012).
Mental Models in Environmental decision making.	The framework investigated distinctive use and the role of mental models in decision making, the role of modelling in adaptive management and fuzzy logic cognitive mapping.	Mental Modeler architecture modeling software.	Improved stakeholder centered and participatory modelling software called “Mental Modeler”. The software facilitates user centered model construction, promotes learning in disparate stakeholder communities through knowledge sharing and allows flexibility for users to refine and test their models intended to facilitate adaptive management planning.	(Gray et al., 2013).
Adaptive environmental	A study was carried out to	IK was collected using a focus	Based on the indicators that were formed, the	

management system: participatory approach through Fuzzy cognitive maps.	verify how the Lejweleputswa and its surrounding communities have been adopting the use of IK, to protect themselves against environmental conditions brought on them by mining operations.	group setting at Nyakallong.	study confirmed how water indicators of the analysis affected one another, for example acid-mine drainage increases acidity in domestic water and decreases wildlife and human health.	(Mbele et al., 2016b)
Local-scale dynamics in conservation biology.	To assess how local community members such as hunters, sellers, and consumers perceive the dynamics of the bush meat trade in relation to the presumptions that underpin externally developed policies for managing the bush meat trade.	Fuzzy-logic cognitive mapping and formation of of participatory modeling. Nine workshops in four Tanzanian villages that border the Serengeti National Park were used to gather data.	The three logic models that were produced from the various bush meat management strategies, revealed structural parallels, but also some minor variations in their methods, connections, and presumptions regarding the dynamics of the bush meat trade.	(Nyaki et al., 2014).
Social learning and facilitated community disaster planning.	To foster adaptation to environmental changes, build social trust and empower diverse stakeholders, by offering opportunities for groups of individuals to challenge, negotiate and propose new norms, policies, or programs.	Regional Focus groups and the use of Fuzzy cognitive maps to collect ideas.	A three-phase social learning framework was present, to facilitate stakeholder driven scenario-based modeling, to inform community disaster planning in relation to the potential impacts of a tsunami.	(Henly-Shepard et al., 2015).

In the process of IK analysis, several methods can be used to train fuzzy cognitive maps, depending on the available data and expert knowledge. Manual, expert-driven techniques are suitable for qualitative systems, as they rely on human knowledge and skills to refine and improve the model. On the other hand, data-driven approaches automatically modify the FCM models weights according to historical data, which qualifies them for quantitative modelling. To capitalise on each strategy's advantages, hybrid tactics that blend the two are also frequently employed Henly-Shepard et al., (2015). The requirements of the system that is being modelled, and the availability of data, therefore, determine which FCM training method to use.

2.6. Related Systems using Artificial Intelligence and Internet of Things.

Environmental monitoring and appropriate correspondence have been a vital task to guarantee safe working conditions and increase productivity in underground coal mineshafts (Singh et al., 2018). Every worker in the industry, particularly in the mining industry, faces harsh conditions (Henriques et al., 2017). This is a result of the generation of a large amount of humidity, heat, poisonous and explosive gases (Wu et al., 2016).
Mining operations are inherently associated with the generation of various poisonous and inflammable gases, such as hydrogen sulphide, sulphur oxide, carbon dioxide, nitrogen oxide, carbon monoxide, and methane (Muduli et al., 2018a). There is a need to create a dynamic correspondence and data system to identify underground conditions and precisely provide area references rapidly. Utilisation of ad hoc networking topology with the wireless sensor network (WSN) to establish connectivity and communication between multiple nodes, extended across an area of interest in a mesh topology structure network, to flexibly extract those poisonous gases (Honghui et al., 2017).

The most commonly used communication and monitoring systems in underground mines are cable-based (Chen et al., 2018). The cable may be inconvenient due to factors such as damage to communication cables, a high fault rate, and the inability to provide the most relevant results when the system operates offline (Prasad and Pillai, 2019). Osunmakinde (2013) further urges that one of the significant drawbacks of wired networks is the total disconnection that occurs when cables are damaged, particularly during a disaster. To alleviate this problem, a wireless sensor-based network would be the most applicable in monitoring underground mines.

2.6.1. Examples of Existing Artificial Intelligence and IoT-Based Systems

South African Air Quality Information System

The South African Air Quality Information System (SAAQIS) is a web-based interactive air quality information system, which aims to deliver the state of the air quality data to South African Citizens (Priority et al., 2017). It is also a research portal for solidifying policy development related to air quality issues in the country. Ambient air quality and meteorological parameters monitoring stations are installed across the country, specifically in provinces that are identified as pollution hotspots. Each station is equipped with wireless sensor nodes that collect PM_{10}, $PM_{2.5}$, SO_2, NO, NO_2, NO_x, O_3, CO, and benzene, as well as meteorological data for wind speed, ambient temperature, relative humidity, wind direction and speed, rainfall, solar radiation, and barometric pressure (Gwaze and Mashele, 2018). The system is regarded as a smart and interoperable system. There is an established wireless communication among all the monitoring stations across all the stations. Monitoring stations can transfer data to each other and to the central monitoring station. Furthermore, data is retrieved from the server to be disseminated through a mobile application and a web portal that is available to the public. Table 2.6 displays air quality information from SAAQIS.

Table 2.6: Air quality information from SAAQS

When the AQI is...	At-risk individuals such as children, the elderly, and persons with health concerns should...	The general population should...
Low (Green)	Enjoy your usual outdoor activities.	Enjoy your usual outdoor activities.
Moderate (Yellow)	Adults and children with lung problems, and adults with heart problems, who experience symptoms, should consider reducing strenuous physical activity, particularly outdoors.	Enjoy your usual outdoor activities.
High (Orange)	Adults and children with lung problems, as well as adults with heart problems, should reduce strenuous physical exertion, particularly outdoors, and especially if they experience symptoms. People with asthma may find they need to use their reliever inhaler more often. Older people should also reduce physical exertion	Anyone experiencing discomfort such as sore eyes, cough or sore throat should consider reducing activity, particularly outdoors

Very High (Red)	Adults and children with lung problems, adults with heart problems, and older people, should avoid strenuous physical activity. People with asthma may find they need to use their reliever inhaler more often	Reduce physical exertion, particularly outdoors, especially if you experience symptoms such as cough or sore throat
Hazardous (Purple)	Adults and children with lung problems, adults with heart problems, and older people, should avoid strenuous physical activity. People with asthma may find they need to use their reliever inhaler more often	Reduce physical exertion, particularly outdoors, especially if you experience symptoms such as cough or sore throat

Mobile Tracker System

This is an Australian system, where wireless sensor nodes are mounted on the mine personnel when they go underground for block caving. The system seeks to monitor the underground environmental conditions (Dawn, 2019). Among the special attributes of the system is the robust Cave Tracker System, which monitors real-time cave flow and detects any rockfall, and a networked smart marker that cannot be affected by ground movement and is equipped with a global positioning system (GPS) for precise location reference. The system aims to ensure the safety of mine workers while they work underground and in underground infrastructure, while maintaining the highest level of ore body production. In the event of any casualties, the system is configured to send alerts to the base station, which will then be viewed by the systems administrator, and necessary actions will be taken based on the received alert (Dawn, 2019). The base station monitors the underground sensor nodes to ensure data integrity and accuracy. The wireless sensor nodes are battery-powered. The only disadvantage of the system is the long time to restore connection from the sensor node to the base station when there is any form of interruption, like internet disconnection.

Mobile Gas Sensing Robot

In Malaysia, a microcontroller-based, real-time mobile gas-sensing robot was developed by (Ghazali et al., 2012) to automatically detect a liquefied petroleum gas leakage in the gas transporting pipeline. In addition, the microcontroller-based mobile gas sensing robot is mounted on a track that is constantly moving back and forth along the gas transporting pipeline to detect any gas leakage (Singh et al., 2018). The microcontroller-based mobile gas sensing robot system features a buzzer that vibrates and emits a sound upon detecting gas leakage, alerting people to a potentially hazardous situation. The Global Positioning System (GPS) has capabilities to track down the exact location of the incident, and a liquid crystal display (LCD) screen displays both the readings of the leaked gas and the precise location of where the robot is currently situated (Singh et al., 2018).

Environmental Air pollution and Water quality monitoring system using IoT

This project was conducted in China, where the researcher proposed an air quality and water quality monitoring system. The system monitors and detects air and water quality in real-time using IoT devices. Wireless sensor networks were used to detect the presence of harmful gases in the air and continuously transmit the collected data to the microcontroller. Also, the system was configured to measure water parameters, such as turbidity, temperature, PH value, and flow, and report it to the cloud-based server through the interconnected IoT devices (Li et al., 2020). The cloud-based server enabled researchers and system users to monitor air pollution in various areas and take action against it. Moreover, water pollutant detection was used to monitor the water quality in or near hospitals, schools, and neighbourhood areas. Standard threshold limits for both datasets (water and air) were set to be constantly compared against the new readings, and notifications were sent out to the relevant stakeholders and system users (Li et al., 2020). The notification aimed to alert the people so that they could take control of the situation by taking all necessary precautionary measures to avoid being affected by contaminated air or water. Monitoring was performed using an Arduino, an open-source hardware platform. The Arduino boards were loaded with the programme code via a serial connection to and from a computer.

Recognition of human activities using multimodal Sensing devices and deep learning approach

In this research, a system that uses a convolutional neural network and a deep machine learning method to recognise multiple human activities was introduced. The system was also integrated with recurrent bidirectional networks capable of extracting features and using temporal dependencies (Ihianle et al., 2020). The system demonstrated that the deep learning model can be effectively applied (Ihianle et al., 2020) to different sensor modalities to recognize multiple human activities, irrespective of the variability in body movements. The model was validated using publicly available datasets obtained for testing purposes. Based on the published results, the deep learning models used outperformed other commonly used deep learning models for human movement detection.

Arduino-based Distributed System as Internet of Things for low- cost, Air Pollution Wireless Monitoring in Real Time

The system was developed using wireless sensor networks and integrated with the Global System for Mobile Communication. The system delivered high-resolution atmospheric and air pollution data in real time, including temperature, wind speed and wind direction, rainfall, humidity, and concentration of other toxic gases (Fuertes et al., 2016). The primary contribution of this study to the body of knowledge is the development of a comprehensive system that integrates software, hardware, and firmware for real-time air quality monitoring and wireless data distribution. The primary objective of this work was to establish a cost-effective solution capable of determining the concentrations of various greenhouse gases and air pollution (Fuertes et al., 2016). This system enables direct observation and evaluation of critical environmental processes, including global warming and climate change.

From a software engineering perspective, this project integrated two agile methods of software development (Fuertes et al., 2016), namely Extreme Programming and Scrum, consecutively. They were deployed for modelling infrastructure, and the development of a software component suitable for transferring real-time data from the hardware (Arduino) to the software (Java enterprise edition) by means of an application programming Interface written in C++ language. During the development of this project, Arduino was used as an open-source hardware platform for detecting poisonous gases in the air and Java libraries using a MySQL database as a back-end. Tzeng and Wey (2011) agrees that the ZigBee standard communicates with the microprocessor of the Arduino board through the serial port. This is a Wi-Fi module built by Roving Network that implements the IEEE 802.11 standard to establish communication with wireless networks. The system was specially designed in a way that simplifies the integration process and reduces development time.

Table 2.7 provides a summary of other existing smart multi-pollutant monitoring systems utilising IoT.

Table 2.7:Summary of smart multi-pollutant monitoring systems using IoT

Research Area	Research purpose	Research Findings and Challenges	Devices used	Sources

IOT-Based Smart Monitoring	Farming soil monitoring	Efficient vegetable crop monitoring. Greenhouse gases affect the health and growth of vegetables like tomatoes	Wireless sensor nodes	Shinde and Siddiqui, (2018)
Environmental Monitoring System using IoT	Monitor the amount of dust and humidity. on air	Scalable and high-density air quality monitoring with the interconnection of heterogeneous sensors. computational complexity due to huge data captured and processed	Mobile sensor Network and WSN and IOT	Zrelli and Ezzedine, (2018)
Environmental pollution monitoring using IoT	Air pollution monitoring System	Real-time data monitoring. Results were not 100% accurate	MQ3 Sensors Model, Raspberry Pi and IOT	Balasubramaniyan and Manivannan, (2016)
Aqua farming and energy conservation	Aqua farming monitoring system	Water quality and quantity control. Higher carbon emissions and energy requirements	Temperature sensor and Odour, pH, conductance	Tranca et al., (2017)
Multi-agent supervising system	e-health monitoring system	Detection of emergencies	Supervising machine learning algorithms and AI	Zhou and Li, (2021)
ZigBee based environment monitoring for smart industrial mining	Smart industry environment monitoring system	Identification of hazardous effects in mining industries	ZigBee and Wireless Sensor Networks	Ugya AY et al., (2018)
The use of LoRa technologies in climate Monitoring	Climate and ecology Change monitoring	Identification of emissions that contaminate the climate	LoRa technologies Sensor network	Duisebekova et al., (2019)

2.7. Conclusion

This chapter has reviewed the existing body of knowledge on air pollution monitoring and prediction, combining insights from a bibliometric analysis with a thematic exploration of the literature. The bibliometric analysis highlighted the global rise in research on air quality monitoring using AI, IoT, and machine learning over the past five years, while also showing that contributions from Africa remain limited. These findings underscore the need for locally relevant studies that integrate advanced technologies with context-specific approaches such as Indigenous Knowledge (IK).

The literature review further established that pollution, particularly from underground mining, poses serious health and environmental risks in South Africa. Mine workers and surrounding communities remain especially vulnerable to particulate matter, sulphur dioxide, nitrogen dioxide, and related pollutants, which are strongly linked to silicosis, tuberculosis, and respiratory conditions. These risks underscore the need for developing adaptive monitoring systems that surpass conventional methods.

The review also demonstrated the potential of Fourth Industrial Revolution (4IR) technologies, including IoT-enabled sensors, machine learning algorithms, and big data analytics, to transform how air quality is monitored and predicted. IoT systems can provide continuous, real-time measurements of multiple pollutants, while machine learning enhances forecasting accuracy by identifying hidden patterns and trends in historical datasets. However, challenges such as the high cost of infrastructure, limitations in sensor reliability, and gaps in technical skills remain barriers to their widespread deployment in South Africa.

Integrating 4IR technologies with Indigenous Knowledge emerges as a promising strategy to bridge these gaps. IK provides locally relevant indicators of environmental change, improves community engagement, and enhances the interpretability of scientific predictions. A hybrid approach can therefore deliver systems that are not only scientifically rigorous but also socially inclusive and context-sensitive.

Overall, this chapter demonstrates both the opportunities and the limitations in current research and practice. It identifies apparent gaps in the South African context, particularly the lack of frameworks that combine advanced technology with IK for underground mining environments. Addressing these gaps formed the basis for this study’s contribution: the design and evaluation of an innovative, adaptive air pollution monitoring system tailored to the Free State Province. The next chapter outlines the methodology used to develop, implement, and evaluate this system.

CHAPTER THREE: RESEARCH METHODOLOGY

3.1. Introduction

This chapter sets out the methodological foundation of the study. It explains how the research was designed, conducted, and validated to address the research aim of developing a smart and adaptive air pollution monitoring system for the Free State Province. The chapter first outlines the research philosophy that underpins the investigation, followed by a description of the research design and approaches adopted. It then details the specific methods applied for data collection and analysis, including both quantitative and qualitative techniques.

To ensure clarity and alignment with best practices, the chapter has been structured into distinct sections:

a) **Research Philosophy**: situating the study within a pragmatic stance that values both scientific rigor and community relevance.

b) **Research Design**: justification of the mixed-methods, multi-phase strategy adopted.

c) **Methods**: procedures for collecting and analysing datasets, Indigenous Knowledge (IK), and machine learning outputs.

d) **Data Collection and Analysis**: step-by-step account of how sensor, ground station, and questionnaire data were obtained, processed, and validated.

e) **Framework Development**: presentation of the conceptual and practical framework that integrates IoT, machine learning, and IK.

f) **Ethical Considerations**: description of how ethical integrity was ensured throughout the study.

3.2. Research Philosophy

This study was guided by a **pragmatic research philosophy**, underpinned by elements of **constructivism**. Pragmatism was selected because it emphasises practical solutions to real-world problems, rather than adhering rigidly to a single methodological tradition (Creswell and Poth, 2018). In the context of air pollution monitoring in mining communities of the Free State, this orientation was essential: the study had to deliver not only theoretical insights but also a prototype that could be understood and applied by semi-literate or illiterate mine workers and surrounding communities.

The constructivist orientation complemented this pragmatic stance by recognising that knowledge is socially constructed. In particular, the Indigenous Knowledge (IK) collected from community members and mine workers was not treated as anecdotal but as an equally valid way of knowing. Constructivism, therefore, provided a foundation for integrating lived experiences and community indicators of pollution into the design of the monitoring framework, ensuring the system was responsive to both scientific data and local realities.

The philosophical position justified the use of **mixed methods,** bringing together quantitative and qualitative approaches. Quantitative methods were necessary for collecting, processing, and modelling datasets from monitoring stations and sensors, as well as for training machine learning algorithms. At the same time, qualitative approaches allowed the study to capture the lived experiences and IK of local communities, giving depth and cultural grounding to the technical framework. Mixed methods thus provided a holistic approach to address the research questions, ensuring that both scientific rigour and community perspectives were fully incorporated.

3.3. Research Design

Research design provides the overarching strategy that guides the collection, analysis, and interpretation of data to address the research questions (Donnette et al., 2016). This study combined **implementation research design** and **descriptive research design** within a multi-method framework, thereby ensuring both scientific rigor and responsiveness to lived realities.

3.3.1. Implementation Research Design

The implementation research design was selected due to its problem-solving and practice-oriented focus. It employs a multi-method inquiry that integrates quantitative and qualitative approaches to evaluate how interventions perform in real-world settings (Alonge et al., 2019). This was particularly relevant for a study aimed at developing an adaptive air pollution monitoring system, where scientific models had to be tested alongside Indigenous Knowledge indicators and community feedback.

The design is also participatory in nature, encouraging collaboration between researchers, communities, mine workers, and decision-makers. Such inclusiveness ensured that the system was not only technically sound but also contextually appropriate and socially acceptable. The multi-phase character of the design enabled iterative learning, where insights gained in early phases (e.g., exploratory engagement and pilot questionnaires) were incorporated into later phases (e.g., system evaluation and framework refinement).

3.3.2. Descriptive Research Design

A descriptive research design was employed to capture and document the lived experiences of communities and mine workers who face daily exposure to underground mining-related air pollution. It focused on systematically gathering information through questionnaires, observations, and case inputs without manipulating variables. This approach was well-suited to understanding the subjective experiences of participants and integrating IK as a legitimate source of knowledge. By applying descriptive design, the study was able to record nuanced patterns of how communities perceive and respond to pollution, thereby creating a richer foundation for validating machine learning (ML) outputs against lived experiences.

3.3.3. Justification for Combined Designs

The combination of implementation and descriptive designs provided a balance between **action-oriented problem-solving** and **experience-based understanding**. Implementation design enabled the development and testing of a functional system that integrated IoT, ML, and IK, while descriptive design anchored the study in the realities of the affected communities. Together, they ensured that the research was scientifically rigorous, contextually grounded, and responsive to both technical and social dimensions of air pollution monitoring.

3.4. Research Methods

The research methods define the specific tools and processes applied in order to implement the research design and achieve the study objectives (Smith et al., 2020). In this dissertation, a **mixed methods approach** was adopted to capture both quantitative and qualitative dimensions of air pollution monitoring. This integration allowed the research to leverage the strengths of statistical modelling and machine learning while incorporating the lived experiences and Indigenous Knowledge of local communities.

3.4.1. Mixed Methods Approach

Mixed methods research was chosen because of its ability to bridge objective measurement with subjective interpretation (Brevik, 2017). Quantitative methods (machine learning models, statistical correlations, and compliance checks with National Ambient Air Quality Standards) provided measurable outputs on pollution levels and trends. Qualitative methods (IK collection, surveys, and participatory inputs) captured community experiences and locally relevant indicators of air pollution. The combination enabled triangulation, enhancing validity by ensuring that results from one method could be cross-verified through the other.

This approach was particularly important in this study, as the end-users of the proposed monitoring system were largely semi-literate to illiterate mine workers and community members. Reliance on technology alone risked excluding them, while reliance on IK alone risked a lack of scientific rigour. Mixed methods, therefore, ensured balance, inclusivity, and robustness.

3.4.2. Surveys and Questionnaires for IK Collection

To gather IK, structured questionnaires were designed and administered to selected participants, including mine workers, local residents, and other stakeholders knowledgeable about environmental conditions. The questionnaires included both closed-ended questions (to provide quantifiable indicators) and open-ended questions (to allow participants to describe their lived experiences).

This method was appropriate because it systematically captured locally relevant pollution indicators such as visible dust, smells, seasonal patterns, and health symptoms. The responses were later translated into a structured knowledge base that could be modelled using Fuzzy Cognitive Maps (FCMs).

3.4.3. Correlation Analysis of IK and Scientific Data

Correlation analysis was used to examine the relationship between IK-based indicators and ground-based measurements of pollution. Data from the South African Weather Service (SAWS) Pelonomi air monitoring station provided authoritative scientific readings, which were compared with the qualitative insights collected through community surveys.

This step was necessary to establish whether IK signals, such as dust levels or perceived respiratory impacts, aligned with recorded scientific pollutant concentrations (e.g., $PM_{2.5}$, PM_{10}, SO_2). Demonstrating convergence between the two enhanced the credibility of IK within the monitoring framework and justified its integration into the prototype system.

3.4.4. Supervised Machine Learning Algorithms

Supervised machine learning (ML) algorithms were employed to train predictive models using labelled historical air quality datasets. These datasets, drawn from the Pelonomi monitoring station, contained pollutant readings alongside temporal and meteorological variables. The algorithms selected for experimentation included **Support Vector Machine (SVM), Decision Tree Regression, Random Forest Regression, and Gradient Boosting Regression**.

The rationale for selecting these algorithms was their proven effectiveness in handling non-linear relationships, categorical and numerical inputs, and time-series forecasting of environmental parameters. By training the models on past datasets, the system was designed to forecast air quality levels up to four days in advance, in line with the thresholds established by the National Ambient Air Quality Standards (NAAQS).

The use of ML ensured that predictions were data-driven and adaptive, while the integration of IK indicators further contextualised these outputs. Together, these methods provided the technical backbone of the smart and adaptive monitoring framework.

3.5. Data Collection

Data collection for this study combined qualitative and quantitative approaches in order to establish a comprehensive dataset that supported the development and testing of the smart and adaptive air pollution monitoring system. The process was structured around three key streams: Indigenous Knowledge (IK) collection, scientific datasets from ground monitoring stations and installed wireless sensors, and a consolidated summary of all data sources.

3.5.1. Indigenous Knowledge Collection

The collection of Indigenous Knowledge (IK) was carried out through **structured questionnaires** administered to selected participants within the Lejweleputswa district of the Free State Province. The participants included mine workers, local community members, and IK experts with long-standing experience of living in mining-affected environments.

- **Sampling**: A probability sampling strategy was applied to ensure representativeness, while balancing the study's time and logistical constraints. Participants were selected on the basis of proximity to mining activities, lived experience of environmental impacts, and willingness to engage.
- **Questionnaire Design**: Two phases of questionnaires were developed. The first phase (July–September 2022) focused on problem exploration, documenting lived experiences and identifying local indicators of air pollution (e.g., visible dust, seasonal odours, respiratory symptoms). The second phase (June–October 2024) concentrated on system evaluation, seeking community feedback on the usability and effectiveness of the developed monitoring prototype.
- **Administration**: Questionnaires were administered digitally using the Survicate survey platform. Survicate is a web-based, Interactive, and integrated questionnaire tool (Albuainain and Choe, 2023). Survicate survey allowed remote distribution and automatic aggregation of responses. The mix of closed and open-ended questions facilitated both quantifiable patterns and rich descriptive insights. This participatory method ensured that IK

indicators were systematically documented and could be formalised for integration into the monitoring framework.

3.5.2. Sensor and Secondary Data Collection

To complement IK, the study also employed scientific datasets from established monitoring stations and locally installed sensors.

- **Pelonomi Station Data**: Secondary datasets were sourced from the **South African Weather Service (SAWS) Pelonomi air quality monitoring station** in Mangaung. These data covered 42 months (September 2020 — March 2024), including particulate matter ($PM_{2.5}$ and PM_{10}), sulphur dioxide (SO_2), and associated meteorological variables. This dataset served as the baseline for training and validating machine learning models.
- **Installed Sensors**: In addition to secondary datasets, a small number of **Arduino-based air quality sensors** were configured and deployed in selected community environments for limited testing. These sensors were used to assess the feasibility of localised data capture, with placements chosen to represent residential areas affected by mining dust dispersion. The testing period was constrained but provided proof-of-concept validation for integrating IoT-enabled monitoring into the broader framework.

The combination of datasets allowed the research to balance official long- term measurements with real-time experimental sensor data, supporting both robustness and adaptability of the system.

3.5.3. Data Sources Summary

Table 3.1 summarises the different data sources used in this study, indicating their type, origin, collection period, and purpose.

Table 3.1: Summary of data sources used in the study

Data Type	**Source / Instrument**	**Period of Collection**	**Purpose in Study**
Indigenous Knowledge	Surveys and questionnaires (Survicate tool)	Phase 1: Jul–Sep 2022 Phase 2: Jun–Oct 2024	Identifying IK-based pollution indicators; evaluating system usability.
Secondary Scientific Data	SAWS Pelonomi Air Quality Monitoring Station	Sep 2020 – Mar 2024	Training and validation of ML predictive models

Installed Sensor Data	Arduino-based wireless air quality sensors	Pilot period (Dates limited)	Proof-of-concept testing of IoT data collection

This triangulated data collection strategy ensured that both quantitative and qualitative evidence were available to design, implement, and evaluate the integrated monitoring framework.

3.6. Data Analysis

The analysis of collected data followed a structured process to ensure consistency, reliability, and integration between scientific datasets and Indigenous Knowledge (Bag et al., 2024). Both qualitative and quantitative techniques were employed, with emphasis on pre-processing, algorithmic modeling, qualitative mapping, and integration of diverse data sources into a unified monitoring framework.

3.6.1. Pre-processing

Prior to analysis, all datasets underwent **pre-processing** to enhance data quality and ensure comparability across sources.

- **Normalisation**: Historical air quality datasets from the Pelonomi monitoring station and sensor readings were normalised to a standard scale to eliminate distortions caused by differing measurement ranges.
- **Handling Missing Values**: Gaps within the Pelonomi datasets were addressed through the *FillMissing (nearest neighbour)* approach, which substitutes absent values with the closest valid observation. This ensured temporal continuity for model training without introducing artificial biases.
- **Data Cleaning**: Redundant fields were removed, and data fields were standardized to conform with the requirements of machine learning models and statistical analysis.

  This step provided a consistent dataset suitable for predictive modelling and integration with IK indicators.

3.6.2. Machine Learning Algorithms

Supervised machine learning techniques were selected to generate short- term forecasts of air pollution levels. The choice of algorithms was guided by their ability to handle non-linear data patterns, interpretability, and track record in environmental modelling.

a) **Support Vector Machine (SVM)**: Applied for its strength in binary classification (e.g., polluted vs. non-polluted), using hyperplanes to separate classes.

b) **Decision Tree Regression**: Adopted for its simplicity and interpretability, enabling straightforward visualization of predictor–outcome relationships.

c) **Random Forest Regression**: Used to improve robustness and accuracy by combining multiple decision trees. This ensemble approach was particularly effective in handling noisy environmental datasets.

d) **Gradient Boosting Regression**: Selected for its ability to iteratively correct errors from weak learners and model complex, non-linear dependencies in air pollution datasets.

e) Together, these algorithms provided complementary strengths. Their outputs were benchmarked against National Ambient Air Quality Standards (NAAQS) to evaluate compliance thresholds and forecasting reliability.

3.6.3. Indigenous Knowledge (IK) Analysis

IK data collected through surveys was analyzed using **Fuzzy Cognitive Maps (FCMs)** and the **Mental Modeler tool**.

Fuzzy Cognitive Maps (FCMs): These were employed to capture relationships between local pollution indicators (e.g., dust levels, odours, visible haze) and perceived pollution outcomes. FCMs allowed participants' qualitative insights to be translated into weighted, semi-quantitative models, reflecting the perceived strength and direction of relationships.

Mental Modeler: This participatory software was used to structure IK responses into standardized diagrams and matrices, facilitating comparative analysis between respondents and integration with scientific datasets.

The use of these tools ensured that IK, despite its qualitative nature, could be modelled in a format compatible with data-driven systems, enhancing the adaptability of the monitoring framework.

3.6.4. Integration Strategy

Integration of IK and scientific datasets was central to this research. The strategy involved cross-verification of indicators on **shared timescales**:

a) Pollution forecasts from ML algorithms (e.g., $PM_{2.5}$, PM_{10}, SO_2 predictions) were compared with IK indicators reported for the same period.

b) Convergences (e.g., high particulate readings coinciding with reports of dusty conditions) were flagged as validation points, while divergences informed refinement of both data processing and IK indicator interpretation.

c) Integration was facilitated through a central processing platform (ThingSpeak), which aggregated quantitative datasets and enabled cross-referencing with qualitative IK outputs.

This strategy ensured that both scientific and local experiential evidence informed the adaptive framework, enhancing both reliability and community relevance.

3.7. Framework Development

The central outcome of this study was the development of a **hybrid air pollution monitoring and prediction framework** that integrates Fourth Industrial Revolution (4IR) technologies with Indigenous Knowledge (IK). The framework was practice-driven, designed to respond directly to the environmental and socio-economic realities of the Free State province, particularly mining-affected communities in Lejweleputswa. Its construction involved systematic steps that translated methodological choices into a functional prototype.

3.7.1. System Architecture

The system architecture consists of five interconnected layers, through which data flows from one layer to another. Some relationships between the layers are bi-directional, while others feed directly into one layer and serve as a gateway between two layers. However, at the user interface level, it appears to be a bi-directional flow for ease of use.

The system was structured around five interconnected modules as shown in Figure 3.1.

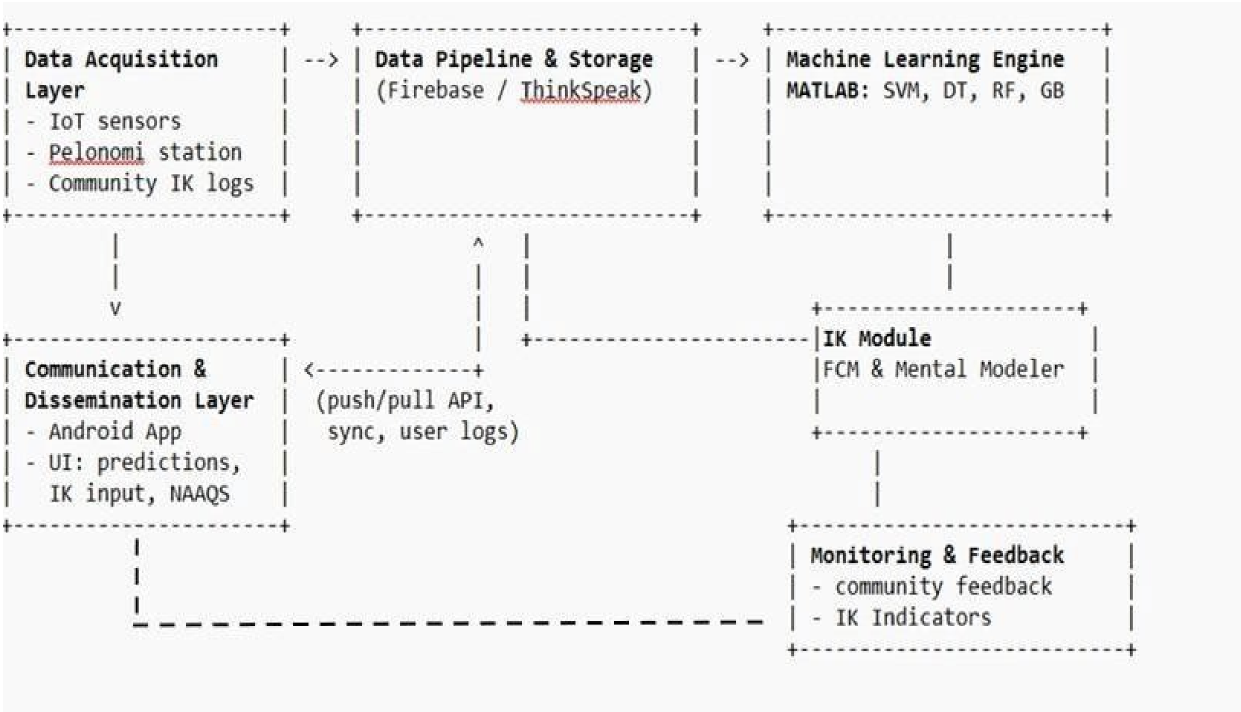


Figure 3.1 Conceptual architecture of the smart and adaptive air pollution monitoring framework.

