An intelligent Internet of Things air pollution monitoring system: case of a local municipality in South Africa

Ntombikayise Koyana

Dissertation submitted in fulfilment of the requirements for the degree:

Master of Engineering in Electrical Engineering

in the

Department of Electrical, Electronic and Computer Engineering
Faculty of Engineering, Built Environment, and Information Technology

at the
Central University of Technology, Free State

Supervisor: Prof. E.D. Markus
Co-Supervisor: Prof. A.M. Abu-Mahfouz

Bloemfontein
2023

Central University of Technology, Free State

DECLARATION

I, Ntombikayise Koyana, Student Number __________, hereby declare that this research project which has been submitted to the Central University of Technology, Free State for the degree of MASTER OF ENGINEERING: Electrical Engineering, is my own work which I completed independently. It complies with all applicable Central University of Technology, Free State, policies, procedures, rules, and regulations, and has not previously been submitted by another person in fulfilment (or partial fulfilment) of the requirements for the attainment of any qualification.

______________________ **2023/06/01**

Signature **Date**

DEDICATION

I dedicate this dissertation to the memory of my mother Nomelikhaya Koyana and to my father Sonwabile Goodluck Koyana, who have always been a source of support to me. I also dedicate it to my husband and family (Moyane), who have shown me how to stand firm in the face of difficulty whenever I want to achieve greater rewards. I will always be thankful, honoured, and appreciative of Jesus Christ, my Lord and Saviour, for the numerous blessings he has placed upon me.

ACKNOWLEDGMENTS

First and foremost, I want to express my gratitude to our Almighty Heavenly Father, who has enabled me in every way and supported me throughout this study. Glory to our Almighty Father, His Son Jesus Christ, and the Holy Spirit for all eternity! Amen!

My sincere thank you goes to the following individuals, to whom I would like to express my deep gratitude. To my supervisor, Professor Elisha Didam Markus, I appreciate the confidence you have placed in my abilities and the drive you have shown me throughout this challenging course. Your encouragement was unquestionably essential to my commitment to this study. I am also grateful to my Co-Supervisor Prof Adnan M. Abu-Mahfouz for his guidance and direction on Artificial Intelligence and Machine Learning. I would also like to acknowledge Dr Malusi Sibiya, from CUT for his valuable guidance during the study of Artificial Intelligence and Machine Learning. I also want to thank the following individuals: Gift and Kelebogile Rapoo, for their guidance and direction during my industrial visit at the “Bojanala Platinum District Municipality (BPDM)”, in the Northwest Province. I would also like to acknowledge Nelvia Phala for her guidance and direction. Furthermore, thanks to my colleagues from the department of electrical, electronic, and computer engineering who have provided ideas, feedback, and guidance. Finally, I want to express my gratitude to the other members of my thesis committee, FRIC, for their support, wise criticism, and challenging yet helpful questions. Thank you to the CSIR for giving me the chance and funding to conduct this research and to the “Central University of Technology, Free State” for allowing me to carry out this study.

I would like to take this opportunity to honour my partner and my children, Michael M Moyane and Siyamthanda L Koyana, Amogetswe G.A Koyana, and Tshepiso A.E Moyane - your unconditional affection, thoughtfulness, and unwavering support meant a lot to me during this crucial period of study in my life. I am so happy to have you. I want to express my gratitude to my closest friends for supporting me. You create happiness when it is most required.

ABSTRACT

Air pollution is fast becoming a major challenge in the 21st century, with its impending threats to the atmosphere and societal health at large. Recently, there have been extensive studies on the monitoring of air pollution and air quality. This interest is owing to the numerous issues and dangers encountered globally because of air pollution and poor air quality. As a fast-developing country, South Africa has been faced with the problem of air pollution that needs to be addressed to avoid the devastating effects on its populace. This dissertation therefore presents the results of a study designed to develop an intelligent air monitoring technique to verify the quality of air pollution among the public. This development was driven by a detailed analysis of existing air quality monitoring methods with a focus on the use of artificial intelligence that has become a part of today's society. The final product is an AI-based air quality surveillance system that is user-friendly for both society and consumers.

The Northwest region of South Africa is known for its mining industries. This is where agriculture and the provincial primary economic activities take place. As a result, air pollution has been a growing concern, causing diseases like bronchitis, asthma exacerbation, cardiovascular diseases in underweight infants, and even mortality. The evidence of negative health impacts and mortality from these mining activities and other transport emissions cannot be disputed.

The primary activity within the Northwest Province is confined to the main routes and mining industries. These include the Platinum Highway N4, N12 and N14, and Northwest province are focused on the Rustenburg and surrounding areas, which is over 90% of mining activities in the province. These activities have resulted in high levels of air pollution and the area is often tagged as an air pollution hot spot. The National Ambient Air Quality Standards (NAAQS) have sought to address the effects of air pollution on human health, but it is unclear what effects air quality has on terrestrial life. This research focuses on areas with high concentrations of air pollution, namely, particulate matter (PM), carbon dioxide (CO2), ozone (O3), nitrogen dioxide (NO2), and sulphur dioxide

(SO2). Therefore, this study explains the need to implement easy, low-cost methods to monitor and predict early hazardous symptoms. By utilising machine learning methods, the internet of things, and artificial intelligence systems more frequently to reduce air pollution, we can help the public (both commercial and residential), who are worried about their welfare by taking preventative measures to avert potential dangers. These technological advances will be particularly useful in predicting the state of air quality to quickly identify the probable source of expected problems, such as machine learning and the Internet of Things (IoT). As part of this research, a cost-effective IoT-based system for real-time monitoring of air quality and control was designed.

The designed hardware is equipped with five sensors, including a QM9, MQ2, Boost Voltage regulator, PM2.5 Air Quality Sensor, Adafruid Feather 32u4 FRM9x LoRa. A LoRa radio module is also added to facilitate radio transmission of sensor information from the Rustenburg Air Quality Station to the data collection facility. The innovation of technologies, such as the Internet of Things (IoT) and Machine Learning, has made it possible to monitor the quality of air contaminated in near real-time.

This dissertation has also explored the use of Random Forest Regressor and Multi-layer perceptron Neural Networks for predicting and monitoring the quality of air pollutants. Furthermore, algorithms such as Support Vector Machine, K-means clustering, DecisionTree, Regression Tree, and Linear regression are discussed. The study sought to use Fourth Industrial Revolution (4IR) tools and technologies like IoT, Artificial, and computational intelligence to develop an intelligent air pollution monitoring system. Random Forest outperforms other models with a prediction error of 5% less than the next best-performing model for the first prediction, 12% less for the second prediction,7% less for the fourth prediction,34% less for the fifth, and 36% less for the sixth prediction. Although it came second in the Neural Network Time Series (NNTS) in the third prediction test with an error percentage of 10.41% as compared to NNTS an error percentage of 6,71% it had the most stable accuracy with a maximum prediction error of less than 15%.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AIR	Atmospheric Impact Report
AI	Artificial intelligence
ANN	Artificial Neural networks
AQA	Air Quality Act
AQMP	Air Quality Management Plan
AQMNS	Air Quality Monitoring Network Station
BPDM	Bojanala Platinum District Municipality
BPLM	Bojanala Platinum Local Municipality
CO	Carbon monoxide
CO2	Carbon dioxide
CNN	Convolutional Neural Network
DT	Decision Tree
DEA	Department of Environmental Affairs
IoT	Internet of Things
LPWAN	Low-power wide-area network
LoRa	Long Range
LSTM	Long Short-Term Memory
MLP	Multilayer Perception
MSE	Mean Squared Error
MAE	Mean Absolute Error
RMSE	Root Mean Squared Error
MLR	Multiple Linear Regression
NAAQS	National Ambient Air Quality Standards
NO	Nitrous oxide
NO2	Nitrogen dioxide
NOx	Oxides of nitrogen
O3	Zone
PM10	Particulate matter with an aerodynamic diameter of less than 10 $\mu g/m^3$

PM2.5	Particulate matter with an aerodynamic diameter of less than 2.5 $\mu g/m^3$
Ppb	Parts per billion
Ppm	Parts per million
Pb	Lead
RLM	Rustenburg Local Municipality
RELU	Rectified Linear Unit
RF	Random Forest
SAAQIS	South African Air Quality Information System
SANS	South African National Standards
SVR	Support Vector Regression
SO2	Sulphur dioxide
SANAS	South African National Accreditation Services
WHO	World Health Organisation
WBPA	Waterberg–Bojanala Priority Area
WDM	Waterberg District Municipality
WHO	World Health Organisation
4IR	Fourth Industrial Revolution

TERMINOLOGIES

N_i	Input parameters
N_o	Output parameters
N_{mlp}	Hidden neurons
R^2	Correlation coefficient
P_i	Predicted value.
O_i	Actual value
n	Number of output nodes
Σ	Sigma/ sum up.
w	Neural network's weight
b	Neural network's biases
$\mu g/m^3$	Micrograms per cubic meter

CHAPTER I

1. INTRODUCTION

1.1 BACKGROUND

As a developing country, South Africa is faced with the problem of air pollution that needs to be addressed to avoid the devastating effects on its populace. Air pollution is one of the biggest challenges affecting the environment and public health in the world today. Although there are many techniques to address this challenge, the use of artificial intelligence and wireless sensors remains inadequate Over the years, numerous methods for tracking air pollution have been used. Some of them include the presentation by Adriana Simona Mihait; a mobile air pollution monitoring framework, along with a data-driven modelling approach for forecasting the quality of the air in urban areas, the data-driven investigation showed that data produced by the mobile sensing units when used outdoors can be accurately used to predict future Nitrogen oxide levels in urban using WIFI evaluation of the data accuracy. The study explored various air quality monitoring techniques, placing particular emphasis on the application of artificial intelligence and machine learning techniques that need to be developed to aid the process. These air pollutants have grown to be a major issue because they have a variety of negative effects on human health, including chronic obstructive pulmonary illnesses [1]. Many studies, for example, have shown that the air we take in affects our health and the indication of adverse health effects due to ambient air pollution has grown dramatically [2]. Exposure to outdoor and interior air pollution is now generally acknowledged to be a factor in a plethora of health problems, including minor physiological effects to death from respiratory and cardiovascular complications. As a result, air pollution has negatively affected human health significantly [3].

Furthermore, according to estimates from the World Health Organization (WHO), many children worldwide consume toxic air. According to the study [4], air pollution-related acute lower respiratory infections claim the lives of six thousand children each year. In South Africa and throughout the world, lower respiratory tract illnesses were the main factor in undetermined deaths. The report further asserts that pollution has caused many indescribable mortalities. Air pollution is also a driving force for non-communicable diseases which have become a global phenomenon particularly in Africa. These include, strokes, ischaemic heart disease, lung cancer, and other acute respiratory diseases, all associated with air pollution. According to the World Health Organization (WHO), 9 out of 10 individuals breathe air that contains high amounts of pollutants, and both indoor and outdoor air pollution are responsible for 7 million deaths annually. According to recent estimates, outdoor air pollution contributed to 4.2 million deaths in 2016, including 25% of adult deaths from strokes, 43% chronic obstructive pulmonary disease, and 29% from lung cancer [5]. The situation is tormenting since worldwide, 93% of the world's children under 15 years old, are exposed to particles with an aerodynamic diameter smaller than 2.5 $\mu g/m^3$.

Air pollution has become a growing concern in Rustenburg. This is a developed city in South Africa with many activities particularly in the mining and agriculture industry. As a result, it is known to contribute to illnesses like bronchitis, asthma exacerbations, cardiovascular diseases in underweight infants and mortality. The Waterberg–Bojanala Priority Area (WBPA) is in South Africa and covers an area of 67 837 km. To ensure that the air pollution levels in this region stay within the National Ambient Air Quality Standards (NAAQS), this area needs an air quality management action plan [7]. The management of air quality and pertinent parties need to be informed about possible risks and effects. The NAAQS addresses how air pollution affects health, but the impact of air quality on the terrestrial is not known.

Consequently, the focus of this research is artificial intelligence air pollution monitoring within South Africa, in the case of the local municipality in the Waterberg-Bojanala Priority Area. It is undeniable that transportation emissions, which are the root cause of these negative impacts, and mining industries have adverse effects on health and mortality [8]. The main activity within the Northwest Province is confined to the main routes and mines industries. These include Platinum Highway N4, N12, and N14, and Northwest province is focused on Rustenburg and surrounding areas, which has over 90% of mining activities in the province. Hence, this explains the need to implement easy, low-cost methods to monitor and predict early hazardous symptoms. Designing an air pollution monitoring system using artificial intelligence principles can be a valuable tool to assist the public, both industrial and residential, in addressing their well-being concerns and taking preventative measures to avoid imminent dangers. Here are some crucial areas in table 1.1 where the system can be implemented.

Table 1.1: Case study of areas with industrial air pollution problems

AREA	SOURCE OF POLLUTION	CRITICAL AREAS IDENTIFIED	SOURCE OF INFORMATION
Northwest	Windblown dust and Vehicular fumes.	Boitikong, Marikana	Rustenburg AQMP

This dissertation presents various intelligent air monitoring techniques, with an evaluation of the air quality. Although Okokpujie, K.O also developed a smart air pollution monitoring system that continuously tracks the levels of NO2 and CO2 in the air, raising awareness of the daily air quality [9]. This is meant to log data to a remote server, keep it updated online, and give real-time measurements of air quality while simultaneously monitoring and analysing air quality.

Furthermore, Swati Dhingra developed an Android application termed the Internet of things "(IoT) – Mobair", so that users can obtain pertinent cloud-based air quality data [10]. Poonam Pal also developed an IoT-based system for tracking air pollution that will sound an alarm when it drops below a certain threshold. The advantage of the system is cost efficient and less complex circuitry [11].

Despite these advanced techniques in monitoring air pollution, more research needs to be done for instance the comparison of these techniques to evaluate their efficiency and performance still requires more attention. Furthermore, these techniques fail to provide early warning signals necessary to warn residents of impending air quality problems.

1.2 PROBLEM STATEMENT

Rustenburg is a highly polluted suburb, in the Northwest province of South Africa. For the past ten years, residents have been complaining about toxic poisonous air that has affected their health [12]. The problem of air pollution has been exacerbated by industrialisation and modernisation and hence has become an increasing health hazard. Furthermore, South Africa is a fast-developing nation with a lot of mining, agriculture, and other industrial activities. As a result, air pollution-related illness has become increasingly prominent. Many systems and techniques have been applied to curb the threat. Given historical and present data, a variety of Artificial Intelligence (AI) algorithms could be used to identify air quality problems. Nevertheless, AI is limited in and of itself because it only makes it possible to recognize patterns and trends that may improve the quality of the air. The residents of Rustenburg, need easy, low-cost methods to monitor and predict early hazardous symptoms. Therefore, there is a need for the continuous Intelligent Internet of Things and Machine Learning to monitor air quality parameters in real time.

1.3 MOTIVATION FOR THE STUDY

Data mining is a computational technique that is frequently used to analyse large data sets, identify trends, extradite illegal information, and forecast the outcomes of upcoming or ambiguous events. These techniques, which combine a variety of computational fields, are used in the process, and include Artificial Intelligence, Machine Learning, Mathematics, and Database Systems. Participating in a post-processing phase, they are typically used to envision the results of the analysis by recognising patterns or retrieving information in an instinctive and easy-to-communicate manner. This model can quickly and effectively narrow the area of study and concentrate on fundamental data analysis methods as they have been used in the study of air pollution epidemiology. Artificial intelligence is an influential technique extensively used in multi-disciplinary production because of its adaptability to address real-world challenges of today.

Artificial Neural networks (ANNs) have a solid regression analysis data processing device that benefits from the capacity of neural networks to map nonlinear interfaces, which is trickier and less cost-effective equated to utilizing customary period series analysis. The model parameters for air pollution are the most important. The conformist procedures of modeling count on datasets that contain a substantial number of vague or indefinite recorded data and typically consist of time-consuming processes. The environmental model, forecasting developments in air quality performs an important role in society. Even though numerous research, artificial neural networks have been implicated in predicting air quality, the forecast regression is greatly influenced by the inputs of the model and the alignment of the neural network [13].

The most important impact of this study is the development of the artificial neural model to predict the quality of pollutants measured in the local municipality in South Africa. The evident limited Air Quality Management (AQM) capacity has an overarching effect on the aspects of current quality management techniques embarked upon by the municipality Data for this study were gathered through interviews conducted during industrial visits, questionnaires, and a review of the literature, the proposed monitoring system is faster, more efficient, and cost-effective.

1.4 AIMS AND OBJECTIVES

The aim of this work was the development of an air pollution monitoring system for early detection and warning of harmful substances in the air for the residents of Rustenburg and the surrounding area.
The objective is to develop a system that will enhance and capacitate the current system used at the “National Ambient Air Quality Monitoring Network (NAAQMN)”, for air pollution. This system should have the following characteristics [13-14].

- To be able to alert authorities and residents of impending dangers because of air pollution.
- Ensuring sustainable monitoring and continuous improvements in data quality.
- To use collected data to predict impending air quality issues using machine learning.

- To assess the prevailing AI algorithms and identify the suitable algorithms to use for learning trends based on the data gathered.
- To implement a real-time system for monitoring air quality, using the IoT and evaluate the performance of the industrialised real-time system-based pollutant sensors and AI learning algorithms.

1.5 EXPECTED OUTCOME OF THE STUDY

Scientific outcomes: The study is documented in publications and a master's dissertation. Machine learning (ML) methods can provide insights into the accuracy and effectiveness of these models in predicting air quality, as well as identify the best model for accurate air quality index (AQI) prediction. ML approaches studied for air pollution prediction include Support Vector Machine, Random Forest, and Artificial Neural Network.

Social impact: Increase awareness of air pollution to save South African communities from its impacts.

1.6 RESEARCH METHODOLOGY

To achieve the objectives of the study, the methodology followed was as follows:

1.6.1 Case Study Analysis

There are different methods of evaluating pollutants, but the purpose of this dissertation is to explore how Northwest normalises pollution and the volumes of pollutants in the different districts. Altogether current air quality trends will be directly studied, and data will be compiled for training using the Machine Learning (ML). The research at Bojanala Platinum Local Municipality (BPLM) was taken to see how issues handled there in the station and what kinds of current technology are being used to ensure good air quality. All air parameters, such Sulphur dioxide SO_2, Particulate matter $PM_{2.5}$, Particulate matter PM_{10}, Nitrogen dioxide NO_2, Oxides of nitrogen NO_x, HUMIDITY, WIND DIRECTION, TEMPERATURE, etc., are there for the intended data will be obtained. The data collected and stored on a database of

SAAQIS would make it easier to produce profound information to the beneficial and residences for the analysis. Due to the maintenance and lack of equipment for the air monitoring station being identified, various AI algorithms will be considered for further use and implementation.

1.6.2 Data Gathering

Gathering data is a time-consuming task that demands a variety of evaluation methodologies to be identified. To gather the requisite data for this research study, a quantitative research approach is used. Some data is gathered from the SAAQIS database, questionnaires, and informal interviews in the Local Municipality [16]. The Data Analysis and Prediction Models were selected, the data was understood, and machine learning algorithms with optimised hyper parameters may be achieved. In this study, three machine learning algorithms were used on a data set of thirteen gas input features to predict the concentration of the hazardous PM 2.5 gas in the air and it was carefully interpreted and standardized to avoid overfitting.

1.6.3 Data Pre-Processing

Historical data for the air quality has been carried from SAAQIS and the pre-processed dataset with 13 predictors and one target feature which is PM2.5. The data was interpreted, analyzed, standardized, and added Seaborn and Matplotlib libraries were used, and the histograms and scatter plots were generated to understand the feature distribution in the data set.

1.6.4 Forecasting Of Air Quality

The data was collected over a four-year cycle of historical data. 90% of the data was used to train the network, and 10% was used for testing and validation. The ANN model is then trained using a subset of the test data and validation of the parameters is carried out as shown in Figure 1.1 below.

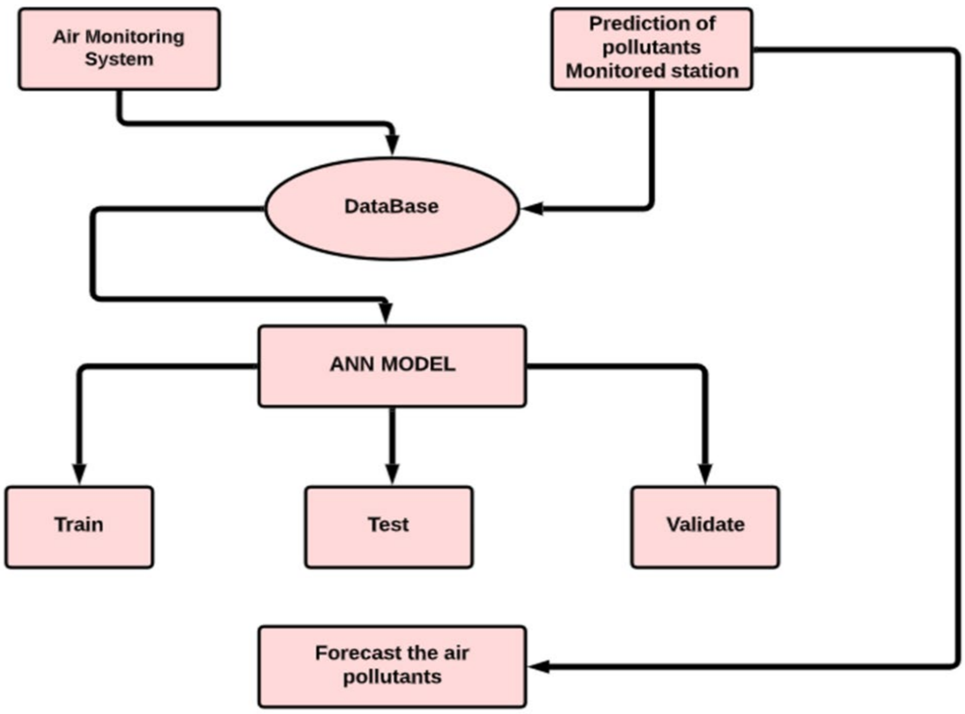


FIGURE 1.1: BLOCK DIAGRAM OF THE MODELLING PROCESS FOR FORECASTING AIR POLLUTANTS

Artificial neurons, which are essentially neurons or nodes, are the fundamental processing components of neural networks. In the simplest ANN model, learning is accomplished by modifying the weights in accordance with the learning algorithm. The weights reflect the effects of the synapses and capture the impact of various input signals. As shown in Figure 1.2, the standard ANN model employs a multi-layered configuration. A stage output can be used to provide feedback and modify the weights.

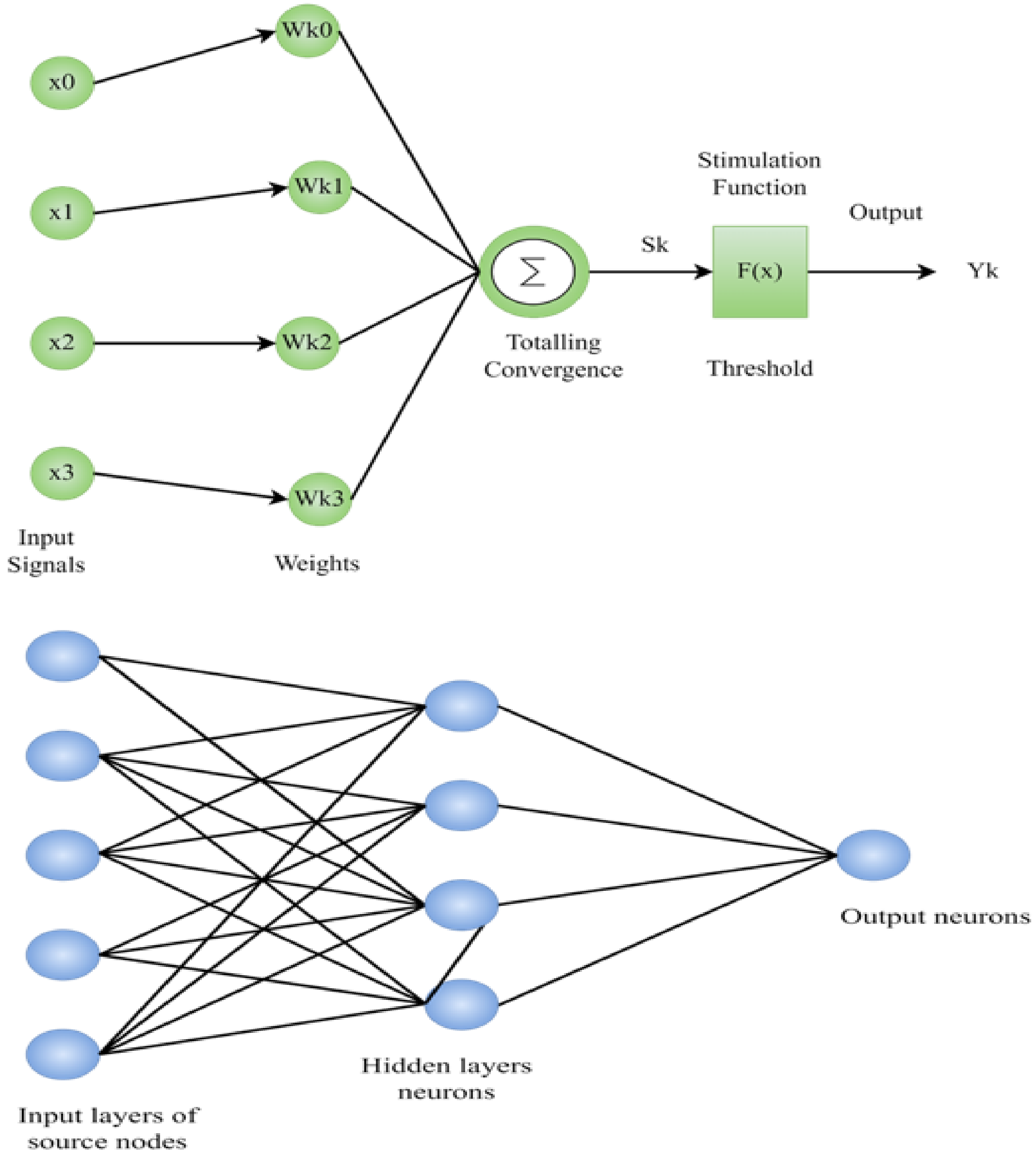


FIGURE 1.2: A FEEDFORWARD NEURAL NETWORK WITH MULTI-LAYERS AND A SINGLE OUTPUT

1.6.5 ALGORITHM ANALYSIS AND DESIGN

Respective AI algorithms will be discovered based on acquired data and analysis. Concerning the data collected, these algorithms can be used to learn trends and can be constantly altered and enhanced to keep up with emerging trends. Random Forest Regressor, Multiple Linear Regression, and Neural Network Predictions are among the forms of AI algorithms to be studied. Based on the data obtained from Bojanala Platinum Local Municipality, each algorithm will be tested to find useful gaps that can be addressed by that algorithm.