1. **Data Acquisition Layer** — comprising installed IoT-based sensors, secondary datasets from the Pelonomi air quality station, and community-reported IK indicators. This is the layer where all data gathering for the study took place; it's the combination of primary (limited Arduino-based datasets) and secondary data (Pelonomi monitoring station datasets).

2. **Data Pipeline and Storage** — including Firebase for structured storage and Create, Read, Update, and Delete (CRUD) operations and ThingSpeak (for cloud-based aggregation and real-time analytics). This layer directly feeds the communication and dissemination layer. This layer had a bidirectional relationship with the machine learning engine as the interchange datasets. It is also directly connected to the Data acquisition layer.

3. **Machine Learning Engine** — implemented in MATLAB, hosting predictive algorithms (SVM, Decision Tree, Random Forest, Gradient Boosting) trained on Pelonomi air quality datasets. This is where all the development, training, and testing of the learning algorithms took place. This involved formulating predictions using datasets from the data acquisition layer through the data pipeline and storage layer. This layer uses the Data Pipeline and Storage layer as the gateway to the Data Acquisition layer.

4. **IK Monitoring and Feedback** — developed using Fuzzy Cognitive Maps (FCMs) and Mental Modeler to formalize and analyze community knowledge. This layer is directly connected to the Communication and Dissemination layer, enabling the end-user to communicate and allowing IK experts to capture new pollution incidents. This layer uses its instance (IK Modelling Class) as the gateway to the Data Pipeline and Storage layer, as well as the Machine Learning Engine Layer.
5. **Communication and Dissemination Layer** — an Android mobile application designed to deliver predictions, allow logging of IK observations, and display compliance against NAAQS thresholds. This layer is primarily for end-users, who interface with the entire system's backend through the Android application, serving as a dissemination tool. This layer is directly connected to the Data Pipeline and Storage layer to receive data. However, it is not directly connected to the Machine Learning Engine layer; it uses the Data Pipeline and Storage layer as the gateway.

3.7.2. Step-by-Step Prototype Construction

1. **Sensor Deployment and Testing**

Low-cost Arduino-compatible sensors were configured to measure particulate matter ($PM_{2.5}$ and PM_{10}) and sulphur dioxide (SO_2).

Sensors were tested in controlled conditions to verify calibration and reliability before integration. Although full-scale deployment inside mines was restricted, sensors were piloted in accessible community settings to validate real-time data capture.

2. **Data Pipeline Setup**

Sensor data and Pelonomi station data were uploaded to **ThingSpeak**, enabling visualization of time-series patterns.

A **Firebase SQL database** was used to store all components of the system, including processed datasets, IK indicators, and model outputs.

3. **Machine Learning Model Development**

Using **MATLAB R2024b** (licensed through the Central University of Technology), supervised ML models were trained and validated on 42 months of historical data.

Forecasting functions were designed to generate predictions up to four days in advance, aligned with NAAQS thresholds for Free State conditions.

4. **Indigenous Knowledge Module**

IK indicators (e.g., visible haze, strong sulfurous smell, dust carried by seasonal winds) were translated into FCMs using **Mental Modeler**.

These indicators were weighted and linked to scientific datasets, ensuring community observations informed system adaptability.

5. **Integration and Cloud-Based Synchronization**

All datasets and outputs were synchronized through the **ThingSpeak API**, enabling seamless updates between local sensors, Firebase storage, and MATLAB models.

Data consistency was ensured through normalization and timestamp alignment.

6. **Mobile Application Development**

A lightweight **Android application** was created to serve as the user interface.

The app disseminated forecasts, issued early warnings (e.g., "air quality unsafe in 2 days"), and allowed users to log IK-based observations.

Limited pilot tests in Welkom demonstrated feasibility, confirming the app's usability for semi-literate community members.

7. **Framework Adaptation to the Free State Context**

The framework was deliberately tailored to the realities of Free State communities. The list below highlights how the framework was tailored and delivered to meet the specific needs of the Free State Province context.

a. **Data Scarcity**: With restricted access to mining companies, the system relied on Pelonomi station data while still incorporating tested sensors for validation.
b. **Community-Centricity**: IK integration ensured that the lived experiences of semi-literate mine workers and residents were embedded into system predictions.
c. **Regulatory Alignment**: Predictions were benchmarked against South Africa's
d. NAAQS thresholds, ensuring compliance with national policy.
e. **Accessibility**: The Android app was designed with simple visuals and bilingual (English–Sesotho) support to ensure inclusivity.

3.7.3. Operationalization

Once fully integrated, the prototype demonstrated how 4IR technologies and IK can be combined to provide:

a. Real-time monitoring of pollutants.
b. Short-term forecasts (up to four days).
c. Community engagement through mobile logging.
d. A feedback loop where scientific and experiential knowledge validate one another.

This framework not only produced a functioning prototype but also served as a scalable model for other mining-affected regions in South Africa and beyond.

3.8. Study Area Description

The study was carried out in the Free State Province of South Africa, a region situated on the central plateau and characterized by a semi-arid climate with hot summers, seasonal rainfall, and cold winters. These climatic conditions directly influence air quality and pollutant dispersion across the province.

Mangaung Metropolitan Municipality, whose name means *"Place of Cheetahs"* in Sesotho, forms the administrative hub of the province. Centered around Bloemfontein, which is South Africa's judicial capital, Mangaung has a diverse economy encompassing government services, education, commerce, agriculture, and mining. Agriculture— particularly livestock farming and crop production—remains vital to surrounding rural communities, while Bloemfontein continues to serve as a center of cultural and administrative importance (Marais et al., 2021).

Welkom, located in the Lejweleputswa District, is the second-largest town in the province and the heart of South Africa's underground gold mining industry. All major mining shafts in the Free State are concentrated in and around Welkom, making it a strategic but also pollution-prone area. The city's economic fortunes have historically depended on mining, but this reliance has also resulted in significant environmental challenges, particularly those related to air quality. Communities living in proximity to mine shafts face daily exposure to airborne pollutants such as particulate matter and sulfur dioxide, making Welkom a relevant case study for this research.

The Free State, therefore, presents a complex environmental and socio-economic context: a blend of administrative centres, agricultural activity, and intensive mining operations. These overlapping pressures heighten the vulnerability of both mine workers and surrounding communities to air pollution. This makes the province—specifically Mangaung and Welkom—an ideal location for testing a hybrid pollution monitoring system that integrates Fourth Industrial Revolution (4IR) technologies with Indigenous Knowledge (IK).

Figure 3.2 presents a map of the Free State Province, showing the districts and municipalities. Mangaung Metropolitan Municipality is highlighted in green, and Lejweleputswa District, where Welkom is located, is highlighted in blue.

Figure 3.2: Free State Districts and their municipalities (source: Municipalities of South Africa 2024)

3.9. Ethical Considerations

Research ethics is a matter of scientific integrity, human rights, dignity, and collaboration between science and society. These principles ensure that participation in studies is voluntary, informed, and safe for research subjects. Even if a research idea is valuable to society, it does not justify violating the human rights or dignity of your study participants (Wexler et al., 2023). Therefore, this study, maintained anonymity as no personally identifying information was collected from the participants and no responses can be linked to a certain participant. All the participants were briefed all the information they need to decide whether they want to participate or not after they had understood the information. This includes information about the study's benefits and institutional approval to conduct the research. The researcher explained to the participants that there are no negative consequences or repercussions for their refusal to participate, and it is permissible for them to stop participating at any time they no longer wanted to be part of the study. The ethics application was signed and approved on 07 March 2023 by the chairperson of the ethics committee, Central University of Technology, Free State. The signed and approved ethical clearance document is attached as Appendix B.

.

3.10. Chapter Conclusion

This chapter presented the methodological foundation of the study, outlining the research philosophy, design, methods, and procedures employed to achieve the study's objectives. A pragmatist orientation guided the use of a mixed methods approach, enabling the integration of both quantitative and qualitative evidence. Specifically, implementation and descriptive research designs were adopted to balance problem-solving in practice with capturing lived experiences from the study population.

The chapter detailed the processes of data collection, including the use of surveys for Indigenous Knowledge (IK), secondary datasets from the Pelonomi monitoring station, and sensor-based measurements. It also explained the rationale for employing supervised machine learning algorithms and fuzzy cognitive mapping, both of which were central to modelling air pollution patterns and integrating IK into the hybrid system. Pre-processing steps, algorithm selection, and the data integration strategy were systematically discussed to ensure methodological rigour.
In addition, the development of the hybrid framework was presented as a practice-oriented process, incorporating sensors, data pipelines, predictive models, an IK module, and a mobile application, all tailored to the Free State context. Ethical safeguards were also articulated to demonstrate adherence to institutional and professional research standards.

Collectively, these methodological choices ensured that the study was grounded in both scientific rigour and contextual relevance. The framework described in this chapter lays the foundation for subsequent chapters, where its performance, functionality, and contribution to air pollution monitoring and prediction will be evaluated.

CHAPTER FOUR: FRAMEWORK ARCHITECTURE, DESIGN AND IMPLEMENTATION

4.1. Introduction

This chapter presents the design and development of the smart and adaptive air pollution monitoring system. The system was constructed to address the research objectives outlined in Chapter 1, and it builds directly on the methodological choices and framework development described in Chapter 3. In line with best practice for integrated environmental monitoring, the system is structured into distinct but interrelated **layers**, each responsible for a core function. This layered architecture ensures modularity, scalability, and adaptability to the Free State context.

The three layers are as follows:

1. **Data Collection Layer** — responsible for acquiring air quality data from multiple sources, including ground station records, installed sensor nodes, and Indigenous Knowledge (IK) indicators obtained through community engagement. This layer ensures that the system captures both quantitative measurements and qualitative lived experiences.
2. **Monitoring and Prediction Layer** — focused on processing, analysing, and forecasting air quality. This includes database management, pre-processing of datasets, application of supervised machine learning algorithms, and modelling of IK through fuzzy cognitive maps. The integration of these approaches allows for robust and context-sensitive predictions.
3. **Communication and Dissemination Layer** — designed to translate the technical outputs into actionable insights for end-users. This involves the Android mobile application, which provides forecasts, alerts, and a platform for community reporting, thereby ensuring that scientific and local knowledge are accessible to decision-makers, mine workers, and residents.

By following this layered approach, the chapter details how each component was developed, integrated, and operationalized into a functional prototype. The resulting system demonstrates the practical feasibility of combining Fourth Industrial Revolution (4IR) technologies with Indigenous Knowledge to create a community-centered environmental monitoring tool for the Free State Province.

4.2. Data Collection Layer

4.2.1. Data Source 1: Wireless Sensor Data Collection

Due to some unforeseen reasons, access to the mine was not granted for the collection of air pollution data at close range. The study used Pelonomi air quality data, collected over 42 months from September 2020 to March 2024. This data was used as a secondary dataset for training and testing of the ML algorithm. The study utilized the National Ambient Air Quality Standards (NAAQS) as a validation tool and compared them against air pollution forecasted values to determine whether the environment was safe or not.

Wireless Sensors – Arduino wireless sensors were configured to collect air pollutants on a real-time basis. Initially, the plan was to plant them at a mining shaft for real-time and instant data collection. However, as mentioned above, access was denied. The Arduino-generated wireless sensor data was used for exploring and functionality testing, in its limited state.
Figure 4.1 below illustrates how data moved from one phase to the next and was refined, with the primary aim of producing new air pollution predictions for days ahead.

Figure 4.1 Arduino Wireless Sensor Datasets Pipeline

Datasets were permanently stored on Firebase. The study utilized Firebase as the system's SQL database, and all the system's components were stored there. All the CRUD (Create, Read, Update, Delete) operations were also done on Firebase, and each component of the system had a direct link to the Firebase.
Data Source 2: Pelonomi Monitoring Station: As mentioned above, secondary datasets from the Pelonomi Ground Station were used as primary datasets of the study.

Figure 4.2 below demonstrates how Pelonomi datasets were imported into MATLAB as an Excel file (“dataquality.xlsx”) for further processing to obtain the required results.

```
% Load the data
% Read the data from the Excel file
[data, headers] = xlsread(filename);
filename = 'dataquality.xlsx';
[data, cleaned] = ʝ:ilmissin(ilename);

% Handle missing values (NaNs) with mean imputation
pm25_cleaned = fillmissing(dataquality(, 2), 'nearest';
pm10_cleaned = fillmissing(dataquality(, 3), 'nearest';
cleaned_dataquality = [dataquality(, 1), pm15_cleaned]
```

Figure 4.2 Import Pelonomi datasets into MATLAB

4.2.2. Data Source 3: Indigenous Knowledge Collection

Population and Sampling Method

Acharya et al. (2013) define a sample as a subset of the population, selected to represent the larger population, since it is not feasible to study the entire population. Based on the study objectives and how the research was carried out, a specific sampling technique was followed. Ahmed (2024) broadly classifies sampling techniques into two types: “Probability and Non- Probability” samples. Probability is a simple, random sampling technique that allows for generalizing the findings of the sample to the target population. To draw a valid conclusion from the IK results, the study employed a probability sampling technique. Every member of the population was given a chance to be part of the respondent (the sample), based on their availability and willingness to participate in the study.

In this study, the Probability sampling technique ensured representation of the entire population by using the sample. Probability sampling enabled the timely and feasible collection of IK results while maintaining the reliability, validity, and overall generalizability of the study findings. Considering the size of the Free State province, the scope and time frames of the study, the adoption of the probability sampling technique enabled the presentation of a quantifiable, comprehensive sample that represents all people affected by pollution in the Free State.

IK Administrative Method

To gather IK, a questionnaire was designed and distributed to the people. A free version of Survicate survey was used as an online service survey. Survicate survey is a web-based, Interactive and integrated questionnaire tool. This study used the free subscription survicate survey, which takes about 80 responses per 1 questionnaire. This study consisted of two questionnaires for phases one and two. Phase 1 is for understanding the research problem in depth, and phase 2 is for the entire system evaluation and assessment.

All the questions were sent from the web portal through a clickable link. The targeted people were mainly the IK experts. Upon finishing answering the questionnaire, all the responses were submitted directly back to the survey backend portal for further analysis and extraction of meaningful data. The survey portal generated themes, word art, sorted, and analyzed easy-to-understand data.

After designing and administering the online questionnaire, it was used to reach out to the targeted group of people remotely. The surveys were configured to allow each participant to answer the questions to the best of their knowledge. The first phase of the online questionnaire was made available to the participants from July 2022 to September 2022. The first phase of the questionnaire was specifically designed to understand the broader spectrum of the challenges that the mining and local communities face due to mining operations. The second phase of the questionnaire was made available from June 2024 to October 2024. This phase of the questionnaire was primarily designed to evaluate the entire system, to confirm whether the study had achieved its objective.

Sample size and Structure

The following participants were targeted to form the study sample for the collection of IK. The sample included four different personnel or groups of people as listed below.

1. **Mine workers:** They provided information on how they have been and are still surviving the harsh environmental conditions using their own local and perceived knowledge that they acquired over the years. The use of any other existing pollution monitoring system, provided they have access to it.

2. **Local Communities:** The residents responded to the questionnaire on how they had been making their living conditions as comfortable as possible, as they were affected by the pollution

generated as a result of underground mining activities around them.

3. **Medical doctors:** Even though they were not the system beneficiaries, based on their experience of working with mine workers and people who might present with symptoms associated with being constantly exposed to pollution, they provided the relevant information that might help in predicting the negative health risk factors resulting from pollution.

4. **Guest users:** This group of people showed interest and contributed to the study by raising awareness about pollution. The pollution ranges from air, water, to soil pollution (especially those they have experienced in their respective locations). These people also shared how they had been protecting themselves against many pollution harms.

The free version of the questionnaire can only take a maximum of 80 responses, out of which 62 responses were received. 15 responses were extracted from each population group. The remaining 2 responses were reserved for equal distribution purposes.

4.3. Monitoring and Prediction Layer

4.3.1. System Design

System design and architecture are the foundation of system implementation. Based on the conceptual framework developed, the system design was formulated. Based on the objectives, the system design focused on how the study integrated 4IR technologies, namely, IoT and machine learning, along with IK to develop a smart and adaptive air pollution monitoring system. Figure 4.3 below illustrates the components of the integrated system.

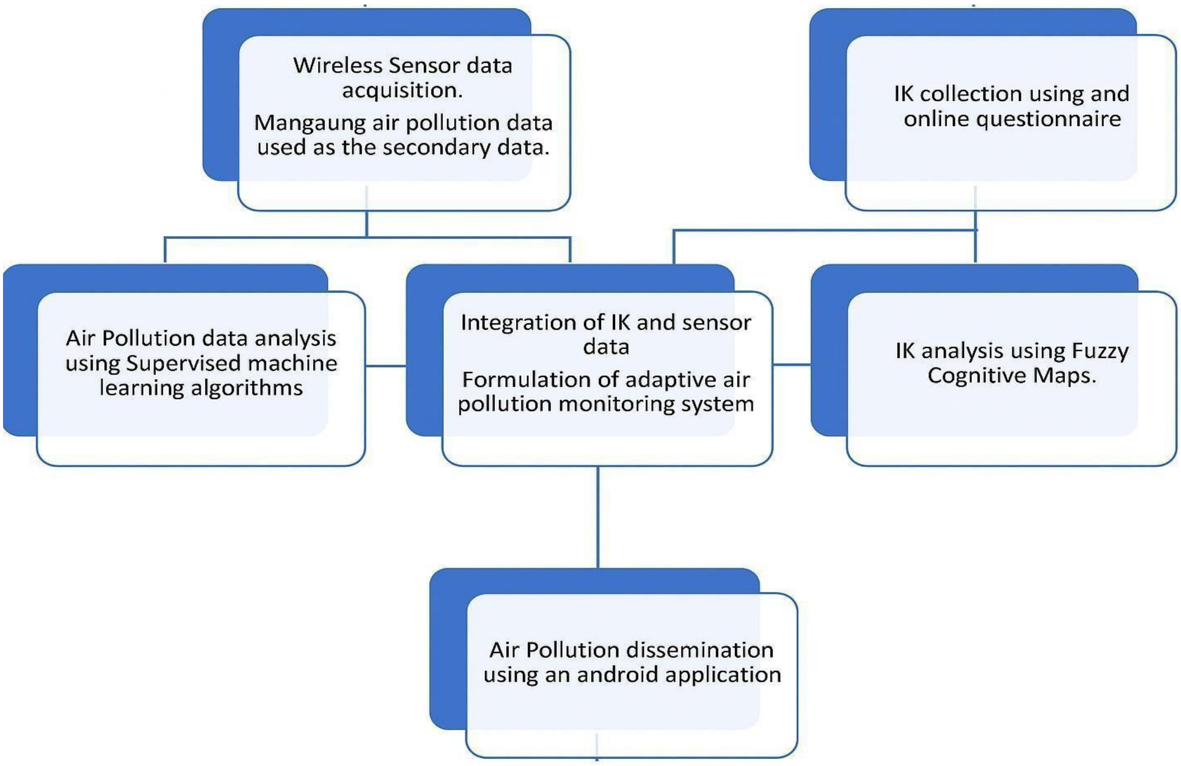


Figure 4.3 Components of the integrated system

Brief description: Components of the Integrated System

a. Wireless sensor Data Acquisition: Consists of both the primary (Arduino- based datasets) and the secondary (Pelonomi datasets). These are the datasets that were used to train and test machine learning models.

b. IK Collection: This also forms part of the data collection; however, this mainly focuses on Indigenous Knowledge.

c. Three Middle components (Dataset's analysis, IK analysis, and integration of both): After both IK and datasets were analyzed, they are integrated to form an adaptive air pollution monitoring system that is tailor-made.

d. Air Pollution dissemination: This is an End-user interaction point; it streams down the information comprehensively to the public. This component consists of the Android application, which serves as both the dissemination tool and deliverable.

4.3.2. Integrating IK with Datasets

On MATLAB, the ML models were trained and tested on the datasets and produced air pollution predictions, which forecast air pollution levels four days ahead. The predictions were made based on historical data. On the other hand, FCM analyzed the questionnaire responses where IK experts described air pollution based on their personal experiences. From the analysis, IK indicators such as dark dirty clouds, constant gaseous smell, and green water ponds IK indicators were formulated. The IK indicators generated from the IK expert have a positive correlation with the air pollution levels from the ML models. Therefore, qualitative and quantitative data agree with each other regarding the current and future state of air pollution. The Integration of IK and sensor-based datasets confirms that these 2 distinct kinds of knowledge are complementing each other. Three of the driver components: **Polluted, Not Polluted, and Moderate,** had their corresponding positive relationship IK indicators, which confirmed the state of the air, respectively.

4.3.3. Database & Data Processing Firebase Realtime Database

Firebase is a Backend-as-a-Service (Baas) platform that can host databases. In this study, a Firebase Realtime Database was designed to accommodate and store all the entities of the study, including internal components such as model development and external elements such as user registration. Figure 4.4 shows the Firebase Console.

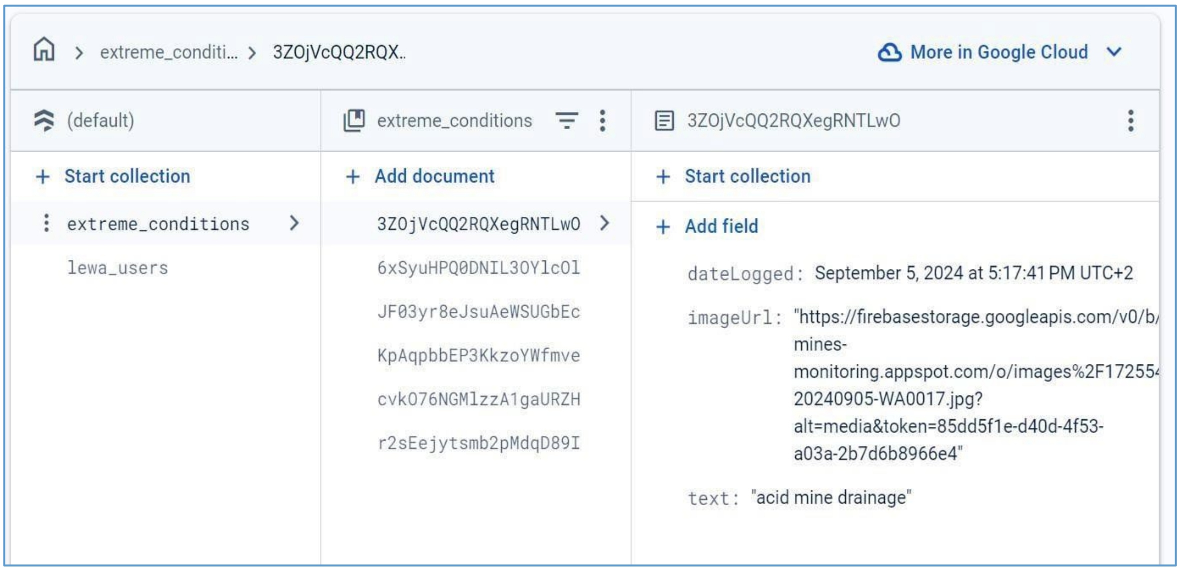


Figure 4.4 Cloud Firestore Console Dashboard.

Each dataset entry in the database was captured as a separate, using predicted dates as the data headers. This part of the database was connected to MATLAB through the database response name: "- O8Z4aslYh3w8-oROHFL". In MATLAB, the URL serves as a unique identifier to locate the designated space for air pollution predictions within the Firebase database. All the air pollution forecasting was performed in MATLAB; however, this section focuses solely on the storage of those air pollution predictions. Figure 4.5 illustrates how pollutants are stored in the database, organized by their respective predicted dates.

Figure 4.5 Firebase Console Per Predicted Date Entry.

ThingSpeak

After both qualitative and quantitative data have been gathered, it must be merged to form a single, integrated air pollution dataset that can be disseminated to people. The study adopted ThingSpeak, a cloud-based IoT processing platform, and it serves as the primary processing cluster.

ThingSpeak is a platform where data analytics can be performed and an open-source, internet of things application programming interface used to store and retrieve data from interconnected things using the hypertext protocol over the internet or via a local area network. It also provides access to a wide range of embedded devices and web services.

A ThingSpeak channel was created for the air pollution monitoring system, where all the study's datasets were transmitted using Hypertext Transfer Protocol (HTTP). The channel's API was called, and the necessary libraries were imported to upload sensor data from the local directory to the ThingSpeak channel. Datasets were synchronised between ThingSpeak and the database

(Cloud Firestore).

Figure 4.6 shows how datasets are displayed on the ThingSpeak channel.

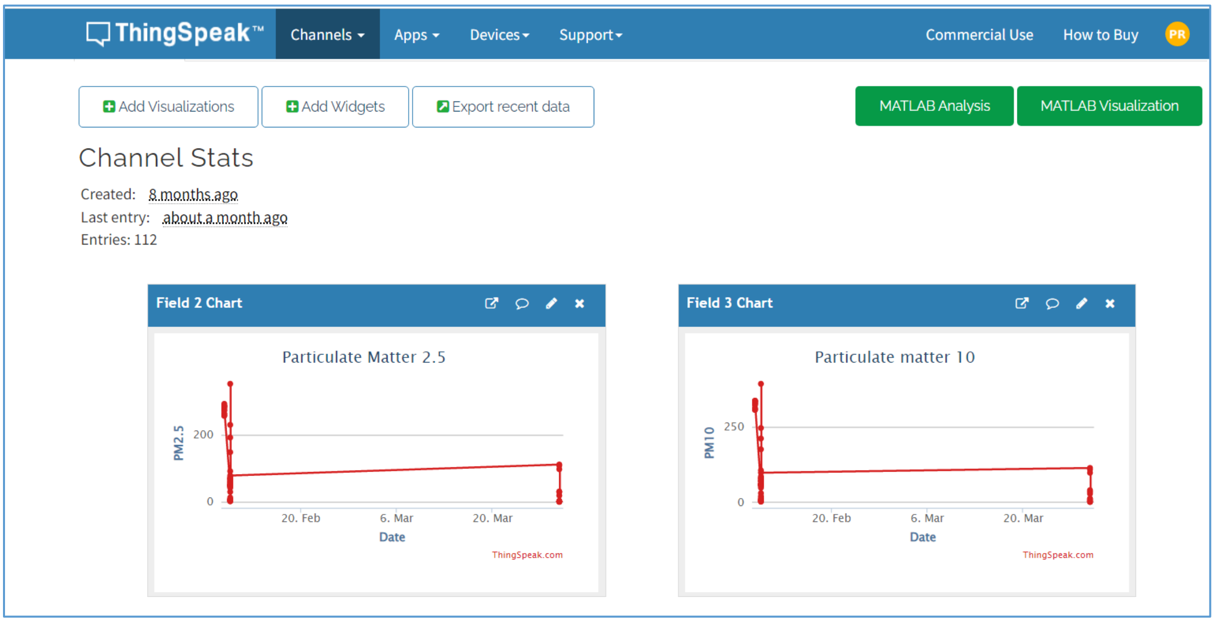


Figure 4.6 ThingSpeak Channel for The Project

4.3.4. Machine Learning Models Support Vector Machine (SVM)

After splitting the air quality data into training and testing sets, the study trained ML models to predict pollutant concentrations. This study explored the resilient and flexible SVM algorithm for supervised learning. SVM is used for regression and classification. SVM can also manage complex, non- linear correlations between input data and the target variable, making it suitable for air quality prediction. The ability of SVM to characterize complicated patterns makes it possible for this system, based on the desired output as well as the proposed system. Figure 4.7 illustrates the development of SVM.

```
%Support Vector Machine (SVM)
svm_model = fitrsvm(X_train, y_train);
% Train the SVM model
svm_model = fitrsvm(X_train, y_train);
% Predict PM2.5 values for testing data
y_pred = predict (svm_model, X_test);
```

Figure 4.7 Support Vector Machine

Before employing the SVM model, input characteristics for range similarity were adjusted. To predict target pollutant concentrations, the study employed SVM regression to learn patterns from the training data. SVM model hyperparameters, including the regularization parameter and kernel- specific parameters, were tuned to balance model complexity with generalization performance. After training the SVM model, it was tested on another dataset. To make the air quality prediction system practical, the ability of the model to project unknown data accurately was assessed. The study used MSE, R-squared, and the coefficient of determination to evaluate the SVM model's prediction accuracy. To assess its merits and downsides, the SVM model was compared to linear regression and decision tree regression. Importantly, SVM is tolerant to outliers and can handle multidimensional feature spaces (Peng et al., 2022). The interpretability of the SVM model was also assessed by evaluating feature relevance and decision boundary design. The SVM model accurately predicted air quality. It correctly interpreted complicated non-linear patterns and project testing data. Feature engineering, hyperparameter tuning, and domain-specific air quality dynamics information may improve the SVM model.

Decision Tree Regression

Decision trees are a popular ML method that models options and their consequences as trees. This makes them ideal for classification and regression. In air quality prediction, Decision Tree Regression can capture non-linear correlations, handle numerical and categorical variables, and provide interpretable models that can be easily understood and communicated (Asha et al., 2022). Figure 4.8 shows how Decision Tree Regression was defined.

```
% Fit Decision Tree Regression model
tree_model = fitrtree(X_train, y_train);
% Predict PM2.5 values for testing data
y_pred_tree = predict(tree_model, X_test);
% Calculate performance metrics (e.g., Mean Absolute Error)
mae_tree = mean(abs(y_test - y_pred_tree));
fprintf('Decision Tree Regression - Mean Absolute Error (MAE): %.2f\n',
mae_tree);
```

Figure 4.8 Decision Tree Regression

The study pre-processed the air quality data to resolve missing values and scale or convert input characteristics before using the Decision Tree Regression model. After that, the training dataset was utilized to extend the Decision Tree by iteratively splitting the data using the feature that reduced the mean squared error (MSE) or another regression-specific measure the most. To minimise overfitting, the study monitored the model validation set performance as the tree grew. When the tree becomes too complex, it may overfit by remembering training data (Whitaker, 2014) Instead of generalising new observations. This involved cutting trees and limiting depth and complexity.

The study also examined ensemble methods, such as Random Forest Regression and Gradient

Boosting Regression, which use several Decision Tree Regression models to improve air quality forecasting accuracy and robustness. Decision Tree Regression and ensemble methods also construct a flexible air quality prediction system (Whitaker, 2014).