1.7 HYPOTHESIS

The ANN can be used to analyze a large dataset for prediction applications based on instant data with good accuracy. To reduce the nonlinear and inefficient existence of trends in the quality air pollution mandate, prognostication models with multi-variable pollutant parameter input and data organization structure from SAAQIS have been used to enhance the quality of air forecasting. The ANN can be used for this design to predict large datasets provided by SAAQIS. A custom-designed ANN technique based on a standard error backpropagation algorithm and a suitable learning method will be further developed.

1.8 SCOPE AND LIMITATION OF THE STUDY

In summary, the study's objective is to concentrate on a subset of pollutant forecasts, specifically SO_2, PM2.5, PM10, NO_2, CO, NO_x, humidity, wind direction, and temperature. The study generates real-time output utilizing ANN-based models constructed in Jupyter Notebooks with Python. The study's limitation is that it solely focuses on a certain collection of contaminants and does not consider other pollutants that may also affect air quality. Furthermore, the study solely employs ANN-based models, but alternative machine learning methods may be more accurate in forecasting air quality.

A low-cost IoT-based system for real-time air quality monitoring and control was implemented in this study. Nevertheless, it was unable to be tested at the Boitikong air pollution station at Bojanala Platinum Local Municipality due to COVID-19 lockdown restrictions.

1.9 DISSERTATION OUTLINE

The rest of the chapters provide deeper insights into the study. The background on air pollution was explored in Chapter 2. Chapter 3 will present a detailed perception of the work done, along with intelligent techniques. Chapter 4 will focus on the implementation approach, which is the methodology used to model the problem and the process followed behind the achieved results. Chapter 5 will explore the results obtained from the simulations and the final chapter will conclude by reviewing the research objectives as well as noting the future work.

1.10 PUBLICATIONS DURING THE STUDY

Conference papers:

Koyana, N., Markus, E.D. and Abu-Mahfouz, A.M., 2019, November. A Survey of Network and Intelligent air pollution monitoring in South Africa. In 2019 International Multidisciplinary Information Technology and Engineering Conference (IMITEC) (pp. 1-7). IEEE.

Journal papers:

Koyana, N., Markus, E.D. and Abu-Mahfouz, A.M., 2022, Early Air-Pollution detection in the Smart Cities using Machine Learning: A Case of Bojanala Platinum District Municipality in South Africa. Accepted for publication in *International Journal of Environmental Science and Development (IJESD).*

1.11 SUMMARY

In this chapter, the background of the study was presented. A motivation for the study was also highlighted stating the need for new air pollution monitoring techniques for smart cities. The problem statement was stated with a methodology of how this was approached. A hypothesis on the study was also included, publications during the study and a brief outline of the entire dissertation. In conclusion, a brief narrative of each chapter is presented.

CHAPTER II

BACKGROUND

2. INTRODUCTION

Chapter 2 of the dissertation discusses the background of air pollution surveillance and control schemes, as well as general air pollution monitoring methods. A thorough control of AQI station systems has been reviewed. Overall, Chapter 2 provides a comprehensive overview of the different methods and techniques used for air pollution monitoring and the pollutants most frequently monitored.

2.1 AIR POLLUTION SURVEILLANCE AND CONTROL SCHEME

The issue of air pollution has become serious in every major city in the world. Globally, respiratory health is negatively impacted by air pollution, a serious environmental problem. The high concentrations of fine particulate matter (PM) have resulted in productivity losses and human mortality. It has detrimental effects on human health; climate and the ecosystem in terms of air quality index (AQI). There are also a few alternative techniques for reducing air pollution that are mentioned. A further discourse is included to highlight the negative effects of air pollution and how different studies have attempted to address the issue. There is a particular focus on South Africa, where pollution levels are increasing because of radiation from industry and traffic. Studies demonstrate that South Africa continues to experience an infrastructure crisis [15]. Additionally, this study considers the applications and benefits of intelligent air pollution monitoring. According to estimates from the World Health Organization (WHO), many children worldwide consume toxic air. According to a 2016 WHO study, air pollution-related acute lower respiratory infections claimed the lives of 6 000 (six thousand) children. The main cause of death worldwide and in South Africa was lower respiratory tract infections [16].

The report further stated that pollution is a major contributor to mortality that cannot be determined, such as deaths from strokes, ischaemic heart disease, and lung cancer. Air pollution was also linked to other severe respiratory illnesses. Particulate matter, nitrogen dioxide (NO2), and Sulphur dioxide (SO2) are air pollutants that have grown in importance as a cause for worry due to their numerous negative effects on human health, including chronic obstructive pulmonary diseases [17]. Thus, the air we breathe has an impact on our wellbeing. The evidence for the harmful health impacts of ambient air pollution has significantly increased over the last 20 years. It is also acknowledged that there are significant and pervasive impacts on human health. In addition, air quality-related fatalities in Africa are among the continents top five killers. Results also indicate that air quality and human health in Africa are significantly impacted by natural sources of wind-blown dust pollution [18].

Although the ambient particulates were the most frequently used pollution index in these studies, the effects of gaseous pollutants like nitrogen dioxide (NO2), sulphur dioxide (SO2), ozone (O3), and carbon monoxide (CO) have also been extensively discussed in the literature [19]. These pollutants have a direct impact on the increased morbidity and mortality rate.

2.2 AIR POLLUTION AND INADEQUACIES IN LEGACY SYSTEMS

Even though everyone seems to recognise the seriousness of the problem with air pollution, it is challenging to find reliable, up-to-date information about it in the city. This makes gathering evidence and providing feedback for future policy more difficult. In addition, according to some studies, air pollution in an urban region of South Africa was estimated to be responsible for 5.1% of mortality due to trachea, bronchus, and lung cancers and 3.7% of national mortality from cardiopulmonary disease. In South Africa, the effects of air pollution on public health are underappreciated. In South Africa, air pollution from transportation and emissions from fossil fuel combustion continue to be major public health concerns [20-22].

The World Health Organization (WHO) developed the following comparative risk assessment (CRA) methodology in most urban areas annual mean concentration of Particulate matter (PM) [23-25].

In a separate recent study, it was revealed that analytical chemistry is a quickly developing branch that focuses on classifying air pollutants and figuring out their levels, which is a time-consuming and expensive process.

The study also provides details on the equipment used in different mobile laboratories. These mobile laboratories can be used for detecting and determining harmful substances in air pollutants. Several studies have shown the use of on-road mobile laboratories for measurements of gaseous pollutants and particles. The study also provides details on the equipment used in different mobile laboratories. Future developments in mobile air monitoring devices are discussed, as well as novel analytical instrumentation techniques that can be applied to air monitoring [26]. Other authors have proposed several advanced air pollution monitoring systems for evaluating foremost air pollutants like CO2, CO, O3, SO2, Particulate matter (PM) with better results [27]. Similarly, Kang et al. aimed to create systems for real-time monitoring and assessment of air quality that could support assessment and analysis of air quality at various levels. They investigated different big data and machine learning-based forecasting methods for air quality. By enhancing the accuracy and reliability of low-cost sensing technologies for air quality assessment, the study explicitly highlights the promise of low-cost sensor technology as a cost-effective alternative for monitoring air quality [28].

2.2.1 MAJOR IMPACTS OF AIR POLLUTION IN THE MUNICIPALITY

The creation of the Air Quality Management Plans (AQMPs) outlined what needed to be done to meet the established air quality requirements. An AQMP had to be a component of every municipality Integrated Development Plan. The philosophy of monitoring air quality in South Africa was stimulated by the implementation of the Air Quality Act (AQA), shifting from a single source demeaning strategy to a more holistic method based on effects on the receiving environment (human, plant, animal, and structure). The concept is based on proactive planning (air quality management plans for all municipal areas), licensing of specific industrial processes, and the identification of priority areas where the air does not meet the air quality standards for specific air pollutants.

In prior years, the BPDM was noted as having poor air quality. This statement was supported by the National Framework and the baseline assessment, which highlighted the industrial area that includes the Merensky Reef and extends from west of Pilanesberg through the Bafokeng neighbourhood and parallel to the Magaliesberg toward Marikana, Boitikong, and Brits in the east [29].
Co-located in the region are several auxiliary manufacturing industries. The BPDM district is home to Anglo Platinum and Impala Platinum, the two biggest platinum mines in the world. Slate, marble, granite, chrome, and lead are also produced in the area. The eastern boundary was discovered to be affected by air quality impacts from Rustenburg up to Brits. This region contains most of the household and industrial fuel burning sources. It was discovered that the other nearby municipalities had minimal pollution loads.

Small boiler sources and businesses like galvanising works and autocatalytic manufacturing were blamed for industrial emissions. The main pollutants emanating from these processes are PM2.5, PM10, CO, SO2, NOx, and Pb.
Agriculture is the primary land use in the area, and this activity is linked to PM10 and PM2.5 pollution. The biggest source of PM10, PM2.5, CO, NOx, and SO2 emissions from biomass burning was identified as wood, coal, and paraffin only used in low-income, heavily populated rural regions and squatter camps.

The N4 highway was found to be the primary source of concern for vehicle emissions. As the most industrialised municipalities in the region, Madibeng and Rustenburg LMs reported having high vehicle emissions. Several active landfills have been found to be sources of heavy metal, dioxin, and furan emissions in the area. Although pollutants were not listed, transboundary emissions were recognised as a source of pollution in the district [30-31].

2.2.2 Measuring air pollution in the local municipality

The Air Quality Act sought to implement Section 24 of the Constitution by establishing sensible safeguards against air pollution. According to the air quality act, municipal governments must include AQMPs in their integrated development plans. Several South African municipalities, including Rustenburg Local Municipality, have handled

their duties, and created AQMPs. This material was based on the Northwest provincial AQMP.
Additionally, the Air Quality Act provides for the designation of priority regions where the air quality is deemed to be poor and harmful to both the human condition and environment [32]. The Minister of Environmental Affairs and Tourism designated South Africa's Vaal Triangle as a region of significance. However, in the case of the Vaal Triangle Airshed Priority Area, the initiatives are intended to manage poor air quality in air pollution hotspot areas that span the Gauteng and Free State provincial borders (VTAPA).

The Highveld Priority Area was the second to be designated as a priority location in South Africa. Parts of the provinces of Gauteng and Mpumalanga are included in the Highveld Priority Area (HPA), which also contains a metropolitan municipality (Ekurhuleni) and three district municipalities (Sedibeng, Gert Sibande and Nkangala). Currently the Highveld Priority Area Air Quality Management Plan [33] is being created.

As a result, the Minister designated the Waterberg-Bojanala Priority Area (WBPA), which spans the Northwest and Limpopo Provinces, as the third National Priority Area (DEA). BPDM has been noted as having bad air quality by the Republic of South Africa’s National Framework for Air Quality Management. Since Rustenburg is in the Waterberg-Bojanala Priority Area, this research focuses on the deposition of particulate matter (PM10, PM2.5), and Sulphur dioxide (SO2) [34].

2.2.3 Inadequacies in the Legacy Systems

South African Air Quality Information System (SAAQIS) indicated that, all verified ambient monitoring data will be transferred to a central database being developed at the South African Weather Services (SAWS). It will be possible to easily access and manipulate data from any area of government, because of the incorporation of some source and emissions data recorded within each Municipality and Provincial, into a national electronic database. The BPDM will have to make sure that their database of current emissions is consistently updated; therefore, maintaining the stations will also help to incorporate with the SAAQIS. Figure 2.1, Illustrates a typical air quality monitoring station, indicating analysers for different pollutant. Figure 2.2 Illustrates a typical air quality monitoring station, indicating some of the systems are missing either for maintenance taking place or stolen.

FIGURE 2.1: ON-SITE AIR QUALITY MONITORING STATION

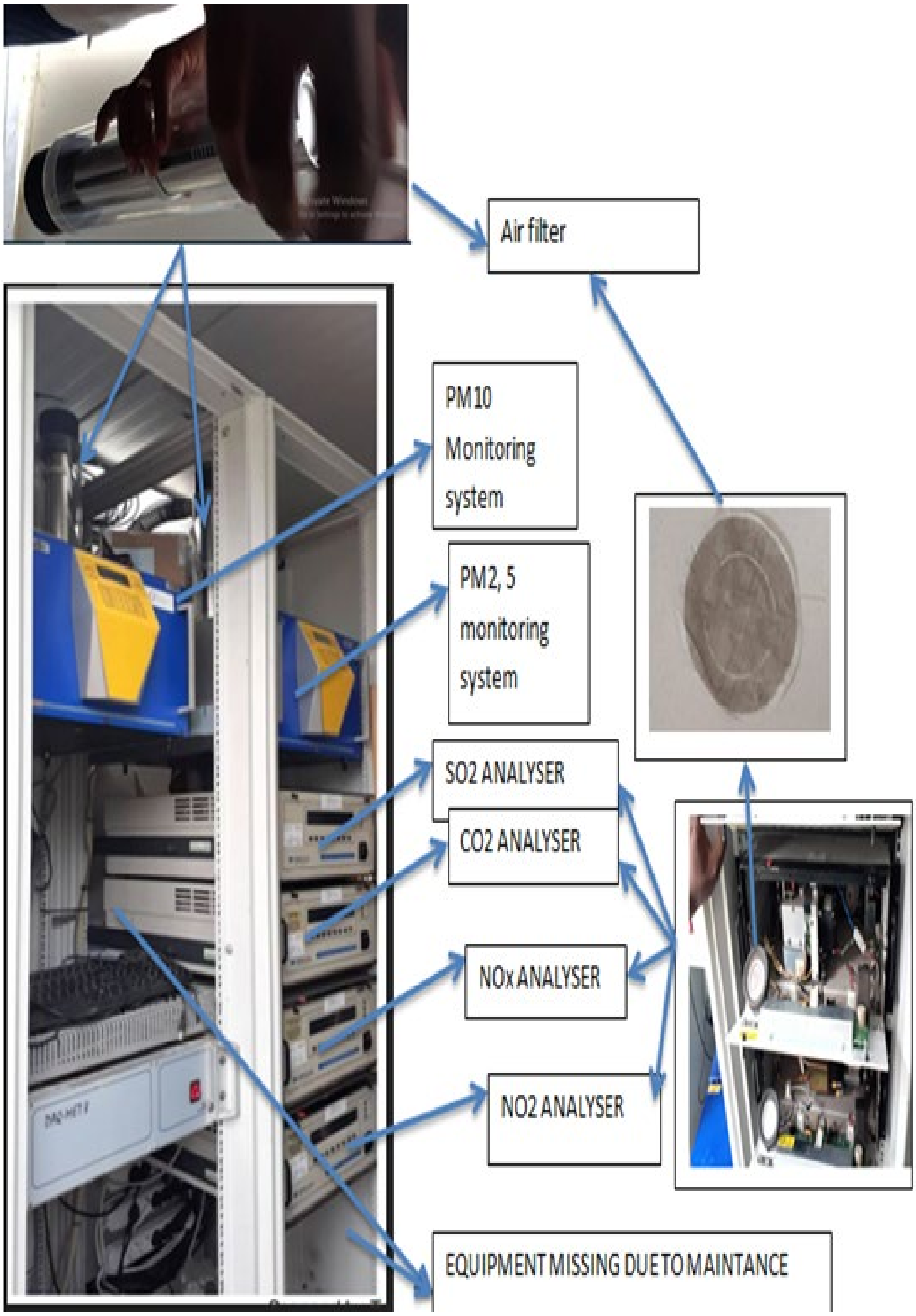


FIGURE 2.2: ANALYSERS FOR DIFFERENT POLLUTANTS

2.3 PRIMARY POLLUTANTS

A volcanic eruption produces and disperses primary pollutants, which are the dominant pollutants. Another is carbon monoxide gas, which is mainly made by a variety of vehicles exhaust fumes, while the industrial processes generate a large quantity of Sulphur dioxide. The following are primary contaminants.

2.3.1 CARBON DIOXIDE (CO2)

One of the most notable pollutants among others is carbon dioxide. It is generated quickly by a variety of factors and is more harmful than other pollutants. Usually, when CO2 is discussed, CO2 is referred to as the leading pollutant and the worst climate impact. Carbon dioxide naturally occurs in excess in the atmosphere and is essential for plants.

2.3.2 SULPHUR OXIDES (SO2)

It is formed by burning high-Sulphur fuel and by natural disasters such as volcanic eruptions. It is present in coal and petroleum it is of great concern to the environment.

2.3.3 NITROGEN OXIDES (NO2)

Nitrogen dioxide is a common example. This chemically toxic gas is produced because of elevated temperatures and other problems. It contributes significantly to the creation of air pollution. Fuel ignition can occur in commercial boilers, electric services, wood burning stoves, and vehicles. It is the root of illnesses that affect the lungs.

2.3.4 VOLATILE ORGANIC COMPOUNDS (VOC)

One of the main environmental pollutants is VOCs. Methane and non-methane components of these contaminants are separated. Methane (CH4), according to researchers, is an effective chemical for accelerating global warming. The fragrant non-methane (NMVOCs) benzene, toluene, and xylene are suspected cancer-causing agents that, upon prolonged exposure, may also cause leukemia.

2.3.5 CARBON MONOXIDE (CO)

This pollutant is a toxic gas that has no color or smell. Natural gas, coal, or wood combustion all result in the creation and dissemination of carbon monoxide. Nevertheless, a lot of carbon monoxide is released into the atmosphere by automobile exhaust. People who are affected by it experience headaches, fatigue, and a sense of drowsiness.

2.3.6 PARTICULATES MATTER (PM)

Particulate matter is a general term for tiny solid or liquid bits suspended in a gas. Now, a mixture of gas and particles is referred to as an aerosol. The elements have some natural origins. Examples include volcanic eruptions, sandstorms, forest fires, marine spray, etc. In addition, some human actions like burning fossil fuels, operating power plants, and conducting industrial processes all produce large amounts of aerosols. These particulates are spreading quickly and posing health risks like asthma, lung cancer, and heart disease.

2.4 SECONDARY POLLUTANTS

When primary pollutants interact, or react, they produce these pollutants in the air. These toxins are not immediately released. Interestingly, one of these contaminants is ground-level ozone. Not to mention that some pollutants can be classified as both primary and secondary. Three types of secondary contaminants are distinguished.

- Particulate matter from mixes and vaporized essential contaminants in photochemical smog. One form of atmospheric pollution is smog. An illustration of an exhaust plume is brought on by a region heavy coal use, which is brought on by a smoke and Sulphur dioxide mixture.

- NOx and VOCs combine to create ground-level ozone (O3). The troposphere most significant element is ozone (O3). Additionally, it is a crucial component of some regions of the atmosphere known as the ozone layer.
- Peroxyacetyl nitrate (C2H3NO5) is formed equally from NOx and VOCs.

2.5 AIR POLLUTION SOURCES

There are several elements believed to be responsible factors for the release of pollutants into the air. Anthropogenic (man-made) and natural sources are the two categories into which these factors or sources are divided.

Man-made sources:

- Static sources, including dung, wood, crop waste, factories, industries, and waste incinerators.
- Movable sources like aircraft, marines, and vehicles.
- Controlled Burn in agriculture
- Fumes released by paint, aerosol mists, varnish, etc.
- Methane affects waste being dumped in landfills.
- The military uses toxic gases, nuclear weaponry, rocketry, and germ warfare.

Fertilized agricultural soil that emits NOx. Natural Sources:

- Earth main sources of dust.
- Methane, the food for animals.
- Radium gas, which results from radioactive decay.
- Wildfire smoke and carbon monoxide.
- Vegetation.
- Volcanic activity.

The safe and unsafe levels of the above-mentioned air quality parameters are given, as a municipality issues the licenses to the facility for enforcement of complying with the standard National Ambient Air Quality Standards (NAAQS) which the facility needs. Because different pollutants have various impacts, there are various NAAQS standards, some of which are listed in Table 2.1. In this research, the Artificial Neural Network is used to predict air quality actual data collected in the air database. The parameters were included in the model to have the plan to prevent their emission. Ambient air needs to be controlled.

TABLE 2.1: NATIONAL AMBIENT AIR QUALITY STANDARDS

Parameter	**Averaging Period**	**Concentration** $\mu g/m^{33}$	**Frequency of exceedance**
SO2	10 minutes	191	526
	1 hour	134	88
	24 hours	48	4
O3	8 hours	61	11
NO2	1 hour	106	88
CO	1 hour	26	88
	8 hours	8,7	11
PM10	24 hours	75	4
PM2, 5	24 hours	40	4

CHAPTER III

LITERATURE REVIEW

3. INTRODUCTION

This chapter provides the theoretical outline of the work done. It starts by exploring air quality and the use of intelligent air monitoring, focusing the study on air pollution monitoring and control using machine learning. A thorough survey of the literature on air pollution control methods, air pollution intelligent systems design, operation, and control of AQI station systems has been reviewed. Conclusively, the chapter ends with reviewing the findings obtained.

3.1 USE OF MODERN TECHNOLOGY IN AIR POLLUTION MONITORING SYSTEM

South Africa, a fast-developing country, that is rapidly growing, has seen levels of air pollution increase due to vehicle and industrial emissions as well as dust from natural phenomena [38]. The statistics below provide an overview of how air pollution affects South Africa and the rest of the world.

- In 2014, 92% of people on the planet were residing in areas where the WHO air quality guidelines were not being fulfilled.
- In 2012, it was estimated that air pollution in both urban and rural regions contributed to 3 million premature deaths globally.
- It is estimated that ischaemic heart disease and strokes caused 72% of the premature deaths attributed to outdoor air pollution in 2012.
- For the roughly 3 billion people who use biomass fuels and coal to heat their homes and cook, indoor smoke poses a significant health risk. In 2012, 4.3 million premature fatalities were linked to household air pollution [39]. Low-middle-income nations like South Africa were the ones who brought about all this burden.

Reviewing and categorising the most pertinent papers on air pollution and its control according to the technologies used and contributions made, each of these systems' advantages and drawbacks have been explored and compiled. Focusing on the problem, a low-cost air quality monitoring system was designed using IoT, Machine Learning (ML), and Cloud Computing. This was proposed as a replacement for conventional methods of quality tracking. A literature review conducted by Koyana et al. discussed the current state of air pollution monitoring in South Africa and the potential of intelligent techniques to address the challenges. The study showed that there is a need to reassess the existing monitoring stations and to measure a complete set of pollutants, rather than focusing on specific ones [35]. The conclusion of this section recommends an intelligent air monitoring in real time such as ML in South Africa.

The main routes are where most of the activity occurs in the Northwest Province, including the mining sectors. These include the Platinum mining area. Studies show that 90% of mining activity is in the Northwest province. The mining and metallurgical sectors are the potential producers of air pollutants, along with landfills, haul roads used by mining and heavy transport vehicles, and fugitive emissions.
This has motivated an evaluation of the air quality regarding the particulate matter, the study will be undertaken in the areas of mining town in North-West province of South Africa. Hence, this explains the need to implement easy, low-cost methods to monitor and predict early hazardous symptoms [36]. This is by designing a system of air pollution control using artificial intelligence systems that will assist the public, both industrial and residential, who are concerned about their well-being, for them to take precautionary steps to preclude impending hazards.

Additionally, several techniques have been developed to tackle the issue of air pollution, such as South African Air Quality Information Systems, which occasionally display the air quality danger. As shown in Figure 5 below clearly shows that in the Northwest and the surrounding region, PM10 levels frequently exceeded the national ambient air quality standard [37-38].

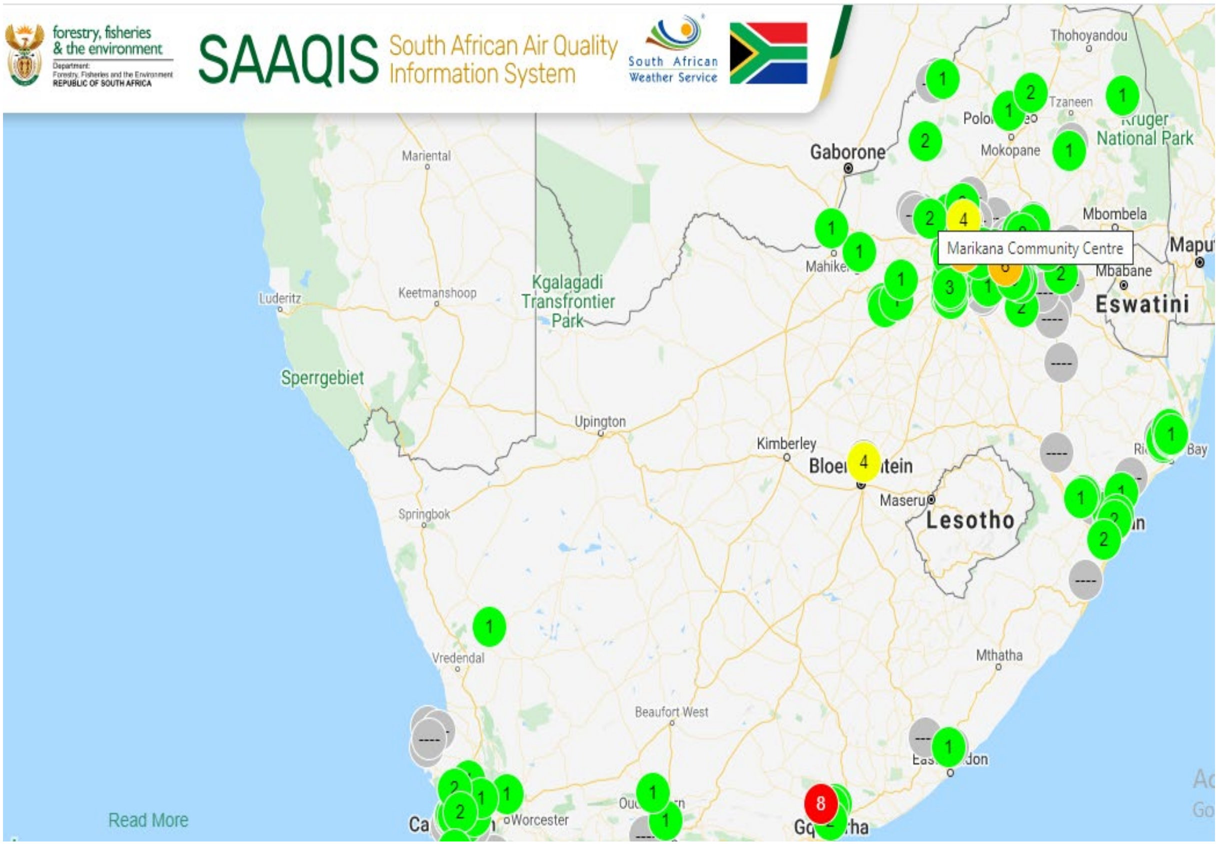


FIGURE 3.1: MONITORING STATION GRAPH MAP

However, the number of monitoring sites tends to be quite small in most areas due to the high costs associated with installing and maintaining such stations. As a result, the artificial intelligence-based air pollution monitoring system is intended to help the public, including businesses and homes. Thus, the need to handle the pollution issue must be addressed to prevent catastrophic effects on the populace.

3.2 USE OF INTELLIGENT TECHNIQUES

Several authors have discussed some intelligent techniques for air pollution monitoring. For instance, the authors suggested that Air-Pollution Prediction through Machine Learning Methods in Smart Cities should be developed. Their primary objective was to evaluate various machine learning approaches to forecast ozone levels. The aim of their research was to find the most accurate model for ozone forecasting [39 -40]. As indicated by Chiwewe the use of machine learning and deep learning to improve air pollution monitoring and environmental health in South Africa requires more attention [41]. Chiwewe et al. and Shikwambana *et al*. have been conducted in South Africa. Chiwewe *et al.* proposed a big data modeling approach to predict the ground-level Ozone levels based on the cross-correlation bandwidth readings for a station and the three-dimensional correlation among multiple stations in Gauteng province in South Africa. Similarly, Shikwambana *et al*. introduced a reliable method to monitor the temporal patterns of PM10 concentration using an observation camera and a regression algorithm calibrated by the atmospheric reflectivity and the measured air quality data [42]. However, Kumar demonstrated a real-time personal air quality monitoring system that considers several constrains, PM 2.5, carbon monoxide, carbon dioxide, and air pressure. Systems for tracking air quality now heavily rely on Internet of Things (IoTs). With the convergence of IoTs and cloud computing, a novel method for better controlling data from various sensors is proposed. This data is gathered and transmitted by the low-cost, ARM-based Raspberry Pi minicomputer [43].

Wei et al suggested an approach for implementing a fuzzy clustering strategy in parallel on heterogeneous servers with a CPU and several GPUs, specifically for classification issues. Sensors, actuators, and embedded computers are used in intelligent transportation systems (ITS), which are pioneering contributions to the application of information and communication technologies in vehicles to enhance safety. The concept of The Next Generation Air Pollution Monitoring System (TNGAPMS) was further developed, with emphasis on the need for the development of low-cost, real-time air pollution monitoring systems with high spatiotemporal resolution, and that further research is needed to improve the accuracy and reliability of these systems [44].

Established on the carriers of the sensing components, the Next Generation Air Pollution Monitoring System (TNGAPMS), divides sensor networks into three categories: static sensor networks (SSN), community sensor networks (CSN), and vehicular sensor networks (VSN). An air pollution monitoring system with high spatiotemporal resolution, cost and energy efficiency, deplorability and maintenance viability, and appropriate access for the public or professional users are all verified to be feasible [45-49]. Jenab, therefore suggests a fuzzy Bayesian air quality monitoring model that is capable of simulating cognitive behaviour like that of humans in environmental analysis. They created a fuzzy intelligent air quality model that employs workstation airborne particle density signals to assess whether the air quality is normal or abnormal. The model fuzzy Bayesian foundation makes it a viable candidate for classifying the state of air quality [50].

The design and execution of a safe, affordable, real-time air pollution monitoring system are presented in Guanochanga. Thus, a three-layer architecture scheme was put into practice. The Raspberry Pi, a data processing node connected to an Arduino platform via a wireless network, transmits messages using the Message Queuing Telemetry Transport (MQTT) protocol. The first layer of the system includes sensors. Additionally, a WEB service was provided to send the data for the subsequent analysis. On a computer, a smartphone, or through a web browser, you can reach the client layer [51].

The findings show that the suggested message architecture is capable of translating JSON strings sent by Raspberry Pi gateway node and Arduino-based sensor nodes. Information about various air contaminants has also been successfully displayed using web services. Genikomsakis also discusses the on-field testing of the proposed low-cost air pollution monitoring system (APMS) implementation using roadside measurements from a mobile laboratory outfitted with a calibrated instrument as the basis of comparison and demonstrates its accuracy on characterising the PM2.5 concentrations on 1 min resolution in an on-road trial. By contrasting the measurements obtained from the Mons fixed monitoring station with those obtained from the low-cost APMS mounted on an e-bike, this paper case study emphasises the need for obtaining high-resolution geolocation data of PM2.5 concentrations in an urban environment [52].

3.3 USE OF INTERNET OF THINGS

Doni et al. proposed establishing multi-sensor air quality monitoring using IoTs. Studies have shown that human data computation using laboratory techniques takes time and that current systems are ineffective at producing reliable results. Additionally, they do not support long-distance communication; however, this initiative makes use of a GPRS module, allowing for extremely long-distance data transmission [53-54]. The IoT-based low-cost air pollution monitoring device was also proposed by Parmar et al. [55]. Additionally, Chen describes an IoT-based system for tracking and forecasting air pollution. They demanded that technology costs could be cut by up to a tenth by utilising the Internet of Things (IoT) [56]. Souvik and Voguish examined the carefully planned monitoring of air pollution on the roads and track vehicles in their study. They suggested utilising IoTs to solve this issue. In providing an overview of the most relevant IoT architectures and protocols for air monitoring, crowd sensing has been presented as an authoritative key to address environmental monitoring, allowing to control air pollution levels in congested urban areas in a collective, low-cost, and inaccurate manner [57].

The authors developed a distributed energy-efficient algorithm known as the Hilbert-order Collection Strategy (HCS), which gathers data from a mobile wireless sensor network (WSN) and recognises environmental events using a mobile sink (such as a drone). To analyse various justification strategies to reduce pollution, Miles et al. proposed a decision support system (DSS) to address the issue of identifying phenomena like air pollution in the IoT environment. Consequently, they designed a sophisticated model that could be used to evaluate techniques and their impact on lowering crucial atmospheric pollution levels. By using the Internet of Things to monitor air quality and combining data from various heterogeneous data sources, urban environments can be improved as well as their surroundings and potentially dangerous areas [58].

To immediately monitor air pollution values and analyse air pollution, a much denser sensor network is needed. Smart sensing of large regions is made possible by Low-Power Wide-Area Networks (LPWAN), in particular LoRa, Sigfox, or NB-IoT. Liu et al., also propose a plan for wireless sensor networks-based air pollution tracking. The sensors are now placed in the environment to track the air quality constantly. These instruments include CO sensors as well. The ZigBee protocol is used to enable contact between the sensors. Data is gathered from the servers by the control centre, which also notifies the users via SMS. Data gathering and storage are done using MySQL [59].

Rushikesh presents a monitoring system based on the Internet of Things (IoT) that can identify vehicles polluting city streets and gauge the number of various pollutants in the air. When necessary, the paper updates the environmental authorities on the state of the air quality [60]. Luo utilises “a.com”, a popular low-cost air quality monitoring system that offers continuous and affordable air quality monitoring capacity to larger communities, to illustrate these problems. We use the name “a.com” for the business of interest to protect the air quality monitoring network under investigation [61].

The use of Single Board Computers for the integration of IoTs with WSN for Air Quality Monitoring System (AQMS) is revealed in Balasubramaniyan's (2016) study. Single Board Computers can complete even complicated tasks with enriched speed and summary convolution. The alerting process is intelligent and real-time, thanks to the merging of cloud services with Single Board Computers [60]. By using a Wi-Fi built plug and sensing device for dedicated IoT air pollution monitoring, Kumar et al. proposed a model to employ the IoTs devices to measure air pollution levels in a specific area. Since the air pollution monitoring system is an IoT application that needs devices to be placed in various locations, all regions are divided into grids and resources are allotted according to priority for each grid.

The results demonstrate that a prioritised planned deployment optimises the amount of devices to be deployed, thereby lowering the deployment cost [61]. Mahajan et al., developed a PM2.5 prediction system using IoT technology and creative imitation. They developed an exponential levelling forecasting algorithm that produces hourly PM2.5 forecasts based on real-time data obtained from the IoT devices installed all over Taiwan. Results were analysed using variables like mean error, accuracy, and computation time. The model was evaluated using data from 132 monitoring stations. The technique was to be created using exponential smoothing with drift.

The real-time PM2.5 data acquired from widespread IoT device deployment was used for experiments and evaluation [62]. Although Saha et al., primary aim was to use IoTs to find pollutants in nature by creating observance devices to monitor air pollution, they ultimately chose to use a Raspberry Pi and a few sensors for this project because it allows the network administrator to monitor and control the desired parameters [38]. The IoTs objective was to facilitate conservative sensing devices to communicate with other devices and to support intelligent service provision. For instance, IoT can be used to monitor air quality in a community to control air pollution and provide real-time information and warnings to residents [63-65].

3.4 USE OF LORA WAN TECHNOLOGY

There are presently a wide range of wireless technologies available to link city sensors and actuators to the Internet, according to Latre et al (e.g., LoRa, Wi-Fi, ZigBee). Regarding range, data rates, and energy usage, each technology has a unique collection of advantages and disadvantages. At both the technological and user levels, it enables the setup and validation of fresh trials in smart cities. The City of Things consists of a network infrastructure with multiple wireless technologies, the ability to conduct data experiments quickly on top of it, and a living lab approach to verify and include other features in the experiments [66-67]. It will also compare popular IoT communication protocols, focusing on the key characteristics and actions of different metrics of power usage security spreading at data rate. The benefit of this technique is that the sender and receiver offset in time and frequencies are equal, greatly reducing the receiver complexity.

Some advantages of Lora WAN technology are:

- Utilises the ISM frequency band without an authorisation.
- It is scalable.
- It supports bi-directional communication.
- Due to encryption algorithms, it offers a high degree of security.
- Provides energy efficiency.

A system that can be spread out widely in the monitoring region to create a sensor network has been envisaged by some authors. The outdoor range tests that were conducted demonstrate that LoRa can communicate over long distances, up to 50 kilometres, while using TX power levels that are lower than those of other wireless technologies. Another option for wireless communication is a LoRa modem, which has advantages like low cost, long range, wide coverage, and long battery life [68-70]. While the star-network topology used by LoRa achieved the settings on the table below, Lee and Ke's proposed LoRa mesh networking system achieved an average 88.49% packet delivery ratio as shown in Table 3 below [71].

TABLE 3.1: COMPARISON OF VARIOUS COMMUNICATION PROTOCOLS

Characteristics	**Bluetooth**	**Zigbee**	**LPWAN**
Frequency band	2.4GHz	2.4MHz	Sub-GHz ISM band
Data Rate	1Mps	250kps	27kps
Rate	Short-range (10-30) m	Short-range (10-100) m	Long- range (10-50) km
Topology	Star-bus Network	Star, Mesh, Cluster	Star
Modulation Type	GFSK	BPSK/ 0-QPSK	FSK-CSS
Power Consumption	12.5mA	40mA	30mA

Consequently, the IoT industry has found use for machine learning technology, particularly in sensor network solutions [72]. To provide information on areas with high pollution levels using IoT monitoring systems with low cost and significant area monitoring data to analyse and report on long term trends in the province, it is necessary to re-access stationary monitoring stations with the full set of pollutants monitoring as opposed to using pollutant specific in this station and passive sampling.

3.5 NEURAL NETWORK TECHNIQUES

Artificial neural networks (ANNs) are frequently regarded as tools that can aid in the analysis of cause-effect relationships within a big-data framework multifaceted system. In this field, large data sets are seldom available therefore it can obtain nonlinear function capabilities to our network and can learn a mapping from inputs to output for a dataset, which means modernizing in detail values of targets starting from input data, and many neurons in the hidden layer are enabled [73]. The design of this system, which has thirteen hidden levels, uses Rectified Linear Units (ReLU) as the activation function for each individual model within the ensemble had a consistent architecture, featuring this: An LSTM (Long Short-Term Memory) layer with 128 units and a ReLU activation function. LSTMs were the selected choice for modelling sequential data like time series.

A dropout layer with a 20% dropout rate. This dropout layer aimed to combat overfitting by randomly deactivating neurons during training.

A dense layer with a single unit, responsible for generating the PM2.5 prediction for the next time step, as shown in Figure 6 below.

.

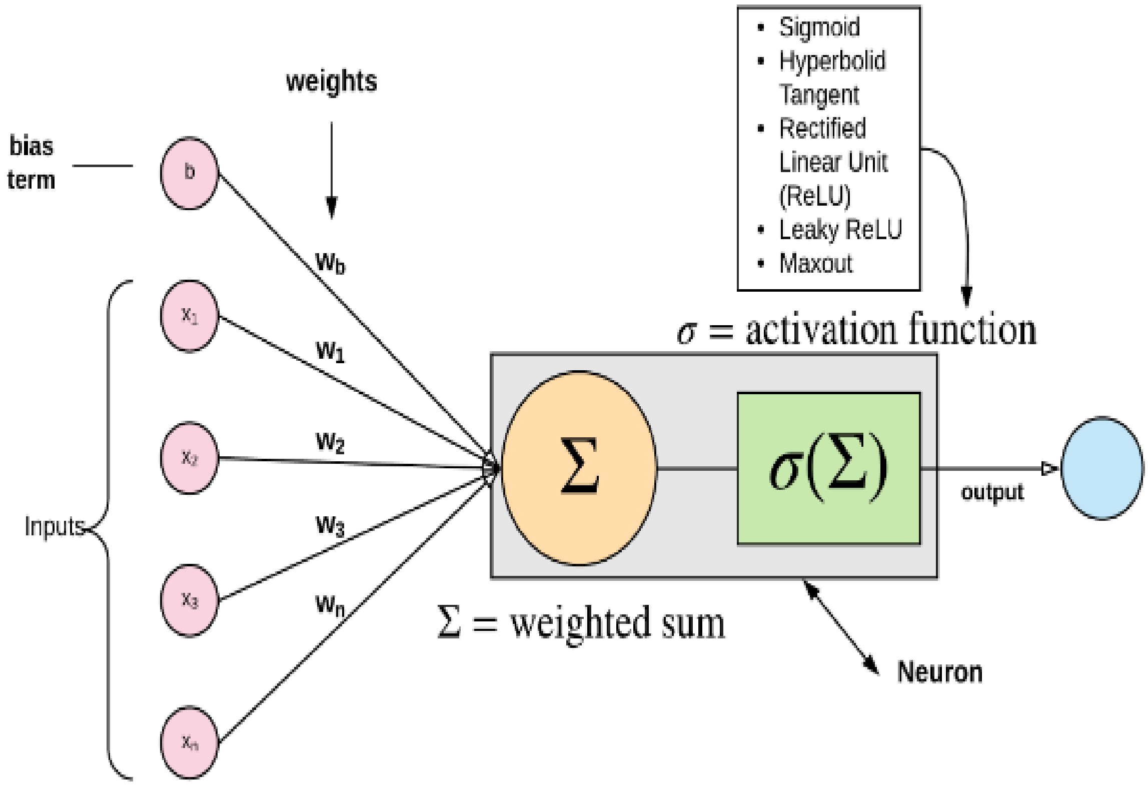


FIGURE 3.2: AN ARTIFICIAL NEURON [71].

These nodes are heavily interconnected by weighted link lines, and weights are changed during a "formation" phase when data is fed into the network. Effective training may lead to the Artificial Neural Networks performing tasks such as forecasting an output value, categorizing an object [74].

Artificial intelligence-based algorithms are frequently used for prediction due to technological advancements, particularly for forecasting air quality. As opposed to a pure statistical model, a machine learning method considers multiple predictive parameters. It appears that Artificial Neural Networks (ANNs) are the method most frequently used to forecast air quality [75].

Other researchers have shown how hybrid or mixed neural-based models can be used for forecasting. Models like the Adaptive- Neuro Fuzzy Interface System have been developed using artificial intelligence algorithms, like fuzzy logic, evolutionary algorithms, Principal component analysis (PCA), and ANNs [76]. Support Vector Machine-based model is one of many machine learning models that have been realised. The concentrations of SO2, NO2, and PM10 have also been forecasted using a modified Wavelet method and Back Propagation Neural Network model, where the Back Propagation Neural Network is modified using Wavelet-Transform Technique.

One of the many machine learning models that have been achieved is the Support Vector Machine-based model. A modified Wavelet method and Back Propagation Neural Network model, where the Back Propagation Neural Network is modified using Wavelet-Transform Technique, have also been used to predict the concentrations of SO2, NO2, and PM10 in the air as stipulated earlier [77].

Bellinger proposed a systematic review of the literature on the use of data mining and machine learning techniques in epidemiology of air pollution; a survey reveals the popularity of the investigation, which was carried out in the USA, China, and Europe, and that data mining is becoming a more widely used tool in environmental health. Since artificial neural networks are preferred over decision trees, support vector machines, and k-means clustering, it has been widely recognised that deep learning can be used to predict air pollution and generate hypotheses [78].

Forecasting and prediction issues have been addressed using classifiers and regressors from machine learning techniques. Knowledge finding and source appropriation have both used clustering algorithms like K-Means and hierarchical clustering. The outcome demonstrates that data mining is becoming more and more beneficial in air pollution research. Kang studied various artificial intelligence, decision trees, and deep learning-based big data and machine learning approaches for air quality evaluation and forecasting. Future smart cities will require the creation of IoT substructures, big data technologies, and machine learning techniques for real-time air quality monitoring and evaluation [79].

To predict the daily API data or any other type of time series, Alyousifi described a novel Markov weighted fuzzy time series model based on the choice of the optimum divider method. In the first stage grid partition technique, two phases were measured to determine forecasting air pollution using the air pollution index. Although there were various partitioning techniques used in the second stage, the grid technique, and the appropriate partition number from the first stage were still in use. Among those partitioning strategies, solid was the best option. Due to the ability of both stage partition methods to prevent the random picking of intermissions, the model precision was significantly improved that better option for air quality forecasting for managing air pollution [80].

For instance, the authors suggested that deep learning and time series models be combined to assess the forecasting of No2 pollutant in the air. The advancement includes neural networks, TBAT, ARIMA, and SARIMA in operation [81]. Therefore, it was evident from the results that the neural network with stacked LSTM outperformed every other model. Another recent research also displays the historical analyses, methodologies, and results of time-series readings that estimate the health risks related to intermittent exposure to particulate matter (PM) [82-83]. The authors suggested using machine learning techniques to create air pollution prediction in smart cities. Their primary objective was to evaluate various machine learning approaches to forecast ozone levels.

Another author recommends a parallelisation methodology of a fuzzy clustering method on heterogeneous servers based on CPU and several GPUs, tailored to classification problems, the aim of their analysis was to obtain the best model for predicting ozone. Sensors, actuators, and embedded computers are used in intelligent transportation systems (ITS), which are pioneering contributions to the application of information and communication technologies in vehicles to enhance safety [84].

In time series of air pollution, Dominici proposes semi-parametric regression that is immediately applicable to risk estimation. Semi-parametric regression models are used to analyse time-series data on air pollution and effect on mortality. This is done under the condition that flat functions of time and temperature considered while the linear component is of interest is consistent with the effects of air pollution. However, the study demonstrates various real-time air pollution modelling through deep learning architectures [85-86].

Kok suggested using a deep learning model to analyse data from IoT-enabled smart cities. To forecast future air quality measurements in a smart city, this model is built on long short-term memory networks. The anticipated model estimation findings are encouraging, and they demonstrate that the model can be applied to other predictions for smart cities, such as air pollution [87].

Hu therefore proposes a machine learning model that integrates data from fixed stations with dense data from mobile sensors to assess the air pollution surface over a specific time span of several days. Tenfold cross-validation was used along with seven regression models. The findings demonstrate that support vector regression (SVR) estimate exactitude is higher than exciting gradient boosting, multi-layer perceptron in linear regression, and adaptive boosting regression, as well as like decision tree regression and random forest regression. The system ability to generate accurate air pollution assessments, greatly strengthen the air pollution map, and will be helpful for long-term research on diseases related to air pollution [88].

To forecasting SO2 awareness in the air, the performance will compare Multiple Linear Regression (MLR) and Multi-layer Perceptron (MLP). To forecast SO2 day-to-day attention, atmospheric parameters, urban road traffic, greenhouse information, and time parameters were chosen. Regression and MLP models are used to compare the two models to create an air pollutant that is precise, efficient, and simple to predict. Therefore, the multi-layer perceptron method produced significantly better results than the regression model. Furthermore, the studies demonstrated that the ANN model was completely applicable to environmental research, including the forecasting of air pollution, given that both the environment and human health are affected by air pollution [89].

The author's reviewed air pollution prediction using machine learning algorithms based on sensor in the smart cities instead of using machine learning techniques. They chose to use progressive and cultured methods to monitor particulate matter of 2.5 micrometres and prediction target, (4) in 41% of the publications the authors conducted the prediction for the following day. As a result, Mahalingam, suggests using (LSTM) machine learning to predict air quality during the COVID-19 outbreak in Jakarta. Because of Adam optimiser ability to distribute the prediction results closer to the data set used, the LSTM model estimation in the review is the root mean square error (RMSE) [90].

Samal, recommends assembling an air pollution data-based framework for circulation prediction. It is imperative to conduct research on air pollution to improve road traffic prediction through air quality because the reasons for excluding the data on air pollution amounts are frequently linked to traffic congestion. To find out which model provides greater accuracy, first decide to conduct a virtual analysis of seven regression models. The results will demonstrate that Multi-Layer Perceptron provided the best outcome. Another author presented a structure using regression models, in which the initial result of the regression model is increased using boosting collaborative technique, and the regression model demonstrates that the proposed structure provides more significant results than the preceding 7 regression models [91].

Other studies propose a review of the monitoring of air pollution and the steps taken to lessen its impacts. The analysis reveals that it is still necessary to re-access these monitoring stations with a full range of pollutant monitoring in contrast to the use of precise pollutant monitoring in those stations. Using a machine learning model to monitor air pollutants will do this at a lower cost and provide information on long-term developments in the province. To produce accurate PM2.5 forecasts for multiple locations at once, Kow, proposed a hybrid model that utilised a convolutional neural network and a back propagation neural network. The results of the evaluation show that the hybrid model uses three different types of machine learning models (static BPNN and RF, dynamic LSTM, and hybrid model with three types of machine learning models) and that it achieves the best results when compared to other models in terms of the smallest RMSE.

For the CNN-BP model, PM2.5 forecast accuracy and consistency greatly improved. As a result, the anticipated CNN-BP model can significantly improve the accuracy and consistency of long-term PM2.5 forecasting [92].

3.6 MACHINE-LEARNING PREDICTION MODELS

The branch of computer science known as machine learning (ML) enables machines to perform tasks without being unambiguously designed to do so. There are numerous studies that classify air quality evaluations using machine learning algorithms and forecast air quality using a variety of scientific techniques. A few models are explored below.

3.6.1 USE OF ARTIFICIAL NEURAL NETWORK

The aim of an artificial neural network (ANN) model is to replicate both the processes and the structural elements of the human brain. An artificial neural network system used for more in-depth pollutants classification and concentration assessment was used to evaluate a sizable amount of investigational data. The architecture of neural networks comprises of nodes that produce signals and evaluate the effects of those signals on the environment. In fact, the ANN can be used to approximate the objective function in optimisation problems so that later methods, like regression, can be used to address the problem.

3.7 REGRESSION

3.7.1 RANDOM FOREST REGRESSION

Random forest is an ensemble learning approach based on bootstrap aggregation and decision tree learning [93]. It can be applied to jobs like classification and regression, among others. The fundamental idea behind random forest is to use an averaging technique to increase prediction accuracy and prevent over fitting, while, fitting multiple decision trees on random subsets of all the features and subsamples in the dataset. To train decision or regression trees frequently with random feature subsets and sample subsets, it specifically employs a bootstrap aggregating (also known as bagging). Then, by averaging all the predictions from trained trees, it guesses unknown input samples. Random forests have the benefit of preventing high variance and high bias, which frequently occur when using a single decision tree. A method for solving optimisation problems that approaches the objective function is random forests. The derivative of the new objective function must be a polynomial to calculate the answer to the optimization problem.

3.7.2 Decision Tree Regression

A popular machine learning technique for categorisation and regression is the decision tree. By learning decision rules from input features, decision tree learning aims to develop a model that can categorise target values. Whether the outcome variable is categorical or numeric determines whether the decision tree is a classifier or regressor.

3.7.3 Multi-Layer regression (MLP) structure

A feed-forward artificial neural network model called a multi-layer perceptron [94], can learn a non-linear function and map a collection of input features to a target. It is made up of perceptron units arranged in tiers. There may be one or more hidden layers between the intake and output layer. Back-propagation is also used to change weights and reduce the loss function.

3.7.4 Multi-Layer regression

In this study, neural network data is categorized into three sets, the training set and the test set and validation respectively. For evaluating the modified weights and biases of a network, the learning set is used. For calibration, the test set is used, to prevent networks from being over-trained. In this research, the test set was allocated 10% and the training set was allocated 90% of the entire data on each model to achieve the best results.

Since there is no universal rule for deciding the properties of the hidden layers and the neurons, the hidden layers and the neurons were built using trial-and-error procedures. The number of hidden layer neurons greatly affects network efficiency. If the number is minimal, the network may not reach the desired accuracy level, but training may be longer if there are too many, and the model may often over-fit the data. Each individual model underwent training for a predetermined number of epochs. Specifically, models were trained for 50 epochs. Unlike some variations that employ early stopping, this approach involved training each model for a fixed number of iterations to maintain consistency within the ensemble. The trial-and-error approach was used in this study to achieve the best models possible for each predicted parameter. Figure 7 shows the structures that were used to predict each parameter.