Random Forest Regression

Along with SVM and Decision Tree Regression, Random Forest Regression for air quality prediction was also tested. Multiple Decision Trees in Random Forest increase model accuracy and durability. Random Forest Regression can predict air quality with multiple input variables, non-linear correlations, and model interpretability. Single Decision Trees may overfit and perform poorly on new data, while Random Forest employs many prediction trees. Each Decision Tree is trained using random data, it improves model robustness and dependability. Before using Random Forest Regression, the study preprocessed the datasets, converted input properties, and fixed missing values. Training dataset boosted decision tree ensemble. The study adjusted hyperparameters to improve model complexity-generalization. This helped Random Forest Regression to predict pollutant concentrations. After training, the study tested the Random Forest Regression model on another dataset. MSE, R-squared, and coefficient determination were used to evaluate the model's prediction accuracy against SVM and Decision Tree Regression models. Figure 4.9 shows Random Forest Regression.

```
% Fit Random Forest Regression model
forest_model = TreeBagger(50, X_train, y_train);
% Predict PM2.5 values for testing data
y_pred_forest = predict(forest_model, X_test);
% Calculate performance metrics (e.g., Mean Absolute Error)
mae_forest = mean(abs(y_test - str2double(y_pred_forest)));
fprintf('Random Forest Regression - Mean Absolute Error (MAE): %.2f\n',
mae_forest);
```

Figure 4.9 Random Forest Regression

Duisebekova et al., (2019) mention that the Random Forest Regression system can handle various input sources and automatically choose the most essential air quality forecast parameters. Most researchers and scientists used feature significance scores to understand how climate, emission sources, and land use affect air pollution. For the interpretability of the Random Forest Regression model, each ensemble Decision Tree was evaluated. The model is more complicated than a Decision Tree, but it is possible to learn from its structure and principles.

Random Forest Regression accurately predicted air quality. It controlled huge feature areas, recorded complex, non-linear patterns, and explained air pollution levels. To enhance Random Forest Regression, most data scientists recommend feature engineering, hyperparameter tweaking and domain-specific air quality dynamics data (Baraneetharan, 2020). In this research work, the goal was to design a comprehensive air quality forecasting system employing the Random Forest Regression model and other ML technologies.

Gradient Boosting Regression

To forecast air quality, the study incorporated Gradient Boosting Regression. Gradient Boosting builds a powerful prediction model from weak learners, usually Decision Trees (Joseph et al., 2022). Gradient Boosting Regression may predict air quality better than other regression approaches because it can capture complicated, non-linear correlations between input data and the target variable (Ihianle et al., 2020). In addition, Gradient Boosting constructs Decision Trees successively to fix past mistakes. Figure 4.10 displays Gradient Boosting Regression.

```
% Fit Gradient Boosting Regression model
boosting_model = fitensemble(X_train, y_train, 'LSBoost', 200, 'Tree');
% Predict PM2.5 values for testing data
y_pred_boosting = predict(boosting_model, X_test);
% Calculate performance metrics (e.g., Mean Absolute Error)
mae_boosting = mean(abs(y_test - y_pred_boosting));
fprintf('Gradient Boosting Regression - Mean Absolute Error (MAE): %.2f\n',
mae_boosting);
% Create a table to display MAE values for all models
model_names = {'SVM', 'Decision Tree', 'Random Forest', 'Gradient
Boosting'};
mae_values = [mae_svm; mae_tree; mae_forest; mae_boosting];
mae_table = table(model_names', mae_values, 'VariableNames', {'Model',
'Mean_Absolute_Error'});
% Display the table
disp('Mean Absolute Error (MAE) for all models:');
disp(mae_table);
```

Figure 4.10 Gradient Boosting Regression

The study developed a Gradient Boosting Regression model using pre- processed datasets, addressed missing values, and converted input attributes as necessary. The study also examined learning rate, ensemble size, maximum tree depth, and regularization parameters in the Gradient Boosting Regression model during training, as well as optimizing hyperparameters for model complexity and generalization. This would ensure Gradient Boosting Regression predicts valid pollutant concentrations. The Gradient Boosting Regression model automatically prioritises the toughest training data observations to forecast. Air quality prediction may benefit from this since data may include anomalies or complicated patterns that a single model cannot explain. Using feature significance scores and partial dependency graphs to examine the Gradient Boosting Regression model's interpretability. Moreover, feature significance ratings identified the most relevant air quality forecasting input factors, while partial dependency graphs illustrated how predictors interact with target pollutant levels. The study evaluated the Gradient Boosting Regression model on another dataset after training.

4.3.5. IK Analysis: Fuzzy Cognitive Maps Fuzzy Cognitive Maps Modelling

The main objective was to understand the Indigenous Knowledge (IK) domain through a statistical analysis of responses used to forecast air pollution. Once the domain was understood, the next step was to formulate the inputs, processing methods, and outputs that define the mathematical logic behind the FCM-based air pollution monitoring system. Figure 4.11 illustrates the process followed in this research study to predict air pollution using FCMs.

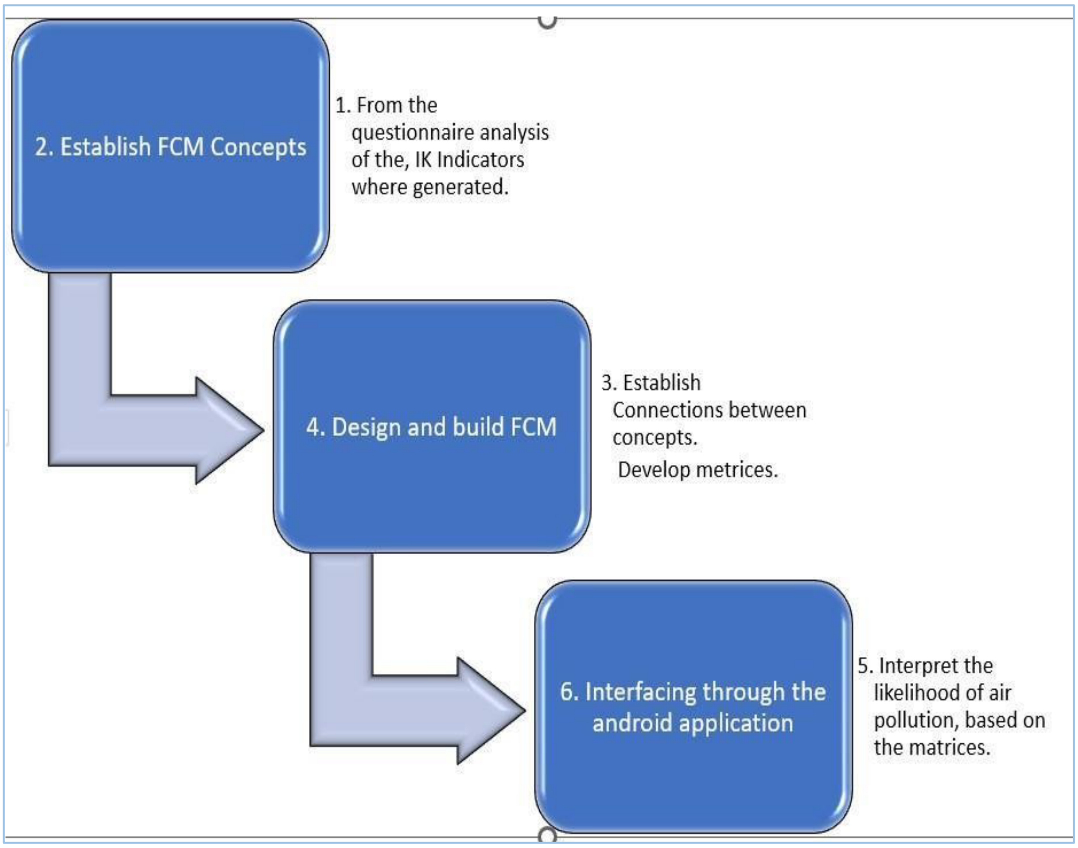


Figure 4.11 The Process used to Predict Air Pollution Using FCM

To make sure that all the data collected from the study participants is handled, the system employed AI technique called fuzzy cognitive maps (FCM). FCM groups the indicators as nodes based on the responses and makes decisions based on the scenarios. FCM consists of nodes, which is the total number of concepts the model has. These concepts have interactions and interaction strength, which signify the weight between the concepts. The relationship strength values range from -1, 0, and +1, where -1 indicates a negative relationship, 0 indicates no causal relation, and +1 indicates a positive relationship (Cleveland et al., 2024).

Figure 4.12 presents an example of FCM of local knowledge on air pollution. The relationship between the components varies, which may be due to seasonal changes such as high and low temperatures, wind pressure, and other climate-related factors. Examining the representation, it becomes easy to identify which concepts influence other concepts by showing the interconnections between them. FCM allows updating the initial structure of the presentation, such as the addition or deletion of an interconnection whenever necessary, making FCM adaptive to change.

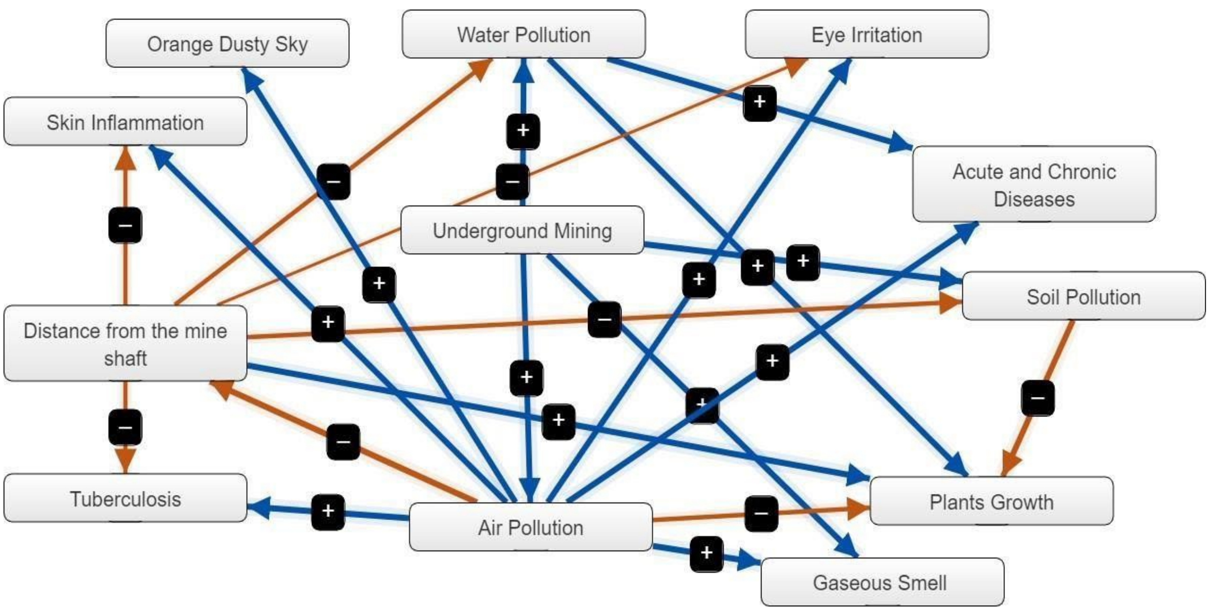


Figure 4.12 : A Simple Fuzzy Cognitive Map Representation Between Nodes

This study adopted the Manual expert-driven technique of FCM because it was found to be appropriate for qualitative systems, as it relies on human knowledge and skills to build the FCM matrix. The construction of this FCM requires the input of human experience and knowledge on air pollution and then integrates the accumulated experience and knowledge concerning the underlying causal relationships amongst the nodes, characteristics, and components of the entire system.

FCM Component's Description

The FCMs were built on 12 components, and each component had a certain number of connections based on its specific influence (positive or negative) on the other.

Table 4.1 demonstrates the generic meaning of the FCM components.

Table 4.1: FCMs Components Description

Tuberculosis	A respiratory infectious disease that can cause infection in lungs or other tissues.
Air pollution	Is the contamination of the indoor or outdoor environment by any chemical, physical or biological agent that modifies the natural air.
Water pollution	Is the release of substances into bodies of water that makes it to be unsafe for human use and disrupts the aquatic ecosystem
Gaseous smell	A colourless and flammable gas that a strong smell concentration levels in the air.
Plant growth	Is the increase of plants volume and normal speed at which plants are growing at.
Acute and Chronic diseases	A medical condition that occurs suddenly and gradually spreading across the body, pollution is amongst contributing factors towards the development.
Distance from the mine shaft	The measure of the distance between the residential area, farmlands, and the mining shaft.
Orange dusty skies	Atmosphere that is filled with dust and pollutants.
Skin Inflammation	Is the swelling and irritation of the skin caused by exposure to harsh environment
Eye irritation	Is the eye discomfort, itchiness or dryness of the eye caused by exposure to harsh environment?
Soil pollution	Soil pollution is the presence of toxic chemicals in the soil, in high enough concentrations that poses a risk.
Underground mining	Is the process of extracting ore from below the surface of the earth?

4.3.6. Framework Integration

Cross-validation

Machine learning results (predicted values from 05 September 2024 to 09 September 2024) were cross-compared with the actual data from Pelonomi- NAQI that were collected on the same time scale, which is 05 September 2024 to 09 September 2024

Figure 4.13 below displays the ML-predicted values for the speculated dates in September 2024.

Date	SO2	PM2,5	PM10
24:00 05/09/2024	10,262	72,977	46,115
24:00 06/09/2024	13,553	64,811	28,783
24:00 07/09/2024	5,6106	95,331	85,415
24:00 08/09/2024	19,979	104,66	76,939
24:00 09/09/2024	15,752	69,093	30,535

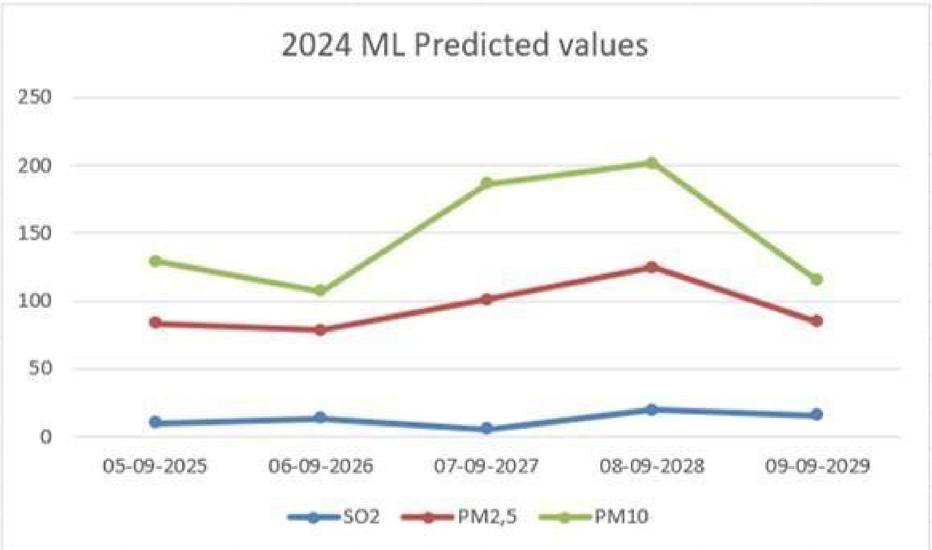


Figure 4.13: ML-Prediction From 05 - 09 September

Figure 4.14 below displays the Pelonomi-NAQI dataset's values for the speculated dates in September 2024.

Date	SO2	PM2,5	PM10
24:00 05/09/2024	1,466	28,57	111,462
24:00 06/09/2024	1,569	46,279	123,019
24:00 07/09/2024	1,499	36,42	111,761
24:00 08/09/2024	1,482	32,533	94,928
24:00 09/09/2024	1,675	6,613	30,407

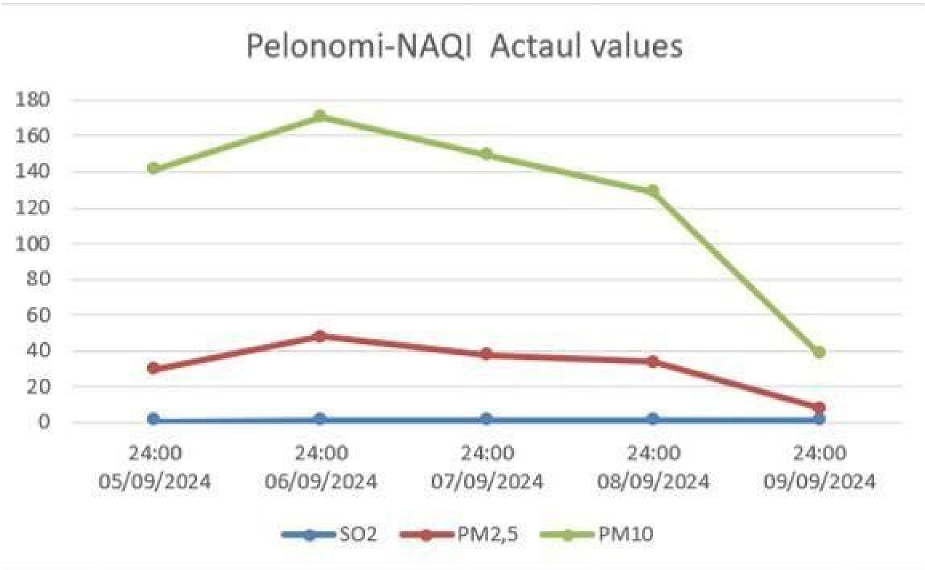


Figure 4.14: NAQI Database From 05-09 September

Cross-validation Results

Based on the above two figures (Figures 4.25 and 4.26), a Cross-validation between the two datasets reveals a positive correlation, indicating external generalization. This suggests strong robustness in the estimation of the developed ML models. The numbers are not precisely the same, which might be because the monitoring stations where the datasets were gathered are not in the same location, but under the Free State province.

4.4. Communication and Dissemination Layer

4.4.1. Android Application Class Diagram

At a larger scale, all the components modeled through Fuzzy Cognitive Maps are integrated with air pollution data from MATLAB, and all system entities are stored together in a single NoSQL database, each under a declared class. Additionally, classes were interconnected with one another. The relationships indicate classifiers that are associated with each other, as well as those that represent generalisations, realisations, or dependencies on other classes and classifiers. Figure 4.15 illustrates the overall relationship between all the classes that make up the system and how they interlink to fulfil the system's objective.

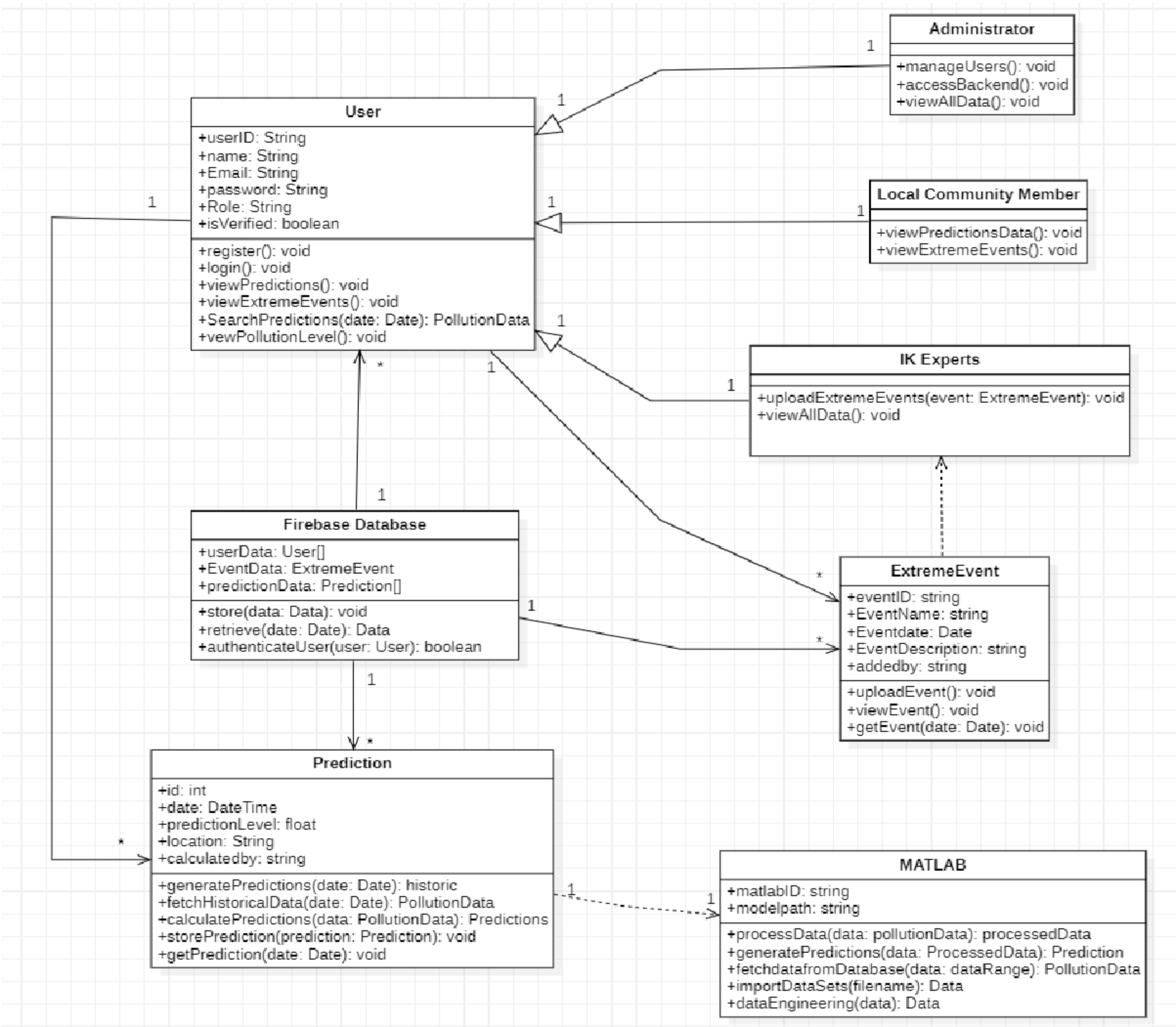


Figure 4.15: System Class Diagram

System Use Case

The system use case demonstrates an interaction between a user and the air pollution monitoring system. The use case describes how the system interacts with an external entity, such as end-users and the NoSQL database. Figure 4.16 presents the use case.

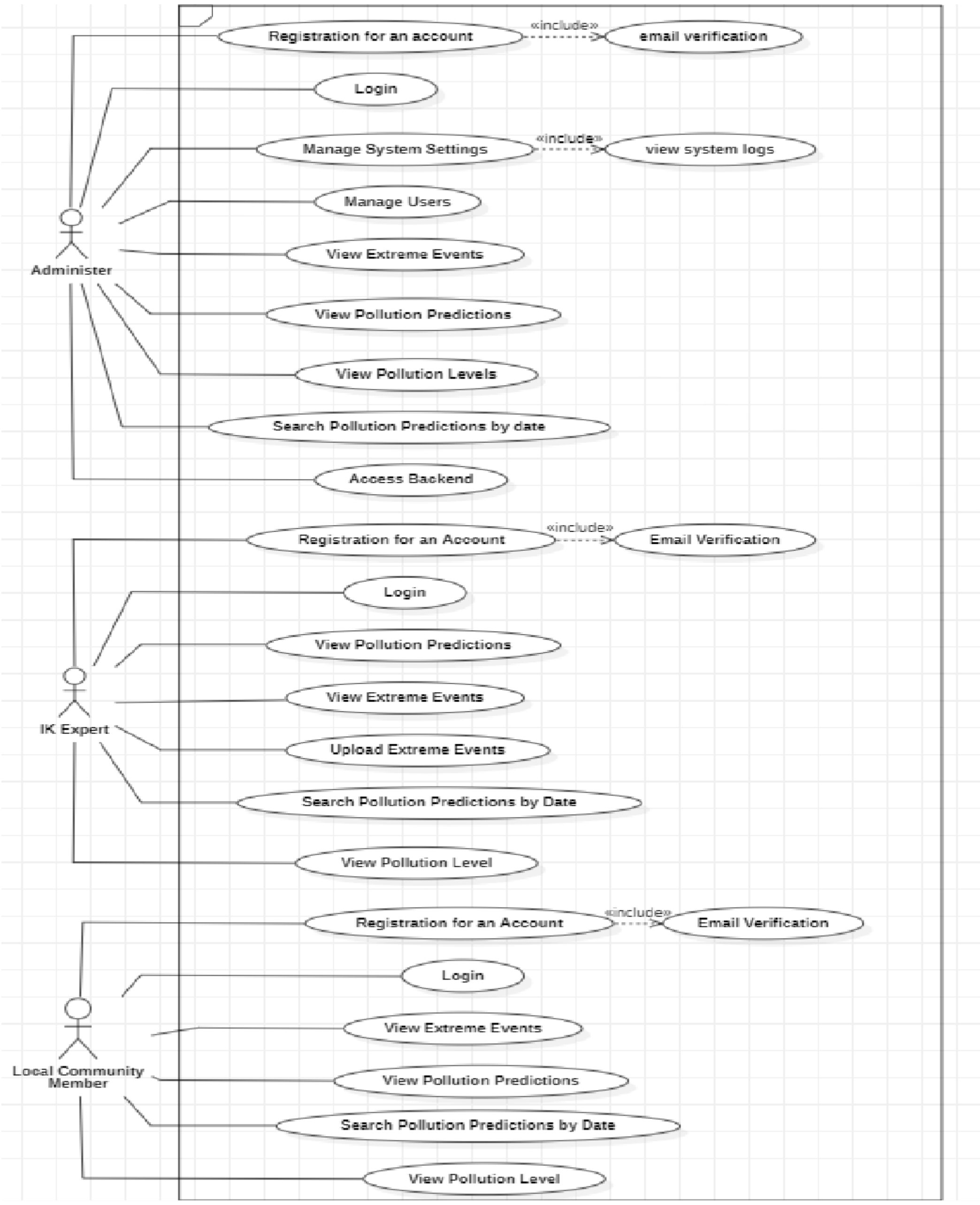


Figure 4.16: Overall System Use Case.

Sequence Diagram: Data Dissemination from Android Application

A sequence diagram was used to visualise the interactions between objects and other components of the system. The sequence diagram focuses on how objects communicate with each other through messages, showing the order in which, these interactions occur. Data is transmitted from one point to another, with entity headers serving as the system's callout points from the external part of the system. Users do not have access to the system's backend. Only system administrators and researchers have access to the system's backend.

Figure 4.17 presents the system sequence diagram of data from one hop to the other

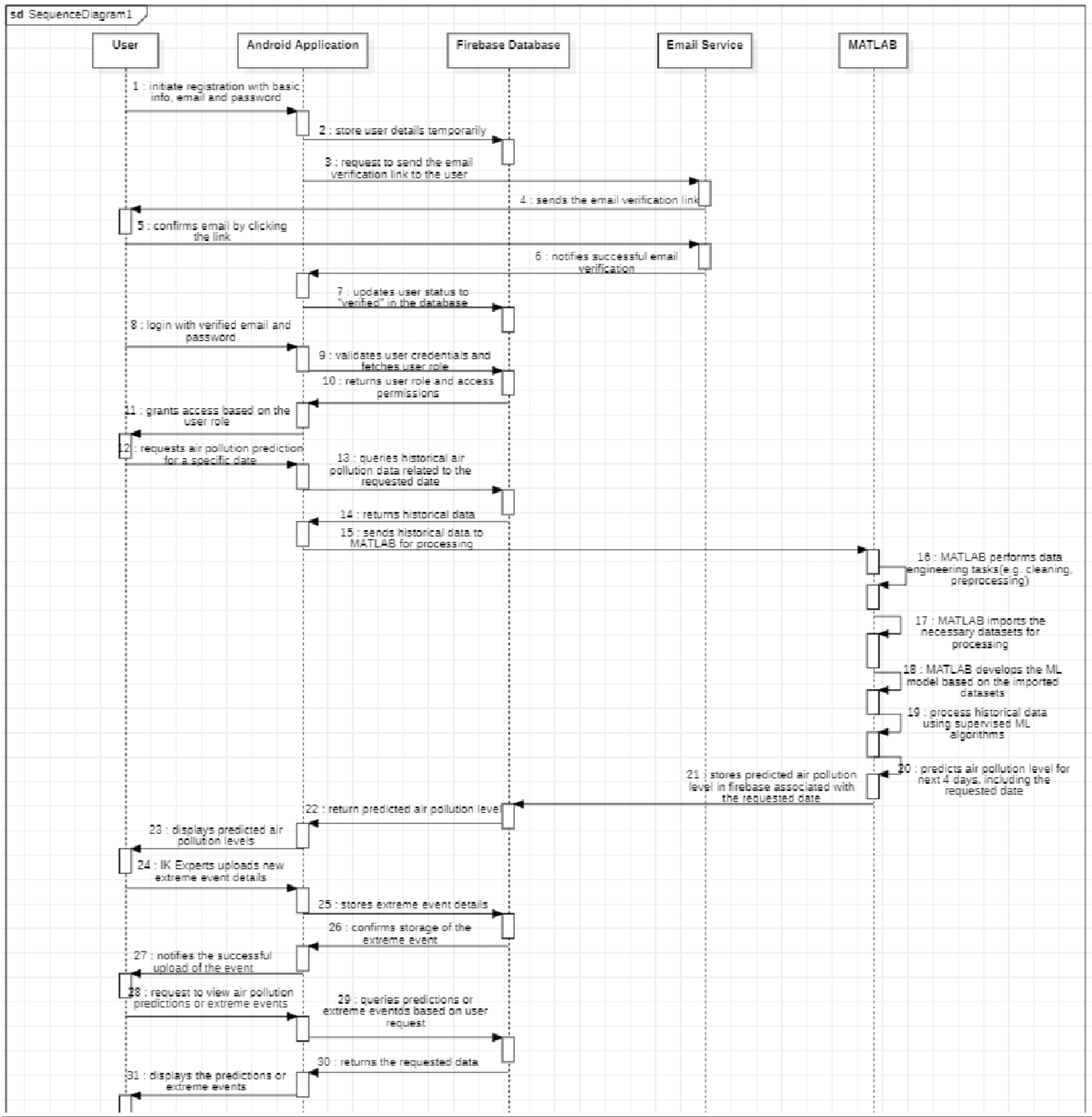


Figure 4.17: Sequence Diagram.

In general, here is what a sequence diagram does.

1. **Sequence interactions**: It shows the flow of messages between objects and processes of the

systems.

2. **Depict messages**: Arrows between the lifelines represent function calls between objects.

3. **Highlights system** behavior: It models complex interactions, making it easy to understand how a system behaves in different scenarios or use cases.

4. **Documents logic**: It documents the logic of how specific processes are carried out.

4.5. Generic System Prototype

4.5.1. Iterative development process.

Until the desired state was produced, the system evolved through an iterative process. Figure 4.18 illustrates the processes involved throughout the development phases that culminate in the final mobile Android application.

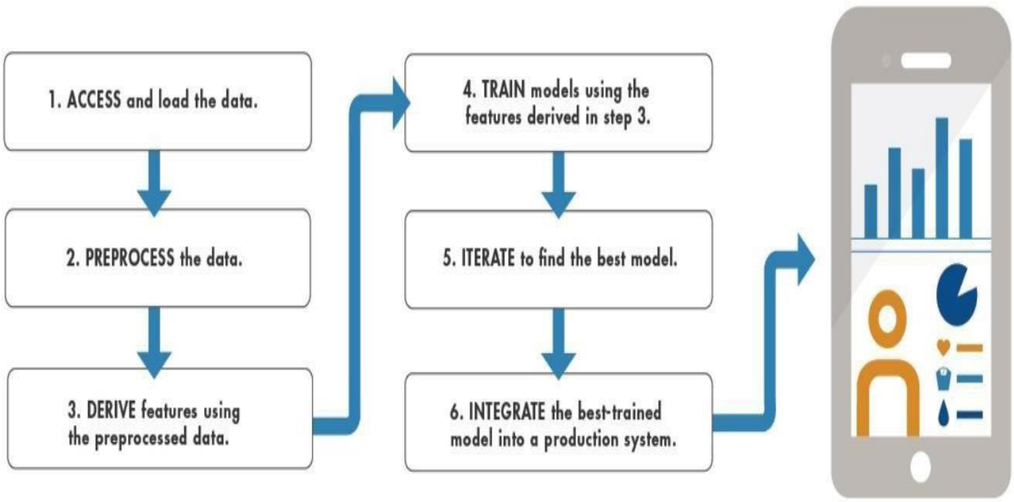


Figure 4.18: Generic Overview and Prototype

Overview of prototype functionality and readiness for evaluation.