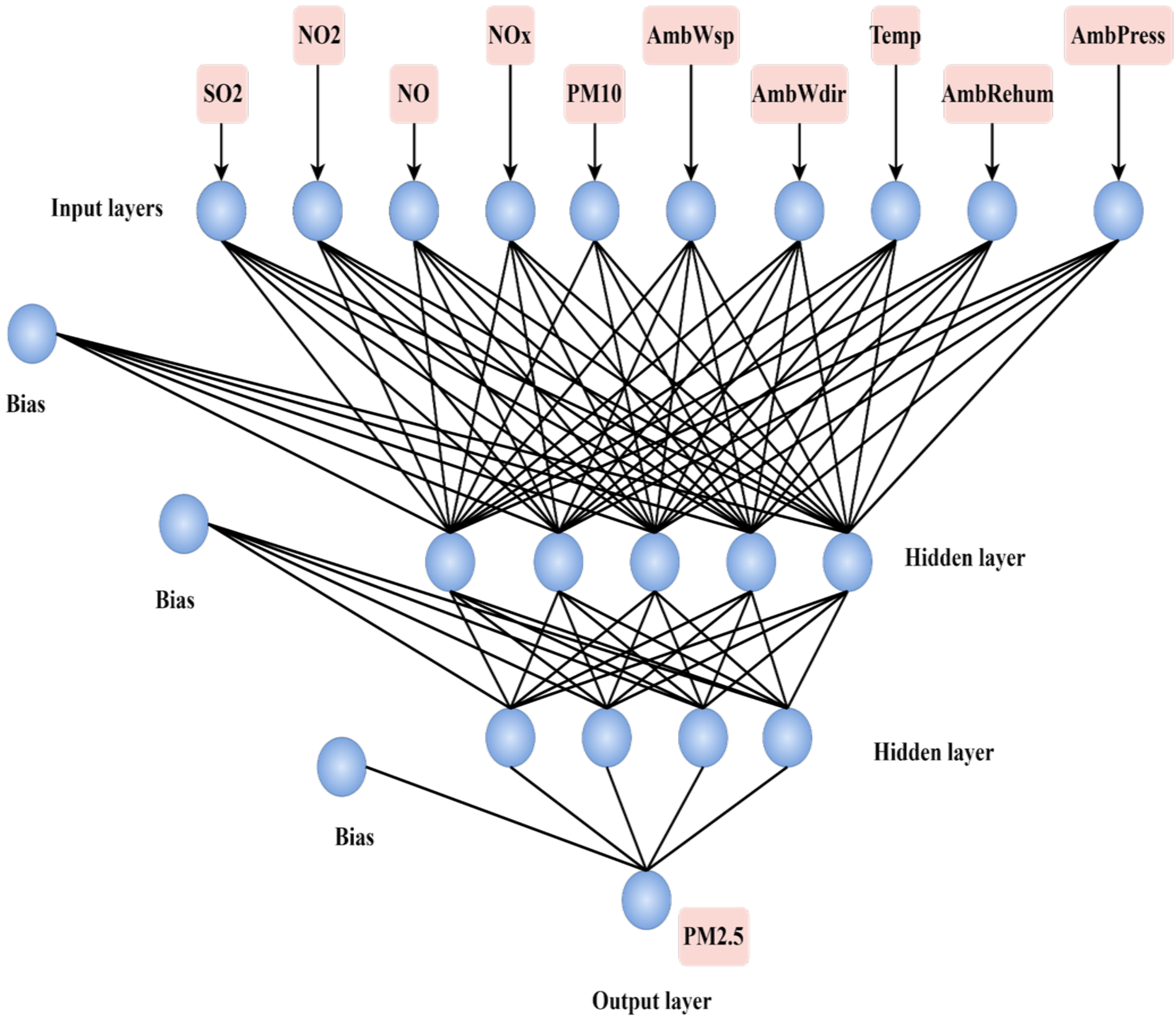


FIGURE 3.3: PROCESS FOR ESTIMATING AIR QUALITY

3.7.5 MULTILAYER PERCEPTRON NEURAL NETWORK

The MLPNN model is the most popular form of data modeling methodology used by ANN. The feed-forward neural network is the general class structure of the ANN that includes the MLP model. A basic form of neural network that can simulate both continuous and combinable functions is the neural feed-forward network. The MLP network design is made up of layers of neurons grouped together. The multilayer perceptron has a wide range of classification and regression applications, including the recognition of patterns, speech, and classification issues. In the MLPNN model, all the input nodes are in one layer and the hidden layer is distributed into one or more hidden layers in various ways. But the choice of architecture has a significant influence on the convergence of these networks.

Table 4 provides a summary of the analysis of the review. Various communication networks used for pollution monitoring are highlighted and how pollution was alleviated. It is clear from the study that intelligent monitoring of air pollution is still at its inception, especially in South Africa.

TABLE 3.2: SUMMARY OF KEY FINDINGS

Ref.	Pollutants Parameters	Network Communication	Air Pollution Monitoring Technique	Pollution alleviation	Gap
[44]	NO2, CO	WSN	Real-time and log data to a remote server,	Monitoring system that continuously tracks the level of air quality in a particular location	Less accurate data is generated
70	PM2.5, PM10,	WSN, LPWAN	Low-Power Wide-Area	Much denser sensor	Developed sensors are

	ozone (O3), NO2, SO2 and CO.		technologies as the foundation for intelligent sensors used to assess air quality	network is used to directly monitor air pollution values	not miniaturized, efficient and precise enough to produce information that is helpful.
[94]	NO2, CO2, SO2	Wireless Sensor Network (WSN)	Vehicular IOT Based	The vehicle user is warned of the extent of pollution from his vehicle. Hence, can control emissions	No appropriate measure taken for air pollution.
[72]	CO, PM (2.5, 10), NO2	Zig-Bee	Monitoring information is transferred from the city fixed nodes to the mobile nodes on public transportation vehicles, including buses and cars, connected through IoT technology	Impact on the city's pollution control because when it is possible to monitor precisely, it is also possible to prevent pollution.	A better communicati on mechanism is required between coordinator nodes for more accurate results

[97]	CO_2, NO_2, and SO_2	GSM	An IoT-based air quality monitoring device for ambient assistance environments	System provides a reliable assessment of interior air quality so that technical interventions can be anticipated to improve indoor air quality.	The security of data transmission and authentication process needs to be improved
[10]	CO_2, CO and HS4	GPS	Android application termed IoT-Mobair enables users to obtain pertinent cloud-based data on air quality.	Detect, monitor, and test air pollution in each area. In cases where the pollution amount is too high, a warning is displayed.	Signals may be lost where licensed band signals are lost
[99]	CO_2, NO_2, and SO_2	Zig-Bee	IOT-based hierarchy-based air quality monitoring device	Analyses real-time information through IOT and information is stored in the database after being gathered	The largest source of air pollution is from automobile emissions.

				from the sensors.	
[100]	CO2, NO2, SO2	LPWAN	Air pollution monitoring system using LoRa as transceiver systems	Prevents the level of air pollution from getting worse	Low cost, it is based on the Internet of things, so remote access is possible
[101]	PM-10	Wireless Sensor Network (WSN)	Integrated pollution monitoring system	Government and policy makers as basic information to take further actions reducing pollution levels	N/A
[1]	NO2	Neural networks	Data-driven Mobile air pollution monitoring	Future NO2 levels in urban areas can be predicted with accuracy using data produced by mobile sensing devices when used outdoors.	Data accuracy is compromise d
[102]	CH_4, NO_2, H_2S	Intelligent DSS with distributed WSN	Autonomous remote monitoring framework	Remote pollution monitoring in the mines is satisfactory. It	The challenge of detecting sensor

				is also done in real-time	nodes and noise
[103]	NO_x, H_2S, CO_2	ANFIS	Using meteorological factors to predict ozone concentration levels	Ozone level components are adequately estimated and assessed	N/A
[104]	SO_2, NO_x, O_3 NH_3, CO, PM10	Machinelearning; RF, DT, KNN	Using data analysis and prediction	Predicting pollution variables enables mitigation activities to be initiated	Prediction is not 100% accurate hence there is room for improvement
[105]	NO_x, CO_x	Intelligent transport system using fuzzy technique	Cooperative intelligent transportation systems using to communication	Pollution system correctly identifies highly polluting traffic areas and drivers	Hardware features included may require different run-time strategies
[106]	NO2	Air Quality Operational Centre's	Machine learning	Algorithms in air quality monitoring conditions that forecast a major photochemical pollutant's daily peak concentration	Data accuracy is compromised

[92]	SO_2, NO_2	Genetic Algorithm was used to choose a portion of the initial set of factors to predict air quality	Algorithm- ANN Model	Air quality monitoring data And meteorological data	Different run-time strategies were applied

3.8 DISCUSSION OF KEY FINDINGS

The recommendations for additional optimisation studies have been made because of the review findings. Alternative air pollution sources like fossil fuel power plants, factories, industries, mobile sources like vehicles, marines, and aircraft, and controlled burns in farmland require further optimising emissions from mobile sources is essential to reduce air pollution. These alternative sources of air pollution through optimisation studies, it becomes possible to identify and implement effective measures to reduce emissions and improve air quality. Such efforts contribute to the overall goal of protecting human health, mitigating climate change, and preserving the environment.
To gain a better understanding of how the smart city model operates and how neural network intelligence techniques can effectively address air pollution, further studies are indeed necessary, especially in developing countries. These studies can shed light on the specific challenges and opportunities faced by developing countries in implementing smart city initiatives and utilizing AI techniques for air pollution management.

3.9. CONCLUSION

Furthermore, comparative study of several individual models, especially in South Africa has been presented. The was a Comparative study of several individual models in this chapter to assess the performance and effectiveness of different models. Several smart technologies used to monitor air quality were discussed. Based on the literature reviewed, it can be concluded that AI could be used to address the complex issue of maintaining the quality of air pollution consumed every day. This dissertation therefore aims to fill the gaps in the application of smart air pollution monitoring techniques by paying particular attention to how smart techniques are applied to air pollution monitoring.

CHAPTER IV

METHODOLOGY

4. INTRODUCTION

This Chapter includes the theoretical and the modelling methods of the operation to solve the problem with air pollution providing through standards. The model is developed and presented in an aim to reduce costs, by using an intelligent monitoring sensor to collect the data and be able identify air quality irregularities when improving the proposed system.

In Section 4.2, the relevant components are discussed, and in Section 4.3, an Intelligent Monitoring System model formulation and limitations are covered. This chapter also summaries how the problem defined in the study resolved and how the decisions and optimizations were made as shown in Figure 4.1 below. The next stage is to search for the appropriate machine learning solution, to recognize the best model based on both input and output data in machine learning solution, which was followed by an algorithm implementation. However, the results and implications of these simulations will be discussed more in detail in Chapter 5.

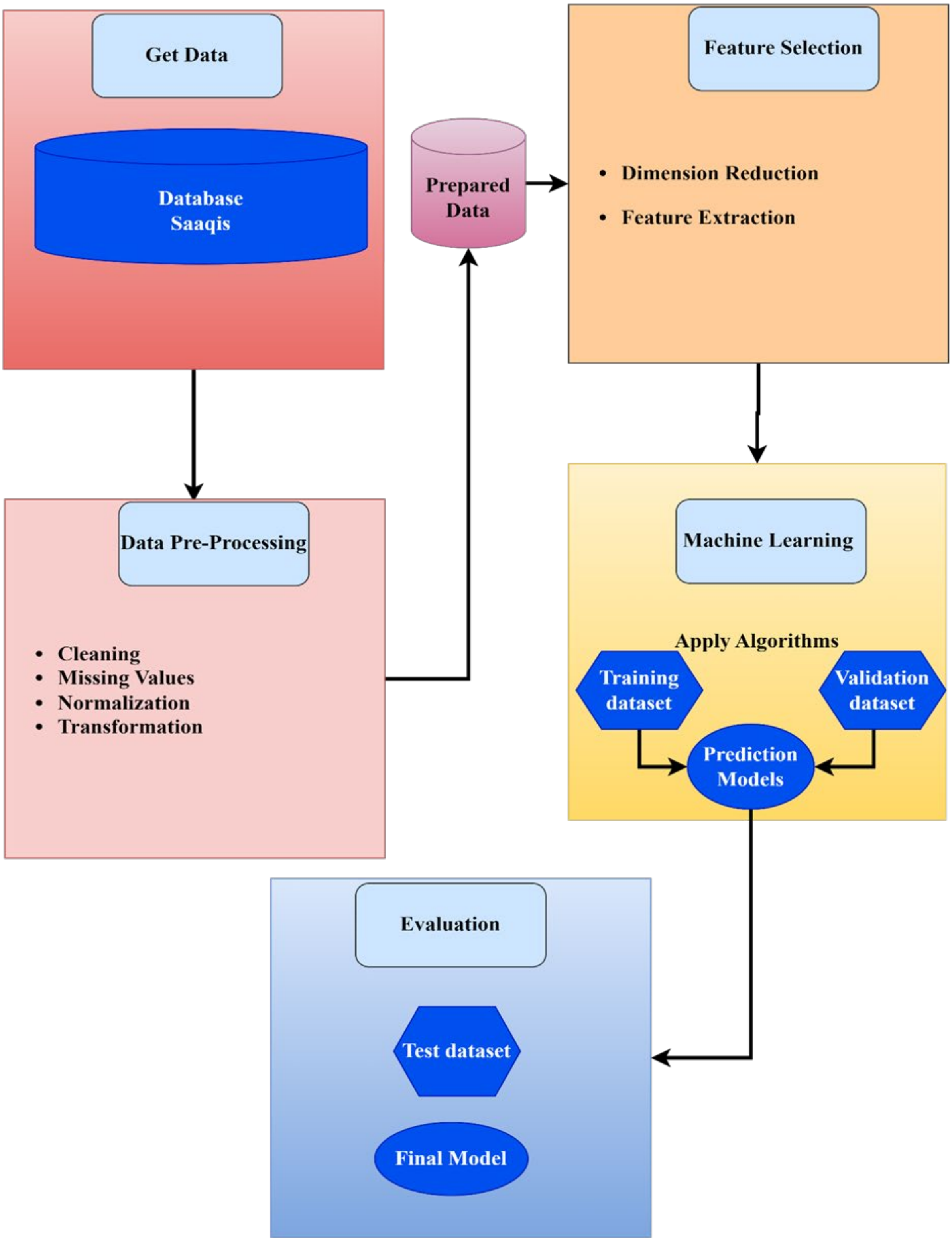


FIGURE 4.1: FLOW DIAGRAM OF METHODOLOGIES FOLLOWED

4.2. BOJANALA PLATINUM DISTRICT MUNICIPALITY CASE STUDY

As shown on Figure 4.2, the Bojanala Platinum District Municipality (BPDM) is in South Africa, Northwest province. It is made up of five local municipalities, including Kgetlengrivier, Madibeng, Moretele, Moses Kotane, and Rustenburg District Management Area, and has an estimated area of 18 333 km^2.

The National Framework for Air Quality Management in the Republic of South Africa (2007) has classified Rustenburg as an area with poor air quality within the district. The main areas where pollution has an effect are where industrial and domestic fuel is burned, which is from the Rustenburg area eastward towards Brits and up the eastern border of BPDM. Pollution levels are generally low throughout the remainder of the municipality [107].

According to the BPDM Growth and Development Strategy 2005–2020, BPDM drives economic growth in the Northwest Province refer figure 9 and accounts for much of the province economic production and employment possibilities. Therefore, the district economic performance is essential to achieving the province's overall growth and development objectives [108].

FIGURE 4.2: THE BOJANALA PLATINUM DISTRICT MUNICIPALITY WITHIN THE NORTHWEST PROVINCE

4.2.1 Analysis Of Algorithms

Prior to determining the modeling techniques and algorithms to be implemented using the data obtained, it was necessary to better understand the actual problem and the type of solution needed to meet this challenge. A comprehensive study was undertaken to understand existing approaches to identify the type of problem and the required solution. In this study, this took a considerable amount of time, following the model training during the simulation phase, which is discussed in more detail in the next Chapter.

Based on the data obtained, the initial phase of the research was to determine what type of learning technique was required to address the issue. During this phase, the appropriate learning approach between supervised and unsupervised learning had to be determined. Since the data were arranged to contain known and named parameters, the supervised learning strategy was selected as the best option based on the available data. Supervised learning involves the use of detailed, column-based data.

4.2.2 Identification Of The Learning Technique

It was necessary to decide between a regression model and a classification model to choose the appropriate model type to explore. Regression was chosen as the approach to be studied in this research after a strict process of acquiring a more thorough understanding of both modeling techniques. The selection was made mainly because the solution to this problem was to predict the actual value of a parameter for a certain range. If the only principle was to evaluate whether the air quality was good or poor, the classification method would have been chosen. Because regression is a supervised learning model that attempts to anticipate an output that changes based on identified input variables as a result, regression was considered the most appropriate model for solving the problem.

After careful selection of the appropriate modelling technique, the final step was to explore a variety of options, Machine Learning algorithms to determine which is best suited to the data obtained. Before their utilisation was verified, the following algorithms were examined: Random Forest Regression, Decision Tree, Logistic Regression, K-Means Clustering, Tree Support Vector Machine, and Linear regression methods were developed, and output was evaluated.

4.3 MODEL SELECTION PROCEDURE

The neural network was initially proposed for this study due to their ability of handling any data with non-linearities. Since a “No free lunch theorem” is highly considered in machine learning, the two other machine learning algorithms were proposed after they passed a machine learning model selection test [109]. These two additional models that are available in the scikit-learn library are “Linear Regression” and “Random Forest Regressor”. Figure 4.3 shows the model selection procedure that was used for selection of the two latter mentioned models. The two models were used with their default parameters. The parameter settings of the neural network which formed the third model are summarized in section 4.3.1.

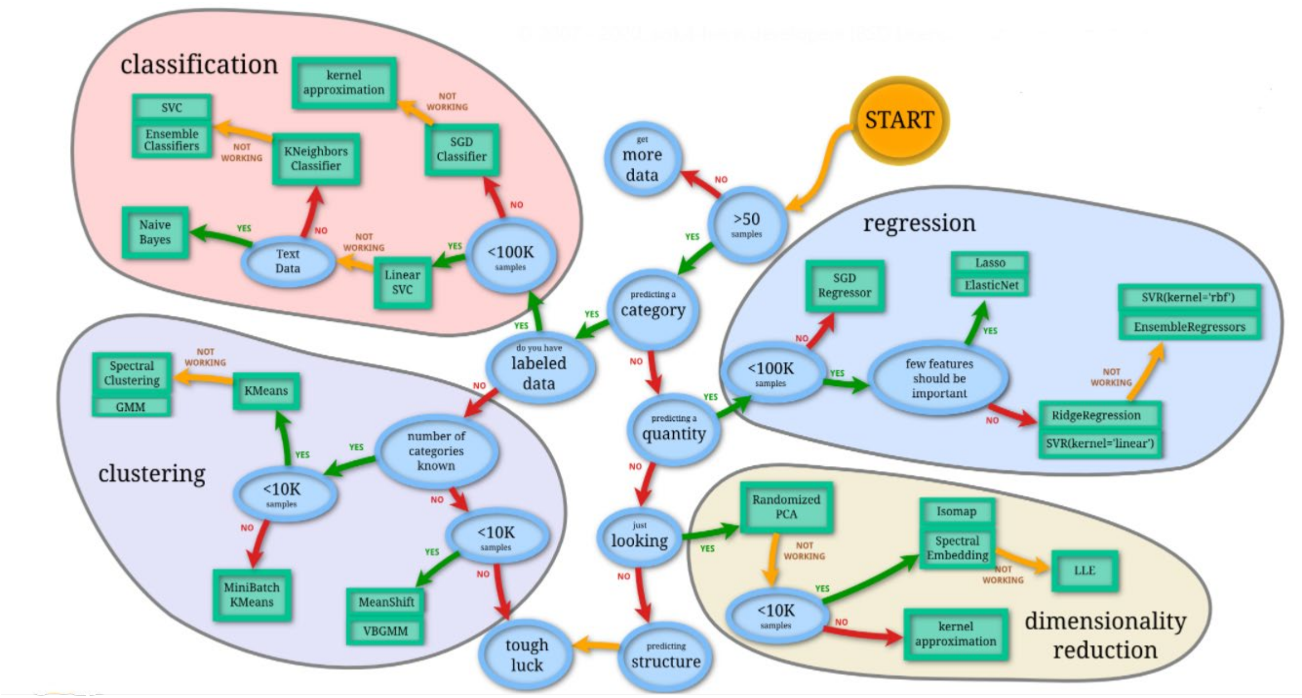


FIGURE 4.3: MODEL SELECTION PROCEDURE

4.3.1 NEURAL NETWORK MODEL PARAMETER SETTINGS

The neural network model was built on Tensorflow – Keras library with the model architecture as shown in Table 4.1.

TABLE 4.1: EXCERPT FROM THE PYTHON – NEURAL NETWORK MODEL ARCHITECTURE

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense (Dense)                (None, 13)                182
_________________________________________________________________
dense_1 (Dense)              (None, 7)                 98
_________________________________________________________________
dense_2 (Dense)              (None, 7)                 56
_________________________________________________________________
dense_3 (Dense)              (None, 1)                 8
=================================================================
Total params: 344
Trainable params: 344
Non-trainable params: 0
_________________________________________________________________
```

The establishment of municipal air quality standards by local authorities that are legally binding is not provided for by the Air Quality Act. It is acknowledged that regional officials may set more stringent ambient air quality standards, like those in Table 4.1. The input layer consisted of 13 input dimensions since there were 13 predictors in the dataset. One output node was used because there was only 1 output target feature. The input layer and two hidden nodes used rectified linear unit (Relu) activations. Each hidden layer consisted of 7 neurons that were obtained by using the general rule of thumb presented by equation 3.1.

$$No\ of\ neurons\ in\ each\ hidden\ layer = \frac{(Number\ of\ input\ nodes)+(Number\ of\ output\ nodes)}{Number\ of\ hidden\ layers} \qquad 4.1$$

Since the nature of the problem dealt with in this study is a regression problem, the 'linear' activation function was used in the output node of the neural network as shown in Figure 4.3.

4.4 PREDICTION MODELS AND DATA ANALYSIS

4.4.1 DATA GATHERING

This study outlines an ANN technique for real-time air quality prediction. The SAAQIS data repository for pollutants served as the source of the information. Particulate matter with an aerodynamic diameter less than 10 µm or 2.5 µm, Sulphur Dioxide, Nitrogen Oxide, Humidity, and Temperature are among the parameters used. The South African Air Quality Information System (SAAQIS), air quality parameters were evaluated. Data for the Boitikong area that were gathered between 2018 and 2021 was received in the form of an electronic data set.

The data were recorded hourly for the purpose of monitoring selected pollutant parameters. The data was then retrieved electronically and transferred to Excel to facilitate analysis and machine learning. After all data was imported into Excel, data processing, analysis and transformation commenced, Jupyter notebooks were used for that technique. The data has been carefully examined in this procedure with the aim of eliminating glitches that could hinder the performance of Machine Learning.

4.4.2 DATA PRE-PROCESSING

Understanding the structure of data is very important in machine learning projects. If the data is understood, machine learning algorithms with optimised hyper parameters may be achieved. In this study, three machine learning algorithms were used on a data set of 13 gas input features to predict the concentration of the hazardous PM 2.5 gas in the air. Each of the predictors and the target data features originally consisted of 12807 observations before the data set was split into training and testing data sets. After the split, the training data set comprised of 10246 observations while 2562 observations were reserved for the testing data test. Using different pre-processing methods in Python integrated development environment (IDE), the data were pre-processed with the help of Python libraries that were imported to the Python IDE.

The time series data was loaded from a CSV file, comprising various features, including PM2.5 values over time.

The target variable (PM2.5) and relevant feature columns were selected.

Feature data was standardized using the **StandardScaler** to ensure consistent scaling across all features for improved training.

. Table 4.2 lists a few important Python libraries that were used for various purposes in this study of development of machine learning algorithms for prediction of pm 2.5 gas.

TABLE 4.2: A SELECT FEW PYTHON LIBRARIES THAT WERE USED IN THE DEVELOPMENT OF MACHINE LEARNING ALGORITHMS FOR PREDICTION OF PM 2.5 GAS

Name of the library	**Purpose of the library**
1. NumPy	To perform array operations
2. Pandas	To provide a data framework and manipulation
3. Scikit-learn	To provide machine learning algorithms
4. Matplotlib and Seaborn	To perform data plots
5. Keras-Tensorflow	To provide a neural network platform

Data pre-processing or rather data wrangling, is the most important stage of machine learning. Therefore, Figure 4.4. Illustrates how the data were pre-processed before they were ready for use by the three machine learning models.

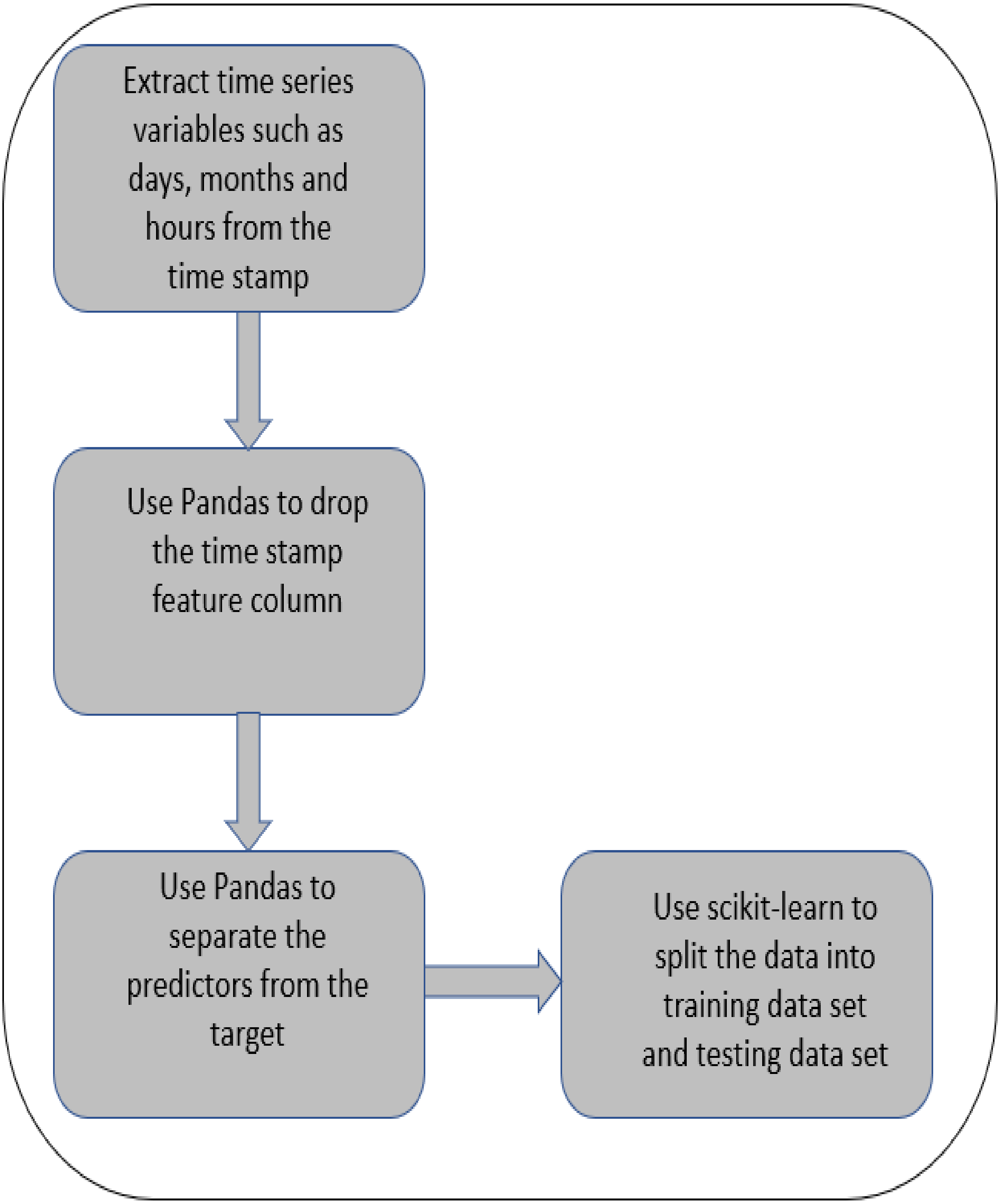


FIGURE 4.4: DATA PRE-PROCESSING STEPS USING APPROPRIATE PYTHON LIBRARIES

TABLE 4.3: FIRST ELEVEN OBSERVATIONS OF THE AIR POLLUTION DATA SET (TIME SERIES) WITH PREDICTORS AND PM2.5 TARGET FEATURE

	SO2	NO2	NO	NOx	PM2.5	PM10	Amb_Wsp	Amb_WDi	Temperatu	Amb_ReHu	Amb_Pres	Months	Hours	Days
0	4	2	0	2	22	36	1.6	76	16	55	882	1	1	10
1	5	2	0	2	19	32	1.9	68	16	56	883	1	2	10
2	3	1	0	2	13	23	2	57	16	57	883	1	3	10
3	2	2	1	2	12	21	2.2	57	15	61	883	1	4	10
4	3	2	0	2	11	20	2.4	66	14	66	883	1	5	10
5	3	2	0	2	10	18	2.3	59	14	68	883	1	6	10
6	2	2	1	2	9	18	2.5	51	14	67	883	1	7	10
7	1	2	1	2	9	20	3	43	17	59	884	1	8	10
8	1	2	1	3	9	24	3.3	43	19	51	884	1	9	10
9	1	1	1	2	10	23	3.5	39	21	42	883	1	10	10
10	1	1	1	2	11	21	3	42	24	33	883	1	11	10

4.5 DATA ANALYSIS

Using Seaborn and Matplotlib libraries, the histograms and scatter plots were generated to understand the feature distribution in the data set. These plots are shown in Figures 4.5 and 4.6 For instance, if information about SO2 feature was to be understood, we would look at the SO2 histogram in Figure 4.5 A closer look at this histogram shows that SO2 has more than 10 000 negative values in the range of 0 - 20.

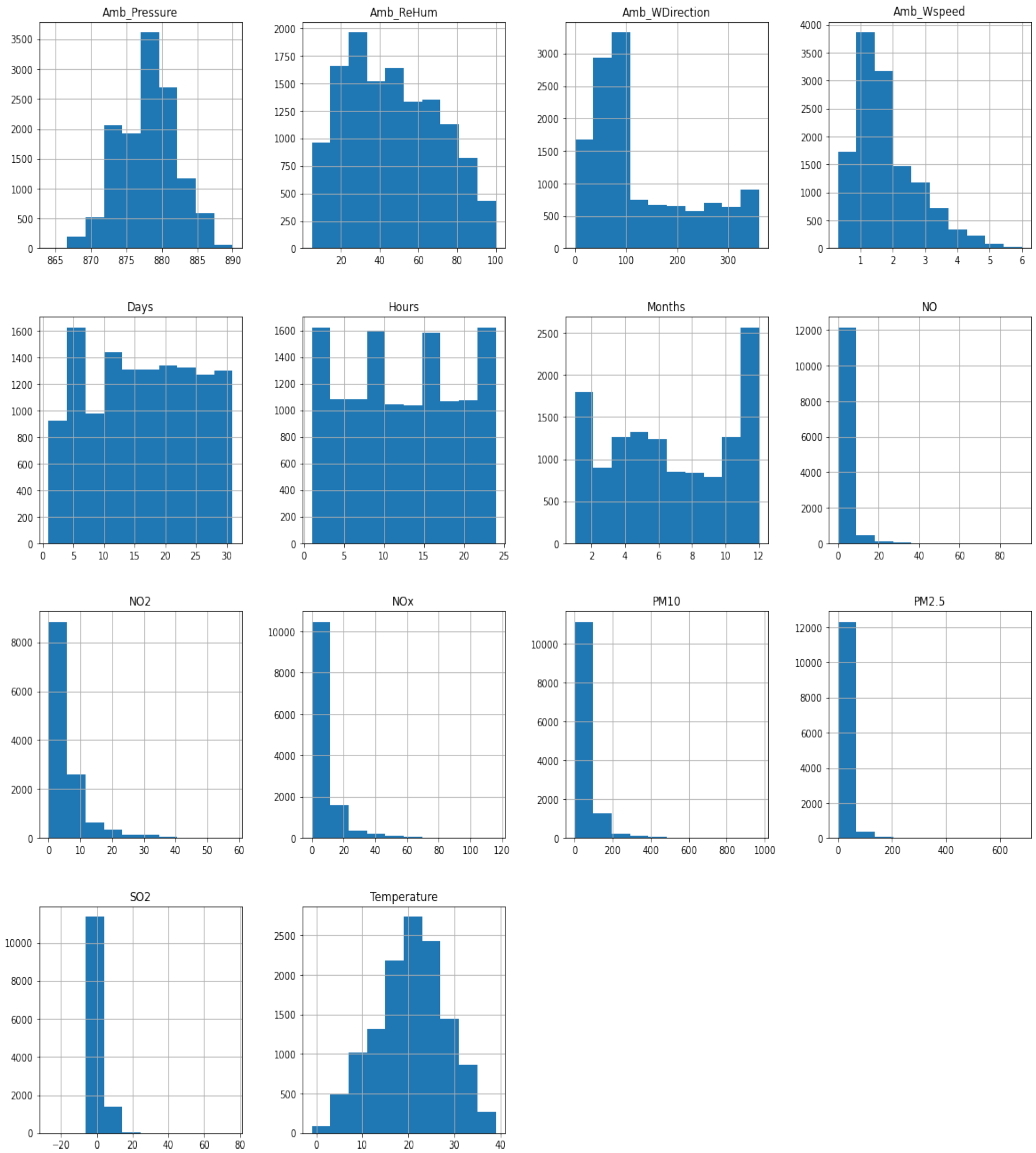


FIGURE 4.5: HISTOGRAMS OF DATA FEATURES

Figure 4.6 shows a scatter matrix plot of the data features. The purpose of this plot is to understand the relationship of data distribution among the data features. For instance, some features are linearly correlated with each other. These linearly correlated features as shown in Figure 4.7, are NOX vs. NO2, NO vs. NOX, and finally PM10 vs. PM2.5.

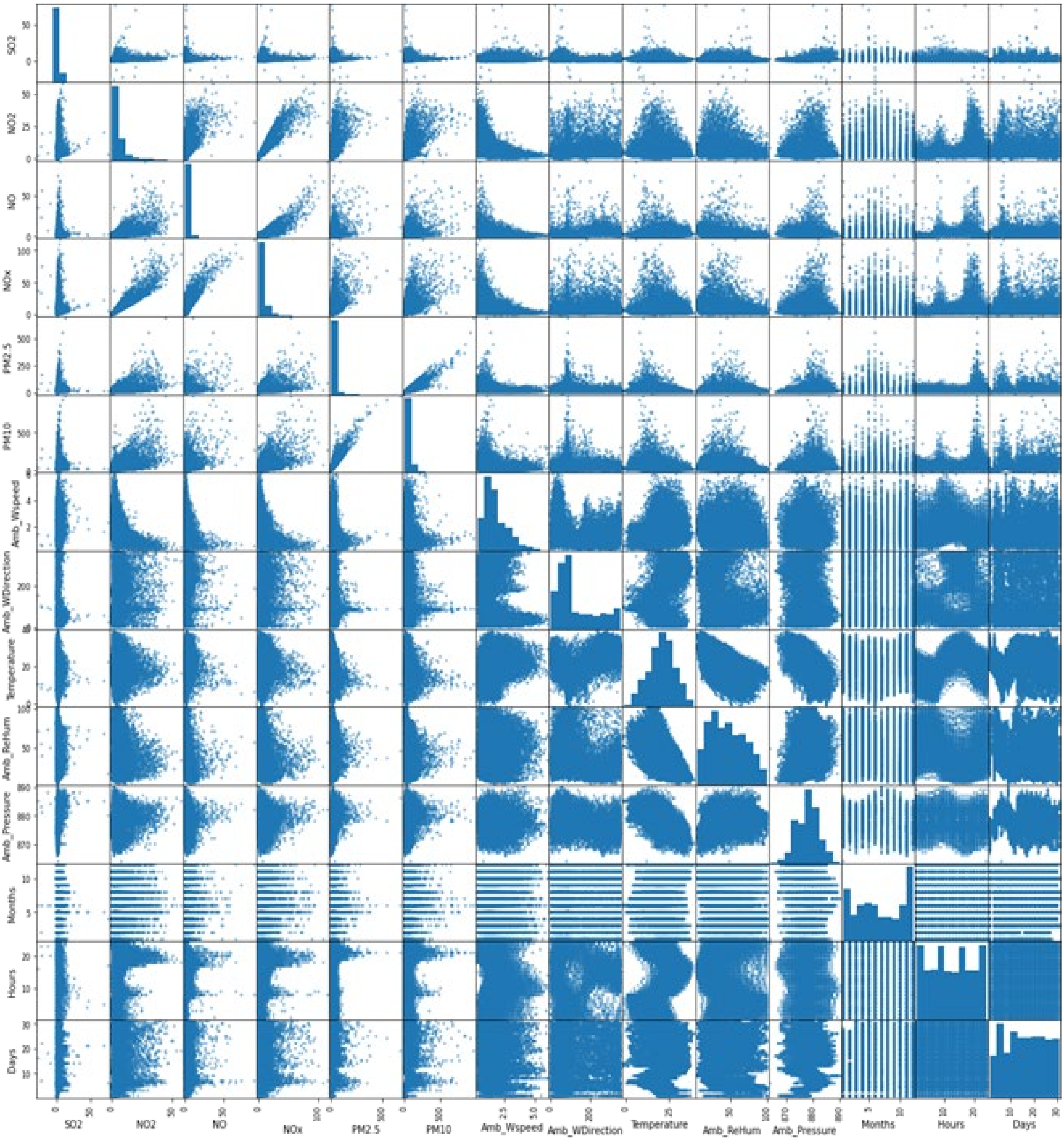


FIGURE 4.6: SCATTER PLOT MATRIX OF THE DATA FEATURES

A variety of plots were generated to assess the ensemble model's performance. These included time series line plots for visualizing actual vs. predicted PM2.5 values, scatter plots for PM2.5 forecasting by days and hours, a residuals plot for understanding prediction errors, and a prediction intervals plot to visualize uncertainty in the predictions. As is evident in Figure 4.6 there are linear correlated features in the data set. It is therefore important to understand the degree to which these features are correlated. In fact, data correlation is a method for comprehending the connection between features and multiple variables in the dataset. Positive correlation states that if feature A increases, feature B increases as well, or vice versa if feature A decreases, feature B decreases as well. As a result, both characteristics are linearly related and move together. The linearly correlated features from the last section all have a positive correlation. If a characteristic A increases, then a feature B decreases, and vice versa, is said to have a negative correlation. In the data set used in this study, the features that tend to have a negative correlation as shown in Figure 4.6 are temperature (Temperature) and relative humidity (Amb_ReHum). Figure 4.7 illustrates the diffidence between negative and positive correlation.

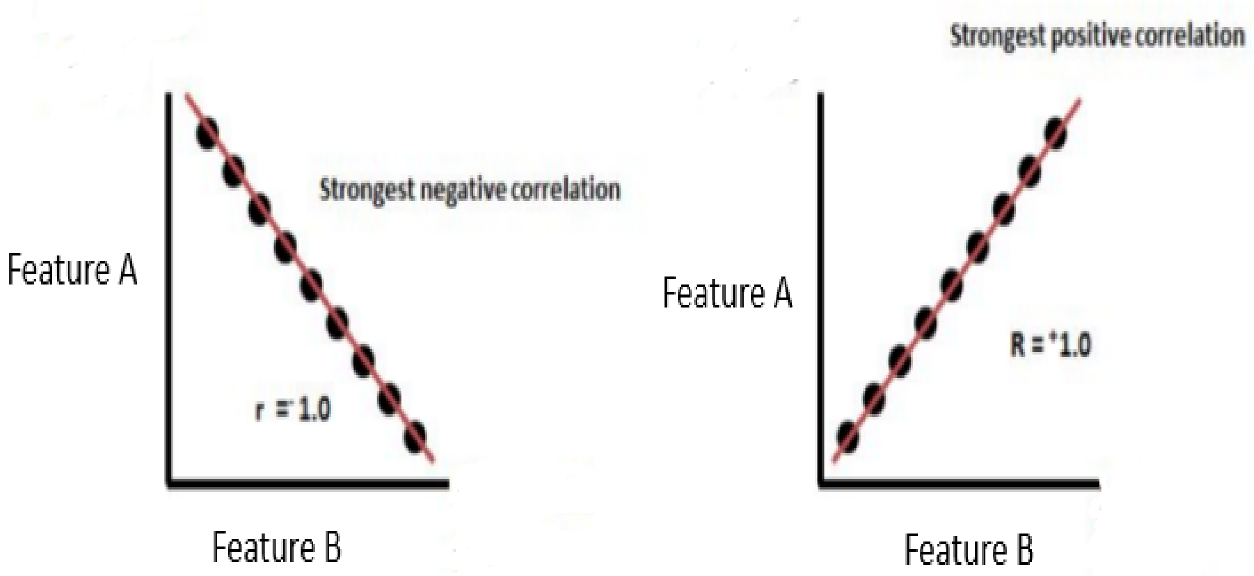


FIGURE 4.7: NEGATIVE CORRELATION (LEFT) AND POSITIVE CORRELATION (RIGHT)

Figure 4.7 illustrates the difference between negative and positive correlation. Each of these correlation types can exist in a range represented by values from 0 to 1 where slightly or highly positive correlation features can be of values in the range of 0.5 or 0.7. If there is a strong and perfect positive correlation, then the result is represented by a correlation score value of 0.9 or 1. If there is a strong negative correlation, then it will be represented by a value of -1. Highly correlated features may lead to skewed or misleading results due to a problem called multicollinearity. To overcome this, among a few available methods, ensemble tree models such as random forests may be used as they are immune to this problem (this was applied in this study).

To understand the degree of correlation between features, a correlation plot was used as illustrated in Figure 4.8. The previously mentioned features that were said to be having a positive linear correlation relationship have yellowish regions in the correlation plot. For instance, using the Pearson Correlation Coefficient in Python IDE, the calculated correlation between NO2 and NOX was 0.92. Among the three machine learning models used in this study, the multiple linear regression model ("Linear Regression") which is not immune to multicollinearity (caused by highly correlated features), scored a variance score of 0.86. This variance score was less than that of the "Random Forrest Regressor" model which scored a variance of 0.93. The variance of the "Random Forrest Regressor" was higher because it is immune to multicollinearity (caused by highly correlated features).

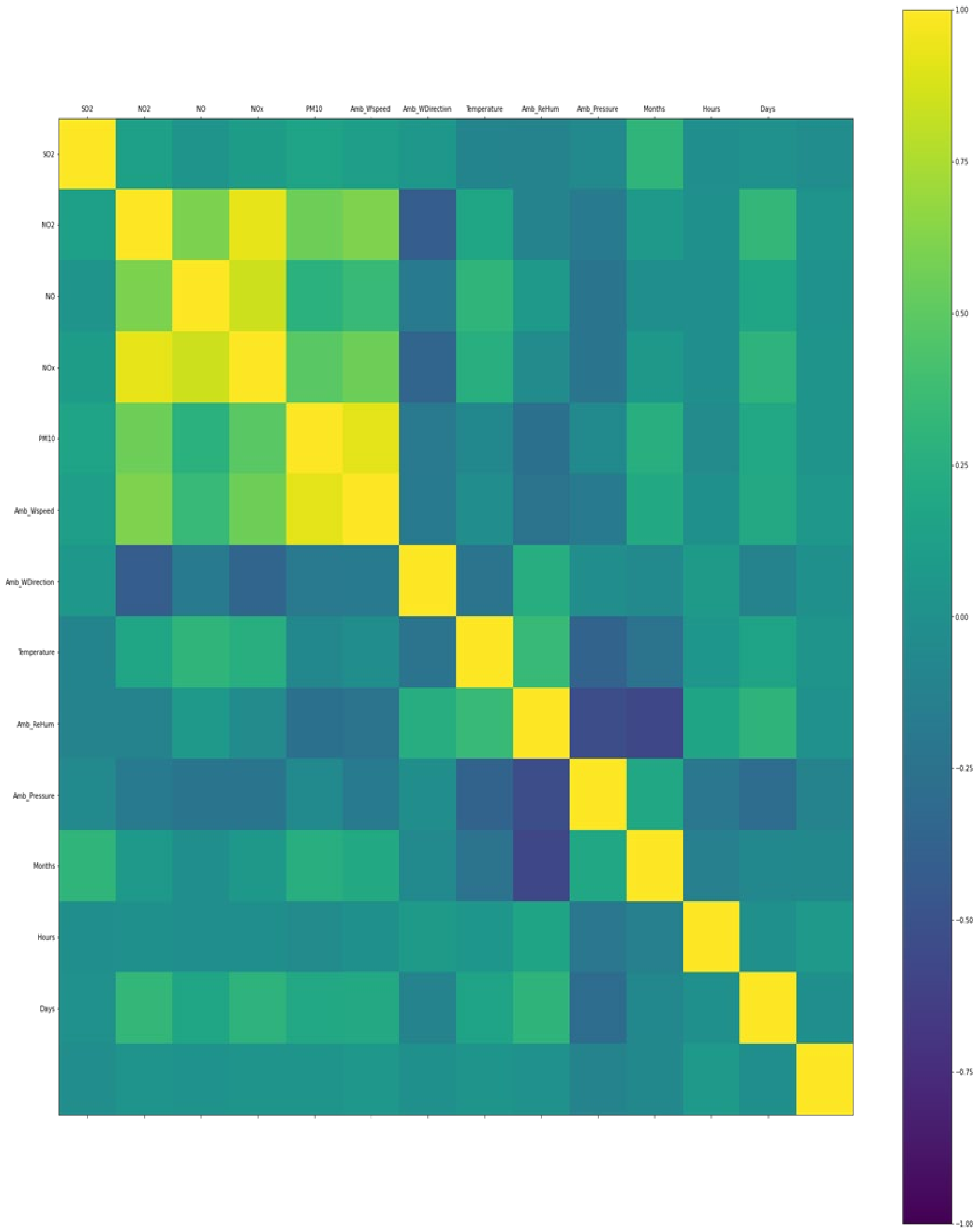


FIGURE 4.8: CORRELATION PLOT OF THE DATA FEATURES

4.6 MATHEMATICAL ANALYSIS

Given that pollutant parameters were specifically monitored over a four-year period, there was a need to review and evaluate the effectiveness of the suggested models. Here, the MLR is used in a combination of several independently trained two-layer MLR networks. The learning algorithm of the scaled conjugated gradient and transfer functions of the sigmoid (hidden layer) and linear (output layer) is used. A network of its own is created and trained for each missing data pattern using an early stopping approach (to avoid over-fitting) with a test set of training in a proportion of 1/5. The number of inputs matches the number of values that are currently accessible, and the number of outputs matches the number of values that is missing from the data rows with the missing pattern. N_{mlp}, the number of hidden neurons is determined through an experimental process.

$$N_o = round(2xI_0) + 1, \tag{4.2}$$

$$N_i \ = round\ (2xI_i) + 1 \tag{4.3}$$

$$\left(IfN_O < N_{i,}then\ N_{mlp} = Ni, then\ Nmlp = No,\right) \tag{4.4}$$

where N_i and N_O are the number of hidden neurons defined from the inputs and outputs (o), I_O and I_i Are the number of neurons in the input and output layers, and round is the rounding towards to nearest integer.

The metric R^2And the root means square error $RMSE$ have been calculated using Equations (5). The most common indicators of imputation ability are the correlation coefficient R and its square: coefficient of determination R^2 n i.e., the variance

explained, which is limited to a range between 0 and 1

$$R^2 = \left[\frac{1}{n}\frac{\sum_{i=1}^{n}[(P_i-\bar{P})-(O_i-\bar{O})]}{\sigma_P\sigma_O}\right]^2 \tag{4.5}$$

Where, N is the number of imputations, O_i the observed data point, P_i the imputed data point, $\bar{O}$ is the average of observed data, $\bar{P}$ as an average of imputed data σ_p standard deviation of the imputed data and σ_0 The standard deviation of the observed data.

The root mean squared error $RMSE$ which summarises the difference between the observed and imputed concentrations was used to provide the average error of the model. The ensemble model's performance was evaluated using established evaluation metrics, including mean squared error (MSE), mean absolute error (MAE), and root mean squared error (RMSE). These metrics served as quantitative indicators of the ensemble's forecasting accuracy.

$$RMSE = \left(\frac{1}{N}\sum_{i=1}^{n}[[P_i - O_i]^2]\right)^{\frac{1}{2}} \tag{4.6}$$

Where n is the number of observations, Oi is the observed parameter, Pi is the calculated parameter, O is the mean of the observed parameter, P is the average of the calculated parameter, is the standard deviation of the observations and is the standard deviation of the calculations. A separate prediction is conducted for each pollutant (PM10 and PM2.5), using Multiple Linear Regression (MLR), Support Vector Regression (SVR) and Random Forest