A generic prototype was built based on the system requirements. It was reworked and reiterated as necessary until the desired system was fully developed and completed as per the proposed Android application design.

4.6. Conclusion

This chapter presented the design of the smart and adaptive air pollution monitoring system, structured around a **layered framework** that integrates diverse sources of knowledge and technology. At the **Data Collection Layer**, information was sourced from wireless sensors, the Pelonomi ground station, and community-derived Indigenous Knowledge indicators, creating a robust foundation of inputs. These datasets were then channelled into the **Monitoring and Prediction Layer**, where supervised machine learning models and Fuzzy Cognitive Maps were employed to generate forecasts and enrich them with local insights. Finally, the **Communication and Dissemination Layer** translated these outputs into accessible formats through a mobile application, ensuring that communities, decision-makers, and other stakeholders could receive timely, actionable information.

By combining Fourth Industrial Revolution (4IR) technologies with Indigenous Knowledge, the system demonstrates a hybrid, context-sensitive approach to environmental monitoring that is both scientifically rigorous and socially relevant. The layered framework offers not only technical robustness but also community ownership and inclusivity, making it a novel contribution to air quality management in the Free State.

The next chapter presents the **evaluation and results** of the implemented prototype. It details the system's performance, assesses its accuracy and usability, and examines the extent to which the framework achieved its objectives in predicting and disseminating air pollution information.

CHAPTER FIVE: RESULTS AND DISCUSSION

5.1. Introduction

This chapter presents and discusses the results of the study, structured around the layered framework introduced in Chapter Three. The outcomes are derived from both scientific methods, namely ground station data analysis and machine learning predictions, as well as Indigenous Knowledge (IK) insights collected from mining communities in the Free State Province. Together, these complementary approaches provide a robust basis for developing and testing the proposed hybrid air pollution monitoring system.

The chapter is organised to reflect the three functional layers of the framework: the **Data Collection Layer**, which integrates inputs from wireless sensors, the Pelonomi ground station, and IK surveys; the **Monitoring and Prediction Layer**, where supervised machine learning algorithms and fuzzy cognitive maps were applied; and the **Communication and Dissemination Layer**, where results are channelled to stakeholders through the mobile application interface.

Importantly, the results are not presented in isolation but are critically examined in relation to the research objectives outlined in Chapter One. This dual focus ensures that the findings are interpreted within their scientific, contextual, and community relevance, highlighting both the achievements and the challenges encountered in operationalising the system for the Free State context.

5.2. Results from the Data Collection Layer

5.2.1. Ground Station Data (Pelonomi)

Pre-processing, cleaning, and exploratory analysis.

Using a supervised learning algorithm requires thorough data engineering to enable models to predict and produce the most accurate air predictions possible. After a careful evaluation of most supervised learning models, four models that met the study's needs performed better and were chosen. For the prediction of air pollution, the study explored and employed SVM, Decision Tree, Random Forest, and Gradient Boosting. Air pollution predictions were a desired output of the study; however, much work was done before arriving at such predictions after it had passed through various layers.

The transformation from raw data to the desired state involved three different layers: Input, Hidden, and Output layers. Figure 5.1 illustrates the ML algorithm prediction layers.

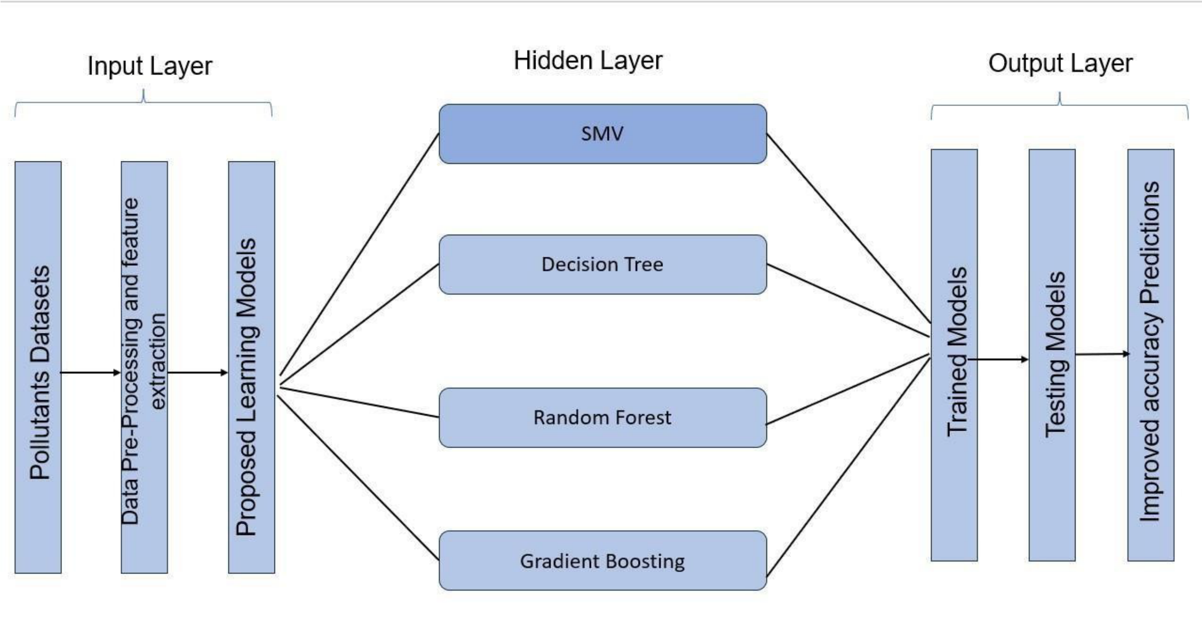


Figure 5.1: Machine Learning Algorithm Layers

Data Engineering

Data engineering is the foundation of contemporary information systems (Alkhatib et al., 2023). It entails creating and maintaining data pipelines that gather, process, and turn data into insightful information (Bhat et al., 2019). In the contemporary landscape of information-driven decision making, it serves as a pivotal discipline that lays the groundwork for practical data analysis and interpretation.

Datasets Preparation

Data normalization was performed to standardize and adjust the datasets, ensuring data consistency, integrity, and scaling. Data normalisation improved the efficiency and effectiveness of data handling, processing, and analysis (Delavar et al., 2019).

Data Cleaning and Processing

Initially, the data used was stored in an Excel file. The file 'dataquality.xlsx' provides air quality data used as the dataset for this project.

On MATLAB, the xlsread() function was used to import data from 'dataquality.xlsx'. The imported data was stored as datasets, while the headers contained column names. The datasets include quantifiable air pollutant concentration levels, whereas the headers define the corresponding variables.

After the data was imported, it had missing values referred to as NaN (Not a Number). For better performance of the models, null values must be handled before working with the datasets. Null values are a result of several factors that might have happened during the data collection phase of the study. These include device (wireless sensors) technical failures, network glitches, noise, or any other general error. These missing values must be addressed for accurate and seamless data analysis and prediction. To address these issues, the fillmissing() MATLAB method was adopted to fill the missing values of the datasets/pollutants: PM2.5, PM10, and SO2, respectively, as shown in Figure 5.2 below.

```
% Handle missing values (NaNs) with mean imputation
pm25_cleaned = fillmissing(dataquality(, 2), 'nearest';
pm10_cleaned = fillmissing(dataquality(, 3), 'nearest';
cleaned_dataquality = [dataquality(, 1), pm15_cleaned]
```

Figure 5.2: Data Cleaning and Filling out the Missing Values

Generally, *fillmissing()* can replace missing values with the closest non- missing values in the dataset. The function uses the nearest non-missing value, using the 'closest' technique. This method was used because it efficiently handles missing data. It also retains data distribution and linkages, ensuring all rows have full data for investigation. A variable called *cleaned_dataquality* was declared, and that held the new data after replacing missing PM2.5, PM10, and SO2 values with the closest non-missing values. The cleansed dataset(*cleaned_dataquality*) was further used for exploration, data analysis, and ML.

Figure 5.3 illustrates how the data was transformed from its raw and unclean state to a clean and prepared dataset for supervised ML modeling.

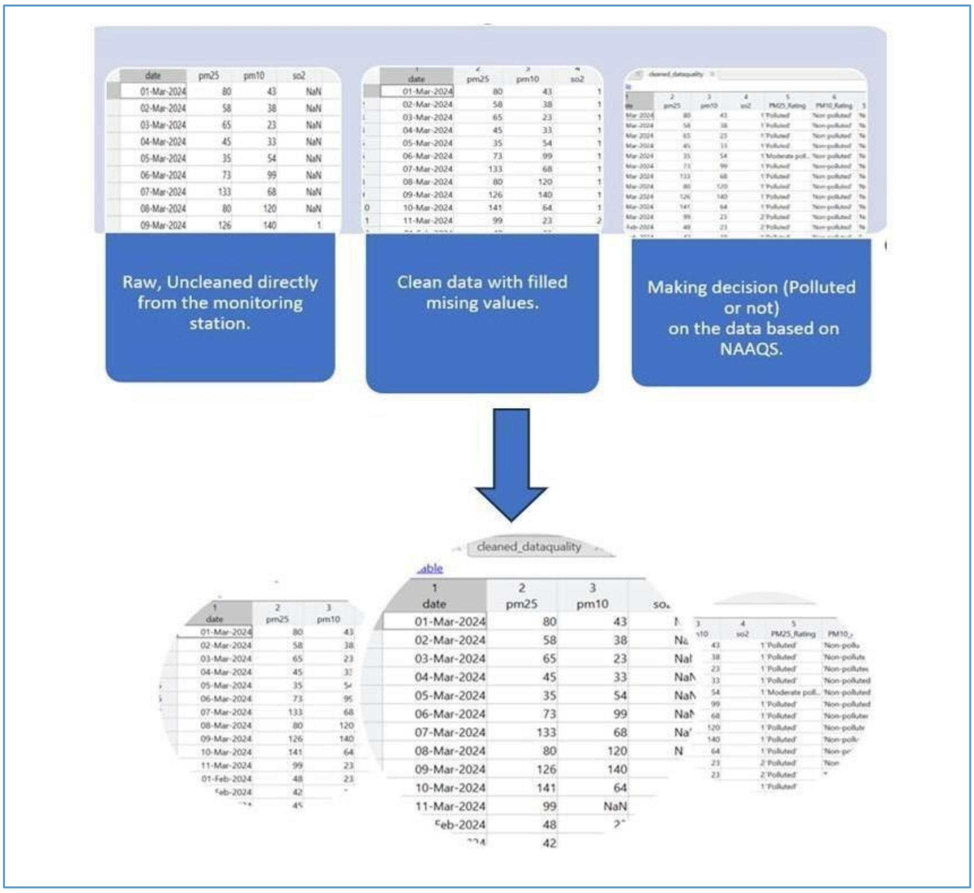


Figure 5.3: Transformation of Datasets to a Desired State

Key descriptive statistics (PM2.5, PM10, SO2)

Following the cleaning and pre-processing of the air quality datasets, a thorough descriptive data analysis was conducted to obtain a better understanding of air quality characteristics and patterns. Summary statistics were generated for the processed data, including concentrations and distributions of PM2.5, PM10, and SO2 in the air. The code in Figure 5.4 shows the summary using MATLAB's summary() command. This function is responsible for summarizing the processed data of all pollutant variables, including minimum, maximum, mean, median, and standard deviation. This helped to identify the data ranges, averages, and dispersion in air quality analysis and interpretation.

```
% Analysis
% Summary statistics
% Compute summary statistics for cleaned data
summary_stats = summary(cleaned_dataquality);
% Display summary statistics
disp('Summary Statistics:');
disp(summary_stats);
```

Figure 5.4:Datasets Summary Statistic

The summary statistics function unveils the air quality levels. Air pollution is depicted by the lowest and greatest pollutant concentrations for best and worst air quality. A mean and median values illustrate the average air quality, while standard deviation shows the data variability. The summary statistics method was of significant help in grasping the air quality, it helped to identify the main trends, deviations, and problematic spots that are worth attention. Besides summary statistics, exploratory data analysis (EDA) relies on histograms for demonstrating data distribution. With the histogram of every pollutant's variable, it was possible to identify the data shape and patterns, find skewness and extreme values, and understand frequency of the air quality. Through this visualization process, it was then possible to understand the nature of air quality in the area and to signify the study's research gap.

Trends and comparison with NAAQS.

The exploration data analysis (EDA) also compared measured air pollution (datasets) with NAAQS for validation purposes. The spread of air quality testing assisted in evaluating frequency and severity. If most observations are good or moderate, air quality is typically satisfactory with occasional NAAQS exceedances. Conversely, lower ratings imply more significant air quality issues that require special attention and swift intervention.

Figure 5.5 illustrates the comparison to NAAQS.

```
% Comparing to NAAQS Values
% Create a new column for PM2.5 air quality classification cleaned_dataquality.PM25_Rating = cell(height(cleaned_dataquality), 1);
cleaned_dataquality.PM25_Rating(cleaned_dataquality.pm25 <= 12) = {'Non- polluted'};
cleaned_dataquality.PM25_Rating(cleaned_dataquality.pm25 > 35) =
{'Polluted'}; cleaned_dataquality.PM25_Rating(cleaned_dataquality.pm25 > 12 & cleaned_dataquality.pm25 <= 35) = {'Moderate polluted'};
% Create a new column for PM10 air quality classification cleaned_dataquality.PM10_Rating = cell(height(cleaned_dataquality), 1);
cleaned_dataquality.PM10_Rating(cleaned_dataquality.pm10 <= 150) = {'Non- polluted'};
cleaned_dataquality.PM10_Rating(cleaned_dataquality.pm10 > 150) =
{'Polluted'};
% Create a new column for SO2 air quality classification cleaned_dataquality.SO2_Rating = cell(height(cleaned_dataquality), 1);
cleaned_dataquality.SO2_Rating(cleaned_dataquality.so2 <= 75) = {'Non- polluted'};
cleaned_dataquality.SO2_Rating(cleaned_dataquality.so2 > 200) =
{'Polluted'}; cleaned_dataquality.SO2_Rating(cleaned_dataquality.so2 > 75 & cleaned_dataquality.so2 <= 200) = {'Moderate polluted'};
% Convert categorical ratings to numerical codes pm25_codes = zeros(size(cleaned_dataquality.PM25_Rating));
pm25_codes(strcmp(cleaned_dataquality.PM25_Rating, 'Non-polluted')) = 1; pm25_codes(strcmp(cleaned_dataquality.PM25_Rating,
'Moderate polluted')) = 2;
pm25_codes(strcmp(cleaned_dataquality.PM25_Rating, 'Polluted')) = 3;
pm10_codes = zeros(size(cleaned_dataquality.PM10_Rating)); pm10_codes(strcmp(cleaned_dataquality.PM10_Rating, 'Non-polluted'))
= 1; pm10_codes(strcmp(cleaned_dataquality.PM10_Rating, 'Moderate polluted')) = 2;
pm10_codes(strcmp(cleaned_dataquality.PM10_Rating, 'Polluted')) = 3;
so2_codes = zeros(size(cleaned_dataquality.SO2_Rating)); so2_codes(strcmp(cleaned_dataquality.SO2_Rating, 'Non-polluted')) = 1;
so2_codes(strcmp(cleaned_dataquality.SO2_Rating, 'Moderate polluted')) = 2; so2_codes(strcmp(cleaned_dataquality.SO2_Rating,
'Polluted')) = 3;
% Count occurrences of each rating for PM2.5  pm25_rating_counts = histcounts(pm25_codes, 'BinLimits', [1 4], 'BinMethod', 'integers');
% Count occurrences of each rating for PM10
pm10_rating_counts = histcounts(pm10_codes, 'BinLimits', [1 4], 'BinMethod', 'integers');
% Count occurrences of each rating for SO2
so2_rating_counts = histcounts(so2_codes, 'BinLimits', [1 4], 'BinMethod', 'integers');
% Plot bar charts for each pollutant figure;
```

Figure 5.5: Comparing to NAAQS

Wireless Sensor Outputs

Data visualizations and summary statistics enhance ML analysis, helping to identify air quality data patterns, correlations, and anomalies that can be used for prediction models. If EDA shows some significant seasonal or meteorological impacts on pollution levels, researchers may use weather characteristics as predictors in machine learning models to increase accuracy and robustness (Xiong et al., 2022). Figure 5.6 illustrates the pollutants visualization.

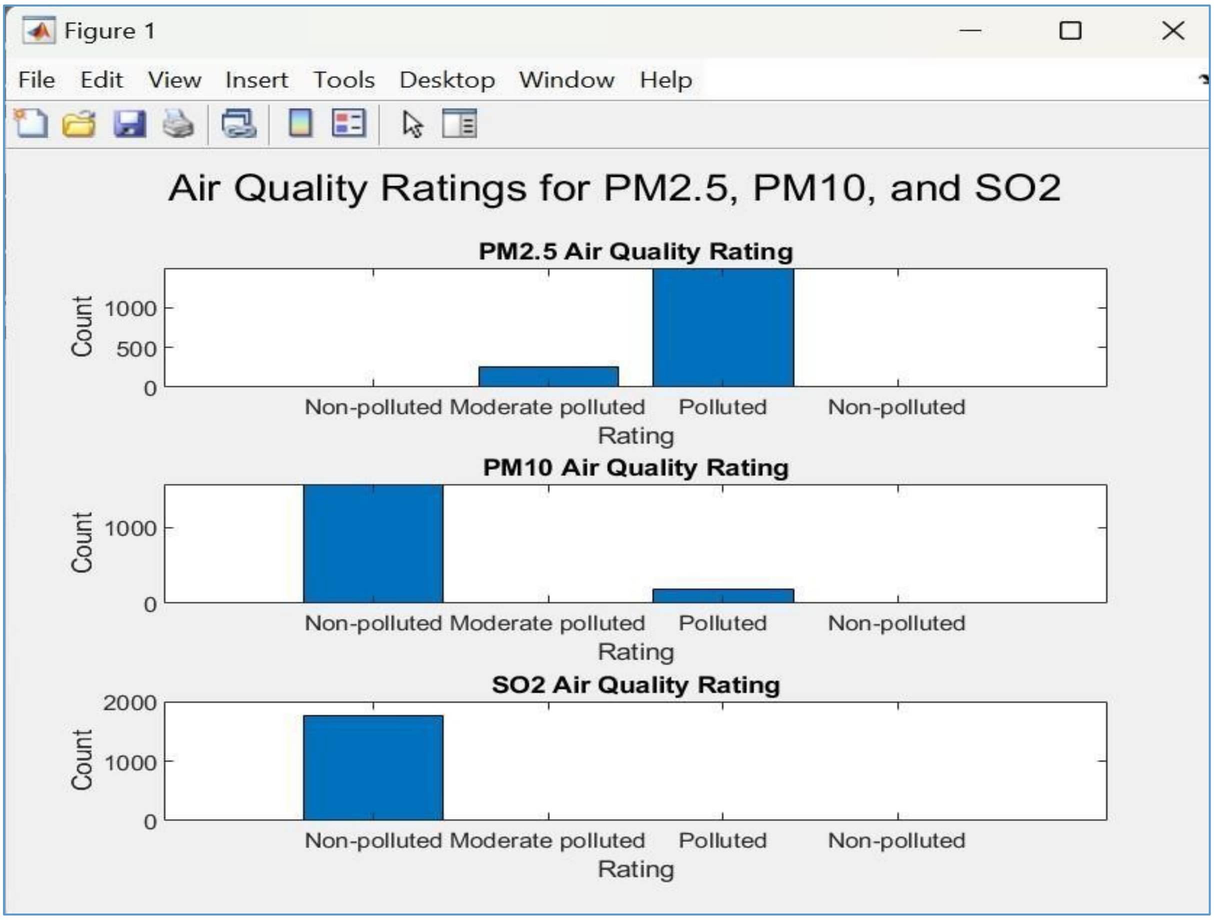


Figure 5.6: Pollutants Visualization

5.3. Indigenous Knowledge Indicators

After engaging with the participants, their responses were analyzed and generated into IK indicators. This section of the study discusses that in detail.

Figure 5.7 is the collection of responses to the question: “In your own words, please share how you have been protecting yourself from pollution”. The responses were selected based on the identified survival method, its relevance, and feasibility.

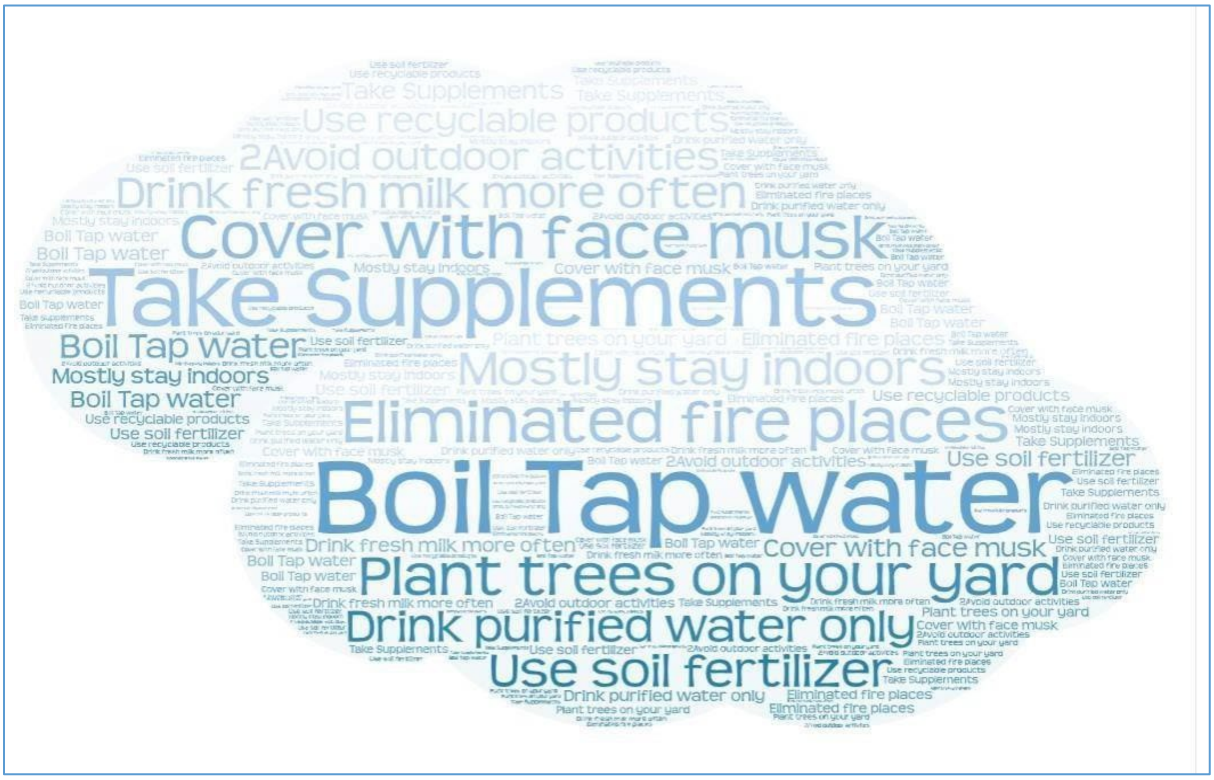


Figure 5.7: Responses on the Survival Method

Figure 5.8 represents the mini graph generated from the question: "Is your method of survival still effective?". This question was derived from the above responses. Accordingly, 54.8% of the responses agreed that these survival tips are always effective. Meanwhile, 38.7% stated that external factors, such as the inability to stay indoors during the day and financial constraints preventing the purchase of purified water, sometimes reduce the effectiveness of these survival tips. Only 6.5% of the Only responses disagreed with their efficacy, possibly due to other external factors, including but not limited to those mentioned above.

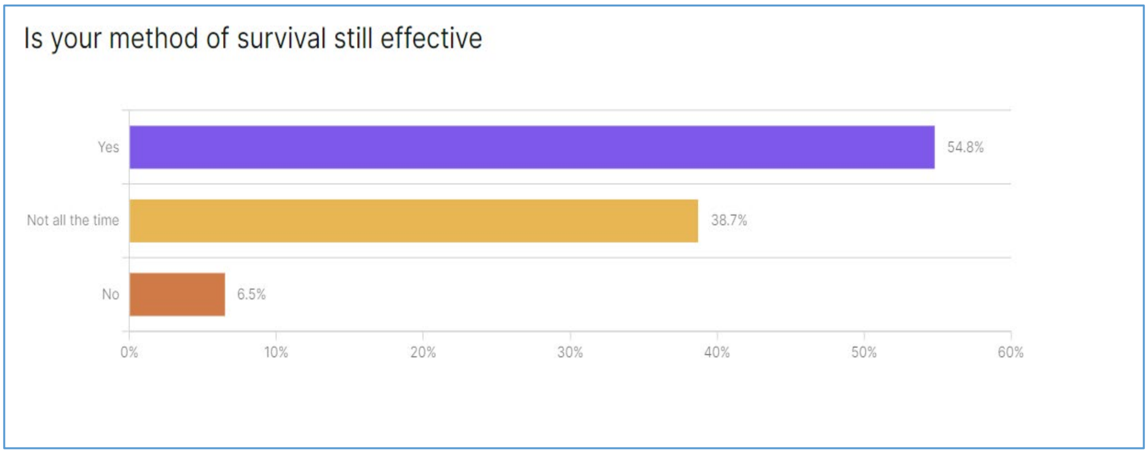


Figure 5.8: Ratings of the Survival Method

Figure 5.9 shows that 67.7% of people would recommend the above survival tips based on their effectiveness. Only 32.3% were unlikely to recommend them due to factors such as the cost of purchasing purified water, lack of backyard space to plant trees, and work constraints.

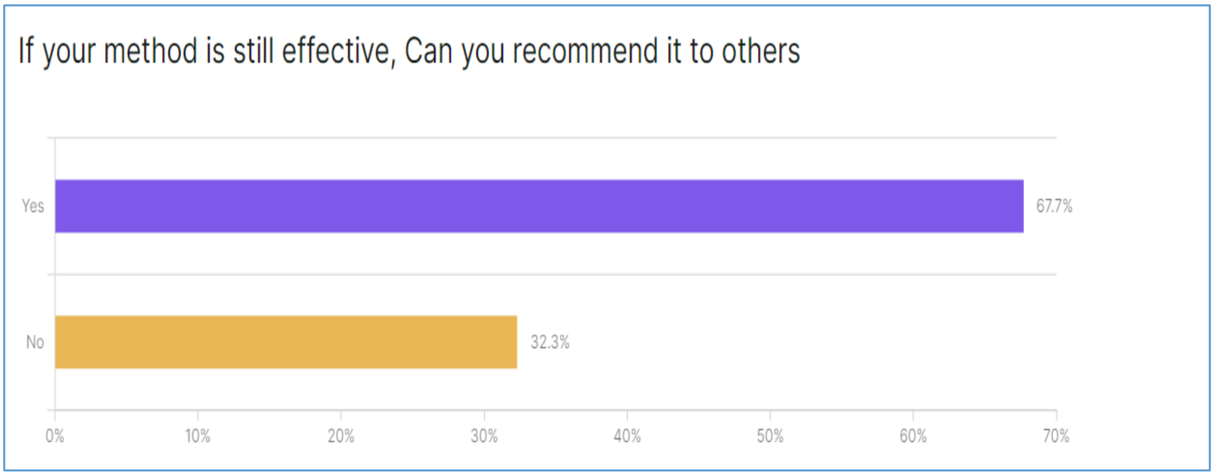


Figure 5.9: Recommendations on Survival Method

The results show that most of the people were aware of what they were facing. They were found to be taking all the possible measures to protect themselves and confirm that their methods are effective. These responses are based on their lived experiences. However, a minority confirmed that these methods do not always fully mitigate the effects of pollution, as certain factors hinder their effectiveness. Although they represent a minority in the data, in practice, a significant number may still fall victim to pollution-related illnesses, which can ultimately lead to death.

Figure 5.10 presents a collection of responses to the question: “What are the effects of pollution in your community?”. For this word art representation, responses were selected based on popularity, focusing on the most likely effects of pollution and common symptoms experienced by individuals exposed to pollution over extended periods. These symptoms primarily manifest in sensitive groups, such as the elderly, infants, and individuals of any age group with underlying chronic conditions.

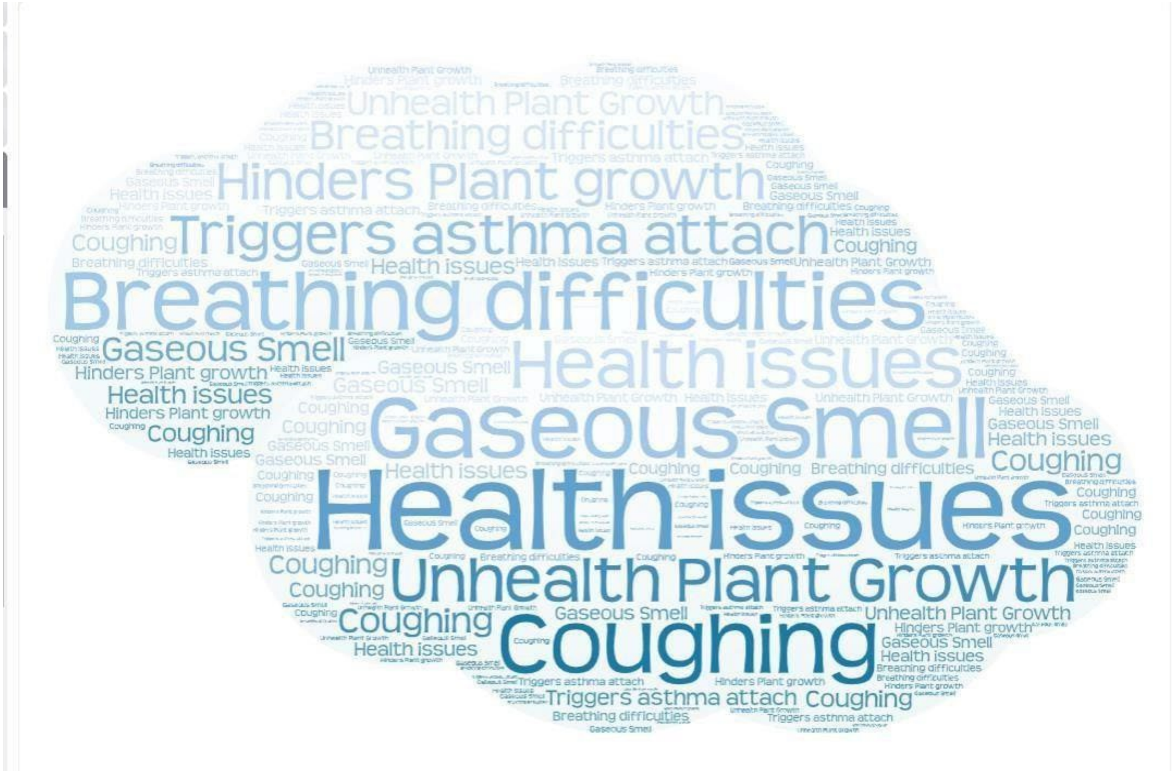


Figure 5.10: Effects of Pollution

5.4. Results from the Monitoring and Prediction Layer Data Segregation

After data exploration and training, meaningful trends were found. Building and testing prediction models required splitting data into training and testing sets. Dividing datasets into training and testing sets enabled the models to match prior data, generalize, and predict new, unknown data. From the training set, models discovered data links and patterns. The testing set was used to assess the trained model's prediction performance impartially. The study carefully examined time-dependent data features, dividing them into training and testing groups.