$$MAE = \frac{1}{N}\sum_{i=1}^{n}[P_i - O_i] \tag{4.7}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{n}\left|\frac{P_i - O_i}{P_i}\right| \tag{4.8}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{n}[[P_i - O_i]^2] \tag{4.9}$$

Where, P_i represent the real values while O_i Represents the predicted values of ambient pollutant to make model results more intuitive and reliable, we performed a comparative analysis by employing MAE, MAPE and MSE with state-of-the-art frameworks such as:

Decision Tree (DT), Random Forest (RF), Support Vector Regression (SVR), Multilayer Perception (MLP), Long short-term Memory (LSTM) using Multiple Linear Regression (MLR) and Stacked Long short-term Memory (SLSTM). The MAE, MAPE and MSE are used to calculate the prediction error, and the lower the value means, the higher the prediction accuracy. The performance metrics are calculated using Equations.

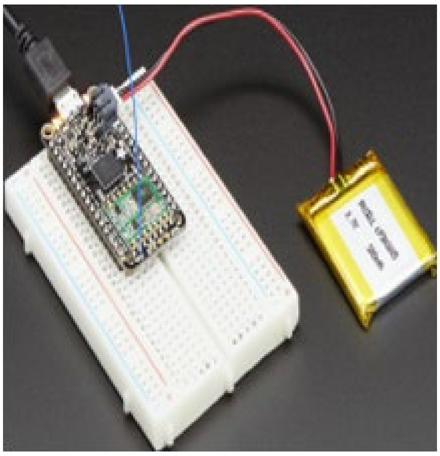
4.7 ARCHITECTURE MODELLING CIRCUIT

To validate the predicted air quality values, a shyness-based prototype smart air monitoring system was built in this study. These devices will also help monitor pollutant parameters and air quality levels. The prototype contains two parameter sensors from the thirteen parameter studies. The developed IoT smart air monitor is comprised of five air parameter sensors namely Particulate matter that is less than 2.5µm in aerodynamic dimension, less than 10µ diameter aerodynamically sized particles, M9, Boost Voltage regulator, PM2.5 Air Quality Sensor, Adafruid Feather 32u4 FRM9x LoRa and battery 3.7V.

The figure below presents a low-cost surveillance system that could radically address some of the current air pollution problems. Sensor nodes for real-time monitoring, along with a LoRa connection for long-range, low power and secure wireless data transmission is included.

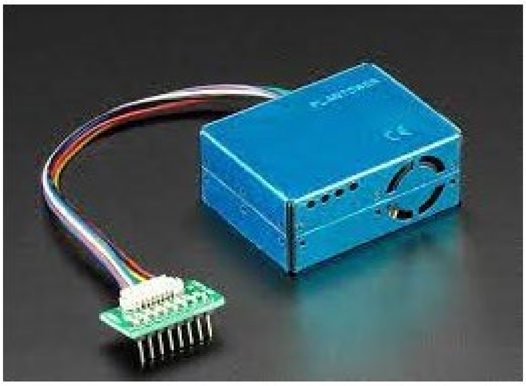
MQ Air Pollution sensors

Environmental combo CCS811



Boost Voltage regulator

Adafruid Feather 32u4 FRM9x LoRa and battery3.7V



PM2.5 Air Quality Sensor

FIGURE 4.9: IOT PARAMETER DEVICE

Figure 4.9 shows the working IoT smart air monitoring device. Figure 4.10 shows the sensors used in the design and construction of the Particle Monitoring System of (PM 2.5). Techniques like those described in the previous research were used in the present study to ensure appropriate use of the sensors throughout the construction of the air monitoring system. It is a handheld smart sensor with a LoRa connection that can monitor air quality and transmit parameter data in near real time.

For several months, the PM sensor was monitored for several types of pollutants to determine the strength of the system and how it reacts to the environment. In the study of the PM2.5 sensor was used in the following manner. The first airborne pollutants were detected by the PM2.5 sensor, the sensor data was then converted into a digital format using an analog-to-digital converter and transmitted to a Raspberry Pi module. The Raspberry Pi module processed the sensor data and transmitted it to the cloud. Finally, through cloud computing, the detected data was made available in the cloud.

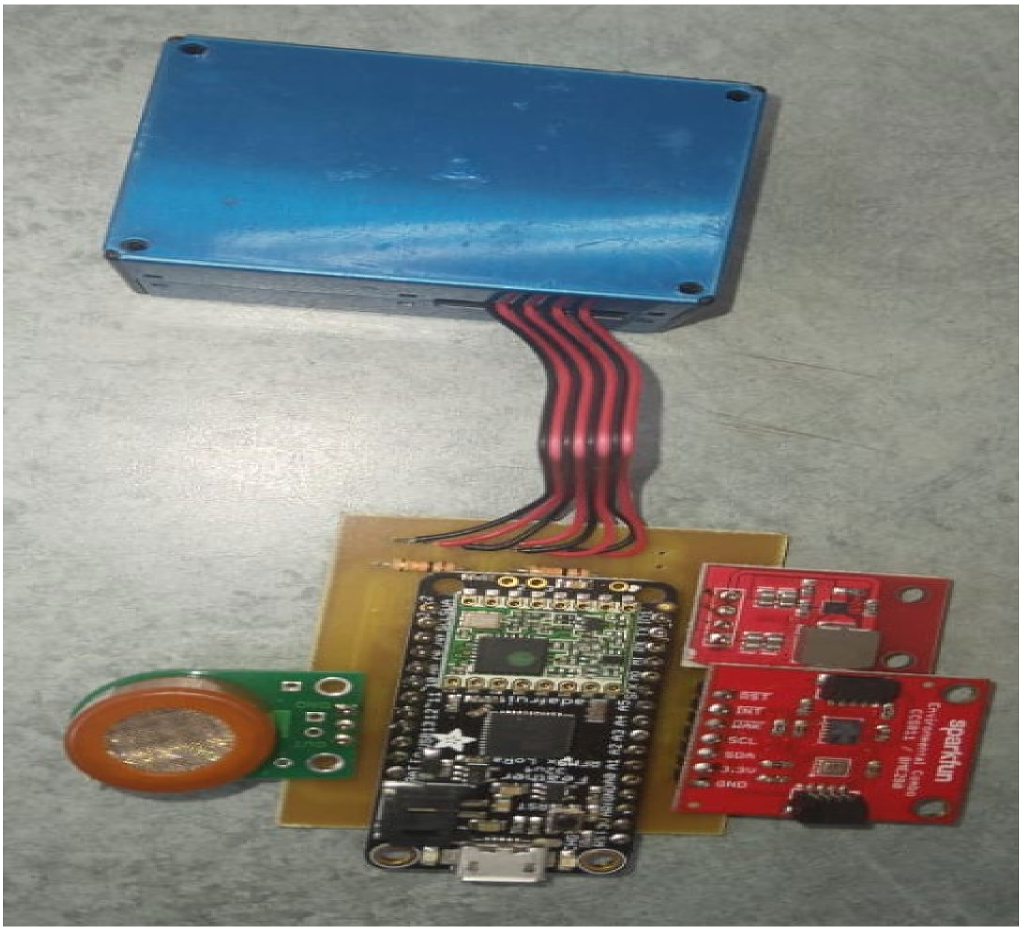


FIGURE 4.10: IOT SMART AIR MONITOR DEVICE

4.8 INCORPORATION OF AN INTELLIGENT IOT AIR POLLUTION SYSTEM

Once the IoT-based intelligent air monitoring system is developed, the next step is to combine the proposed system into the airborne equipment process. Figure 3.11 illustrates how the system is integrated and how each component contributes to the overall air quality monitoring system.

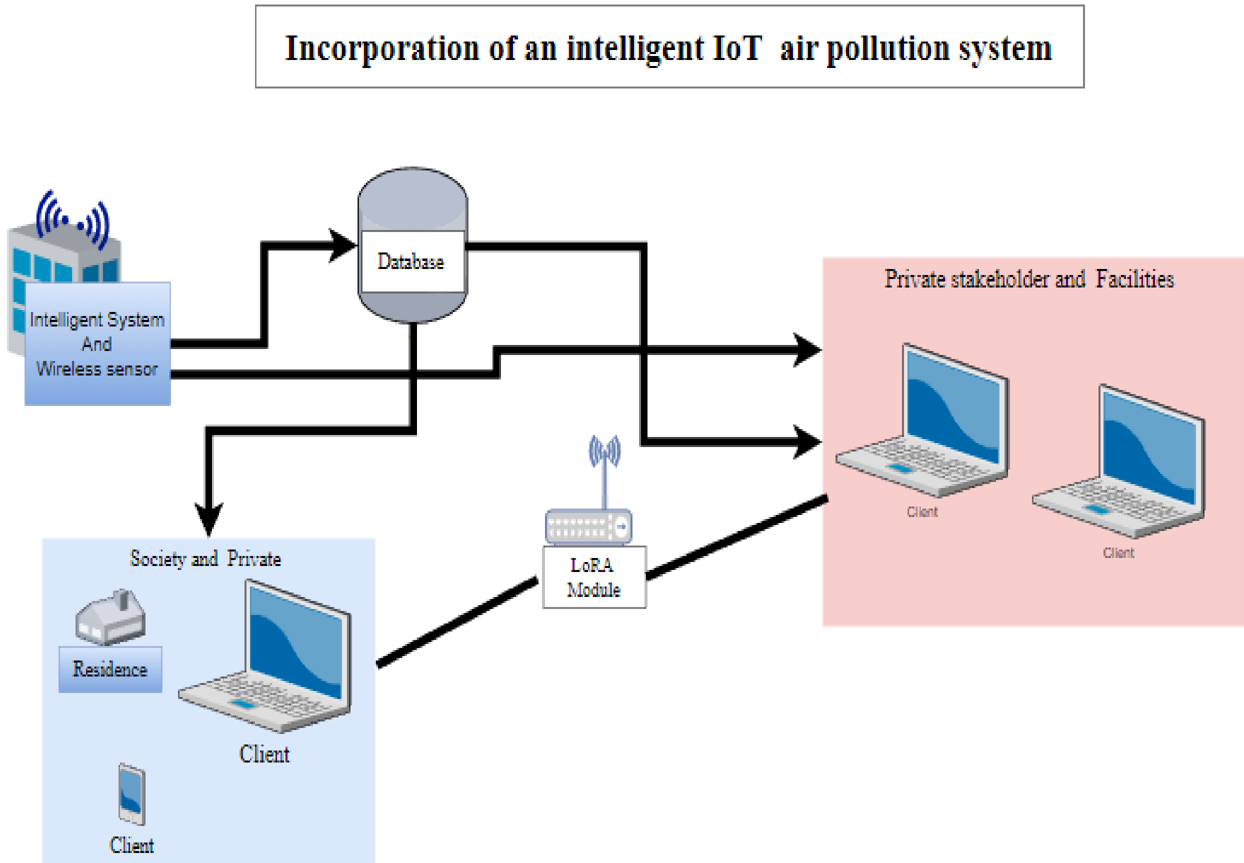


FIGURE 4.11: INCORPORATION OF AN INTELLIGENT IOT AIR POLLUTION SYSTEM

The initial phase of this procedure was to implement the LoRa IoT Gateway, which was used for long distance, low power, and secure wireless data transmission. The second step consisted of connecting the LoRaIoT gateway to a personal computer to provide a complete and clear display. And provide measurements of wireless sensor parameters once these have been configured. The third step was to install and configure the wireless sensors seen in Figure 4.10. These sensors are then installed in the air station facility, especially in the final phase of the pollution process, to ensure that real-time parameters are monitored. The next step was to build the database so that parameter readings could be captured and archived for future reference. These historical data can also be used to improve predictive models because more data is needed to obtain more accurate forecasts. After one year of storage in the database, the acquired data will be deemed historical. The final phase involved the inclusion of developed prediction models that would provide local municipalities with sufficient time to identify and reduce air pollution. These predictive models work in collaboration with real-time reporting, which is an alarm facility and a society for the predicted values.

4.9 CONCLUSION

In this chapter the concept of the different techniques was described, and the techniques used to gather the necessary data as well as the explanations for collecting them. It also reflected the entire process taken in this study to resolve the problem, from data collection to selecting and implementing the algorithm. Furthermore, the chapter offers some background information on the research subject, on which the input data is based. The mathematical modeling used for evaluation metrics against the implemented models, as well as the network analysis and training methods were discussed. Finally, the smart air pollution monitoring system and the use of monitoring sensors were described to provide an understanding of how these sensors work. The next chapter will present the simulation result in ANN.

CHAPTER V

5. RESULTS AND DISCUSSION

5.1. INTRODUCTION

This section discusses the results obtained from the various machine learning simulations that were performed to determine the effectiveness of the models developed for the prediction process. Firstly, the data collected was trained using artificial neural networks for the proposed pollution monitoring system. Conclusions are drawn from these simulations, such as the number of time series and layers to be used in each forecasting model and error analysis. This chapter concludes with a conclusion of the models that were created and how well they performed.

5.2 NEURAL NETWORKS

Below show simulations from the trained neural network. Figure 5.1 presents the training and validation losses for the neural network model which show a low error margin.

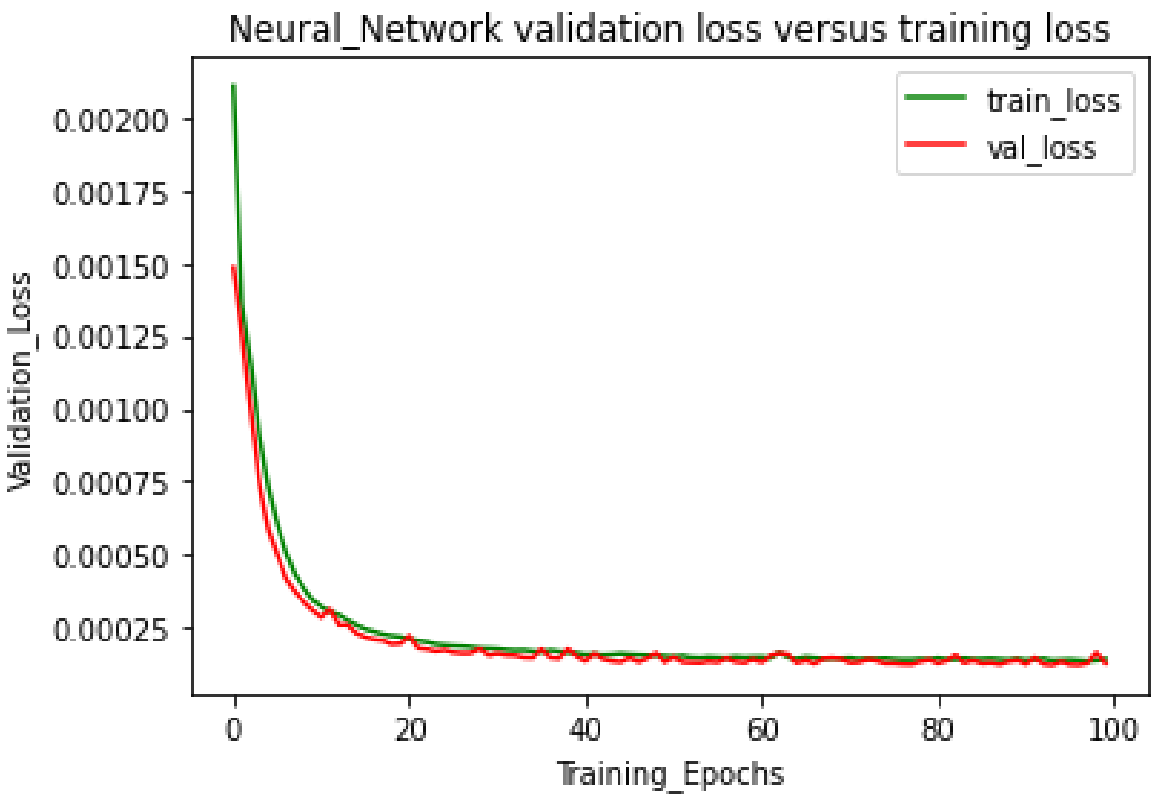


FIGURE 5.1: TRAINING LOSS COMPARED TO VALIDATION LOSS DURING TRAINING

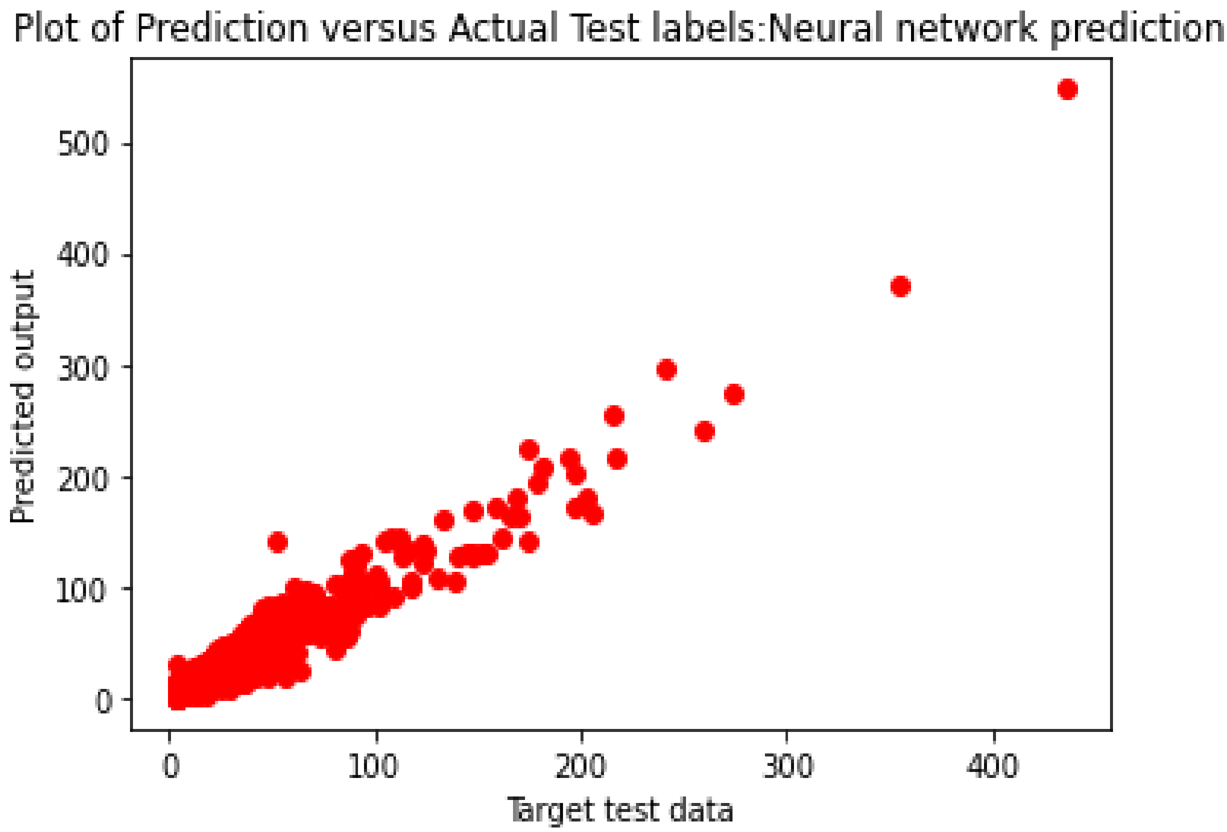


FIGURE 5.2: COMPARISON OF PREDICTED OUTPUT AGAINST THE TARGET OUTPUT FOR THE NEURAL NETWORK MODEL

The validation loss of $23x10^{-5}$ Was achieved, and this was in line with the training loss as shown in Figure 4.1. To check the model performance, the results of the predicted output were plotted against the test data of the target (see Figure 5.2). The linear relationship between the predicted output and the test target data demonstrated a model that was predicted well, because the predicted output results were achieved by using the test data of the input features that were not used during training.

5.3 MULTIPLE LINEAR REGRESSION

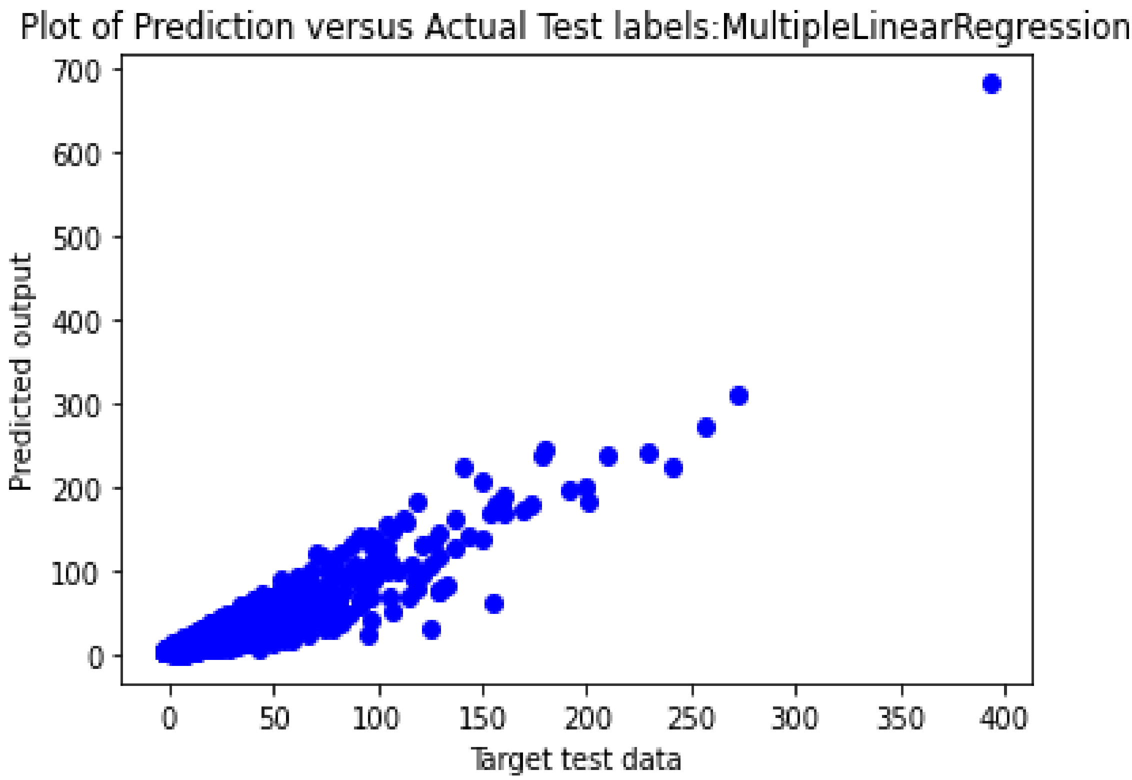


FIGURE 5.3 COMPARISON OF PREDICTED OUTPUT AGAINST THE TARGET OUTPUT FOR THE MULTIPLE LINEAR REGRESSION MODEL

As was evident in Section 3, the dataset consisted of highly correlated features, using linear regression models suggesting that we could not take it as the final model, whereas according to literature there are models that can oversee the multicollinearity (caused by highly correlated features), such as the random forest trees. However, after experimenting with the Linear regression model on the same data set, variance score of 0.86 resulted. For a model that was not hyper parameter tuned, this was not a bad score at all.

Next, check the model performance, the results of the predicted output were plotted against the test data of the target (see Figure 5.3). The linear relationship between the predicted output and the test target data demonstrated a model that was predicted well, because the predicted output results were achieved by using the test data of the input features that were not used during training.

5.4 RANDOM FOREST REGRESSION

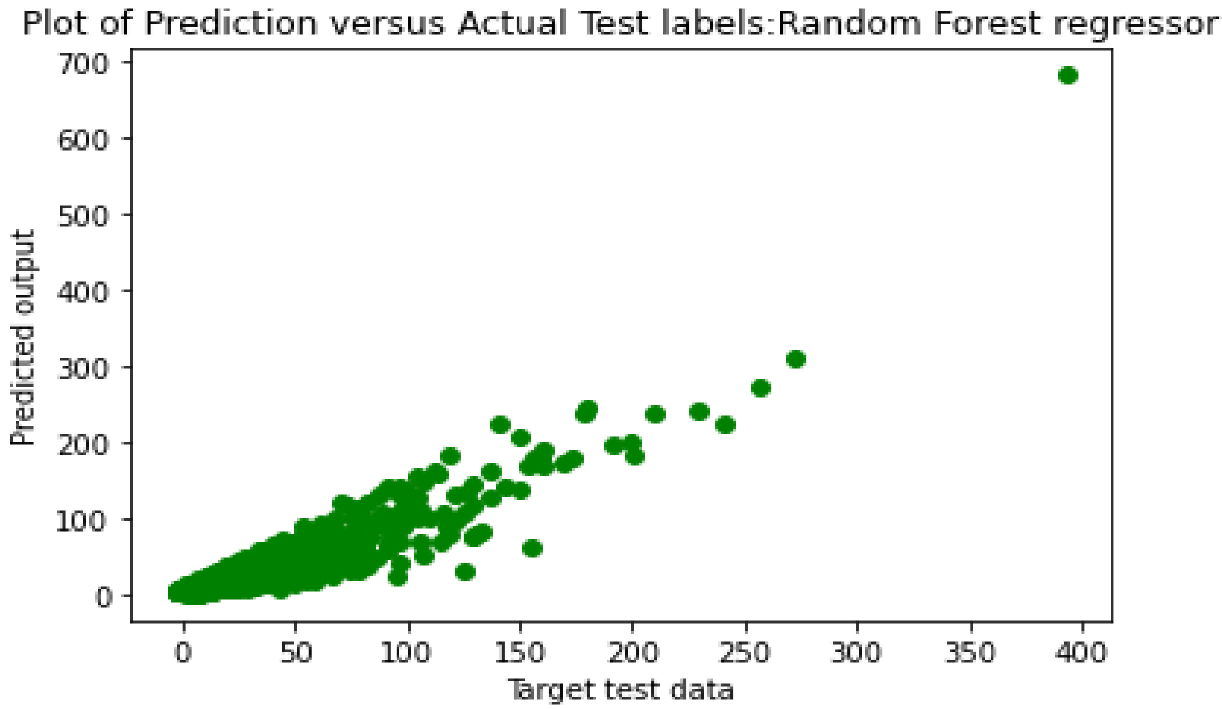


FIGURE 5.4: COMPARISON OF PREDICTED OUTPUT AGAINST THE TARGET OUTPUT FOR THE RANDOM FOREST REGRESSOR MODEL

In comparison with the variance score that was achieved by the linear regression model ("Linear Regression"), the "Random Forest Regressor" model achieved a variance score of 0.93. This proved the fact that the ensemble tree models can oversee the multicollinearity (caused by highly correlated features) problem.

Next, check the model performance, the results of the predicted output were plotted against the test data of the target (see Figure 5.4). The linear relationship between the predicted output and the test target data demonstrated a model that was predicted well, because the predicted output results were achieved by using the test data of the input features that were not used during training.

Central University of
Technology, Free State

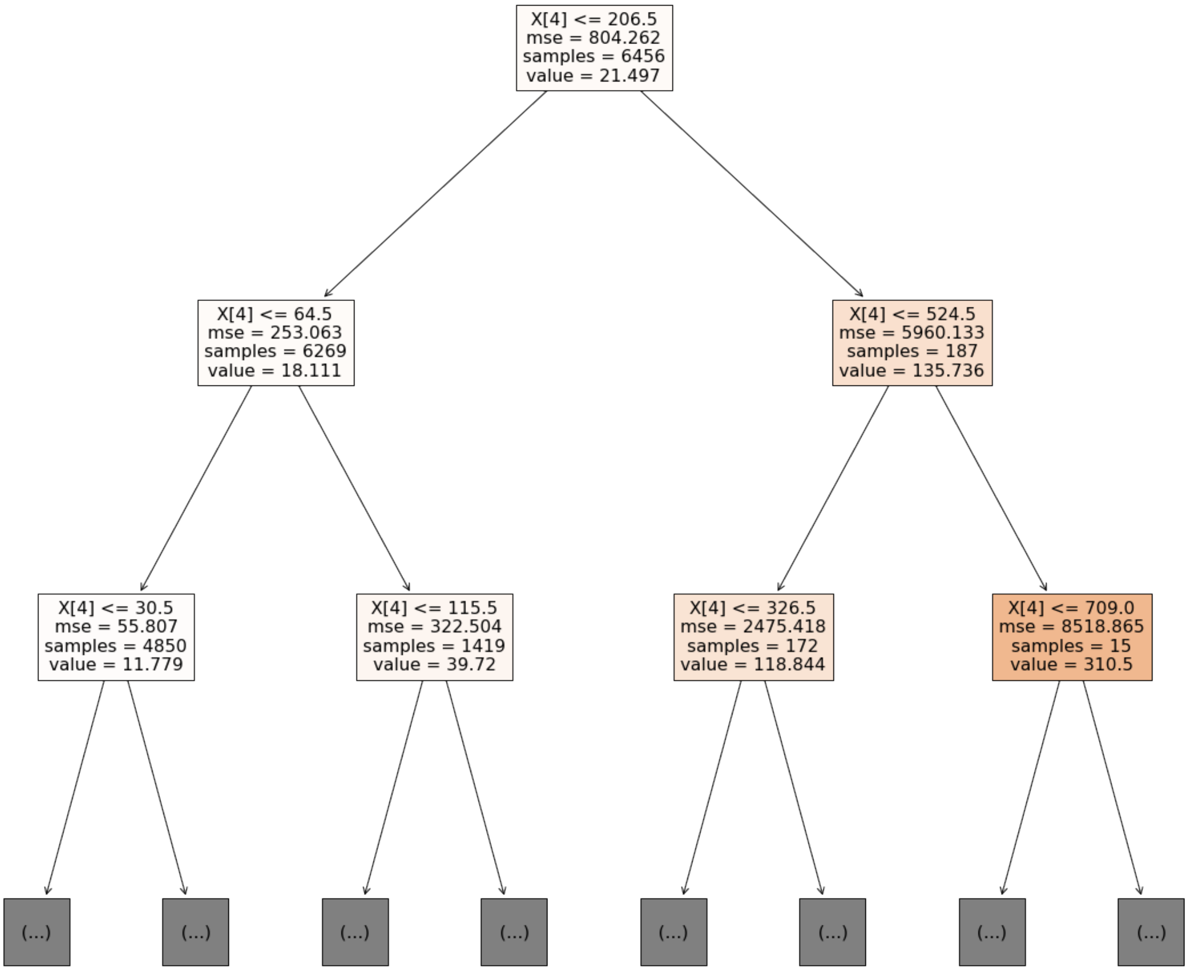


FIGURE 5.5: A FIRST TREE ESTIMATOR EXTRACTED FROM ONE HUNDRED TREE ESTIMATORS USED BY THE RANDOM FOREST REGRESSOR MODEL

The "Random Forest Regressor" model used one hundred tree estimators from the default settings of the model. Figure 5.5 shows the first tree and the predictions made from the root node to the leaf nodes.

4.5 TIME SERIES PREDICTIONS

Based on time series data set, Table 4.1, compares the performance of the three models by comparing the predicted outputs against the actual test target outputs that were reserved during the split of the data for this purpose.

TABLE 5. 1: TIME SERIES MODEL PREDICTIONS USING A SELECT FEW OF THE TEST DATA SETS.

Neural network Performance														
SO2	NO2	NO	NOX	PM10	Amb_Wspeed	Amb_WDirection	Temperature	Amb_RelHum	Amb_Pressure	Months	Hours	Days	Predicted Output	Actual test Output
8	6	1	7	53	1.3	56	13	43	880	8	21	6	25.64	31
8	11	1	12	63	1.1	78	11	52	881	8	22	6	27.53	41
7	5	2	7	38	1.2	76	10	54	881	8	23	6	22.39	24
1	2	1	2	27	1	75	17	79	879	12	24	4	20.66	16
1	1	1	1	12	1.2	73	17	78	878	12	1	4	16.56	7
1	1	1	1	13	1.1	81	16	81	878	12	2	4	16.83	7
Multiple Linear Regression Performance														
SO2	NO2	NO	NOX	PM10	Amb_Wspeed	Amb_WDirection	Temperature	Amb_RelHum	Amb_Pressure	Months	Hours	Days	Predicted Output	Actual test Output
8	6	1	7	53	1.3	56	13	43	880	8	21	6	24.44	31
8	11	1	12	63	1.1	78	11	52	881	8	22	6	29.78	41
7	5	2	7	38	1.2	76	10	54	881	8	23	6	19.09	24
1	2	1	2	27	1	75	17	79	879	12	24	4	18.29	16
1	1	1	1	12	1.2	73	17	78	878	12	1	4	9.50	7
1	1	1	1	13	1.1	81	16	81	878	12	2	4	10.22	7
Random Forest Regressor Performance														
SO2	NO2	NO	NOX	PM10	Amb_Wspeed	Amb_WDirection	Temperature	Amb_RelHum	Amb_Pressure	Months	Hours	Days	Predicted Output	Actual test Output
8	6	1	7	53	1.3	56	13	43	880	8	21	6	28.15	31
8	11	1	12	63	1.1	78	11	52	881	8	22	6	34.94	41
7	5	2	7	38	1.2	76	10	54	881	8	23	6	21.5	24
1	2	1	2	27	1	75	17	79	879	12	24	4	17.07	16
1	1	1	1	12	1.2	73	17	78	878	12	1	4	6.9	7
1	1	1	1	13	1.1	81	16	81	878	12	2	4	7.69	7

The performance results tabulated in Table 5.1 are visually illustrated by means of histograms as shown in Figures 5.6 - 5.8.

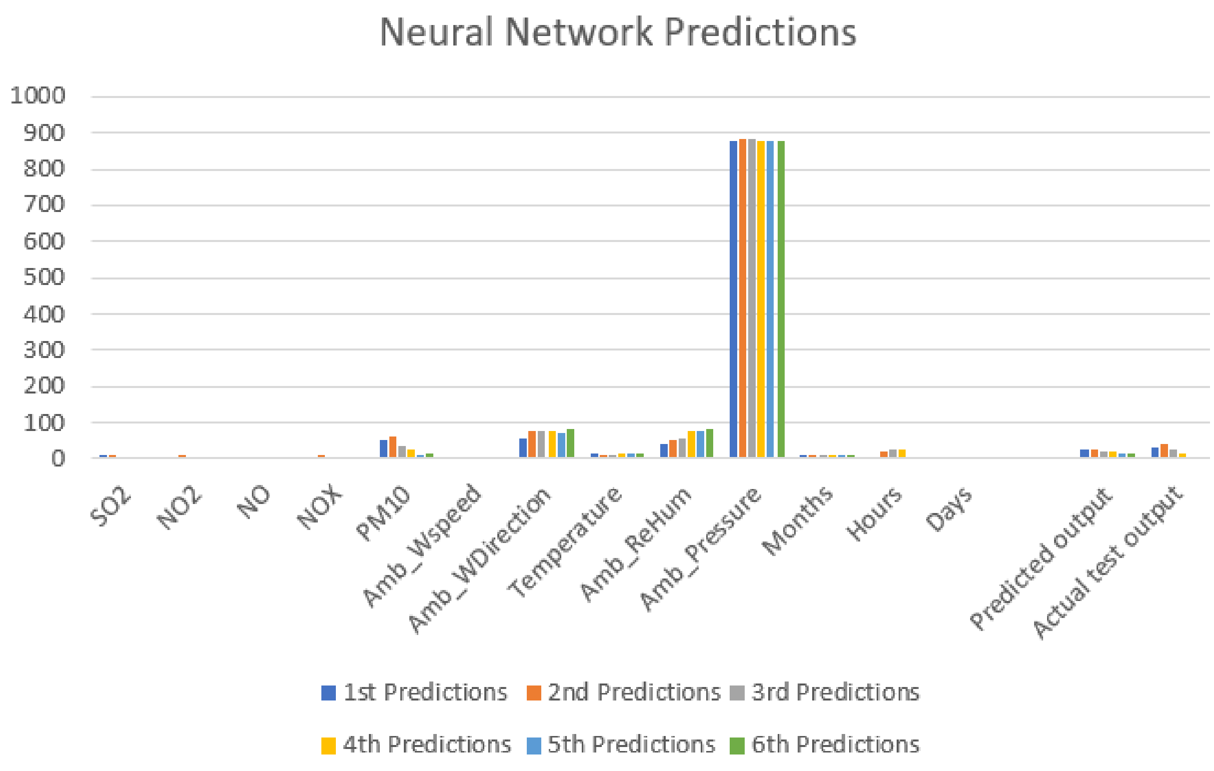


FIGURE 5.6: NEURAL NETWORK PREDICTIONS BASED ON TEST DATASETS OF TABLE 4.1

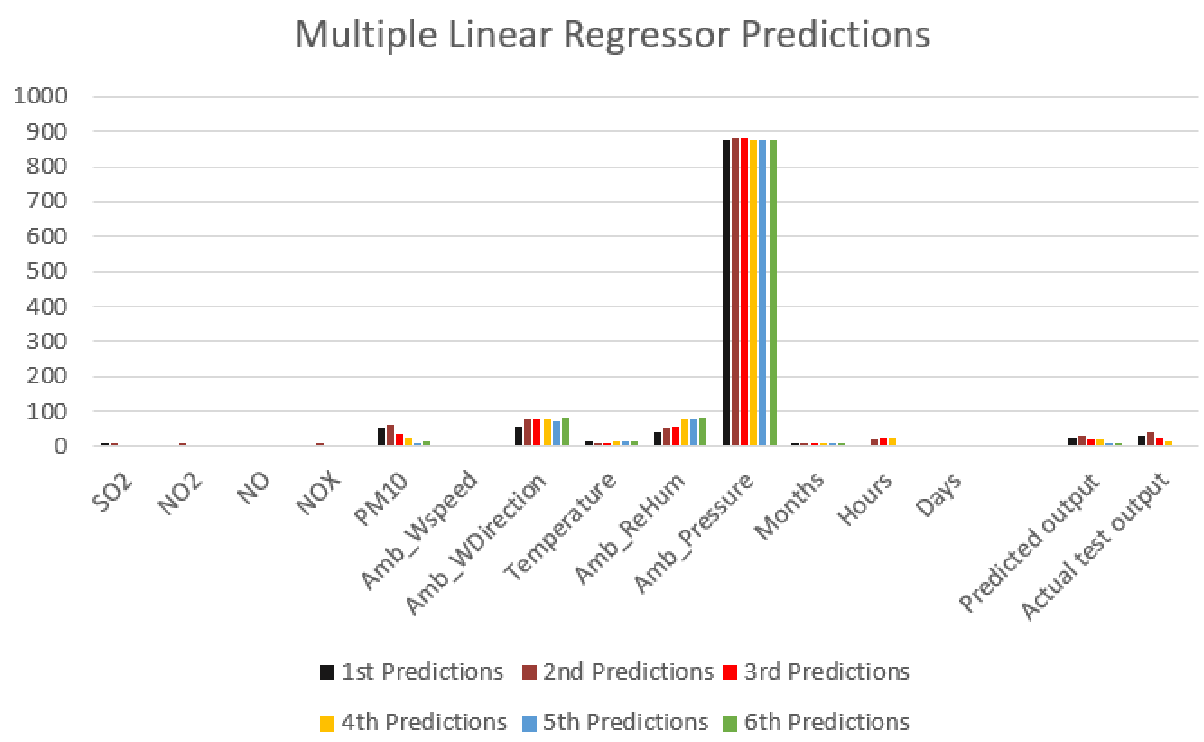


FIGURE 5.7: MULTIPLE LINEAR REGRESSOR PREDICTIONS BASED ON TEST DATASETS OF TABLE 4.1

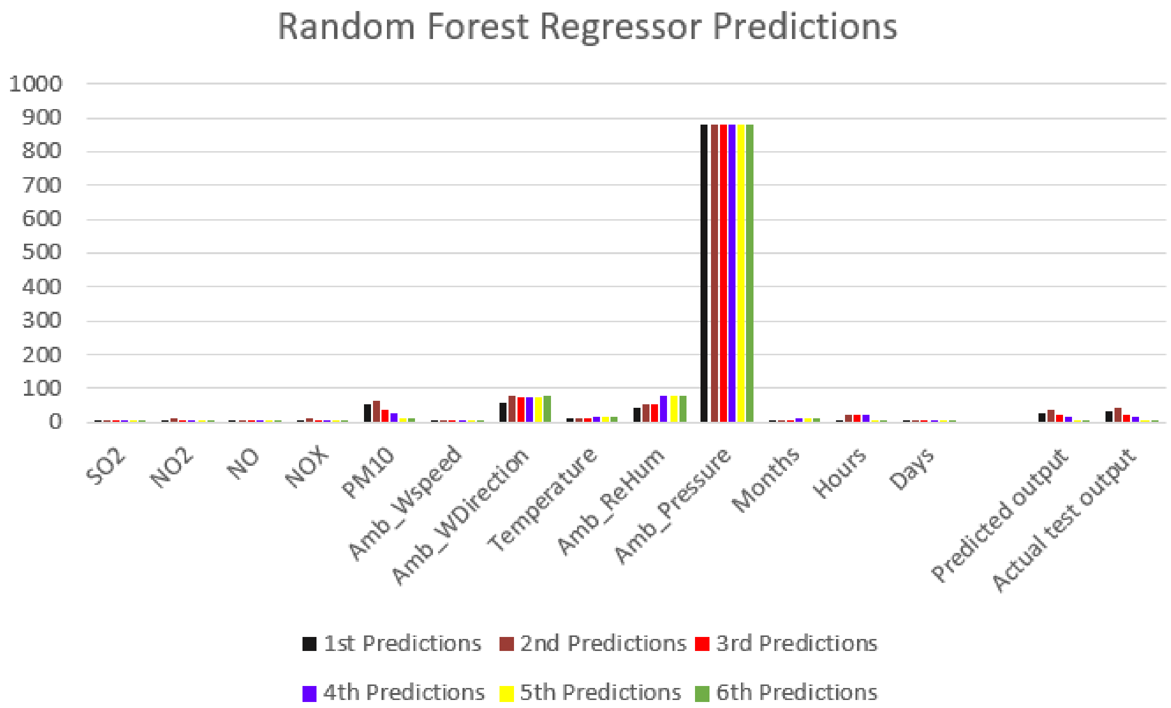


FIGURE 5.8: RANDOM FOREST REGRESSOR PREDICTIONS BASED ON TEST DATASETS OF TABLES 5.1

As shown in Table 5.1, the “Random Forest Regressor” model outperformed the other two models when the same input features of the test data set were used to make predictions. Since the models were operated at their default parameters and no hyper parameter tuning was made, it is assumed that the models would improve in their performance. For instance, in the case of neural network, the improvement would include hyper parameter tuning such as considering the network with additional hidden layers, applying the dropout in the hidden layers, and reducing the learning rate. Figures 5.9 - 5.11, shows the code snippets of the results tabulated in Table 5.1.