Figure 5.11 below shows how datasets were divided into training and testing groups.

```
% Assuming the target variable is PM2.5 and features include PM10 and SO2
X = cleaned_dataquality(:, {'pm10', 'so2'});
y = cleaned_dataquality.pm25;
% Define the training set size (e.g., 80% for training, 20% for testing)
train_percentage = 0.8;
train_size = floor(train_percentage * size(X, 1));
% Split data into training and testing sets
X_train = X(1:train_size, :);
y_train = y(1:train_size);
X_test = X(train_size+1:end, :);
y_test = y(train_size+1:end);
```

Figure 5.11: Splitting Data into Training and Testing Datasets

Meanwhile, 80% of the data as training and 20% as testing. This process trained and assessed the models using the provided dataset, and from them, predictions were formulated. A time-series split was employed to test the model's temporal dynamics and trends of air quality data, which was crucial for accurate forecasts. This method prevented data leakage from the testing set from impacting model training and overestimating performance.

The study tested model generalization using *k-fold* cross-validation. This method divided the dataset into *k* distinct, non-overlapping subgroups and trained and tested the model *k* times using a different subset as the testing set. This prevents the data split into training and testing sets from affecting the performance of the model, resulting in a more precise prediction accuracy estimate. To guard for potential ML modelling challenges, the data separation technique was carefully built.

Predictions

The study focused on predicting air pollution for the next 4 days from the current date. Supervised machine learning algorithms were used to forecast PM2.5, PM10, and SO_2 air concentration levels. To enhance system prediction accuracy, the training data was updated with the latest observations, the machine learning model's hyperparameters were adjusted, and alternative data sources or domain-specific knowledge were considered.

Figure 5.12 bellow shows how models were developed from MATLAB, with PM2.5 as the target variable.

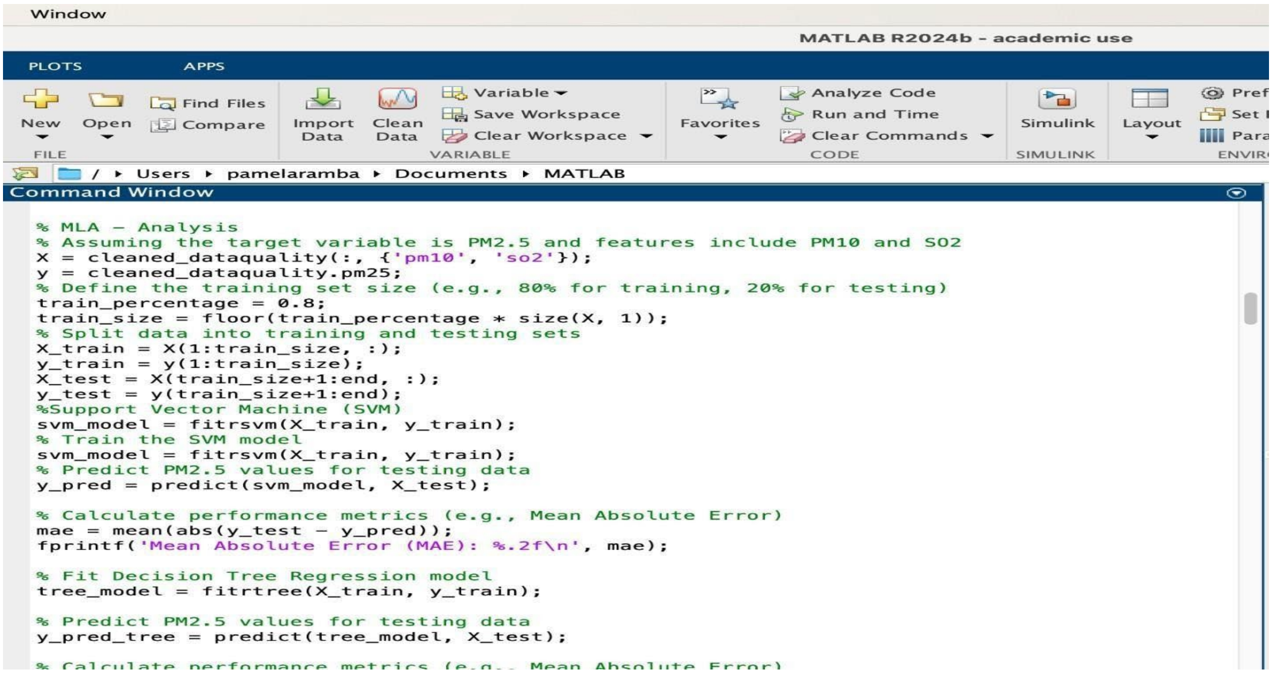


Figure 5.12: Development of Models from MATLAB

Figure 5.13 below mainly focuses on how the mean absolute error and Fit Gradient Boosting were developed.

/ ▸ Users ▸ pamelaramba ▸ Documents ▸ MATLAB

Command Window

```
% Calculate performance metrics (e.g., Mean Absolute Error)
mae_tree = mean(abs(y_test - y_pred_tree));
fprintf('Decision Tree Regression - Mean Absolute Error (MAE): %.2f\n', mae_tree);


% Fit Random Forest Regression model
forest_model = TreeBagger(50, X_train, y_train);

% Predict PM2.5 values for testing data
y_pred_forest = predict(forest_model, X_test);

% Calculate performance metrics (e.g., Mean Absolute Error)
mae_forest = mean(abs(y_test - str2double(y_pred_forest)));
fprintf('Random Forest Regression - Mean Absolute Error (MAE): %.2f\n', mae_forest);


% Fit Gradient Boosting Regression model
boosting_model = fitensemble(X_train, y_train, 'LSBoost', 200, 'Tree');

% Predict PM2.5 values for testing data
y_pred_boosting = predict(boosting_model, X_test);
```

Figure 5.13 : Development of MAE and Gradient Boosting

The training set taught models air quality data patterns and correlations, while the testing set impartially examined the algorithm’s ability to forecast new and unseen data. There were multiple factors that required this strong data separation method, it first reduced overfitting, when models over-adapted to training data and failed to apply their knowledge to fresh observations. After assessing the testing set, the study rectified overfitting flaws to ensure the predictive models could accurately anticipate real-world occurrences. The study compared ML algorithms and modelling approaches using data splitting and assessed model strengths and downsides using a similar testing set.

Future Predictions

Figure 5.14 below illustrates how the next four days were predicted from the current date

```
% Preducting next 5 days stored in future_data
% Here, we generate dummy future data for demonstration purposes
future_dates = datetime('today') + days(0:4)'; % Dates for the next 5 days
future_pm10 = rand(5, 1) * 100; % Dummy future PM10 data
future_so2 = rand(5, 1) * 20; % Dummy future SO2 data

% Create a table for future data
future_data = table(future_dates, future_pm10, future_so2, 'VariableNames', {'Date', 'pm10', 'so2'});

% Train SVM model for PM10 prediction
svm_model_pm10 = fitrsvm(X_train, cleaned_dataquality.pm10(1:train_size));

% Train SVM model for SO2 prediction
svm_model_so2 = fitrsvm(X_train, cleaned_dataquality.so2(1:train_size));

% Make predictions for the next 5 days
X_future = future_data(:, {'pm10', 'so2'}); % Adjust features for prediction
y_future_pred_pm10 = predict(svm_model_pm10, X_future); % Predict PM10
y_future_pred_so2 = predict(svm_model_so2, X_future); % Predict SO2

% Assuming svm_model is trained for PM2.5 prediction as in the previous code snippet
% Make predictions for PM2.5 values for the next 5 days
y_future_pred_pm25 = predict(svm_model, X_future); % Predict PM2.5
```

Figure 5.14: Predicting the Next 4 Days

The study created a full dataset that correctly predicted future air quality by integrating historical data with forecasted input variable values. The variable *future_data* dataset predicted PM2.5, PM10, and SO_2 concentrations during ML model training. The code builds a *future_data* database with PM10 and SO_2 levels and 4 days ahead from the current date. The study used expected future input variable values to generate the future datasets. The code shows how SVM forecasts PM10 and SO_2 values. It also predicts PM2.5 values using the pre-trained SVM model. It also predicts air quality over the next day's using *y_future_pred_pm25*, PM10, and SO_2. Combining the predictions with the *future_data* database created the *future_data_with_predictions* table as shown in Figure 5.15

Date	pm10	so2	pm25_pred	pm10_pred	so2_pred
05-Sep-2024	42.57	10.478	72.977	46.115	10.262
06-Sep-2024	24.767	13.883	64.811	28.783	13.553
07-Sep-2024	82.929	5.6661	95.331	85.416	5.6106
08-Sep-2024	74.194	17.426	104.66	76.939	16.979
09-Sep-2024	26.559	16.157	69.093	30.535	15.752

Figure 5.15: Forecasted Air Pollution Values

Machine Learning: Model Evaluation

Four machine learning models: Support Vector Machine (SVM), Decision Tree Regression, Random Forest Regression, and Gradient Boosting Regression, were evaluated for their ability to predict air quality. The Mean Absolute Error (MAE) performance measure helped in evaluating each model's strengths and weaknesses and choosing the optimal one for this system. Figure 5.16 displays the MAE for each of the models.

Model	Mean_Absolute_Error
'SVM'	30.798
'Decision Tree'	36.31
'Random Forest'	41.511
'Gradient Boosting'	31.279

Figure 5.16: MAE for Each of the Models

Among the tested models, Support Vector Machine (SVM) regression achieves the lowest Mean Absolute Error (MAE) of 30.798, indicating it has the best predictive accuracy. Gradient Boosting Regression yields a close result with an MAE of 31.279. These results demonstrate that both SVM and Gradient Boosting Regression models effectively estimate air quality. In contrast, Decision Tree and Random Forest Regression models exhibit higher MAEs of 36.31 and 41.511, respectively, suggesting they perform worse in forecasting pollutant concentrations.

An SVM model manages complex, non-linear data interactions. SVM key benefit is transforming input data into a higher-dimensional space and selecting the optimum hyperplane to identify classes or forecast the continuous target variable. SVM takes over when linear models fail to describe relationships. Additionally, complex environmental, meteorological, and contaminant interactions make air quality estimates non-linear (Veltin et al., 2019). The SVM model's ability to represent complex data patterns helps it perform well in testing. SVM handle outliers and high-dimensional feature spaces well. SVM model comprehensibility is crucial for prediction models. Consideration of feature relevance and decision restrictions revealed predictor-target pollutant links.

Meanwhile, effective air quality prediction techniques include Gradient Boosting Regression as Gradient Boosting creates a strong prediction model from several weak Decision Trees learners (Rao, 2016). Gradient Boosting Regression automatically selects the most challenging training data observations for the forecast. New ensemble trees learn from boosting to fix previous defects. This may aid in air quality estimates, as data may include anomalies or complex patterns that no model can explain. In this study, GBR predicted better than SVM and others. Boosting helped the model represent complex, non- linear data interactions and improve predictions.

Moreover, Decision Tree Regression and Random Forest Regression fared well but were less accurate than SVM and Gradient Boosting Regression. This may be because these algorithms cannot capture the complex, non-linear correlations in air quality data. Decision trees can represent numerical and categorical variables, but they may overfit and fail to adapt to complicated patterns (Baraneetharan, 2020). Additionally, many Random Forest Regression decision trees can improve ensemble consistency. Decision Trees was unable to capture non-linear correlations in air quality data.

Gradient Boosting: The use of Gradient Boosting yielded the following results for the study:

1. **Robust to Noisy and Incomplete Data**

The nature of both datasets (secondary datasets from Pelonimi monitoring station and the limited Arduino-based wireless sensors collected data) suffered from missing data, outliers, and other noisy signals. Gradient Boosting was relatively robust in handling that kind of imperfect dataset.

2. **High Predictive Accuracy**

Gradient Boosting models were built sequentially, where each new model corrected the errors of the previous model. This was done for validation purposes.

This led to very strong predictive performance, making it effective for forecasting pollutant concentrations.

3. **Resistant to Overfitting**

Through regularization techniques such as subsampling, learning rate, and depth, Gradient Boosting balances bias and variance, making it reliable for continuous monitoring. Hence, the models could still predict and perform exceptionally well with new, unknown data (Pelonomi-NAQI from April 2024 to September 2024) that was later introduced into the learning environment.

4. **Enables short- and long-term prediction**

It was effective for real-time monitoring and for forecasting future air quality trends based on historical data.

In contrast to the above advantages, training was computationally intensive and required very large datasets. Gradient Boosting also requires careful hyperparameter tuning to prevent model overfitting. Table 5.1 below demonstrates the comparison between ML models that were used.

Table 5.1: Comparison of ML Models used in the Study

Model	Advantages	Limitations
Gradient Boosting	High prediction accuracy for pollutant concentration and Air Quality Index. Handled nonlinear, complex relationships well. Performed well with inconsistent and imperfect datasets. Works well with both short-term real- time monitoring and long-term forecasting.	Computationally expensive for training very large datasets Required careful hyperparameter tuning to avoid overfitting.

Random Forest	Easy to implement and interpret. Robust against overfitting compared to single decision trees. Performed well with high-dimensional datasets.	Less accurate predictions than Gradient Boosting in complex, nonlinear pollution patterns. Tends to be slower for real-time applications due to many trees. Provides feature importance, but sometimes less precise than Gradient Boosting.
Support Vector Machine	Suitable for smaller datasets with clear class separations Effective in high-dimensional spaces.	Struggles with large-scale air pollution datasets (computationally expensive). Less interpretable compared to Gradient Boosting and Random Forest. Requires careful kernel selection for nonlinear relationships.

In terms of performance, both Gradient Boosting Regression and SVM models effectively forecast air quality. Both models demonstrate high predictive accuracy, however, Gradient Boosting Regression has a lower MAE. Gradient Boosting Regression improves predictions by assigning higher weights to mis-predicted observations and capturing complex, non-linear relationships in the data. Meanwhile, SVM is well-suited for non-linear, high-dimensional datasets related to air quality.

Additionally, interpretable relationships are shown using SVM. Veltin et al., (2019) says that: when choosing a model for implementation, researchers or data scientists must go beyond correctness. This comprises interpretability, computational resources, and user's needs, or the project scope. To build a reliable air quality forecasting system that improves environmental management and data-driven decision-making, researchers must thoroughly assess these factors and understand each algorithm's pros and cons as they all differ.

5.5. FCM: Indigenous Knowledge Findings

This part of the IK analysis is designated for Mental Modeler - Fuzzy Logic Cognitive Mapping. Fuzzy Cognitive Maps (FCM) were used to collect, verify, and validate indigenous knowledge indicators.

FCM is a graphic technique that describes causal understanding of a wide range of applications. This practice reviews the lived experiences of IK experts. Fuzzy cognitive maps (FCMs) have gained considerable research interest due to their ability TO represent structured knowledge and model complex systems in various fields. Gray et al., (2013) affirm that FCM promotes a comprehensive view of the mind. While traditional techniques may investigate cognitive modules in isolation, FCMs stress the interconnection of these modules (Lusilao-Makiese et al., 2013). This comprehensive perspective can provide deeper insights into how cognitive systems interact to produce complicated behaviours (Gray et al., 2012). Furthermore, FCMs are a useful computational tool for modeling complex systems and processes. They integrate characteristics of neural networks and fuzzy logic, making them particularly ideal for addressing issues with substantial uncertainty and imprecision (Gray et al., 2012). One of the fascinating applications of FCMs is to comprehend and map mental modularity, which posits that the human mind is composed of distinct cognitive functions. Relationships between components are represented using a range of numbers from +1 to -1, where +1 indicates a strong positive relationship and -1 indicates a strong negative relationship.

The FCM system consists of various components and connections. The boxes, commonly referred to as factors or concepts can represent anything expressed as a variable. Importantly, it does not have to be quantifiable or have data behind it, which means it can be used to capture knowledge without empirical data (Gray et al., 2012).

The primary objective of integrating IK with sensor data was to examine the correlation between climatological data collected using technology (IoT devices) and information that evolved over time through lived experiences (IK). In the distributed questionnaire, some responses served as IK indicators to verify whether people's observations align with sensor data. From the analysis, 12 IK indicators were extracted, which are represented as nodes in the FCM models. Three driver components were developed, and the relationships between driver and receiver components are represented with arrows. The strength of these relationships varies across different scenarios. The modeling, representation, and development of FCMs for this study are presented below, along with figures and corresponding explanations of the analysis process.

Figure 5.17 is an FCM model representing indicators for a polluted day. All the positive (+) relationships between the driver component and pollution indicators are what people would see and feel when it is polluted.

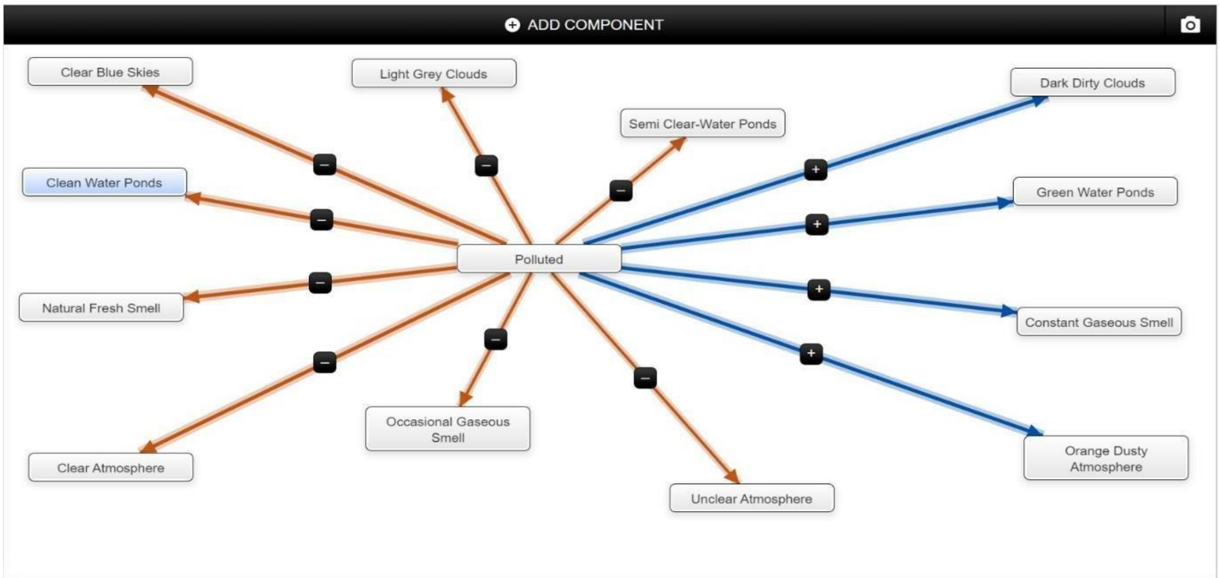


Figure 5.17: FCM Model of Polluted Indicators

Figure 5.18 illustrates an FCM scenario for a polluted day, showing the relevant indicators. The scenario describes an inverse proportional relationship between components that indicate polluted air conditions and those that indicate moderate or non-polluted air conditions.

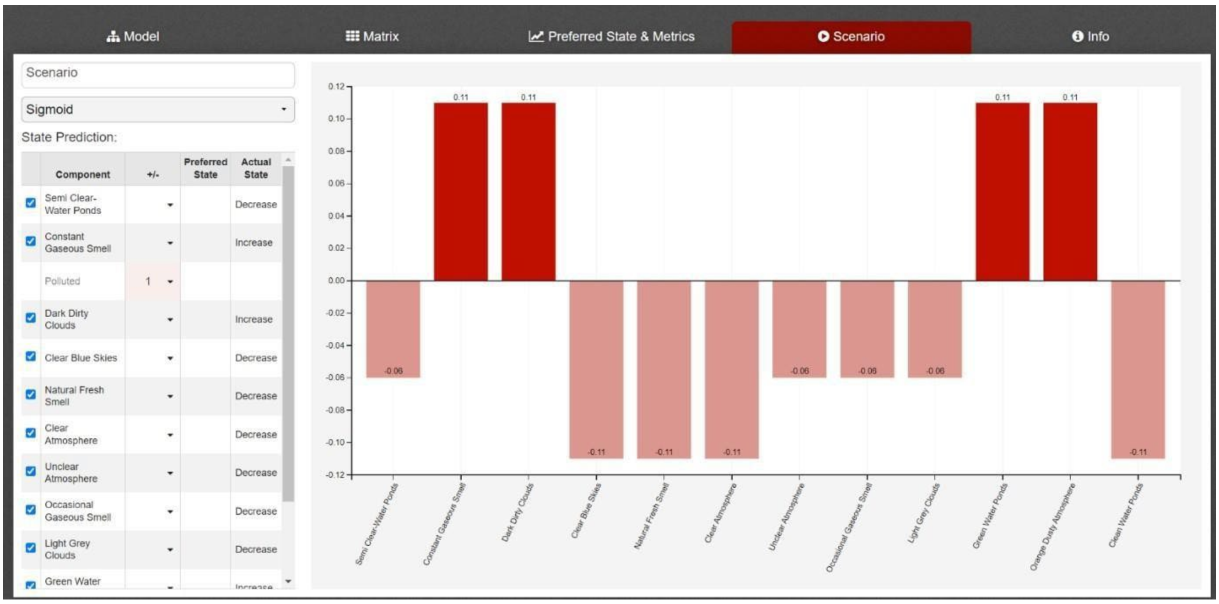


Figure 5.18: FCM Scenario of Polluted Indicators

Figure 5.19 is an FCM preferred state and metrics of a polluted day. It is the presentation of the total number of components, total connections, density percentage, number of connections per component, type of component into the relationship, and their numbers, as well as the total complexity score of the presented scenario.

Model | Matrix | Preferred State & Metrics | Scenario | Info

Total Components	13
Total Connections	12
Density	0.0769230769
Connections per Component	0.9230769231
Number of Driver Components	1
Number of Receiver Components	12
Number of Ordinary Components	0
Complexity Score	12

Component	Indegree	Outdegree	Centrality	Preferred State	Type
Semi Clear-Water Ponds	0.5	0	0.5		receiver
Constant Gaseous Smell	1	0	1		receiver
Polluted	0	10.04	10.04		driver
Dark Dirty Clouds	1	0	1		receiver
Clear Blue Skies	1	0	1		receiver
Natural Fresh Smell	1	0	1		receiver
Clear Atmosphere	1	0	1		receiver
Unclear Atmosphere	0.51	0	0.51		receiver
Occasional Gaseous Smell	0.53	0	0.53		receiver
Light Grey Clouds	0.5	0	0.5		receiver
Green Water Ponds	1	0	1		receiver
Orange Dusty Atmosphere	1	0	1		receiver
Clean Water Ponds	1	0	1		receiver

Figure 5.19: FCM Preferred State and Metrics of Polluted Indicators

Figure 5.20 is an FCM matrix, demonstrating a precise value ranging from +1 to -1, indicating the strength of the relationship between the driver and receiver components. A polluted day matrix, as the driver component has four +1 connections with the receiver components, confirming polluted air conditions, and four -1 connections, confirming non- polluted conditions. And four -0,5 connections that confirm moderate air conditions.

Model | Matrix | Preferred State & Metrics | Scenario | Info

	Semi Clear-Water Ponds	Constant Gaseous Smell	Polluted	Dark Dirty Clouds	Clear Blue Skies	Natural Fresh Smell	Clear Atmosphere	Unclear Atmosphere	Occasional Gaseous Smell	Light Grey Clouds	Green Water Ponds	Orange Dusty Atmosphere	Clean Water Ponds
Semi Clear-Water Ponds													
Constant Gaseous Smell													
Polluted	-0.5	1		1	-1	-1	-1	-0.51	-0.53	-0.5	1	1	-1
Dark Dirty Clouds													
Clear Blue Skies													
Natural Fresh Smell													
Clear Atmosphere													
Unclear Atmosphere													
Occasional Gaseous Smell													
Light Grey Clouds													
Green Water Ponds													
Orange Dusty Atmosphere													
Clean Water Ponds													

Figure 5.20: FCM Matrix of Polluted Indicators

Figure 5.21 is an FCM model representing indicators for a not polluted day. All the positive (+) relationships between the driver component and pollution indicators are what people would see and feel when it is not polluted.

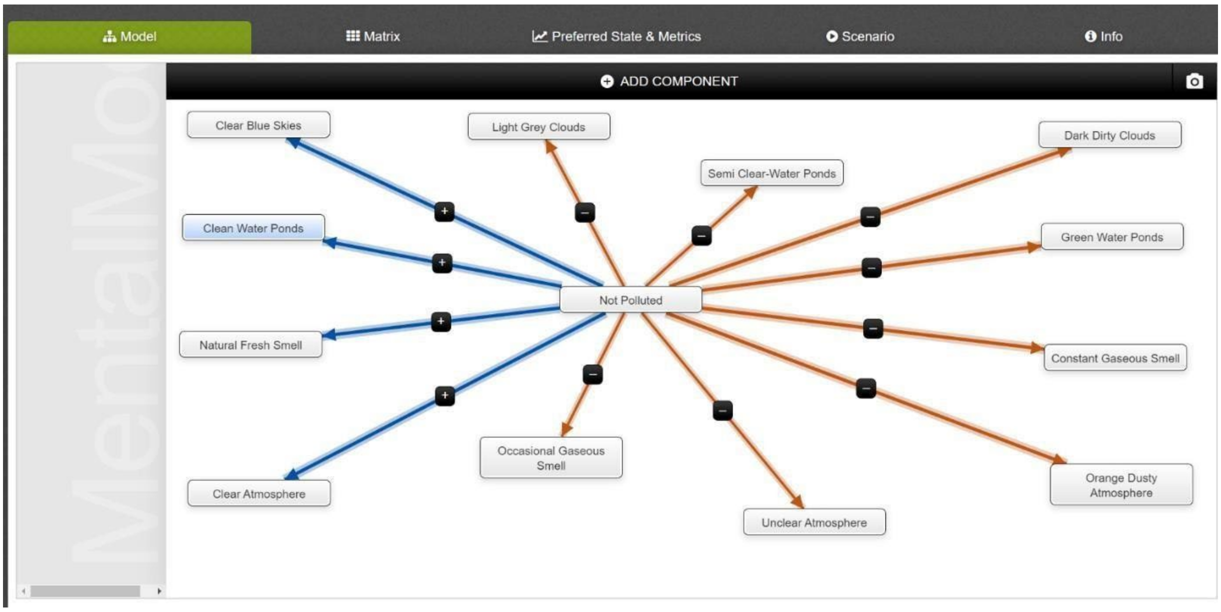


Figure 5.21: FCM Model of Not Polluted Indicators

Figure 5.22 is an FCM scenario, representing indicators for a *Not polluted* day. The scenarios show an inversely proportional relationship between components that confirm Not polluted air conditions to components that confirm Moderate and Polluted air conditions.

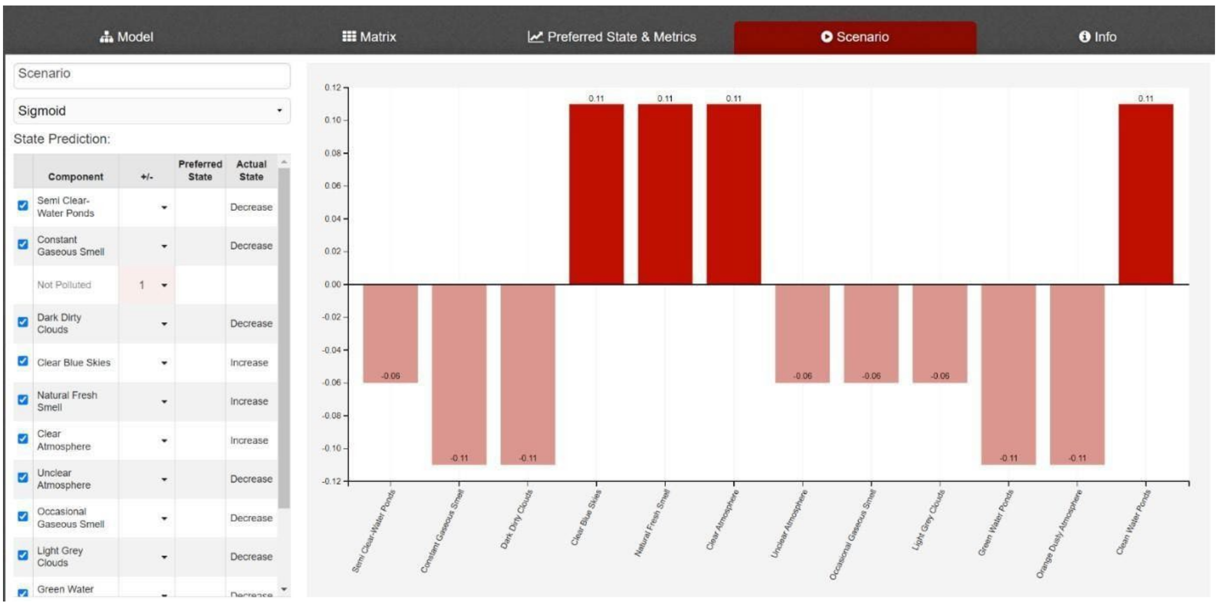
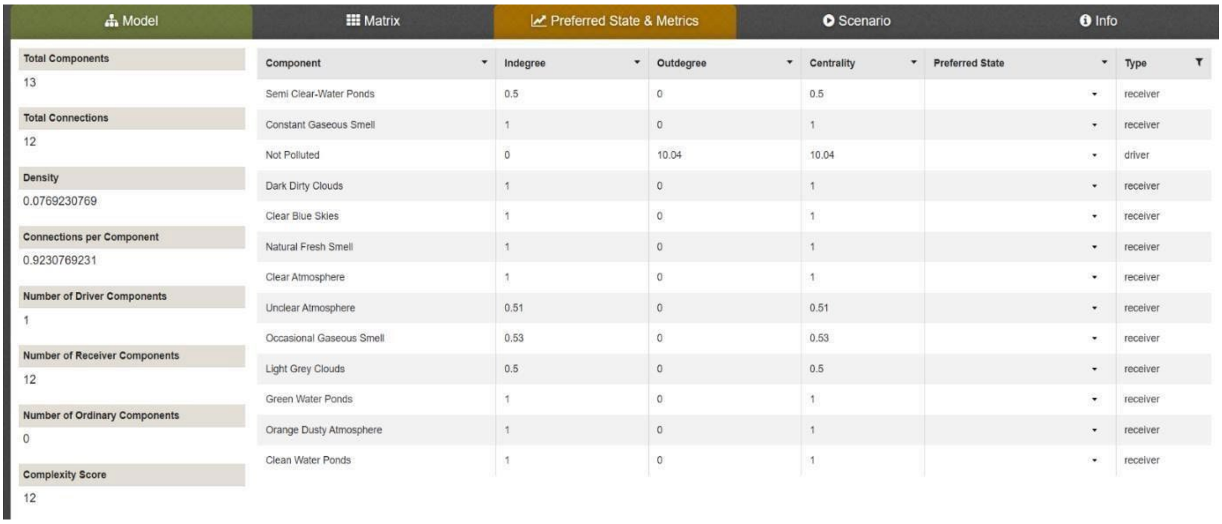


Figure 5.22: FCM Scenario of Not Polluted Indicators

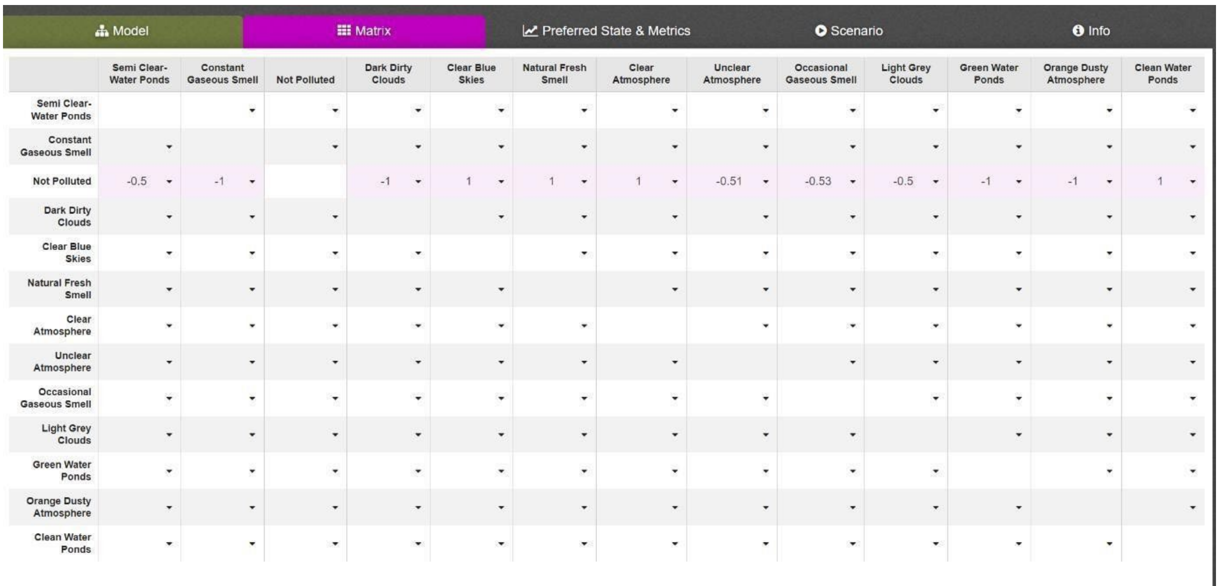
Figure 5.23 is an FCM Preferred state and metrics of a *Not polluted*, which is the presentation of the total number of components, total connections, density percentage, number of connections per component, type of component into the relationship and their numbers, as well as the total complexity score of the presented scenario.



Component	Indegree	Outdegree	Centrality	Preferred State	Type
Semi Clear-Water Ponds	0.5	0	0.5		receiver
Constant Gaseous Smell	1	0	1		receiver
Not Polluted	0	10.04	10.04		driver
Dark Dirty Clouds	1	0	1		receiver
Clear Blue Skies	1	0	1		receiver
Natural Fresh Smell	1	0	1		receiver
Clear Atmosphere	1	0	1		receiver
Unclear Atmosphere	0.51	0	0.51		receiver
Occasional Gaseous Smell	0.53	0	0.53		receiver
Light Grey Clouds	0.5	0	0.5		receiver
Green Water Ponds	1	0	1		receiver
Orange Dusty Atmosphere	1	0	1		receiver
Clean Water Ponds	1	0	1		receiver

Figure 5.23: Preferred State and Metrics of a Not Polluted Indicators

Figure 5.24 is an FCM matrix demonstrating *Not polluted* day, demonstrating a precise value ranging from +1 to -1 on relationship strength between the driver and receiver components. *Not polluted* as the driver component has four +1 connections with the receiver components that confirm its *Not polluted*. Also, four -1 connections confirm Polluted, and four -0.5 connections confirms Moderate air conditions.



	Semi Clear-Water Ponds	Constant Gaseous Smell	Not Polluted	Dark Dirty Clouds	Clear Blue Skies	Natural Fresh Smell	Clear Atmosphere	Unclear Atmosphere	Occasional Gaseous Smell	Light Grey Clouds	Green Water Ponds	Orange Dusty Atmosphere	Clean Water Ponds
Semi Clear-Water Ponds													
Constant Gaseous Smell													
Not Polluted	-0.5	-1		-1	1	1	1	-0.51	-0.53	-0.5	-1	-1	1
Dark Dirty Clouds													
Clear Blue Skies													
Natural Fresh Smell													
Clear Atmosphere													
Unclear Atmosphere													
Occasional Gaseous Smell													
Light Grey Clouds													
Green Water Ponds													
Orange Dusty Atmosphere													
Clean Water Ponds													

Figure 5.24 : FCM Matrix of Not Polluted Indicators

In Figure 5.25, an FCM model represents indicators for a Moderately polluted day. All the positive (+) relationship between the driver component and pollution indicators is what people would see and feel when it is Moderate.

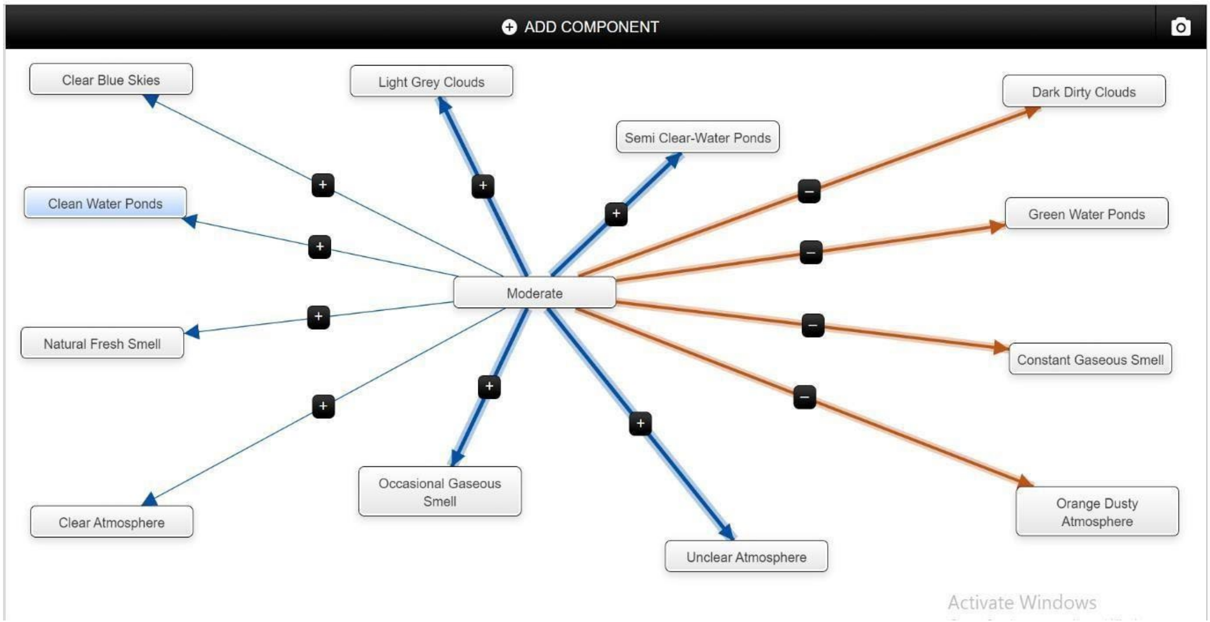


Figure 5.25: FCM Model of a Moderate Indicators

Figure 5.26 is an FCM scenario, representing indicators for a moderate day. The scenarios show an inverse proportional relationship between components that indicate moderately polluted air conditions and components that indicate polluted air conditions, and a direct proportional relationship with indicators that indicate not polluted, even though it is not a strong strength.

Figure 5.26: FCM Scenario of Moderate Indicators

Figure 5.27 is an FCM preferred state and metrics of a *Moderately polluted* day. This is the presentation of the total number of components, total connections, density percentage, number of connections per component, type of component into the relationship and their numbers, as well as the total complexity score of the presented scenario.

Total Components	13
Total Connections	12
Density	0.0769230769
Connections per Component	0.9230769231
Number of Driver Components	1
Number of Receiver Components	12
Number of Ordinary Components	0
Complexity Score	12

Component	Indegree	Outdegree	Centrality	Preferred State	Type
Semi Clear-Water Ponds	1	0	1		receiver
Constant Gaseous Smell	0.54	0	0.54		receiver
Moderate	0	6.71	6.71		driver
Dark Dirty Clouds	0.55	0	0.55		receiver
Clear Blue Skies	0.11	0	0.11		receiver
Natural Fresh Smell	0.09	0	0.09		receiver
Clear Atmosphere	0.09	0	0.09		receiver
Unclear Atmosphere	1	0	1		receiver
Occasional Gaseous Smell	1	0	1		receiver
Light Grey Clouds	1	0	1		receiver
Green Water Ponds	0.51	0	0.51		receiver
Orange Dusty Atmosphere	0.66	0	0.66		receiver
Clean Water Ponds	0.16	0	0.16		receiver

Figure 5.27: Preferred State and Metrics of Moderate Indicators

Figure 5.28 is a FCM matrix demonstrating a moderately air pollution day, with a precise relationship strength value ranging from +1 to -1 between the driver and receiver components. The moderate driver component has four +1 connections with the receiver components that confirm Moderately polluted day, four -1 connections that confirm polluted and four -0,5 connections that confirm *Not polluted* air conditions.

	Semi Clear-Water Ponds	Constant Gaseous Smell	Moderate	Dark Dirty Clouds	Clear Blue Skies	Natural Fresh Smell	Clear Atmosphere	Unclear Atmosphere	Occasional Gaseous Smell	Light Grey Clouds	Green Water Ponds	Orange Dusty Atmosphere	Clean Water Ponds
Semi Clear-Water Ponds													
Constant Gaseous Smell													
Moderate	1	-0.54		-0.55	0.11	0.09	0.09	1	1	1	-0.51	-0.66	0.16
Dark Dirty Clouds													
Clear Blue Skies													
Natural Fresh Smell													
Clear Atmosphere													
Unclear Atmosphere													
Occasional Gaseous Smell													
Light Grey Clouds													
Green Water Ponds													
Orange Dusty Atmosphere													
Clean Water Ponds													

Figure 5.28: FCM Matrix of Moderate Indicators

Throughout the analysis of IK using FCM, the Moderate air pollution driver component had neutral connections, representing a state that was neither polluted nor non-polluted. This represents a grey area that may not have been clearly defined. FCM made it easier to highlight such connections through relationship strengths ranging from -0.5 to +0.5. The relationship between polluted and non-polluted conditions was explicitly represented by values of either -1 or +1, with multiple values illustrating this contrast. Clear blue skies versus dark, polluted skies reflect the real-world perception of pollution levels. Mental Modeler also provided a robust modeling approach that enabled us to capture and standardize responses for scenario analysis.

5.6. IK Validation and Cross-Comparison

Ground Station Pelonomi-NAQI, data that was collected over 8 months (Period: 01/04/2024 00:01 - 31/10/2024 00:00). Reveals a gradual increase in pollution throughout the Winter season through to early spring. Windy seasons experience increased pollution levels, where gaseous air blows towards the residential homes.

Zooming in on Figure 5.29, the minimum values were moderate, indicating conducive conditions. However, from the average values, there is a noticeable spike across all three pollutants. All the pollutants' maximum readings are significantly above the threshold limit, indicating very polluted conditions. Based on the NAAQS, they are deemed to be highly contaminated.

Station Name: Pelonomi-NAQI, TimeBase: 24 Hour,			
Type	**Pelonomi-NAQI**		
	SO2 ppb	PM2.5 µg/m3	PM10 µg/m3
Std	1.9	44.4	71.1
Avg	2,413	56,546	110,871
Num	177	213	213
Maximum	9,843	193,662	280,716
Max Date	29/06/2024	29/06/2024	29/05/2024
Min Date	09/07/2024	30/10/2024	09/04/2024
Minimum	0	3,518	8,426

Figure 5.29: Pelonomi-NAQI (April 2024- September 2024)

Cross-comparing the Pelonomi-NAQI 24-hour time-based average and maximum readings with the IK indicators collected over that period confirms a robust positive correlation. The figure 5.20

below displays a WordArt of the IK indicators that were collected around the same time.

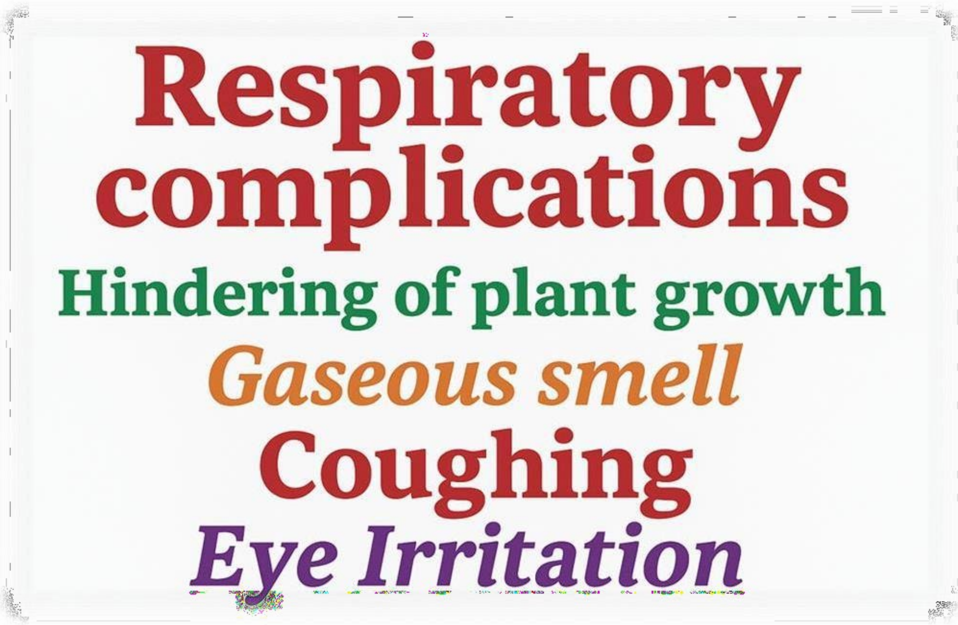


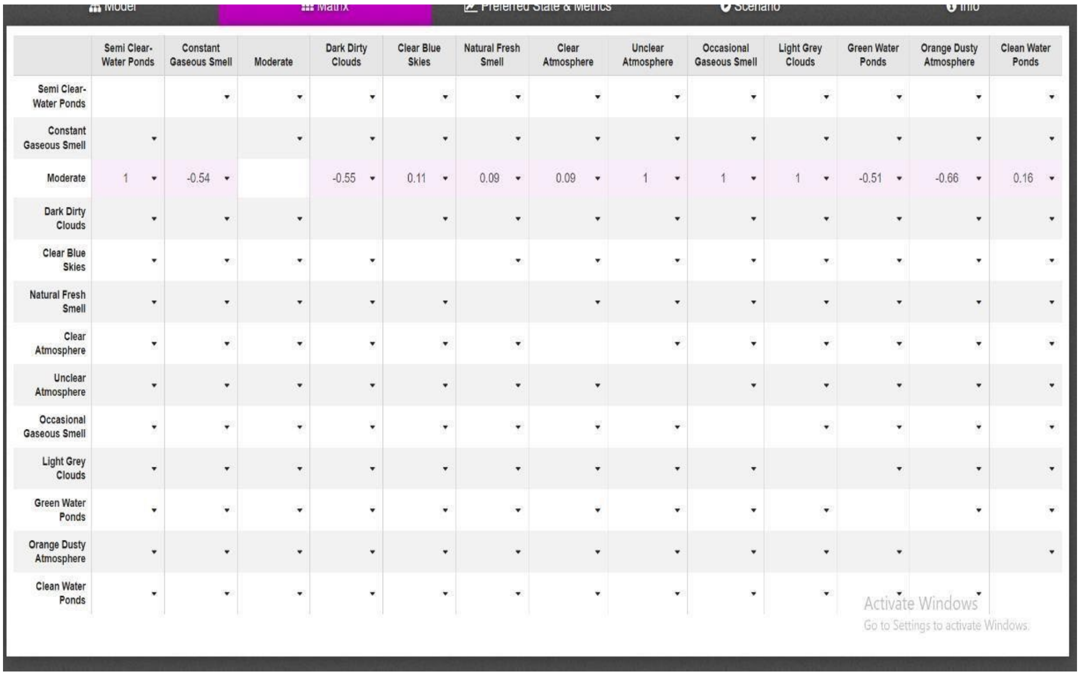
Figure 5.30: IK Cross-Comparing with Pelonomi - NAQI

Strengths and limitations of FCM in representing “grey areas” (moderate pollution)

In a physical setting, people can only confirm whether the air is polluted or not through their local knowledge indicators; it becomes difficult to confirm moderate air pollution conditions, which is the grey area that FCM fully addresses. Throughout the analysis of IK using FCM, the moderate air pollution driver component exhibited neutral connections, indicating neither pollution nor lack of pollution. That is the grey area, which could not have been clearly demonstrated. FCM made it easy to highlight such a connection by relationship strength, ranging from -0.5 to +0.5. The relationship between polluted conditions and unpolluted conditions was explicit, as shown by numerical values of either
-1 or +1, and by clear blue skies versus dark, dirty skies, which is the reality and what people perceive as pollution level indicators. In addition, Mental Modeler offered a robust modelling technique that allowed the study to model and capture responses in a standardized format for scenario analysis.

The relationship between driver and receiver components, their strengths, and the indicators correlating with the sensor datasets were demonstrated using graphs, WordArt, matrices, preferred states, metrics, and scenario analysis.
Figure 5.31 presents an FCM matrix depicting a moderately polluted air quality day. It demonstrates precise relationships between the driver and receiver components, ranging from +1

to -1. The moderate pollution level is indicated by the driver component having four (+1) connections with receiver components, confirming a moderately polluted day, four (-1) connections, indicating polluted conditions, and four (-0,5) connections, representing non-polluted air conditions.



	Semi Clear-Water Ponds	Constant Gaseous Smell	Moderate	Dark Dirty Clouds	Clear Blue Skies	Natural Fresh Smell	Clear Atmosphere	Unclear Atmosphere	Occasional Gaseous Smell	Light Grey Clouds	Green Water Ponds	Orange Dusty Atmosphere	Clean Water Ponds
Semi Clear-Water Ponds													
Constant Gaseous Smell													
Moderate	1	-0.54		-0.55	0.11	0.09	0.09	1	1	1	-0.51	-0.66	0.16
Dark Dirty Clouds													
Clear Blue Skies													
Natural Fresh Smell													
Clear Atmosphere													
Unclear Atmosphere													
Occasional Gaseous Smell													
Light Grey Clouds													
Green Water Ponds													
Orange Dusty Atmosphere													
Clean Water Ponds													

Figure 5.31: FCM Matrix of Moderate Indicators

5.7. Results from the Communication and Dissemination Layer

5.7.1. Mobile Application

Step-by-step description of the mobile Android application

Registration: Registration of actors or users serves as the entry point into the system, and it requires the user's email address as the primary key for the authentication process. The system consists of three actors: the system administrator, the IK expert, and the local community member. User roles and access levels vary depending on the type of registration under which a user is registered.

The systems can register and create records for three types of leading actors:

1. **System Administrator:** Systems administrators have full access to the system, including the system's backend, Android application testing, database administration, and user management.

2. **IK Expert**: This user has access to all the system functionalities, including the ability to log extreme events that occur in their surroundings while using the system.

3. **Local community member**: These are general users with view-only rights. They have access

to all Android application functionalities in a read-only capacity.

Upon successful sign-up, an email verification is sent to the corresponding mailbox for account verification. Once the email is verified, the user can explore all the available and relevant features based on their account types, as shown in Figure 5.32.

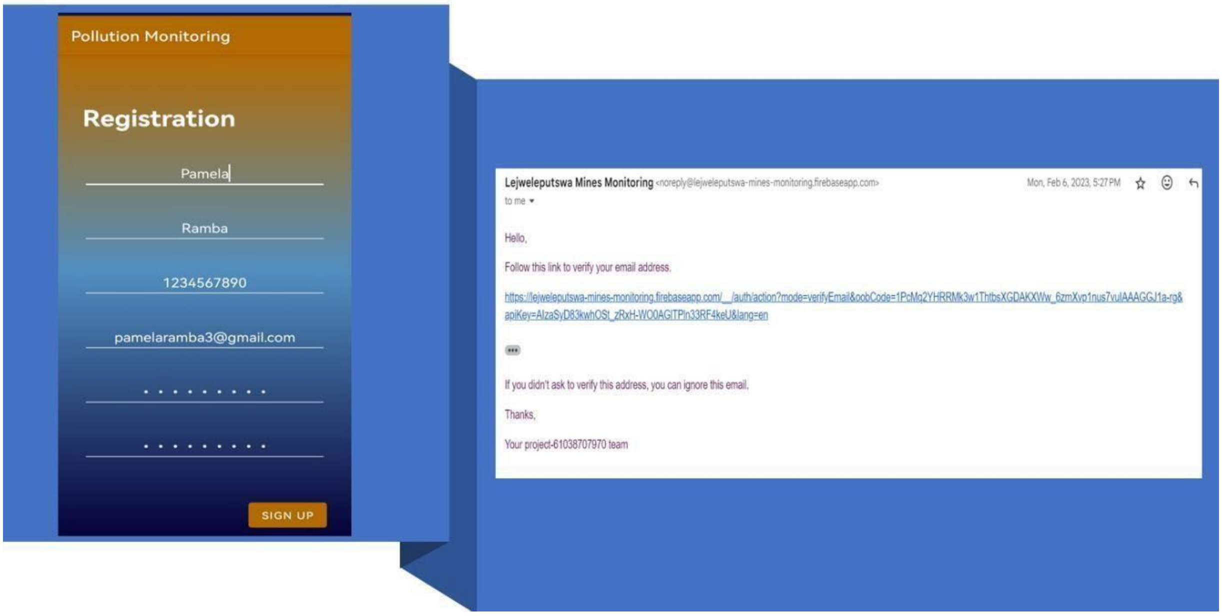


Figure 5.32: Sign Up and Email Verification

Login: To gain access to the Android application, users must log in using a verified email address and password. After an authenticated login, the splash screen is displayed while loading a default landing page of the Android application, as shown in Figure 5.33 below.

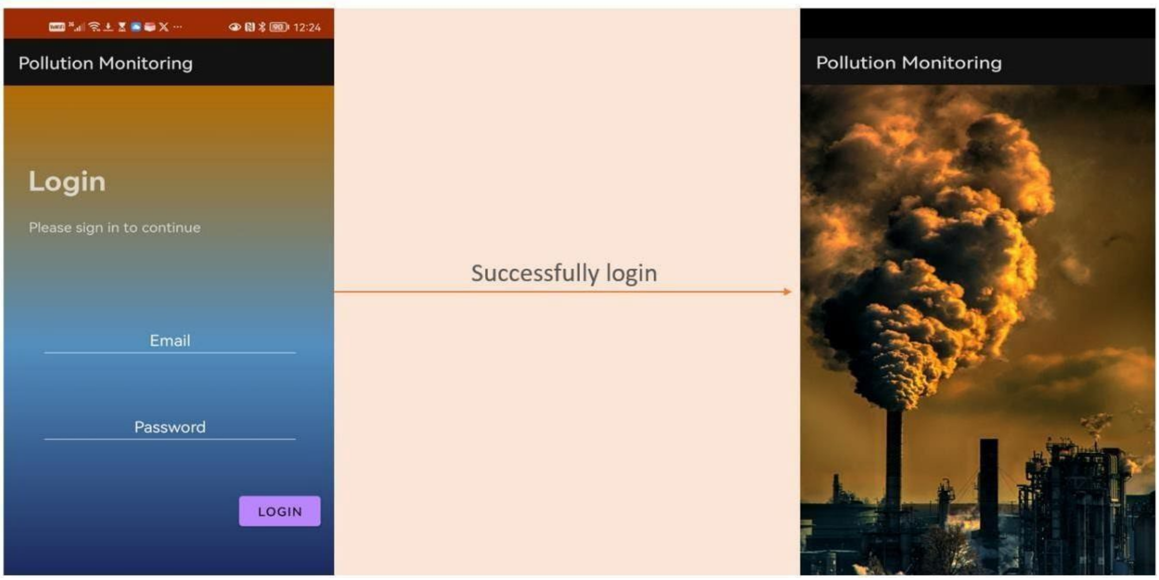


Figure 5.33: Login Screen

View current-day pollution levels: The default landing page has the current- day air pollution predictions, NAAQS threshold limits, and various buttons to navigate to different pages of the Android application. NAAQS threshold limits served as guidelines to determine whether an area was polluted or not, as shown in Figure 5.34

Figure 5.34: Default landing page.

View Extreme Events: This is part of the Android application where system users get to review all the previously logged extreme events. All the users have access to extreme events. This is shown in Figure 5.35.

Figure 5.35: View Previously Logged Extreme Events

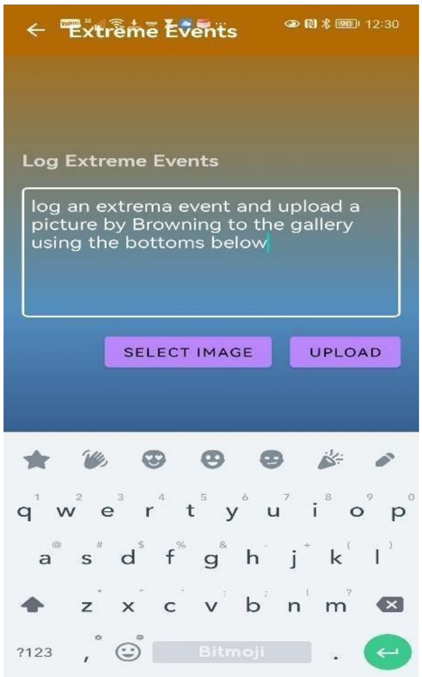


Figure 5.36: System Administrators

Upload Extreme Event: Only IK experts have access to log in and see an extreme event happening around them. The system administrators can also log in for testing purposes. This is shown in Figure 5.36.

Query air pollution per date: This is the part of the Android application where users can query a specific date as per their needs or preferences, such as looking at the previous day's air pollution predictions. All authenticated users have access to this part of the application, as shown in Figure 5.37.

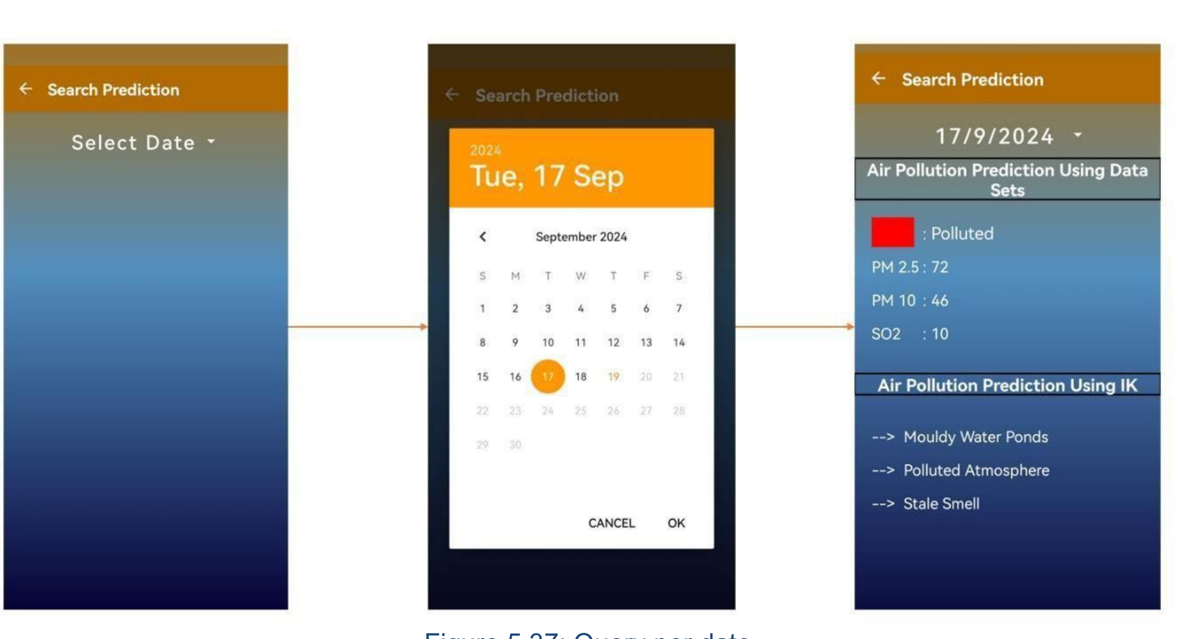


Figure 5.37: Query per date

Integrated system: This part of the system displays an integrated system, where both air pollution predictions agree with the IK indicators. The system was dissected into three parts: Polluted, Not Polluted, and Moderate. The first segment of the screen displays predicted air pollution values, and the last part shows the IK indicators, as illustrated in Figure 5.38.

Figure 5.38: Integrated IK indicators with air pollution predictions

In Figure 5.38 above, three states of air pollution are presented. Polluted air conditions represent air pollution levels that are above the threshold limit. These levels of pollution are considered unhealthy for human health in general. Not Polluted air conditions follow this. These conditions are deemed safe and represent a normal state of the air. The last one is a moderately polluted state of the air. These are said to be unhealthy and a threat to a particular group of people, such as elderly people, small babies, as well as a group of people who might be living with a particular underlying medical condition. All three 3 segments of the Android application displaying the air pollution levels were guided by and measured against the NAAQS. Threshold limits were established and validated in accordance with NAAQS.

The entire Android application was developed throughout the study period, and it evolved and progressed as the system advanced. All the proposed working components were delivered and exhibited in the interface. Additionally, testing of the application was conducted concurrently with the system's development and advancement.

5.8. System Usability and User Feedback

Another important question that needed to be addressed was whether the entire system effectively tackled pollution. The figure below demonstrates that 89% of the responses were 'Yes', while 11 % of the responses were 'No', as displayed in Figure 5.39. Users evaluated the overall system's performance through a questionnaire designed specifically for feedback purposes. Testing and evaluation were conducted to verify whether all the proposed objectives were met and whether the entire system achieved its own objective.

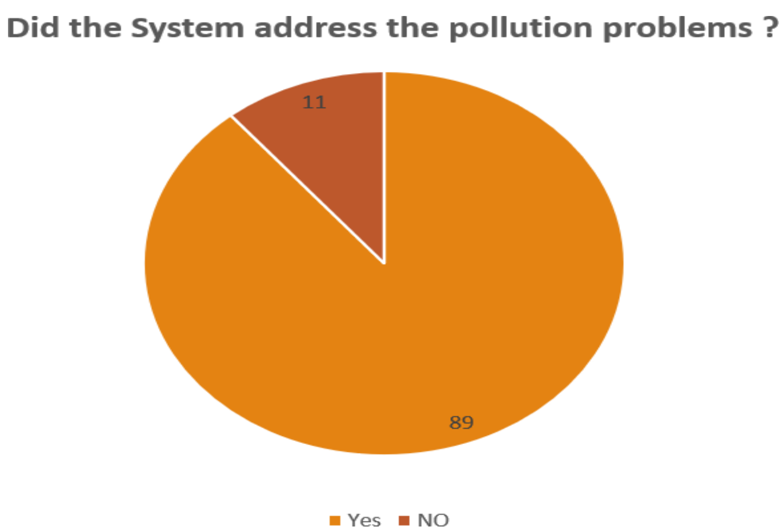


Figure 5.39 : Pollution problem addressed

5.9. Chapter Conclusion

This chapter presented and discussed the results of the hybrid air pollution monitoring framework, organised across its layered architecture. At the **data collection layer**, inputs were drawn from three complementary sources: wireless Arduino-based sensors (functionally tested though spatially constrained), long-term secondary datasets from the Pelonomi ground station, and Indigenous Knowledge (IK) indicators captured through community surveys and questionnaires.

At the **monitoring and prediction layer**, supervised machine learning algorithms, including Support Vector Machine, Random Forest, Gradient Boosting, and Decision Tree, were trained on the ground station data to forecast air quality levels up to four days in advance. The integration of IK indicators through Fuzzy Cognitive Maps added context- specific validation, bridging scientific predictions with lived community insights.