```
In [75]: y1_predict = model.predict(np.array([[8,6,1,7,53,1.3,56,13,43,880,8,21,6]])) # Actual is 31
print("The predicted output is {} and the actual output is 31 ".format (y1_predict))
y2_predict = model.predict(np.array([[8,11,1,12,63,1.1,78,11,52,881,8,22,6]])) # Actual is 41
print("The predicted output is {} and the actual output is 41 ".format (y2_predict))
y3_predict= model.predict(np.array([[7,5,2,7,38,1.2,76,10,54,881,8,23,6]])) # Actual is 24
print("The predicted output is {} and the actual output is 24 ".format (y3_predict))
y4_predict = model.predict(np.array([[1,2,1,2,27,1,75,17,79,879,12,24,4]])) # Actual is 16
print("The predicted output is {} and the actual output is 16 ".format (y4_predict))
y5_predict = model.predict(np.array([[1,1,1,1,12,1.2,73,17,78,878,12,1,4]])) # Actual is 7
print("The predicted output is {} and the actual output is 7 ".format (y5_predict))
y6_predict = model.predict(np.array([[1,1,1,1,13,1.1,81,16,81,878,12,2,4]])) # actual is 7
print("The predicted output is {} and the actual output is 7 ".format (y6_predict))
```

```
The predicted output is [[26.643219]] and the actual output is 31
The predicted output is [[27.536377]] and the actual output is 41
The predicted output is [[22.397385]] and the actual output is 24
The predicted output is [[20.669685]] and the actual output is 16
The predicted output is [[16.566565]] and the actual output is 7
The predicted output is [[16.83355]] and the actual output is 7
```

FIGURE 5.9: NEURAL NETWORK TIME SERIES PREDICTIONS

```
In [87]: y1_predict = regressor.predict(np.array([[8,6,1,7,53,1.3,56,13,43,880,8,21,6]])) # Actual is 31
print("The predicted output is {} and the actual output is 31 ".format (y1_predict))
y2_predict = regressor.predict(np.array([[8,11,1,12,63,1.1,78,11,52,881,8,22,6]])) # Actual is 41
print("The predicted output is {} and the actual output is 41 ".format (y2_predict))
y3_predict= regressor.predict(np.array([[7,5,2,7,38,1.2,76,10,54,881,8,23,6]])) # Actual is 24
print("The predicted output is {} and the actual output is 24 ".format (y3_predict))
y4_predict = regressor.predict(np.array([[1,2,1,2,27,1,75,17,79,879,12,24,4]])) # Actual is 16
print("The predicted output is {} and the actual output is 16 ".format (y4_predict))
y5_predict = regressor.predict(np.array([[1,1,1,1,12,1.2,73,17,78,878,12,1,4]])) # Actual is 7
print("The predicted output is {} and the actual output is 7 ".format (y5_predict))
y6_predict = regressor.predict(np.array([[1,1,1,1,13,1.1,81,16,81,878,12,2,4]])) # actual is 7
print("The predicted output is {} and the actual output is 7 ".format (y6_predict))
```

```
The predicted output is [[24.44692444]] and the actual output is 31
The predicted output is [[29.78587615]] and the actual output is 41
The predicted output is [[19.09284263]] and the actual output is 24
The predicted output is [[18.294313]] and the actual output is 16
The predicted output is [[9.50120114]] and the actual output is 7
The predicted output is [[10.22394335]] and the actual output is 7
```

FIGURE 5.10: MULTIPLE LINEAR REGRESSION TIME SERIES PREDICTIONS

```
In [92]: y1_predict = rf.predict(np.array([[8,6,1,7,53,1.3,56,13,43,880,8,21,6]])) # Actual is 31
print("The predicted output is {} and the actual output is 31 ".format (y1_predict))
y2_predict = rf.predict(np.array([[8,11,1,12,63,1.1,78,11,52,881,8,22,6]])) # Actual is 41
print("The predicted output is {} and the actual output is 41 ".format (y2_predict))
y3_predict= rf.predict(np.array([[7,5,2,7,38,1.2,76,10,54,881,8,23,6]])) # Actual is 24
print("The predicted output is {} and the actual output is 24 ".format (y3_predict))
y4_predict = rf.predict(np.array([[1,2,1,2,27,1,75,17,79,879,12,24,4]])) # Actual is 16
print("The predicted output is {} and the actual output is 16 ".format (y4_predict))
y5_predict = rf.predict(np.array([[1,1,1,1,12,1.2,73,17,78,878,12,1,4]])) # Actual is 7
print("The predicted output is {} and the actual output is 7 ".format (y5_predict))
y6_predict = rf.predict(np.array([[1,1,1,1,13,1.1,81,16,81,878,12,2,4]])) # actual is 7
print("The predicted output is {} and the actual output is 7 ".format (y6_predict))
```