At the **communication and dissemination layer**, a prototype Android mobile application was developed to serve as the interface between the system and its end-users. This platform successfully disseminated forecasts, early warnings, and community-driven observations, thereby demonstrating practical usability.

Overall, the results confirmed the **proof-of-concept**: a layered, adaptive framework that combines 4IR technologies with IK can generate scientifically rigorous predictions while remaining socially relevant and accessible to semi-literate communities. The pilot findings underscore both the novelty and the practical value of the system for Free State mining communities, mine workers, and policymakers.

CHAPTER SIX: EVALUATION, DISCUSSION AND CONCLUSION

6.1. Introduction

This chapter presents the evaluation and discussion of the developed smart and adaptive air pollution monitoring framework. Building on the design and implementation phases outlined in the preceding chapters, the evaluation focuses on how effectively the system met its stated research objectives. The discussion extends beyond the technical assessment to examine the broader contributions of the study, both theoretically and practically, within the context of air quality management in the Free State Province.

The chapter is structured into five main parts. First, the system is evaluated against each of the research objectives, demonstrating the extent to which they were achieved. This is followed by an integrated discussion that situates the findings within wider scientific and applied debates, highlighting the novelty of bridging Fourth Industrial Revolution (4IR) technologies with Indigenous Knowledge (IK). The chapter then outlines the limitations encountered during the research process, proposes recommendations for improving and scaling the framework, and identifies avenues for future work.

Through this layered evaluation, the chapter provides a critical reflection on the strengths and weaknesses of the hybrid framework, while emphasizing its proof-of-concept value as a tool for adaptive, context-sensitive environmental monitoring in South Africa.

6.2. Evaluation Against Research Objectives

The evaluation of the developed framework was conducted against the study's stated objectives to determine whether they were achieved and to assess the overall performance of the system. The evaluation criteria focused on the integration of technologies and knowledge systems, the accuracy of machine learning (ML) predictions, the incorporation of Indigenous Knowledge (IK), and the usability of the mobile application.

Objective 1: To develop a hybrid framework integrating 4IR technologies and Indigenous Knowledge for air pollution monitoring

This objective was successfully achieved. A layered framework was designed and implemented, combining IoT sensors, secondary ground station data, ML models, and IK indicators into a single adaptive system. The modular architecture allowed each layer, data collection, monitoring and prediction, and communication and dissemination, to function independently while contributing to the integrated framework. The hybrid approach demonstrated the feasibility of merging scientific and community-based knowledge for pollution monitoring in a resource-constrained context.

Objective 2: To build and evaluate ML-based models for predicting air quality levels up to four days in advance

This objective was partially achieved. Using supervised ML algorithms, Support Vector Machines, Random Forest, Gradient Boosting, and Decision Tree Regression, the system generated forecasts of particulate matter ($PM_{2.5}$, PM_{10}) and SO_2 levels up to four days in advance. Predictions aligned with the National Ambient Air Quality Standards (NAAQS) thresholds, offering actionable early warnings. While the models demonstrated reasonable accuracy, their performance was constrained by reliance on secondary data from the Pelonomi monitoring station and limited primary sensor inputs. Nevertheless, the results confirmed the potential of ML methods to enhance anticipatory capacity in air pollution management.

Objective 3: To systematically collect and integrate IK indicators of air pollution into the monitoring system

This objective was fully achieved. IK indicators were collected through surveys and questionnaires with mine workers and local communities in the Lejweleputswa district. Using Fuzzy Cognitive Maps (FCMs) and the Mental Modeler tool, these indicators were formally modelled and integrated into the monitoring framework. The process captured lived experiences such as visible dust clouds, unusual odours, and environmental changes, thereby contextualizing scientific predictions. The inclusion of IK enhanced the adaptability and cultural relevance of the framework, making it more acceptable to communities.

Objective 4: To design a mobile application for communicating forecasts and facilitating community reporting

This objective was achieved at a proof-of-concept level. An Android mobile application was developed to disseminate pollution forecasts, provide early warnings, and enable communities to log observations based on IK indicators. Pilot testing confirmed that the app improved accessibility and offered a two-way communication platform between communities and the monitoring system. However, large-scale usability testing was limited, and broader deployment remains a future step.

Objective 5: To evaluate the framework's applicability within the Free State Province context

This objective was substantially achieved. The framework was tested using Welkom in the Lejweleputswa district as a case study, supported by data from Mangaung's Pelonomi station. The system demonstrated that combining secondary datasets, testing IoT sensors, and IK indicators could generate a functional and adaptive prototype suited to the provincial context. While access restrictions in mining shafts limited continuous primary data collection, the evaluation confirmed that the framework is viable and relevant to Free State communities, policymakers, and industry stakeholders.

6.3. Integrated Discussion

The findings of this study demonstrate the feasibility and potential of a hybrid framework that integrates Fourth Industrial Revolution (4IR) technologies with Indigenous Knowledge (IK) to monitor and predict air pollution. By combining supervised machine learning algorithms, IoT-enabled sensors, and community-derived indicators, the framework responds to both the technological and socio-cultural dimensions of environmental monitoring.

6.4. Novelty of the Framework

A major novelty lies in bridging the divide between modern scientific techniques and IK systems. Whereas most air quality monitoring frameworks rely exclusively on high-resolution sensors and computational models, this study demonstrates that IK, captured through community surveys and fuzzy cognitive mapping, can complement scientific approaches. The integration provides a richer, more context-sensitive view of air quality challenges in the Free State, particularly in mining-intensive regions such as Welkom. This dual lens is especially important in environments where access to real-time scientific data is limited.

6.5. Theoretical Implications

From a theoretical perspective, the study advances the discourse on adaptive monitoring systems in resource-constrained contexts. It demonstrates that hybrid approaches, underpinned by pragmatism and constructivism, offer a viable pathway for strengthening environmental governance. By acknowledging multiple ways of knowing, scientific and local, the research contributes to scholarship on participatory system design and adaptive monitoring. It reinforces the argument that environmental systems must be not only technically robust but also socially inclusive to remain sustainable.

6.6. Practical Implications

The practical value of the framework is evident for diverse stakeholders. For **communities**, particularly mine workers and residents of the Free State, the system provides a tool for anticipating air quality risks and adopting preventive measures. For **miners**, the system supports occupational health by forecasting exposure risks up to four days in advance. For **policymakers**, the hybrid framework offers evidence-based insights that can guide regulation, resource allocation, and community engagement strategies. The dissemination of results through a mobile application ensures accessibility, even for semi-literate users, and strengthens the framework's utility as a decision-support tool.

6.7. Comparison with International Literature

Globally, significant efforts have been made to employ machine learning, IoT, and sensor-based monitoring systems in environmental management, for example (Ghazali et al., 2012; Hou et al., 2021; Ding et al., 2021). However, most of these systems are heavily dependent on high-quality, continuous sensor datasets and often overlook the contribution of local communities. In contrast, studies in indigenous and resource-constrained contexts (Fernández-Llamazares et al., 2020 emphasise the centrality of local knowledge in environmental stewardship. This research aligns with the latter perspective but advances it by demonstrating how IK can be systematically integrated with advanced ML models and IoT infrastructure. In doing so, the study positions South Africa at the frontier of hybrid environmental monitoring approaches that both harness 4IR tools and respect community epistemologies.

6.8. Limitations

While the study successfully developed and demonstrated a hybrid framework for air quality monitoring and prediction, several limitations should be acknowledged:

Restricted Access to Mining Sites

The initial plan included installing wireless sensors within underground mine shafts to collect primary air quality data. However, access was denied by mining companies, limiting the ability to capture high-resolution, site-specific data. This restricted the scope of sensor deployment to controlled pilot tests rather than full-scale operational monitoring.

Reliance on Secondary Data
The framework relied heavily on datasets from the Pelonomi air quality monitoring station, which covered 42 months of observations. While these data provided a robust basis for training and validating machine learning models, they may not fully capture the spatial variability of air quality conditions across the Free State's mining areas.

Small-Scale Mobile Application Pilot
The mobile application designed to disseminate forecasts and integrate community input was tested only on a limited scale. Although feedback from pilot users was positive, the restricted sample size means that further testing is needed to evaluate long-term usability, accessibility for semi-literate populations, and integration into community decision-making.

6.9. Computational Constraints

The machine learning models were trained and validated using MATLAB and related platforms under constrained computational resources. While sufficient for proof of concept, these limitations may have reduced the opportunity to test more advanced models or conduct extensive hyperparameter tuning, potentially affecting predictive performance.

6.10. Recommendations

Building on the findings and limitations of this study, several recommendations are proposed to strengthen the hybrid air pollution monitoring framework and ensure its long-term sustainability and impact:

Improving Data Quality
Future work should prioritise the deployment of primary sensors within mining environments to complement secondary datasets from established stations such as Pelonomi. Expanding sensor coverage to multiple monitoring sites across the Free State will increase the spatial resolution and accuracy of predictions. Longer and more consistent data collection periods will also improve the robustness of machine learning models.

Expanding Indigenous Knowledge Participation
While the current study engaged a focused group of community members, future initiatives should broaden participation to include additional communities and stakeholders. This expansion would ensure that the system incorporates a wider range of lived experiences, thereby enhancing

representativeness and community ownership. Regular feedback sessions and participatory co-design workshops could further deepen integration between IK and technological components.

Strengthening User Interface and Adoption Strategies
The mobile application developed in this study served as a proof of concept. For broader adoption, it is recommended that the user interface be refined to be more intuitive, multilingual, and accessible to users with low levels of digital literacy. Training sessions, community demonstrations, and partnerships with local organisations could accelerate adoption and long-term use.

Policy Integration and Partnerships
To ensure sustainability, the framework should be embedded into formal policy frameworks for environmental monitoring and occupational health. Collaboration with government agencies, mining companies, and civil society organisations will be crucial in scaling up the system. Such partnerships could provide both financial support and institutional legitimacy, enabling the system to transition from a prototype to an operational tool for environmental governance.

6.11. Future Work

Scaling the framework to other provinces, expanding the framework outside the borders of the Free State province, as would be ideal, as pollution is still a problem even in different provinces.
Incorporating real-time IoT sensor networks, instead of relying heavily on secondary datasets, wireless-sensor-based collected datasets will be used.
Linking with health datasets, this would be to check the exact correlation between long-term exposure to pollution and the health complexity using real and actual patient data.
Long-term monitoring and sustainability, the project has the potential to be expanded and sustainable for the future.

6.12. Conclusion

This chapter evaluated the developed air pollution monitoring framework against the study's objectives, highlighting the extent to which each goal was achieved and reflecting on the system's contributions, limitations, and implications. The evaluation demonstrated that the proof-of-concept was successfully realised: a layered hybrid framework was built that integrates IoT devices, machine learning algorithms, and Indigenous Knowledge into a cohesive system tailored for the Free State Province.

The results showed that the framework can forecast pollution levels up to four days in advance, incorporate community-based indicators through Fuzzy Cognitive Maps, and disseminate early warnings using a prototype mobile application. While constraints such as limited sensor deployment and reliance on secondary data restricted the system's scope, the findings nevertheless establish its feasibility, adaptability, and relevance for both scientific and community stakeholders.

The study contributes to knowledge by illustrating how 4IR technologies can be combined with lived experiences to strengthen adaptive monitoring systems in resource-constrained contexts. Practically, it offers a pathway for more inclusive and anticipatory approaches to environmental management, directly benefiting mine workers, local communities, and policymakers.

In conclusion, this dissertation has met its overarching aim of demonstrating a context- sensitive, hybrid framework for air pollution monitoring. The work provides a foundation for future scaling, refinement, and policy adoption, and underscores the importance of bridging technological innovation with Indigenous Knowledge to achieve sustainable and locally grounded solutions.

REFERENCES

Abbasi, S., 2018. Defining Safety Hazards; Risks in Mining Industry: A Case-Study in United States. Asian Journal of Applied Science and Technology (AJAST) (Open Access Quarterly International Journal 2, 1071–1078.

Acharya, A.S., Prakash, A., Saxena, P., Nigam, A., 2013. Sampling: why and how of it? Indian Journal of Medical Specialities 4. https://doi.org/10.7713/ijms.2013.0032

Adisa, O.M., Masinde, M., Botai, J.O., Botai, C.M., 2020. Bibliometric analysis of methods and tools for drought monitoring and prediction in Africa. Sustainability (Switzerland). https://doi.org/10.3390/su12166516

Aguirre-Ayerbe, I., Merino, M., Aye, S.L., Dissanayake, R., Shadiya, F., Lopez, C.M., 2020. An evaluation of availability and adequacy of Multi-Hazard Early Warning Systems in Asian countries: A baseline study. International Journal of Disaster Risk Reduction 49, 101749. https://doi.org/10.1016/j.ijdrr.2020.101749

Ahmed, S.K., 2024. How to choose a sampling technique and determine sample size for research: A simplified guide for researchers. Oral Oncology Reports. https://doi.org/10.1016/j.oor.2024.100662

Albuainain, H.K.R.B., Choe, P., 2023. Improving User Experience with Online Learning Using Analytical Comparasions between Instructors & Students of Qatar University. International Journal on Cybernetics & Informatics 12, 19–44. https://doi.org/10.5121/ijci.2023.120402

Alkhatib, R., Sahwan, W., Alkhatieb, A., 2023. applied sciences A Brief Review of Machine Learning Algorithms in Forest Fires Science.

Almetwally, S.A.H., Hassan, M.K., Mourad, M.H., 2020. Real Time Internet of Things (IoT) Based Water Quality Management System, in: Procedia CIRP. Elsevier B.V., pp. 478–485. https://doi.org/10.1016/j.procir.2020.03.107

Alonge, O., Rodriguez, D.C., Reveiz, L., Peters, D.H., Brandes, N., Geng, E., 2019. How is implementation research applied to advance health in low- income and middle-income countries ? 1–10. https://doi.org/10.1136/bmjgh-2018-001257

Aluru, N.R., Aydin, F., Bazant, M.Z., Blankschtein, D., Brozena, A.H., de Souza, J.P., Elimelech, M., Faucher, S., Fourkas, J.T., Koman, V.B., Kuehne, M., Kulik, H.J., Li, H.K., Li, Y., Li, Z., Majumdar,
A., Martis, J., Misra, R.P., Noy, A., Pham, T.A., Qu, H., Rayabharam, A., Reed, M.A., Ritt, C.L.,
Schwegler, E., Siwy, Z., Strano, M.S., Wang, Y.H., Yao, Y.C., Zhan, C., Zhang, Z., 2023.

Fluids and Electrolytes under Confinement in Single-Digit Nanopores. Chem Rev. https://doi.org/10.1021/acs.chemrev.2c00155

Asha, P., Natrayan, L., Geetha, B.T., Beulah, J.R., Sumathy, R., Varalakshmi, G., Neelakandan, S., 2022. IoT enabled environmental toxicology for air pollution monitoring using AI techniques.
Environ Res 205, 112574. https://doi.org/10.1016/j.envres.2021.112574

Atzori, L., Iera, A., Morabito, G., 2017. Understanding the Internet of Things: definition, potentials, and societal role of a fast evolving paradigm. Ad Hoc Networks 56, 122–140. https://doi.org/10.1016/j.adhoc.2016.12.004

Bag, S., Srivastava, G., Cherrafi, A., Ali, A., Singh, R.K., 2024. Data-driven insights for circular and sustainable food supply chains: An empirical exploration of big data and predictive analytics in enhancing social sustainability performance. Bus Strategy Environ 33, 1369–1396. https://doi.org/10.1002/bse.3554

Balasubramaniyan, C., Manivannan, D., 2016. IoT enabled Air Quality Monitoring System (AQMS) using Raspberry Pi. Indian J Sci Technol 9. https://doi.org/10.17485/ijst/2016/v9i39/90414

Barbrook-Johnson, P., Penn, A.S., 2022. Theory of Change Diagrams, Systems Mapping. https://doi.org/10.1007/978-3-031-01919-7_3

Barnhardt, R., Oscar, A., Barnhardt, K., Kawagley, R., 2005. Indigenous Knowledge Systems and Alaska Native Ways of Knowing, Anthropology and Education Quarterly.

Bhat, A., Manek, A.S., Mishra, P., 2019. Machine Learning based Prediction System for Detecting Air Pollution. International Journal of Engineering Research & Technology (IJERT) 8, 155–159.

Bravo, M.A., Fuentes, M., Zhang, Y., Burr, M.J., Bell, M.L., 2012. Comparison of exposure estimation methods for air pollutants: Ambient monitoring data and regional air quality simulation. Environ Res 116, 1–10. https://doi.org/10.1016/j.envres.2012.04.008

Brevik, L.M., 2017. The Multiphase Design: How Mixed Methods Research Adds Value to Research Projects. UiO:Department of Teacher Education and School Research.

Bull, F.C., Al-Ansari, S.S., Biddle, S., Borodulin, K., Buman, M.P., Cardon, G., Carty, C., Chaput, J.P., Chastin, S., Chou, R., Dempsey, P.C., Dipietro, L., Ekelund, U., Firth, J., Friedenreich, C.M., Garcia, L., Gichu, M., Jago, R., Katzmarzyk, P.T., Lambert, E., Leitzmann, M., Milton, K., Ortega, F.B., Ranasinghe, C., Stamatakis, E., Tiedemann, A., Troiano, R.P., Van Der Ploeg, H.P., Wari, V., Willumsen, J.F., 2020. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports

Med 54, 1451–1462. https://doi.org/10.1136/bjsports-2020- 102955
Baraneetharan, E., 2020. Role of Machine Learning Algorithms Intrusion Detection in WSNs: A Survey.
Journal of Information Technology and Digital World 02, 161–173. https://doi.org/10.36548/jitdw.2020.3.004

Bzai, J., Alam, F., Dhafer, A., Bojović, M., Altowaijri, S.M., Niazi, I.K., Mehmood, R., 2022. Machine Learning-Enabled Internet of Things (IoT): Data, Applications, and Industry Perspective.
Electronics (Switzerland). https://doi.org/10.3390/electronics11172676

Carandell, M., Mihai, D., del Río, J., Ganchev, K., Peudennier, J., 2018. Evaluation of Sigfox LPWAN technology for autonomous sensors in coastal applications. Instrumentation viewpoint 20, 36–37.

Carm, E., 2014. Inclusion of Indigenous Knowledge System (IKS) – A Precondition for Sustainable Development and an Integral Part of Environmental Studies Ellen Carm* 4, 58–76.

Chaparro, Mauro A.E., Chaparro, Marcos A.E., Molinari, D.A., 2024. A Fuzzy-Based Analysis of Air Particle Pollution Data: An Index IMC for Magnetic Biomonitoring. Atmosphere (Basel) 15. https://doi.org/10.3390/atmos15040435

Chen, X., Ge, G., Liu, Y., Cheng, A., Wu, J., Fu, J., 2018a. Construct mine environment monitoring system based on wireless mesh network. AIP Conf Proc 1955. https://doi.org/10.1063/1.5033646

Chen, X., Ge, G., Liu, Y., Cheng, A., Wu, J., Fu, J., 2018b. Construct mine environment monitoring system based on wireless mesh network. AIP Conf Proc 1955. https://doi.org/10.1063/1.5033646

Chen, Y., Gu, P., Schulte, N., Zhou, X., Mara, S., Croes, B.E., Herner, J.D., Vijayan, A., 2022. A new mobile monitoring approach to characterize community-scale air pollution patterns and identify local high pollution zones. Atmos Environ 272, 118936. https://doi.org/10.1016/j.atmosenv.2022.118936

Clark, D.B., Tanner-Smith, E.E., Killingsworth, S.S., 2016. Digital Games, Design, and Learning: A Systematic Review and Meta-Analysis. Rev Educ Res 86, 79–122. https://doi.org/10.3102/0034654315582065

Cleveland, R., Trauernicht, C., Bremer, L., Pickett, E., Oleson, K.L.L., 2024. Guiding Fuzzy Cognitive Mapping With Structured Decision Making to Inform Complex Natural Resource Management Problems in Wai'anae Hawai'i. Community Science 3.

https://doi.org/10.1029/2023csj000060
Creswell, J.W.., Poth, C.N.., 2018. Qualitative inquiry and research design. SAGE.
Dahl, P.P.E., Hansen, A.M., 2019. Does indigenous knowledge occur in and influence impact assessment reports? Exploring consultation remarks in three cases of mining projects in greenland. Arctic Review on Law and Politics 10, 165–189. https://doi.org/10.23865/arctic.v10.1344
Dawn, T., 2019. Technologies of ground support monitoring in block caving operations 109–122. https://doi.org/10.36487/acg_rep/1925_06_dawn
Delabrida, S.E., Dangelo, T., Oliveira, R.A.R., Loureiro, A.A.F., 2016. Towards a wearable device for monitoring ecological environments, in: Proceedings - 2015 Brazilian Symposium on Computing Systems Engineering, SBESC 2015. Institute of Electrical and Electronics Engineers Inc., pp. 148–153. https://doi.org/10.1109/SBESC.2015.35
Delavar, M.R., Gholami, A., Shiran, G.R., Rashidi, Y., Nakhaeizadeh, G.R., Fedra, K., Afshar, S.H., 2019. A novel method for improving air pollution prediction based on machine learning approaches: A case study applied to the capital city of Tehran. ISPRS Int J Geoinf 8. https://doi.org/10.3390/ijgi8020099
Ding, Q., Zhu, R., Liu, H., Ma, M., 2021. An Overview of Machine Learning-Based Energy-Efficient Routing Algorithms in Wireless Sensor Networks.
Dilanjani, H.U.K., Marasinghe, S.A., Kumara, H.U.C.S., 2025. A bibliometric analysis of research trends in air pollution and environmental impacts. Pollution Study 6, 3121. https://doi.org/10.54517/ps3121
Donnette, Q., Rent, G., Koreli, N., Nunes, I., Everett, C., Herman, E., 2016. Research Design
Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., Lim, W.M., 2021. How to conduct a bibliometric analysis: An overview and guidelines. J Bus Res 133, 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070
Duisebekova, K.S., Tuyakova, Z.N., Amanzholova, S.T., Sarsenova, Z.N., Duzbayev, N.T., Pyagay, V.T., Aitmagambetov, A.Z., 2019. Environmental monitoring system for analysis of climatic and ecological changes using LoRa technology. ACM International Conference Proceeding Series 1–6. https://doi.org/10.1145/3330431.3330446
Fanoro, M., Božanić, M., Sinha, S., 2021. A Review of 4IR/5IR Enabling Technologies and Their Linkage to Manufacturing Supply Chain. Technologies (Basel) 9, 77. https://doi.org/10.3390/technologies9040077

Fashola, M.O., Ngole-Jeme, V.M., Babalola, O.O., 2016. Heavy metal pollution from gold mines: Environmental effects and bacterial strategies for resistance. Int J Environ Res Public Health. https://doi.org/10.3390/ijerph13111047

Fernández-Llamazares, Á., Garteizgogeascoa, M., Basu, N., Brondizio, E.S., Cabeza, M., Martínez- Alier, J., McElwee, P., Reyes-García, V., 2020. A State-of-the-Art Review of Indigenous Peoples and Environmental Pollution. Integr Environ Assess Manag 16, 324–341. https://doi.org/10.1002/ieam.4239

Fitria, D., Husaeni, A., Bayu, A., Nandiyanto, D., 2021. Bibliometric Using Vosviewer with Publish or Perish (using Google Scholar data): From Step-by-step Processing for Users to the Practical Examples in the Analysis of Digital Learning Articles in Pre and Post Covid-19 Pandemic. https://doi.org/10.17509/ijost.v6ix

Fuertes, W., Carrera, D., Villacis, C., Toulkeridis, T., Galarraga, F., Torres, E., Aules, H., 2016. Distributed system as internet of things for a new low-cost, air pollution wireless monitoring on real time. Proceedings - 2015 IEEE/ACM 19th International Symposium on Distributed Simulation and Real Time Applications, DS-RT 2015 58–67. https://doi.org/10.1109/DS-RT.2015.28

Fujdiak, R., Blazek, P., Mikhaylov, K., Malina, L., Mlynek, P., Misurec, J., Blazek, V., 2018. On track of sigfox confidentiality with end-to-end encryption. ACM International Conference Proceeding Series. https://doi.org/10.1145/3230833.3232805

Ghazali, M.Z., Noor, N.M., Ramly, N., Putit, S., 2012. Development of microcontroller based mobile gas sensing robot. Procedia Eng 41, 1190–1196. https://doi.org/10.1016/j.proeng.2012.07.300

Ghobakhloo, M., Fathi, M., Iranmanesh, M., Maroufkhani, P., Morales, M.E., 2021. Industry 4.0 ten years on: A bibliometric and systematic review of concepts, sustainability value drivers, and success determinants. J Clean Prod 302, 127052. https://doi.org/10.1016/j.jclepro.2021.127052

Gold, H., 2014. CHAMBER OF MINES OF SOUTH AFRICA Applicant ACTING IN ITS OWN NAME & OBO.

Gold Mining in Africa: Maximizing Economic Returns for Countries Ousman Gajigo Emelly Mutambatsere Guirane Ndiaye, 2012.

Gomez, C., Veras, J.C., Vidal, R., Casals, L., Paradells, J., 2019. A sigfox energy consumption model. Sensors (Switzerland) 19. https://doi.org/10.3390/s19030681

Gopichandran, V., Luyckx, V.A., Biller-andorno, N., Fairchild, A., Singh, J., Tran, N., Saxena, A., Launois, P., Reis, A., Maher, D., Vahedi, M., 2016. Developing the ethics of implementation research in health. Implementation Science 1–13. https://doi.org/10.1186/s13012-016-0527-y

Gray, S., Chan, A., Clark, D., Jordan, R., 2012. Modeling the integration of stakeholder knowledge in social-ecological decision-making: Benefits and limitations to knowledge diversity. Ecol Modell 229, 88–96. https://doi.org/10.1016/j.ecolmodel.2011.09.011

Gray, S.A., Gray, S., Cox, L.J., Henly-Shepard, S., 2013. Mental Modeler: A fuzzy-logic cognitive mapping modeling tool for adaptive environmental management. Proceedings of the Annual Hawaii International Conference on System Sciences 965–973. https://doi.org/10.1109/HICSS.2013.399

Gu, X., Li, H., Fan, H., 2023. Spatiotemporal Hybrid Air Pollution Early Warning System of Urban Agglomeration Based on Adaptive Feature Extraction and Hesitant Fuzzy Cognitive Maps. Systems 11. https://doi.org/10.3390/systems11060286

Gwaze, P., Mashele, S.H., 2018. South African Air Quality Information System (SAAQIS) mobile application tool: Bringing real time state of air quality to South Africans. Clean Air Journal 28, 3–4. https://doi.org/10.17159/2410-972X/2018/v28n1a1

Halbrendt, J., Gray, S.A., Crow, S., Radovich, T., Kimura, A.H., Tamang, B.B., 2014. Differences in farmer and expert beliefs and the perceived impacts of conservation agriculture. Global Environmental Change 28, 50–62. https://doi.org/10.1016/j.gloenvcha.2014.05.001

Henly-Shepard, S., Gray, S.A., Cox, L.J., 2015. The use of participatory modeling to promote social learning and facilitate community disaster planning. Environ Sci Policy 45, 109–122. https://doi.org/10.1016/j.envsci.2014.10.004

Henriques, V., Malekian, R., Capeska Bogatinoska, D., 2017. Mine safety system using wireless sensor networks. 2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2017 - Proceedings 4, 515–520. https://doi.org/10.23919/MIPRO.2017.7973480

Hiwasaki, L., Luna, E., Syamsidik, Shaw, R., 2014. Process for integrating local and indigenous knowledge with science for hydro-meteorological disaster risk reduction and climate change adaptation in coastal and small island communities. International Journal of Disaster Risk Reduction 10, 15–27. https://doi.org/10.1016/j.ijdrr.2014.07.007

Honghui, W., Xianguo, T., Yan, L., Qi, L., Donglin, N., Lingyu, M., Jiaxin, Y., 2017. Research of the Hardware Architecture of the Geohazards Monitoring and Early Warning System Based on the Iot. Procedia Comput Sci 107, 111–116. https://doi.org/10.1016/j.procs.2017.03.065

Hou, H., Ding, Z., Zhang, S., Guo, S., Yang, Y., Chen, Z., Mi, J., Wang, X., 2021. Spatial estimate of ecological and environmental damage in an underground coal mining area on the Loess Plateau: Implications for planning restoration interventions. J Clean Prod 287, 125061. https://doi.org/10.1016/j.jclepro.2020.125061

Ibrahim, I., Abdulazeez, A., 2021. The Role of Machine Learning Algorithms for Diagnosing Diseases. Journal of Applied Science and Technology Trends 2, 10–19. https://doi.org/10.38094/jastt20179

Ihianle, I.K., Nwajana, A.O., Ebenuwa, S.H., Otuka, R.I., Owa, K., Orisatoki, M.O., 2020. A deep learning approach for human activities recognition from multimodal sensing devices. IEEE Access 8, 179028–179038. https://doi.org/10.1109/ACCESS.2020.3027979

Joseph, J., Senith, S., Alfred Kirubaraj, A., Jino Ramson, S.R., 2022. Machine Learning for Prediction of Clinical Appointment No-Shows. International Journal of Mathematical, Engineering and Management Sciences 7, 558–574. https://doi.org/10.33889/IJMEMS.2022.7.4.036

Jayasundara, R., Fernando, I., Fernando, A., Ragel, R., Thambawita, V., Nawinne, I., 2025. Inductive transfer learning from regression to classification in ECG analysis.

Kadaifci, C., Karadayi-Usta, S., Yanmaz, O., 2024. An analysis of consumer opinions on waste medicine management utilizing fermatean fuzzy cognitive mapping. Environ Dev 49, 100961. https://doi.org/10.1016/j.envdev.2023.100961

Kumar, K., Pande, B.P., 2023. Air pollution prediction with machine learning: a case study of Indian cities. International Journal of Environmental Science and Technology 20, 5333–5348. https://doi.org/10.1007/s13762-022-04241-5

Lee, A., Taylor, P., Kalpathy-Cramer, J., 2017. Machine Learning Has Arrived! Ophthalmology 124, 1726–1728. https://doi.org/10.1016/j.ophtha.2017.08.046

Li, Y., Wang, X., Zhao, Z., Han, S., Liu, Z., 2020. Lagoon water quality monitoring based on digital image analysis and machine learning estimators. Water Res 172, 115471. https://doi.org/10.1016/j.watres.2020.115471

Lööw, J., 2022. Understanding technology in mining and its effect on the work environment. Mineral Economics 35, 143–154. https://doi.org/10.1007/s13563-021-00279-y

Lusilao-Makiese, J.G., Cukrowska, E.M., Tessier, E., Amouroux, D., Weiersbye, I., 2013. The impact of post gold mining on mercury pollution in the West Rand region, Gauteng, South Africa. J Geochem Explor 134, 111–119. https://doi.org/10.1016/j.gexplo.2013.08.010

Manisalidis, I., Stavropoulou, E., Stavropoulos, A., Bezirtzoglou, E., 2020. Environmental and Health Impacts of Air Pollution: A Review. Front Public Health. https://doi.org/10.3389/fpubh.2020.00014

Marais, L., Owen, J.R., Kotzè, T., Nel, P., Cloete, J., Lenka, M., 2021. Determinants of place attachment among mineworkers: Evidence from South Africa. Extractive Industries and Society 8. https://doi.org/10.1016/j.exis.2021.100943

Matandirotya, N.R., 2021. Research trends in the field of ambient air quality monitoring and management in South Africa: A bibliometric review. Environmental Challenges 5, 100263. https://doi.org/10.1016/j.envc.2021.100263

Mbele, M., Masinde, M., Kgololo-Ngowi, I., 2016. Adaptive environmental management system for lejweleputswa district: A participatory approach through fuzzy cognitive maps. Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST 171, 225–226.