```
The predicted output is [28.15] and the actual output is 31
The predicted output is [34.94] and the actual output is 41
The predicted output is [21.5] and the actual output is 24
The predicted output is [17.07] and the actual output is 16
The predicted output is [6.9] and the actual output is 7
The predicted output is [7.69] and the actual output is 7
```

FIGURE 5.11: RANDOM FOREST REGRESSOR TIME SERIES PREDICTIONS

5.6 CONCLUSION AND FUTURE WORK

This chapter has presented a machine learning technique for early air-pollution detection in a local municipality in South Africa. Real time data were collected from SAAQIS, the local service used for air monitoring within the South African local municipality. The data collected had several discrepancies and the state of air pollution was sparsely represented. Using various machine learning techniques, a section of the data was used to train the models and another set of data was used to predict the occurrence of air pollution in the short term. Furthermore, various validation methods were used to evaluate the efficacy of the predictions. Based on the results, the random forest regressor model provided better predictions and is recommended for deployment. Future work would involve comparing various models with tuned hyper parameters.

CHAPTER VI

6. CONCLUSION

This dissertation has focused on using artificial intelligence elements to predict the air quality index. The study discussed and concluded that artificial intelligence could be better to solve the problem and maintained that air contaminants on the world of today. BPDM region is facing increasing pollution due to more population, mines and daily agricultures activities. However, most research in this area has been done focusing on indoor and outdoor air quality matters. Additional review may perhaps consider on how intelligent procedures could be applied for air quality, especially in developing Africa continents. The hourly statistics, which cover the years 2019 through 2021, were gathered in BPLM. Multilayer regression and multilayer perceptron models were used to design air pollution prediction problems. After training, procedures for validation and testing are carried out to conduct comparisons and choose the best prediction model.

The literature review led to the conclusion that; artificial intelligence could be used to tackle the challenge of air pollution. The phenomenon of air pollution is deserving and crucial to human survival. Numerous problems, including the most significant early air quality warning, can benefit from precise air quality forecasting. It encourages people to exercise greater caution and take safety precautions, such as avoiding close contact with polluted air, the fossil fuel industry, transit, and manufacturing exhaust. Without clean oxygen, living is difficult and rife with illnesses. This algorithm was created to forecast air quality in real-time to address this problem.

A multilayer regression (MLR) was conducted as part of the study to forecast thirteen air quality parameters based on Biotikong dataset. This research focuses on critical parameters for air pollution quality. The studied parameters were Particulate matter with an aerodynamic diameter of less than 2.5 $\mu g/m^3$, Particulate matter with an aerodynamic diameter of less than 10 $\mu g/m^3$, QM9, Boost Voltage regulator, PM2.5 Air Quality Sensor, Adafruit Feather 32u4 FRM9x LoRa and battery 3.7V.

Of all the models constructed as part of this research, the model proved to be the best, given the model's performance for all the evaluation parameters used for this model. The algorithms are trained, validated, and tested using the data. To identify the models that are more accurate, the simulation outcomes of the models after the train and test phases are compared. The root-mean-square error was used for the evaluations. According to the findings, the Learner Regression model produces more accurate outcomes than other models. The most time-consuming procedure in the simulation was training the models. However, after extensive study and testing of various metrics, the models gave reasonably excellent results. The Random Forest Regressor model surpassed both other models when the same input characteristics of the test dataset were used to make predictions. The aim is to develop an intelligence system for predicting air pollution that could efficiently and correctly forecast hourly air quality with a low prediction error.

The models demonstrate that when all the entries for each parameter are within the required ranges, they have a positive influence on the following parameter. These desirable ranges, which result in positive impacts, begin at the parameter level, impacting the next parameter, which, in turn, determines the overall quality of air. The proposed models reveal the principles that determine quality air, implying that each parameter value must be within a specific range to achieve excellent quality air. The importance of this study is to leverage the capacity of Artificial Intelligence and Machine Learning capabilities to assist air pollution facilities in the municipality in improving their operations to ensure the provision of high-quality air is maintained.

Furthermore, in this study, an IoT-based smart air monitoring system was developed. The relevance of this system is to assure real-time monitoring of air parameters, particularly those on which this research focuses, as well as to validate the forecasted parameter values. This system used air pollution sensors for real-time monitoring, as well as a LoRa connection for wireless long range, low power, and secure data transmission. Annexure A shows the system's printed circuit board, a low-cost technology that can significantly improve the environment.

6.1 RECOMMENDATION FOR FUTURE WORK

In general, this research has prospered in integrating several methods of analytical and modeling that will prove to be useful for different organizations that are directly involved in air quality management. In the future, more studies need to be undertaken to define other suitable forecasting models considering modification in the air parameters. A physical deployment of the prototype system is being considered for deployment in the future.

Central University of Technology, Free State

REFERENCES

1. Mihăiţă, A.S., Dupont, L., Chery, O., Camargo, M. and Cai, C., 2019. Evaluating air quality by combining stationary, smart mobile pollution monitoring and data-driven modeling. *Journal of Cleaner Production*, *221*, pp.398-418.
2. Naude, B.; Mandela, N.; Sefike, M.L. Report Title: The Rustenburg Local Municipality Air Quality Monitoring Summary Report-for Compiled for: Rustenburg Local Municipality. vol. 2019
3. Gazette, G.; Notice, G. National Environment Management: Air Quality Act, 2004. *Government Gazette REPUBLIC OF SOUTH AFRICA* vol. *476*, no. 39, February 2005
4. Ghozikali, M.G., Mosaferi, M., Safari, G.H. and Jaafari, J., 2015. Effect of exposure to O3, NO2, and SO2 on chronic obstructive pulmonary disease hospitalizations in Tabriz, Iran. *Environmental Science and Pollution Research*, 22(4), pp.2817-2823.
5. West, J.J., Cohen, A., Dentener, F., Brunekreef, B., Zhu, T., Armstrong, B., Bell, M.L., Brauer, M., Carmichael, G., Costa, D.L. and Dockery, D.W., 2016. What we breathe impacts our health: improving understanding of the link between air pollution and health, pp 4895-4904.
6. Bauer, S.E., U., Mezuman, K. and Gao, C.Y., 2019. Desert dust, industrialization, and agricultural fires: Health impacts of outdoor air pollution in Africa. *Journal of Geophysical Research: Atmospheres*, pp.4104-4120.
7. Pmg.org.za. 2019. Colloquium on Climate, Air Pollution, Energy and Health; Kempton Park residents petition on refuse collection | PMG. [online] Available at: https://pmg.org.za/committee-meeting/27536/ [Accessed 28 May 2019].
8. Theunis Meyer; Dirk Cilliers BOJANALA PLATINUM DISTRICT MUNICIPALITY ENVIRONMENTAL MANAGEMENT FRAMEWORK Final EMF Report. *Centre for Environmental Management* vol. *11*, July 2018
9. Altieri, K., and Keen, S., 2016. The cost of air pollution in South Africa. International Growth Centre Blog
10. Zunckel, M., 2016. Atmospheric impact report in support of the EIA for the proposed Tshivhaso Coal-Fired Power Plant near Lephalale, Limpopo Province.
11. Saaqis.environment.gov.za. 2019. [online] Available at https://saaqis.environment.gov.za/pagesfiles/Strategy_to_address_air_pollution_in_dense_low-income_settlements.pdf [Accessed 7 Jun. 2019].
12. Okokpujie, K.O., Noma-Osaghae, E., Odusami, M., John, S.N. and Oluga, O., 2018. A Smart Air Pollution Monitoring System. *International Journal of Civil Engineering and Technology (IJCIET), 9(9)*, pp.799-809.
13. Dhingra, S., Madda, R.B., Gandomi, A.H., Patan, R. and Daneshmand, M., 2019. Internet of Things Mobile-Air Pollution Monitoring System (IoT-Mobair). *IEEE Internet of Things Journal 6(3),* pp.5577-5584.

14. Pal, P., Gupta, R., Tiwari, S. and Sharma, A., 2017. IoT based air pollution monitoring system using Arduino. *International Research Journal of Engineering and Technology (IRJET), 4(10)*, pp.1137-1140.
15. Ong, B.T., Sugiura, K. and Zettsu, K., 2016. Dynamically pre-trained deep recurrent neural networks using environmental monitoring data for predicting PM 2.5. *Neural Computing and Applications*, *27*, pp.1553-1566.
16. Saaqis.environment.gov.za. 2021. [online] Available at: <https://saaqis.environment.gov.za/Pagesfiles/Rustenberg%20Local%20Municipality%20Air%20Pollution%20Control%20By-Law_03-07-2008.pdf> [Accessed 10 June 2021].
17. Asha, P., Natrayan, L.B.T.J.R.R.G.S., Geetha, B.T., Beulah, J.R., Sumathy, R., Varalakshmi, G. and Neelakandan, S., 2022. IoT enabled environmental toxicology for air pollution monitoring using AI techniques. *Environmental research, 205*, p.112574.
18. Hanif, I., 2018. Impact of economic growth, nonrenewable and renewable energy consumption, and urbanization on carbon emissions in Sub-Saharan Africa. *Environmental Science and Pollution Research*, *25*(15), pp.15057-15067.
19. World Health Organization, 2021. Health benefits of raising ambition in Pakistan’s nationally determined contribution: WHO technical report.
20. Zhao, H., Chen, K., Liu, Z., Zhang, Y., Shao, T. and Zhang, H., 2021. Coordinated control of PM2. 5 and O3 is urgently needed in China after implementation of the “Air pollution prevention and control action plan”. *Chemosphere*, *270*, p.129441.
21. Dhimal, M., Chirico, F., Bista, B., Sharma, S., Chalise, B., Dhimal, M.L., Ilesanmi, O.S., Trucillo, P. and Sofia, D., 2021. Impact of air pollution on global burden of disease in 2019. *Processes*, *9*(10), p.1719.
22. Saaqis.environment.gov.za. (2019). Saaqis. [online] Available at: https://saaqis.environment.gov.za/ [Accessed 5 Apr. 2019].
23. Bruce, N., Perez-Padilla, R. and Albalak, R., 2000. Indoor air pollution in developing countries: a major environmental and public health challenge. *Bulletin of the World Health organization*, *78*(9), pp.1078-1092.
24. Narain, U. and Sall, C., 2016. Methodology for valuing the health impacts of air pollution.
25. Nath, B., Singh, J.V., Awasthi, S., Bhushan, V., Kumar, V., and Singh, S.K., 2007. Urban Health Analysis. Indian J Med Sci, 61(11), pp.598-606.
26. Okello, N.O., Okello, T.W. and Zunckel, M., 2020. Changes in health risk associated with air pollution and policy response effectiveness, Richards Bay, South Africa. *Clean Air Journal*, *30*(1), pp.1-10.
27. Lelieveld, J., Klingmüller, K., Pozzer, A., Burnett, R.T., Haines, A. and Ramanathan, V., 2019. Effects of fossil fuel and total anthropogenic emission removal on public health and climate. *Proceedings of the National Academy of Sciences, 116(15)*, pp.7192-7197.
28. Fried, A. and Richter, D., 2006. Analytical techniques for atmospheric measurement. *Analytical Techniques for Atmospheric Measurement, Blackwell Publishing, Oxford, UK*, pp.72-146.

29. Maciejczyk, P., Chen, L.C. and Thurston, G., 2021. The role of fossil fuel combustion metals in PM2. 5 air pollution health associations. *Atmosphere, 12(9),* p.1086.
30. Shamsipour, M., Hassanvand, M.S., Gohari, K., Yunesian, M., Fotouhi, A., Naddafi, K., Sheidaei, A., Faridi, S., Akhlaghi, A.A., Rabiei, K. and Mehdipour, P., 2019. National and sub-national exposure to ambient fine particulate matter (PM2. 5) and its attributable burden of disease in Iran from 1990 to 2016. *Environmental Pollution*, *255*, p.113173.
31. DeSouza, P., Anjomshoaa, A., Duarte, F., Kahn, R., Kumar, P. and Ratti, C., 2020. Air quality monitoring using mobile low-cost sensors mounted on trash-trucks: Methods development and lessons learned. *Sustainable Cities and Society, 60*, p.102239.
32. Kang, G.K., Gao, J.Z., Chiao, S., Lu, S. & Xie, G. 2018, "Air quality prediction: Big data and machine learning approaches", *International Journal of Environmental Science and Development, vol. 9, no. 1*, pp. 8-16[3]
33. Shackleton, N. and Bird, T., 2020. Air Quality Specialist Report for the Bakubung Platinum Mine TSF Project near Ledig, Northwest Province.
34. Büscher, B., Koot, S. and Thakholi, L., 2022. Fossilized conservation, or the unsustainability of saving nature in South Africa. *Environment and Planning E: Nature and Space*, p.25148486211062002.
35. Magaliesbergbiosphere.org.za. 2021. [online] Available at: <https://magaliesbergbiosphere.org.za/wp-content/uploads/2019/11/BPDM-Final-EMF-report-Rev-11-1.pdf> [Accessed 10 June 2021].
36. Simelane, S.P., 2021. *A Critical Evaluation of Health Risk Assessments of PM 2.5 from Coal-Fired Power Stations*. University of Johannesburg (South Africa).
37. Saaqis.environment.gov.za. 2021. [online] Available at: <https://saaqis.environment.gov.za/Pagesfiles/Rustenberg%20Local%20Municipality%20Air%20Pollution%20Control%20By-Law_03-07-2008.pdf> [Accessed 10 June 2021].
38. Cer.org.za. 2021. [online] Available at: <https://cer.org.za/wp-content/uploads/2005/09/WBPA-Air-Air-Quality-Management-Plan.pdf> [Accessed 21 July 2021].
39. Phala, R.N., 2015. Using an inferential model to estimate dry deposition of SO2 and NOX (as NO2) *in Lephalale in the Waterberg-Bojanala priority area* (Doctoral dissertation).
40. Koyana, N., Markus, E.D. and Abu-Mahfouz, A.M., 2019, November*. A survey of network and intelligent air pollution monitoring in South Africa. In 2019 International Multidisciplinary Information Technology and Engineering Conference (IMITEC)* (pp. 1-7). IEEE.
41. Kumar, P., Singh, A.B., Arora, T., Singh, S. and Singh, R., 2023. Critical review on emerging health effects associated with the indoor air quality and its sustainable management. *Science of The Total Environment*, *872*, p.162163.
42. Subramaniam, S., Raju, N., Ganesan, A., Rajavel, N., Chenniappan, M., Prakash, C., Pramanik, A., Basak, A.K. and Dixit, S., 2022. Artificial

Intelligence Technologies for Forecasting Air Pollution and Human Health: A Narrative Review. Sustainability, 14(16), p.9951.
43. Cer.org.za. (2019). [online] Available at: https://cer.org.za/wp-content/uploads/2010/03/North-West-Environmental-Implementation-Plan.pdf [Accessed 8 April. 2019].
44. Saaqis.environment.gov.za. (2019). Saaqis. [online] Available at: https://saaqis.environment.gov.za/ [Accessed 5 Apr. 2019].
45. Yafouz, A., Ahmed, A.N., Zaini, N.A., Sherif, M., Sefelnasr, A. and El-Shafie, A., 2021. Hybrid deep learning model for ozone concentration prediction: Comprehensive evaluation and comparison with various machine and deep learning algorithms. *Engineering Applications of Computational Fluid Mechanics*, *15*(1), pp.902-933.
46. Pavani, M. and Rao, P.T., 2017. Urban air pollution monitoring using wireless sensor networks: A comprehensive review. *International Journal of Communication Networks and Information Security*, *9*(3), pp.439-449.
47. Sendra, S., Garcia-Navas, J.L., Romero-Diaz, P. and Lloret, J., 2019, June. Collaborative lora-based sensor network for pollution monitoring in smart cities. In *2019 Fourth International Conference on Fog and Mobile Edge Computing (FMEC)* (pp. 318-323). IEEE.
48. Kumar, P., Singh, A.B., Arora, T., Singh, S. and Singh, R., 2023. Critical review on emerging health effects associated with the indoor air quality and its sustainable management. *Science of The Total Environment*, *872*, p.162163.
49. Sahhid, N., Shah, M.A., Khan, A., Maple, C. and Jeon, G., 2021. *Towards Greener Smart Cities and Road Traffic Forecasting Using Air Pollution Data. Sustainable Cities and Society,* p.103062
50. Jenab, K., Seyedhosseini, S.M., Khoury, S. and Sarfaraz, A., 2012. An intelligent air quality monitoring model in manufacturing. *Clean Technologies and Environmental Policy,* 14(5), pp.917-923.
51. Guanochanga, B., Cachipuendo, R., Fuertes, W., Salvador, S., Benítez, D.S., Toulkeridis, T., Torres, J., Villacís, C., Tapia, F. and Meneses, F., 2018, November. Real-time air pollution monitoring systems using wireless sensor networks connected in a cloud-computing wrapped up web services. In Proceedings of the Future Technologies Conference (pp. 171-184). Springer, Cham
52. Genikomsakis, N.K., Galatoulas, N.F., Dallas, I.P., Candanedo Ibarra, L.M., Margaritis, D., and S. Ioakimidis, C., 2018. Development and on-field testing of low-cost portable system for monitoring PM2. 5 concentrations. Sensors, 18(4), p.1056.
53. Doni, A., Murthy, C., and Kurian, M.Z., 2018. Survey on multi sensor-based air and water quality monitoring using IoT. *Indian J. Sci. Res*, *17*(2), pp.147-153.
54. Kumar, Somansh, and Ashish Jasuja*. "Air quality monitoring system based on IoT using Raspberry Pi." 2017* International Conference on Computing, Communication, and Automation (ICCCA). IEEE, 2017
55. Parmar, G., Lakhani, S., and Chattopadhyay, M.K., 2017, October. *An IoT based low-cost air pollution monitoring system.* In 2017 International

Conference on Recent Innovations in Signal Processing and Embedded Systems (RISE) (pp. 524-528). IEEE.
56. Mahajan, S., Chen, L.J., and Tsai, T.C., 2018. *Short-Term PM2. 5 Forecasting Using Exponential Smoothing Method: A Comparative Analysis. Sensors, 18(10),* p.3223.
57. Manna, Souvik, Suman Sankar Bhunia, and Nandini Mukherjee. *"Vehicular pollution monitoring using IoT."* International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014). IEEE, 2014.
58. Abu Safia, A., Al Aghbari, Z. and Kamel, I., 2017. *Efficient data collection by mobile sink to detect phenomena in internet of things. Information, 8(4),* p.123.
59. Lin, Y.C., Chi, W.J. and Lin, Y.Q., 2020. *The improvement of spatial-temporal resolution of PM2. 5 estimations based on micro-air quality sensors by using data fusion technique. Environment international, 134*, p.105305.
60.] Rushikesh, Ramagiri, and Chandra Mohan Reddy Sivappagari. "*Development of IoT based vehicular pollution monitoring system."* 2015 International Conference on Green Computing and Internet of Things (ICGCIoT). IEEE, 2015.
61.] Luo, L., Zhang, Y., Pearson, B., Ling, Z., Yu, H. and Fu, X. 2018*. On the Security and Data Integrity of Low-Cost Sensor Networks for Air Quality Monitoring. Sensors, 18(12)*, p.4451.
62. Balasubramaniyan, C. and Manivannan, D. 2016. *Iot enabled air quality monitoring system (AQMS) using raspberry Pi. Indian Journal of Science and Technology, 9(39),* pp.1-6.
63. Kumar, S., Tiwari, P. and Zymbler, M., 2019. Internet of Things is a revolutionary approach for future technology enhancement: a review. *Journal of Big data*, *6*(1), pp.1-21.
64. Mahajan, S., Chen, L.J., and Tsai, T.C. 2018. *Short-Term PM2. 5 Forecasting Using Exponential Smoothing Method: A Comparative Analysis. Sensors, 18(10),* p.3223.
65. Saha, H.N., Auddy, S., Chatterjee, A., Pal, S., Pandey, S., Singh, R., Singh, R., Sharan, P., Banerjee, S., Ghosh, D. and Maity, A., 2017, August. Pollution control using the Internet of Things (IoT). In 2017 8th Annual Industrial Automation and Electromechanical Engineering
66. Cheng, Y., Li, X., Li, Z., Jiang, S., Li, Y., Jia, J. and Jiang, X., 2014, November. AirCloud: a cloud-based air-quality monitoring system for everyone. In Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems (pp. 251-265). ACM. Conference (IEMECON) (pp. 65-68). IEEE.
67. Latre, S., Leroux, P., Coenen, T., Braem, B., Ballon, P. and Demeester, P., 2016, September. City of things: An integrated and multi-technology testbed for IoT smart city experiments. In 2016 IEEE International Smart Cities Conference (ISC2) (pp. 1-8). IEEE.
68. Haxhibeqiri, J., De Poorter, E., Moerman, I. and Hoebeke, J., 2018. *A survey of LoRaWAN for IoT: From technology to application. Sensors, 18(11),* p.3995.
69. Fadeyi, Johnson, Elisha Didam Markus, and Adnan M. Abu-Mahfouz. "*Technology Coexistence in LPWANs-A Comparative Analysis for Spectrum*

Optimization." 2019 IEEE 28th International Symposium on Industrial Electronics (ISIE). IEEE, 2019.

70. Molefi, M., Markus, E.D. and Abu-Mahfouz, A., 2019, *January. Wireless Power Transfer for LoRa Low-Power wide-area Networks (LPWANs).* In 2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA) (pp. 105-110). IEEE.
71. Lee, H.C. and Ke, K.H. 2018. Monitoring of large-area IoT sensors using a LoRa wireless mesh network system: *Design and evaluation. IEEE Transactions on Instrumentation and Measurement,* 67(9), pp.2177-2187.
72. Liu, Jen-Hao, et al. *"AN AIR QUALITY MONITORING SYSTEM FOR URBAN AREAS BASED ON THE TECHNOLOGY OF WIRELESS SENSOR NETWORKS." International Journal on Smart Sensing & Intelligent Systems* 5.1 (2012).
73. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J., 2011. Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 12, pp.2825-2830.
74. Keskin, G.A., Doğruparmak, Ş.Ç. and Ergün, K., 2022. Estimation of COVID-19 patient numbers using artificial neural networks based on air pollutant concentration levels. *Environmental Science and Pollution Research*, *29*(45), pp.68269-68279.
75. Bell, M.L., Samet, J.M. and Dominici, F., 2004. Time-series studies of particulate matter. *Annu. Rev. Public Health, 25*, pp.247-280.
76. Abba, S.I., Pham, Q.B., Usman, A.G., Linh, N.T.T., Aliyu, D.S., Nguyen, Q. and Bach, Q.V., 2020. Emerging evolutionary algorithm integrated with kernel principal component analysis for modeling the performance of a water treatment plant. *Journal of Water Process Engineering*, *33*, p.101081.
77. Masood, A. and Ahmad, K., 2021. A review on emerging artificial intelligence (AI) techniques for air pollution forecasting: Fundamentals, application, and performance. *Journal of Cleaner Production*, *322*, p.129072.
78. Bellinger, C., Jabbar, M.S.M., Zaïane, O. and Osornio-Vargas, A., 2017. A systematic review of data mining and machine learning for air pollution epidemiology. *BMC public health*, 17(1), pp.1-19.
79. Kang, G.K., Gao, J.Z., Chiao, S., Lu, S. and Xie, G., 2018. Air quality prediction: Big data and machine learning approaches. *International Journal of Environmental Science and Development,* 9(1), pp.8-16.
80. Alyousifi, Y., Othman, M., Faye, I., Sokkalingam, R. and Silva, P.C., 2020. Markov weighted fuzzy time-series model based on an optimum partition method for forecasting air pollution. International Journal of Fuzzy Systems, 22, pp.1468-1486.
81. Heydari, A., Majidi Nezhad, M., Astiaso Garcia, D., Keynia, F. and De Santoli, L., 2022. Air pollution forecasting application based on deep learning model and optimization algorithm. *Clean Technologies and Environmental Policy*, pp.1-15.

82. Parmezan, A.R.S., Souza, V.M. and Batista, G.E., 2019. Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. *Information sciences*, *484*, pp.302-337.
83. Zhang, J. and Li, S., 2022. Air quality index forecast in Beijing based on CNN-LSTM multi-model. *Chemosphere*, *308*, p.136180.
84. Mohammadi, M., Al-Fuqaha, A., Sorour, S. and Guizani, M., 2018. Deep learning for IoT big data and streaming analytics: A survey. *IEEE Communications Surveys & Tutorials*, *20*(4), pp.2923-2960.
85. Dominici, F., McDermott, A. and Hastie, T., 2003. Issues in semi-parametric regression with applications in time series models for air pollution and mortality. *Preprint*.
86. Tec, M., Mudele, O., Josey, K. and Dominici, F., 2023. Causal Shift-Response Functions with Neural Networks: The Health Benefits of Lowering Air Quality Standards in the US. *arXiv preprint arXiv:2302.02560*.
87. Kök, İ., Şimşek, M.U. and Özdemir, S., 2017, December. A deep learning model for air quality prediction in smart cities. In 2017 IEEE International Conference on Big Data (Big Data) (pp. 1983-1990). IEEE
88. Hu, K., Rahman, A., Bhrugubanda, H. and Sivaraman, V., 2017. HazeEst: Machine learning based metropolitan air pollution estimation from fixed and mobile sensors. IEEE Sensors Journal, 17(11), pp.3517-3525.
89. Shams, S.R., Jahani, A., Kalantary, S., Moeinaddini, M. and Khorasani, N., 2021. The evaluation on artificial neural networks (ANN) and multiple linear regressions (MLR) models for predicting SO2 concentration. Urban Climate, 37, p.100837.
90. Iskandaryan, D., Ramos, F. and Trilles, S., 2020. Air quality prediction in smart cities using machine learning technologies based on sensor data: A review. Applied Sciences, 10(7), p.2401.
91. Samal, K.K.R., Panda, A.K., Babu, K.S. and Das, S.K., 2021. An improved pollution forecasting model with meteorological impact using multiple imputation and fine-tuning approach. *Sustainable Cities and Society*, *70*, p.102923.
92. Kow, P.Y., Wang, Y.S., Zhou, Y., Kao, I.F., Issermann, M., Chang, L.C., and Chang, F.J., 2020. Seamless integration of convolutional and back-propagation neural networks for regional multi-step-ahead PM2. 5 forecasting. Journal of Cleaner Production, 261, p.121285.
93. Chiwewe, T.M. and Ditsela, J., 2016, July. Machine learning based estimation of Ozone using spatio-temporal data from air quality monitoring stations. In *2016 IEEE 14th International Conference on Industrial Informatics (INDIN)* (pp. 58-63). IEEE.
94. Wang, Z., Chen, H., Zhu, J. and Ding, Z., 2022. Daily PM2. 5 and PM10 forecasting using linear and nonlinear modeling framework based on robust local mean decomposition and moving window ensemble strategy. *Applied Soft Computing*, *114*, p.108110.
95. Yi, H., Xiong, Q., Zou, Q., Xu, R., Wang, K. and Gao, M., 2019, July. A novel random forest and its application on classification of air quality. In *2019 8th*

International Congress on Advanced Applied Informatics (IIAI-AAI) (pp. 35-38). IEEE.
96. Sahhid, N., Shah, M.A., Khan, A., Maple, C. and Jeon, G., 2021. Towards Greener Smart Cities and Road Traffic Forecasting Using Air Pollution Data. Sustainable Cities and Society, p.103062.
97. Jamal, A. and Nodehi, R.N., 2017. Predicting air quality index based on meteorological data: A comparison of regression analysis, artificial neural networks, and decision tree. *Journal of Air Pollution and Health*, *2*(1).
98. Zheng, K., Zhao, S., Yang, Z., Xiong, X. and Xiang, W., 2016. Design and implementation of LPWA-based air quality monitoring system. *IEEE Access*, *4*, pp.3238-3245.
99. Mumtaz, R., Zaidi, S.M.H., Shakir, M.Z., Shafi, U., Malik, M.M., Haque, A., Mumtaz, S. and Zaidi, S.A.R., 2021. Internet of things (IoT) based indoor air quality sensing and predictive analytic—A COVID-19 perspective. *Electronics*, *10*(2), p.184.
100. Ma, Y., Yang, S., Huang, Z., Hou, Y., Cui, L. and Yang, D., 2014, c December. Hierarchical air quality monitoring system design. In 2014 International Symposium on Integrated Circuits (ISIC) (pp. 284-287). IEEE.
101. [DIRosmiati, M., Rizal, M.F., Susanti, F. and Alfisyahrin, G.F., 2019. Air pollution monitoring system using LoRa modul as transceiver system. Telkomnika, 17(2), pp.586-592.
102. Idrees, Z. and Zheng, L., 2020. Low-cost air pollution monitoring systems: A review of protocols and enabling technologies. *Journal of Industrial Information Integration*, *17*, p.100123.
103. Taylan, Osman. "Modelling and analysis of ozone concentration by artificial intelligent techniques for estimating air quality." Atmospheric environment 150 (2017): 356-365.
104. Martínez-España, R., Bueno-Crespo, A., Timon-Perez, I.M., Soto, J.A., Munoz, A., and Cecilia, J.M., 2018. Air-pollution prediction in smart cities through machine learning methods: A Case of study in Murcia, Spain. *J. Univers. Comput. Sci.*, *24*(3), pp.261-276.
105. Cecilia, J.M., Timón, I., Soto, J., Santa, J., Pereñíguez, F. and Muñoz, A., 2018. High-throughput infrastructure for advanced its services: A case study on air pollution monitoring. *IEEE Transactions on Intelligent Transportation Systems*, *19*(7), pp.2246-2257.
106. Mahalingam, U., Elangovan, K., Dobhal, H., Valliappa, C., Shrestha, S. and Kedam, G., 2019, March. A machine learning model for air quality prediction for smart cities. In *2019 International conference on wireless communications signal processing and networking (WiSPNET)* (pp. 452-457). IEEE.
107. Satrio, P., Mahlia, T.M.I., Giannetti, N. and Saito, K., 2019. Optimization of HVAC system energy consumption in a building using artificial neural network and multi-objective genetic algorithm. *Sustainable Energy Technologies and Assessments*, *35*, pp.48-57.

108. Cer.org.za. (2019). [online] Available at: https://cer.org.za/wp-content/uploads/2010/03/North-West-Environmental-Implementation-Plan.pdf [Accessed 8 April. 2019].].
109. Mphahlele, M. and Zandamela, H., 2021. Local government capacity development: A case study of a South African district municipality. *Journal of Public Administration and Governance*, *11*(2), pp.156177-156177.
110. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J., 2011. Scikit-learn: Machine learning.

APPENDICES

Appendix A

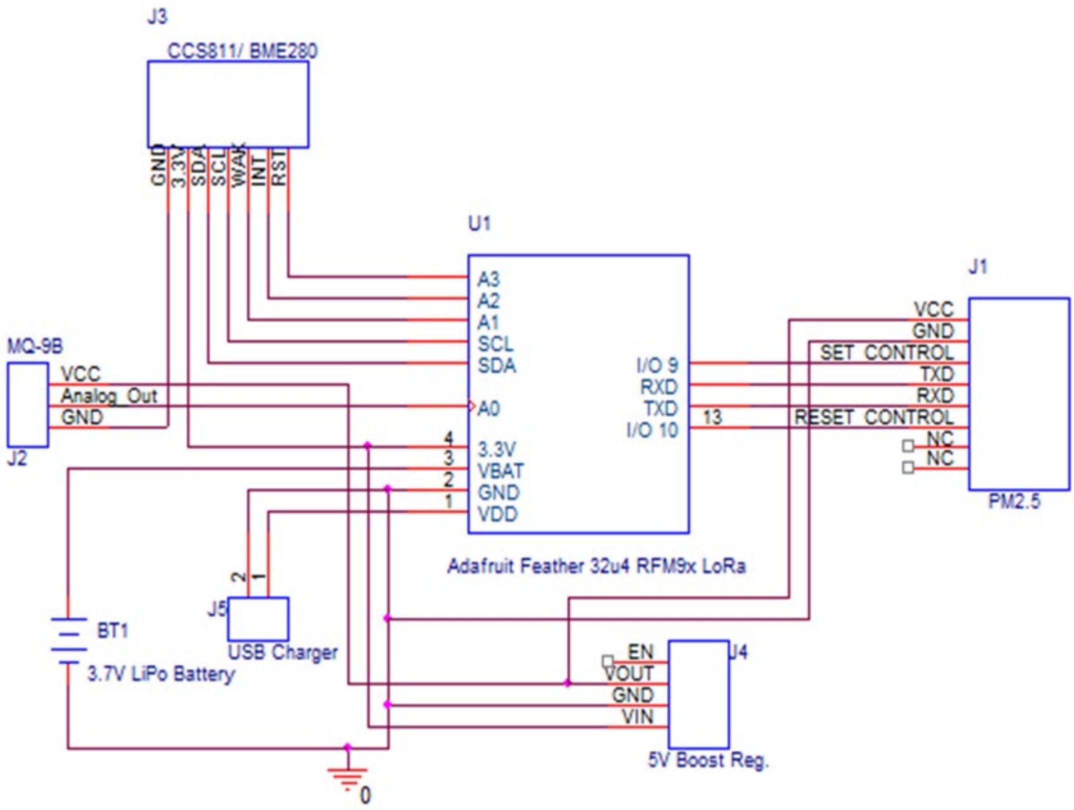


FIGURE 4.11.1 INTELLIGENT OF AIR POLLUTION SYSTEM SCHEMATIC

Appendix B

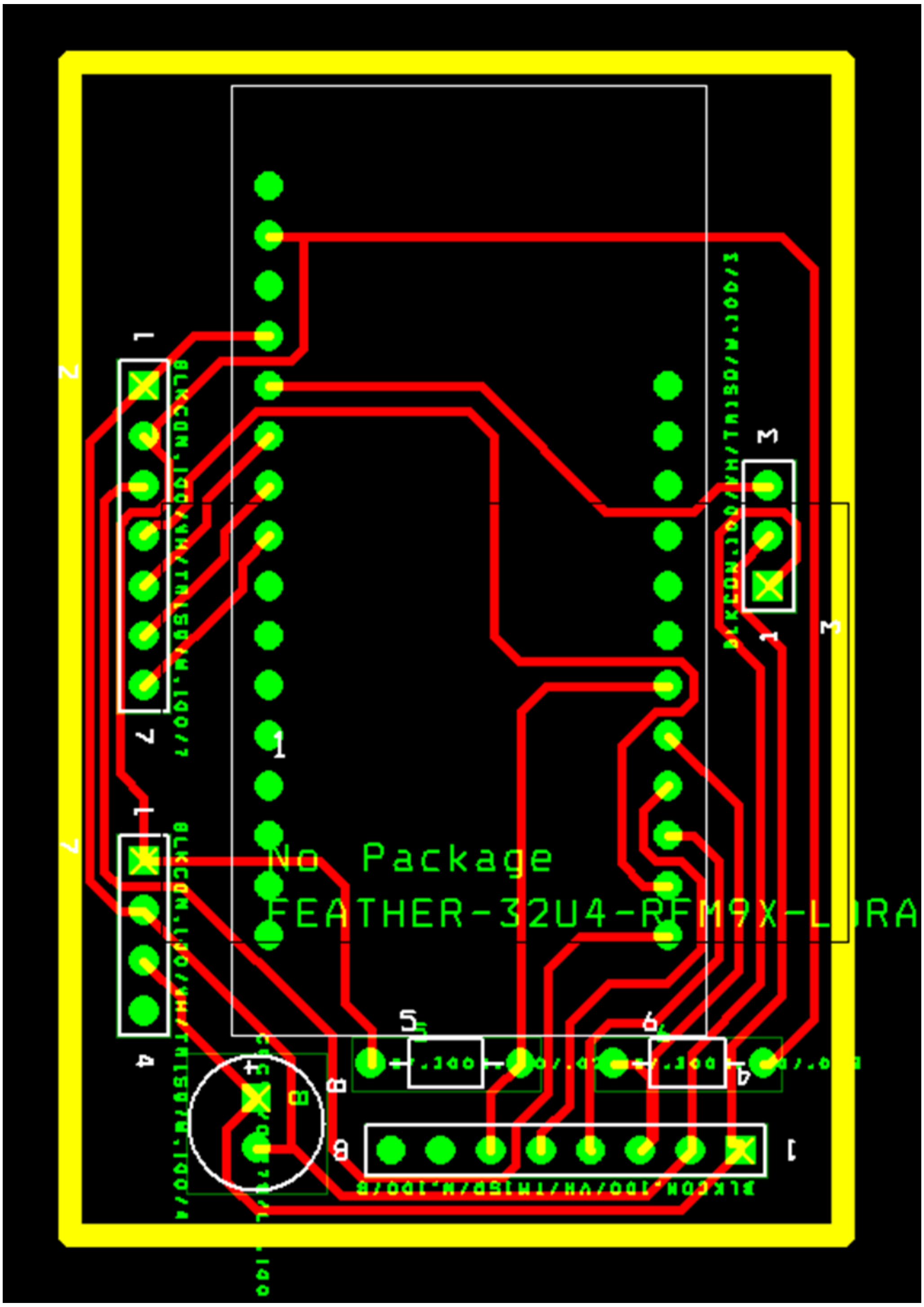


FIGURE 4.11.2 INTELLIGENT OF AIR POLLUTION SYSTEM PCB