Michalski, B., Skonieczka, S., Strzelecki, M., Simiera, M., Kupczyńska, K., Szymczyk, E., Wejner-Mik, P., Lipiec, P., Kasprzak, J.D., 2024. The use of artificial intelligence for predicting postinfarction
myocardial viability in echocardiographic images. Cardiol J 31, 699–707. https://doi.org/10.5603/cj.93887

Mitchell, M., 2019. Artificial intelligence hits the barrier of meaning. Information (Switzerland) 10. https://doi.org/10.3390/info10020051

Mohsin, M., Zhu, Q., Naseem, S., Sarfraz, M., Ivascu, L., 2021. Mining industry impact on environmental sustainability, economic growth, social interaction, and public health: An application of semi-quantitative mathematical approach. Processes 9. https://doi.org/10.3390/pr9060972

Mpanza, T.D.E., Hassen, A., Akanmu, A.M., 2020. Evaluation of Stylosanthes scabra accessions as forage source for ruminants: Growth performance, nutritive value and in vitro ruminal fermentation. Animals 10, 1–15. https://doi.org/10.3390/ani10111939

Muduli, L., Jana, P.K., Mishra, D.P., 2018a. Wireless sensor network-based fire monitoring in underground coal mines: A fuzzy logic approach. Process Safety and Environmental Protection 113, 435–447. https://doi.org/10.1016/j.psep.2017.11.003

Muduli, L., Mishra, D.P., Jana, P.K., 2018b. Application of wireless sensor network for environmental monitoring in underground coal mines: A systematic review. Journal of Network and Computer Applications 106, 48–67. https://doi.org/10.1016/j.jnca.2017.12.022

Muimba-Kankolongo, A., Banza Lubaba Nkulu, C., Mwitwa, J., Kampemba, F.M., Mulele Nabuyanda, M., 2022. Impacts of Trace Metals Pollution of Water, Food Crops, and Ambient Air on Population Health in Zambia and the DR Congo. J Environ Public Health

2022. https://doi.org/10.1155/2022/4515115

Nainggolan, R., Eviyanti Purba, 2020. The Cluster Analysis of Online Shop Product Reviews Using K- Means Clustering. Data Science: Journal of Computing and Applied Informatics 4, 111–121. https://doi.org/10.32734/jocai.v4.i2-2855

Nápoles, G., Espinosa, M.L., Grau, I., Vanhoof, K., 2018. FCM Expert: Software Tool for Scenario Analysis and Pattern Classification Based on Fuzzy Cognitive Maps. International Journal on Artificial Intelligence Tools 27. https://doi.org/10.1142/S0218213018600102

Nor, S.M.C.M., Shaharudin, S.M., Ismail, S., Najib, S.A.M., Tan, M.L., Ahmad, N., 2022. Statistical Modeling of RPCA-FCM in Spatiotemporal Rainfall Patterns Recognition. Atmosphere (Basel) 13. https://doi.org/10.3390/atmos13010145

Nyaki, A., Gray, S.A., Lepczyk, C.A., Skibins, J.C., Rentsch, D., 2014. Local-Scale Dynamics and Local Drivers of Bushmeat Trade. Conservation Biology 28, 1403–1414. https://doi.org/10.1111/cobi.12316

Obiero, K.O., Klemet-N'Guessan, S., Migeni, A.Z., Achieng, A.O., 2023. Bridging Indigenous and non- Indigenous knowledge systems and practices for sustainable management of aquatic resources from East to West Africa. J Great Lakes Res 49, S128–S137. https://doi.org/10.1016/j.jglr.2022.12.001

Orellano, P., Reynoso, J., Quaranta, N., Bardach, A., Ciapponi, A., 2020. Short-term exposure to particulate matter (PM10 and PM2.5), nitrogen dioxide (NO2), and ozone (O3) and all-cause and cause-specific mortality: Systematic review and meta-analysis. Environ Int. https://doi.org/10.1016/j.envint.2020.105876

Osunmakinde, I.O., 2013. Towards safety from toxic gases in underground mines using wireless sensor networks and ambient intelligence. Int J Distrib Sens Netw 2013. https://doi.org/10.1155/2013/159273

Paula, A., Takano, C., Tocci, L., Villa, N., Valeria, M., Afonso, P., André, D., Marcelo, F., Augusto, C., Vizeu, L., Singer, J.M., Diva, C., André, S. De, Hilário, P., Saldiva, N., Matera, M., 2019. Pleural anthracosis as an indicator of lifetime exposure to urban air pollution : An autopsy-based study in Sao Paulo. Environ Res 173, 23–32. https://doi.org/10.1016/j.envres.2019.03.006

Power, A.L., Tennant, R.K., Stewart, A.G., Gosden, C., Worsley, A.T., Jones, R., Love, J., 2023. The evolution of atmospheric particulate matter in an urban landscape since the Industrial Revolution. Sci Rep 13. https://doi.org/10.1038/s41598-023-35679-3

Peng, J., Han, H., Yi, Y., Huang, H., Xie, L., 2022. Machine learning and deep learning modeling and simulation for predicting PM2.5 concentrations. Chemosphere 308, 136353.

https://doi.org/10.1016/j.chemosphere.2022.136353

Peters, B., Trunschke, M., 2021. Benefits of Investing into the Development of 4IR- Related Technologies.

Poola, I., 2017. Automated artificial intelligence powered decision-making in big data.

Prasad, G.S.C., Pillai, A.S., 2019. Role of Industrial IoT in Critical Environmental Conditions. Proceedings of the 2nd International Conference on Intelligent Computing and Control Systems, ICICCS 2018 1369–1372. https://doi.org/10.1109/ICCONS.2018.8663213

Praveen Kumar, D., Amgoth, T., Annavarapu, C.S.R., 2019. Machine learning algorithms for wireless sensor networks: A survey. Information Fusion 49, 1–25. https://doi.org/10.1016/j.inffus.2018.09.013

Priority, H., Air, A., Monitoring, Q., 2017. South African Weather Service - Highveld Priority Area Air Quality Monitoring Network Monthly Report — April 2017 1–55.

Radhakrishnan, S., Erbis, S., Isaacs, J.A., Kamarthi, S., 2017. Novel keyword co-occurrence network- based methods to foster systematic reviews of scientific literature. PLoS One. https://doi.org/10.1371/journal.pone.0172778

Rao, B., 2016. Machine Learning Algorithms: A Review. International Journal of Computer Science and Information Technologies 7, 1174–1179. https://doi.org/10.21275/ART20203995

Rooney, R.C., Daniel, J., Mallory, M., Hedd, A., Ives, J., Gilchrist, G., Gjerdrum, C., Robertson, G., Ronconi, R., Wilcox, K., Wong, S., Provencher, J., 2023. Fuzzy cognitive mapping as a tool to assess the relative cumulative effects of environmental stressors on an Arctic seabird population to identify conservation action and research priorities. Ecological Solutions and Evidence 4. https://doi.org/10.1002/2688-8319.12241

Salinas Salmeron, K.S., Kim, H.J., Seo, K.-S., 2019. Effects site concentrations of propofol using target-controlled infusion in dental treatment under deep sedation among different intellectual disability types. J Dent Anesth Pain Med 19, 217. https://doi.org/10.17245/jdapm.2019.19.4.217

Sarmiento, I., Cockcroft, A., Dion, A., Belaid, L., Silver, H., Pizarro, K., Pimentel, J., Tratt, E., Skerritt, L., Ghadirian, M.Z., Gagnon-Dufresne, M.C., Andersson, N., 2024. Fuzzy cognitive mapping in
participatory research and decision making: a practice review. Archives of Public Health 82. https://doi.org/10.1186/s13690-024-01303-7

Sharifi-Rad, M., Anil Kumar, N. V., Zucca, P., Varoni, E.M., Dini, L., Panzarini, E., Rajkovic, J., Tsouh Fokou, P.V., Azzini, E., Peluso, I., Prakash Mishra, A., Nigam, M., El Rayess, Y., Beyrouthy, M. El, Polito, L., Iriti, M., Martins, N., Martorell, M., Docea, A.O., Setzer, W.N., Calina, D., Cho, W.C., Sharifi-Rad, J., 2020. Lifestyle, Oxidative Stress, and Antioxidants: Back and Forth in the Pathophysiology of Chronic Diseases. Front Physiol. https://doi.org/10.3389/fphys.2020.00694

Shilenge, M., Telukdarie, A., 2021. 4IR integration of information technology best practice framework in operational technology. Journal of Industrial Engineering and Management 14, 457–476. https://doi.org/10.3926/jiem.3429

Shinde, D., Siddiqui, N., 2018. IOT Based Environment change Monitoring Controlling in Greenhouse using WSN. 2018 International Conference on Information, Communication, Engineering and Technology, ICICET 2018 1–5. https://doi.org/10.1109/ICICET.2018.8533808

Singh, A., Kumar, D., Hötzel, J., 2018. IoT Based information and communication system for enhancing underground mines safety and productivity: Genesis, taxonomy and open issues. Ad Hoc Networks 78, 115–129. https://doi.org/10.1016/j.adhoc.2018.06.008

Singh, R.P.B., ; U-Gauhati, ;, 2017. Appraising the Indian Cultural Landscape: Envisioning Ecological Cosmology in the 21st Century, North Eastern Geographer. North East India Geographical Society.

Smith, J.D., Li, D.H., Rafferty, M.R., 2020. The Implementation Research Logic Model : a method for planning , executing , reporting , and synthesizing implementation projects 1–12.

Solana-Gutiérrez, J., Rincón, G., Alonso, C., García-de-Jalón, D., 2017. Using fuzzy cognitive maps for predicting river management responses: A case study of the Esla River basin, Spain. Ecol Modell 360, 260–269. https://doi.org/10.1016/j.ecolmodel.2017.07.010

Stewart, J.M., Malatji, S.M., 2018. Development of the MOSH leading practice adoption system –– a science-based system for managing behaviour change. J South Afr Inst Min Metall 118, 259–277. https://doi.org/10.17159/2411-9717/2018/v118n3a9

Sun, Z., Wu, K.-T., Kruger, S., Levesque, D., Gagnon, D., Quenneville, Y., Lacroix, R., Royer, R., 2019. A new paradigm in ground support monitoring through ultrasonic monitoring of clusters of rockbolts 75–84. https://doi.org/10.36487/acg_rep/1925_03_sun

Tan, H., Hao, Y., 2022. Mapping the Global Evolution and Research Directions of Information Seeking, Sharing and Communication in Disasters: A Bibliometric Study. Int J Environ Res Public Health 19. https://doi.org/10.3390/ijerph192214878

Thomas, S., Ravishankaran, S., Justin, N.A.J.A., Asokan, A., Kalsingh, T.M.J., Mathai, M.T., Valecha, N., Montgomery, J., Thomas, M.B., Eapen, A., 2018. Microclimate variables of

the ambient environment deliver the actual estimates of the extrinsic incubation period of Plasmodium vivax and Plasmodium falciparum: A study from a malaria-endemic urban setting, Chennai in India.
Malar J 17. https://doi.org/10.1186/s12936-018-2342-1

Tibane, L.V., Mamba, D., 2022. Ecological risk of trace metals in soil from gold mining region in South Africa. Journal of Hazardous Materials Advances 7, 100118. https://doi.org/10.1016/j.hazadv.2022.100118

Tranca, D.C., Rosner, D., Curatu, R., Surpateanu, A., Mocanu, M., Pardau, S., Palacean, A.V., 2017.
Industrial WSN node extension and measurement systems for air, water and environmental monitoring: IoT enabled environment monitoring using NI WSN nodes. 16th Networking in Education and Research RoEduNet International Conference, RoEduNet 2017 - Proceedings. https://doi.org/10.1109/ROEDUNET.2017.8123750

Tzeng, C.B., Wey, T.S., 2011. Design and implement a cost effective and ubiquitous air quality monitoring system based on ZigBee wireless sensor network. Proceedings - 2011 2nd International Conference on Innovations in Bio-Inspired Computing and Applications, IBICA 2011 245–248. https://doi.org/10.1109/IBICA.2011.64

Uddin, S., Khan, A., Hossain, M.E., Moni, M.A., 2019. Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inform Decis Mak 19, 1–16. https://doi.org/10.1186/s12911-019-1004-8

Ugya AY, Ajibade FO, Ajibade TF, 2018. Water Pollution Resulting From Mining Activity: an Overview.
Proceedings of the 2018 Annual Conference of the School of Engineering & Engineering Technology (SEET), The Federal University of Technology, Akure, Nigeria, 17-19 July, 2018.

Veltin, K., Preston, R., Pakula, A., Kennard, D., 2019. Utilising database algorithms and three-dimensional visualisation software to optimise ground-control management 593–600. https://doi.org/10.36487/acg_rep/1925_43_veltin

Wexler, B.J., Schultz, C., Biddinger, P.D., Ciottone, G., Cornelius, A., Fuller, R., Lefort, R., Milsten, A., Phillips, J., Nemeth, I., 2023. The 2023 Model Core Content of Disaster Medicine. Prehosp Disaster Med 38, 699–706. https://doi.org/10.1017/S1049023X23006556

Whitaker, J., 2014. Predicting Historical Air Quality - A Zindi Project by Yasin Ayami and 0–7.

Wu, X., Xu, C., Ma, Y., 2016. Coal Mine Gas Early Warning System Based on Fuzzy Data Fusion Technology 928–930. https://doi.org/10.2991/mmebc-16.2016.194

Wu, Y., Chen, M., Wang, K., Fu, G., 2019. A dynamic information platform for underground coal mine safety based on internet of things. Saf Sci 113, 9–18. https://doi.org/10.1016/j.ssci.2018.11.003

Xiong, Y., Luo, J., Liu, X., Liu, Y., Xin, X., Wang, S., 2022. Machine learning-based optimal design of groundwater pollution monitoring network. Environ Res 211, 113022. https://doi.org/10.1016/j.envres.2022.113022

Yazici, M.T., Basurra, S., Gaber, M.M., 2018. Edge machine learning: Enabling smart internet of things applications. Big Data and Cognitive Computing 2, 1–17. https://doi.org/10.3390/bdcc2030026

Yun, S.J., Kwon, J.W., Kim, W.T., 2022. A Novel Digital Twin Architecture with Similarity-Based Hybrid Modeling for Supporting Dependable Disaster Management Systems. Sensors 22. https://doi.org/10.3390/s22134774Zhao, H., Yang, W., 2018. An emergency rescue communication system and environmental monitoring subsystem for underground coal mine based on wireless mesh network. Int J Distrib Sens Netw 14. https://doi.org/10.1177/1550147718805935

Zhou, B., Li, X., 2021. The monitoring of chemical pesticides pollution on ecological environment by GIS. Environ Technol Innov 23, 101506. https://doi.org/10.1016/j.eti.2021.101506

Zrelli, A., Ezzedine, T., 2018. Design of optical and wireless sensors for underground mining monitoring system. Optik (Stuttg) 170, 376–383

APPENDIX A: IK QUESTIONNAIRE

Good day. Thank you for participating in this survey. The aim is to gather participant's views and indigenous knowledge on the issues around mining pollution. The response will contribute towards academic research finding, no income will be generated from this survey. All responses will be kept anonymous

1. Please specify your gender

Mark only one oval.

- Male
- Female
- Prefer not to say

2. Please select your age group

Mark only one oval.

- 18-24
- 25-39
- 40-60
- 60 Plus

3. Have you ever heard about Pollution?

Mark only one oval.

- Yes
- No

4. If yes, on a scale of 1 - 10 How would you rate your knowledge about Pollution

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Very											Very satisfied

5. From the list below, please specify from which platform did you learn aboutPollution

Check all that apply.

- Internet
- News, Radio or TelevisionSocial Media
- Business/Academic conference /WorkshopCommunication Flyer or Billboards

6. From the list below, please select any type of Pollution you are knowledgeable on

Check all that apply.

- Air Pollution Water PollutionSoil Pollution

Mine Tailings Pond

7. In your own words, please share how have you been protecting yourself fromPollution

8. If you have been affected by Pollution, from the list bellow please specify whichpollution you have been mostly exposed to

Check all that apply.

Air Pollution Water
PollutionSoil
Pollution

Mine Tailings Pond

9. Is your method of survival still effective?

Mark only one oval.

Yes
No
Not all the time

10. If your method is still effective, would you recommend it to others

Mark only one oval.

Yes
No

11. Do you agree that Pollution is still a major problem in the mines?

Mark only one oval.

Strongly agree
Agree
Disagree
Strongly disagree

12. In your own views, what can be done to minimize the impact of pollution

13. In your own views, what can be done to minimize the impact of pollution

Mark only one oval.

- 0 - 5 Years
- 6 - 10 Years
- 11- 20 years
- Over 21 years

14. In general, what is most likely to affect your community and other communities around the mining area?

15. Please state which user are you

Mark only one oval.

- Mine Worker
- Medical Doctor
- IK expert
- Local Community member

16. If Doctor, do you think that patients that are exposed to pollution associated with mining are more vulnerable towards any kind of sickness

Mark only one oval.

- Yes
- No

17. If Mine worker, how long have you been working at the mines

18. As an IK Expert, do you sometimes pick up ... aminants?

APPENDIX B: ETHICAL CLEARANCE

FACULTY OF ENGINEERING AND INFORMATION TECHNOLOGY
Department of Civil Engineering

APPLICATION FOR ETHICAL CLEARANCE TO CONDUCT RESEARCH IN THE FACULTY OF ENGINEERING, BUILT ENVIRONMENT AND INFORMATION TECHNOLOGY

The Central University of Technology (CUT) Research Ethics and Integrity Policy applies to all undergraduate and post graduate students, and staff members who conduct research on CUT campuses and outside the campus. CUT policy bounds any person who wishes to conduct research with CUT students and/or staff but is not CUT affiliated to abide by the ethics framework. All CUT members who conduct research take responsibility to implement this, Policy.

1. APPLICANT INFORMATION

1.1.	**Title (Prof Dr /Mr /Mrs /Ms)**	Miss		
1.2	**Name(s) and Surname**	Pamela Ramba		
1.3	**Student / Staff number**	214125793		
1.4	**Department**	DIT		
1.5	**Campus**	Bloemfontein		
1.6	**Postal address**	33505 Bronville, Welkom. 9463		
1.7	**Contact details**	Office		
		Cell	071 059 7671	
		e-mail	pamelaramba3@gmail.com	
1.8	**Supervisor (s)/Project Leader**	Ms Mpho Mbele Prof M Masinde		
1.9	**Qualification registered for/Level of research**	***Please tick relevant option:***		
		Masters qualification		X
		Doctorate		
		Independent research (Non-qualification purposes)		
1.10	**FRIC Approval Number (LS262a) (where applicable)**			
1.11	**Conflict of interest (Please underline/highlight):** 1) Personal relationship Yes/No			

	2) Financial benefit	Yes/No
	If yes, please provide details:	

2. DETAILS OF THE STUDY

2.1	Approved/Proposed title of the study/project /dissertation/thesis

Integration of Fourth Industrial Revolution technologies and indigenous knowledge in developing a smart and integrated pollution monitoring system for Lejweleputswa mines

2.2	Research question(s)

1. What are the indigenous knowledge systems that are applicable for monitoring pollution in South Africa's underground mines?
2. What are the wireless sensor network-based solutions available for monitoring pollution in the mines and which ones are applicable to South Africa's underground mines?
3. To what extent can a sensor network based underground mines pollution monitoring solution be used to complement indigenous knowledge systems-based pollution monitoring?

2.3	Aim and objectives of the study

1. Investigate the extent to which artificial intelligence and internet of things, have been incorporated into pollution monitoring systems in South Africa and beyond.
2. To investigate the indigenous knowledge systems that are applicable in monitoring pollution in South Africa's underground mines.
3. To develop and evaluate a generic framework that integrates artificial intelligence and internet of things with indigenous knowledge systems to achieve a smart and interoperable pollution monitoring system.
4. To evaluate and assess the effectiveness of the model developed in objective (3) above using the case of Lejweleputswa mines and the South Africa's Integrated Climate-driven Multi-Hazard Early Warning System (ICMHEWS).

2.4	Research methodology
2.4.1	Research participants and their age brackets (where applicable, e.g. 10 Students from Civil Engineering Department)

Please list with number of participants

30 Mine workers, 20 Lejweleputswa mining community residents.

All the participants must be older than 18 years of age, there will be no age restriction thereafter (as we are seeking indigenous knowledge), if they are interested in the study and have view in the topic/ want to learn more on the topic.	
2.4.2	**How will participants be selected/sampled?**
Participation to the research work is voluntary. **Participants are chosen based on their occupation (Mine workers) and people that experience mining related pollution daily (Local residents of Lejweleputswa district)**	
2.4.3	**Research site(s) (e.g., Borong Construction Site)** **Please list**
Online	
2.4.4	**Data collection instruments (e.g., questionnaire(s)/interview schedule(s)/observation schedule(s)/artefacts/other)**
List all instruments to be used and attach copies/a copy/ schedule • ***For each data collection instrument, explain the quality measures to be observed*** • ***Please attach a copy of all data collection instruments to be used in the study (where applicable).*** • Online questionnaire using survicate free version. The questionnaire contains both open-ended questions and closed-ended questions. • All the Online questionnaire using responses will be re-routed back to the survicate portal for further analysis.	
2.4.5	**Data collection procedure (Please outline WHEN, WHERE and HOW data will be collected)**
	Indigenous Knowledge data collection **WHEN**: from February 2023 to March 2023 **WHERE**: Online using survicate free version **HOW**: Each participant will receive a link to the questionnaire, follow the link and complete the questionnaire. Response will be automatically submitted to the survicate portal.

3. PROPOSED PLAN OF STUDY/RESEARCH

	Set out your intended plan of work for the research, indicating important target dates necessary to meet your proposed deadlines	
	ACTIVITY	**DATE**
	Indigenous Knowledge data collection	February 2023 – March 2023
	Sensor Data collection	December 2022 – February 2023
	Data Analysis	April 2023 – May 2023
	Data Integration	June 2023 – July 2023
	Dissertation writeup	July 2023 – September 2023

4. ETHICAL ISSUES AND RISK ASSESSMENT

To assess whether your proposed research is ethically compliant, ethics risks are categorised into four categories:

(1) Research involving minor risk
The likelihood of projected harm or inconvenience in the research is not greater than that experienced in daily life.

(2) Research involving low risk
Research in which the only anticipatable risk is one of potential awkwardness or discomfort to the participants.

(3) Research involving medium risk
Research in which there is a possible risk of harm or discomfort, but where appropriate steps can be taken to lessen or moderate overall risk.

(4) Research involving high risk
Research in which there is a real and foreseeable risk of harm and discomfort, which may lead to a serious adverse event if not managed in a responsible manner.

		Yes	No	N/A
4.1	**Will human research participants be used in your study?** *Please mark with an X or √ in the Yes/No/N/A box*	**Yes** X	**No**	**N/A**
4.2	**If yes, does the research study involve any of the following:**			
a) Children or youth under the age of 18 (Attach parental consent letter)			X	
b) Individuals living with disabilities (physical, mental and/or sensory) (Attach consent letter of legal guardian)			X	
c) Individuals that might find it difficult to make independent and informed decisions for socio, economic, cultural, political and/or medical reasons			X	
d) Communities that might be considered vulnerable, thus finding it difficult to make independent and informed decisions for socio, economic, cultural, political and/or medical reasons			X	
e) Individuals who might be vulnerable for age related reasons e.g. the elderly			X	
f) Individuals whose spoken language differs from the language used for the research (Make sure you translate your consent form and participant information sheet in the participants' first language – you should also have an interpreter if you do interviews – describe it below the table)			X	
g) Women considered to be vulnerable (pregnancy, victimisation, marginalised etc.)			X	
h) Other (Please explain):				

4.3	Will data collection involve any of the following:	Yes	No	N/A
a)	Access to confidential data without prior permission of participants		X	
b)	Participants expected to commit an act which might reduce self-respect or cause them to experience shame, embarrassment, or regret		X	
c)	Expose participants to worrying or upsetting questions or to processes which may have disagreeable or harmful side effects		X	
d)	The use of stimuli, errands or procedures which may be experienced as stressful, harmful, or hostile		X	
e)	Any use of materials risky to human beings		X	

4.4 If you answered "Yes", to any of the previously mentioned, explain (attach as an appendix) and justify. Explain, too, what steps you will take to minimise the potential stress/harm. ***(Please indicate if it is not applicable to your study)***
N/A

4.5 Confidentiality of participants' identity				
4.5.1	**Will the identity and privacy of participants be protected through pseudonyms or other forms of identification and the use of an informed consent form, which specifies (in a language that participants will understand):** *Place an 'X or ✓' in the Yes/No box*	**YES** X	**NO**	**N/A**
4.5.2	**Please note that participants should be informed about the following (where applicable)**			
a)	The purpose/s of the research and how it is conducted	X		
b)	The researcher, project leader and supervisor's identity, their institutional association, and their contact details	X		
c)	Voluntary participation of participants	X		
d)	Making sure that participants' responses will be treated in a confidential manner	X		
e)	Be transparent about any possible limits on confidentiality which may apply	X		

f) Ensuring participants that they are free to withdraw from the research at any time without any negative or undesirable consequences to themselves	X		
g) How the findings of the study will have any benefits, or may receive as a result of their participation in the research	X		
4.5.2 Please attach the proposed consent and assent documents prepared to address all the above, if not a full explanation is needed explaining how will participants be respected and protected.			

5. DOCUMENTS TO BE ATTACHED TO THE APPLICATION

The following documents must be attached as a prerequisite for approval to undertake research in the Department (where applicable)	
5.1	LS 262a approved by the FRIC (FEBIT)
5.2	Proof of registration/Funding received and funder reference details
5.3	Data collection instruments as identified under 2.4.4

6. DECLARATION BY THE APPLICANT

I undertake to use the information that I acquire through my research, in a balanced and a responsible manner. I furthermore take note of, and agree to adhere to the following conditions (where applicable):

a) I will schedule my research activities in consultation with the relevant Company or Organisation and research participants (where relevant);
b) I agree that involvement of participants in my research is voluntary, and that participants have a right to decline to participate.
c) I will obtain signed consent forms from participants prior to any engagement with them;
d) I will inform participants about the use of recording devices such as tape-recorders and cameras, and participants will be free to reject them if they wish;
e) I will honour the right of participants to privacy, anonymity, confidentiality, and respect for human dignity at all times. Participants will not be identifiable in any way from the results of my research, unless written consent is obtained otherwise;
f) All interviews (recordings) will be transcribed verbatim and analysed as per conventional data analysis techniques (example(s) of interview transcript to be included in final dissertation)
g) I will adhere to the principles of rigorous data collection, analysis and interpretation consistent with the design of the study;
h) I will keep a data trail for possible auditing purposes as well as the safe keeping of raw data for a period of three years after publication of the results;

i) I will send the draft research findings to research participants before finalisation, in order to validate the accuracy of the information in the report;
j) I will not use the resources of the university when I am conducting my research (such as stationery, photocopies, faxes, and telephones) and
k) I will include a disclaimer in any report, publication or presentation arising from my research, that the findings and recommendations of the study do not represent the views of the Central University of Technology.
l) Aside from laboratory as well as consumables or materials supplied by the university needed to complete practical projects which might be central to my study (dependent on study field), I will not use the resources of the University when I am conducting my research (such as stationery, photocopies, faxes, and telephones).
m) All practical artefacts produced in support of my study using the university's laboratories, consumables, and materials will remain the property of the University.
n) If I supplied my own materials and consumables, I will permit access to all practical projects or artefacts to the University for a period of three (3) years for exhibition purposes.
o) All data collected for the research (including, but not limited to, completed questionnaires; statistical analysis performed on the data; interview audio-files/transcripts; artefacts/audio-visual materials; documents) will be kept safe at a designated space at the university for a period of at least three years. Computer files will be backed-up and password-protected.

I declare that all statements made in this application are true and accurate. I accept the conditions associated with the granting of approval to conduct research and undertake to abide by them.

STUDENT SIGNATURE / PROJECT LEADER SIGNATURE / SIGNATURE OF RESEARCHER	
DATE	06-03-2023

7. DECLARATION BY SUPERVISOR(S) (where applicable)

I/We declare that I/we shall oversee the student's adherence to all statements as set out above.	
SIGNATURE (Main supervisor)	
SIGNATURE (Co-supervisor)	
DATE	**06-03-2023**

FOR OFFICIAL USE
APPROVAL OF FEBIT ETHICAL COMMITTEE (FRIC)

Please tick relevant decision and provide conditions/reasons where applicable			
Decision			***Please tick relevant option***
1.	Application approved		√
2.	Ethical clearance number	FRIC: 2023-02(1)	
3.	Application approved subject to certain conditions. ***Specify conditions below***		

Central University of Technology, Free State

4.	Application not approved***. Provide reasons for non-approval below***	

SIGNATURE: **Chairperson: Ethics committee**	
DATE	**07/03/2023**

Cc Dean: FEBIT

Central University of
Technology, Free State

APPENDIX C: SYSTEM FEEDBACK QUESTIONNAIRE

Voluntary Participation Concerned Form

Study Title:

Integration of forth industrial revolutions technologies and indigenous knowledge in developing a smart and integrated pollution monitoring system for Free State mines.

Dear Participant

My name is Pamela Ramba, currently registered as a full-time master's student at the Central University of Technology, Free State. and I am doing research-based study under the supervision of Mss Mpho Mbele (lecturer, Information Technology department) and Prof. Masinde (Associate professor and HOD, Information Technology).

a) The purpose of the research

The overall aim of this research is to apply at least two 4IR technologies in conjunction with indigenous knowledge in designing a generic pollution monitoring system that is tailor-made for South African mining context.

b) Contact details

Researcher: Pamela Ramba (pamelaramba3@gmail.com)
Supervisor: Mss Mpho Mbele (mmbele@cut.ac.za) - lecturer in the Department of Information Technology at the Central University of Technology, Free State

c) Voluntary participation

Participating in this study is voluntary and you are under no obligation to consent to participation. You are free to withdraw at any time and without giving a reason. You will not be penalized or lose benefits if you refuse to participate or decide to terminate participation.

d) Confidentiality

All information taken from the study will be coded to protect each participant's identity. No names or other identifying information will be used when discussing or reporting data. The researcher will safely keep all files and data collected in a secured locked cabinet.

e) Possible limits on confidentiality which may apply

All data gathered will be regarded as confidential.

f) Potential benefits or risks

As part of the research study, an android application will be given to the participants as a study deliverable, for air quality readings.
There are no known risks associated with the deliverable

I voluntarily agree to participate in this research program
□ Yes
□ No

I understand that I will be given a copy of this signed Consent Form.

Name of Participant (print):______________________________
Signature:____
Date:____

Name of Witness (print):_________________________________
Signature:____
Date:____

APPENDIX D: VOLUNTARY PARTICIPATION CONSENT FORM

Application Feedback Questionnaire

1. Did You used the application

   Mark only one oval.

   ◯ Yes
   ◯ No

2. Please state which type of a user did you sign up as

   Check all that apply.

   ☐ IK Expert
   ☐ Administrator
   ☐ Mine Worker
   ☐ Local Community Member

3. On a scale of 1 -10 , Please rate the level of your satisfaction

   Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Not	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Satisfied

4. Did the application meet your expectations?

   Mark only one oval.

   ◯ Yes
   ◯ No

5. If Yes, Please elaborate how your expectations were met

6. If No, Please elaborate how your expectations were not met

7. In your own words , how was your experience using the application

8. Did the application address the pollution problems ?

Mark only one oval.

- Yes
- No

9. Was the application user friendly?

Mark only one oval.

- Yes
- No

10. For future work, is there anything you would like to see incorporated into the application

11. Do you think e application is expandable and can cater for future innovations?

Mark only one oval.

- Yes
- No