

Predictive Modeling and Machine Learning Approaches for Real-Time Air Pollution Monitoring in Dhaka City

By
Rihan Jamil Rahat
213-15-4517

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the **Degree of Bachelor of Science in Computer Science and Engineering**

Supervised by

Md. Sazzadur Ahamed
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised by

Mr. Golam Rabbany
Lecturer
Department of Computer Science and Engineering
Daffodil International University



**DAFFODIL INTERNATIONAL
UNIVERSITY**
Dhaka, Bangladesh

September 16, 2025

APPROVAL

This Project titled “**Predictive Modeling and Machine Learning Approaches for Real-Time Air Pollution Monitoring in Dhaka City**” submitted by **Rihan Jamil Rahat**, ID: 213-15-4517 to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **16-09-2025**.

BOARD OF EXAMINERS



Dr. Arif Mahmud
Associate Professor & Associate Head
Department of Computer Science and Engineering
Faculty of Science and Information Technology (FSIT)
Daffodil International University

Board Chairman



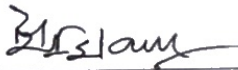
Dr. Md. Ali Hossain
Associate Professor
Department of Computer Science and Engineering
Faculty of Science and Information Technology (FSIT)
Daffodil International University

Internal Examiner



Mr. Amir Soheli
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science and Information Technology (FSIT)
Daffodil International University

Internal Examiner



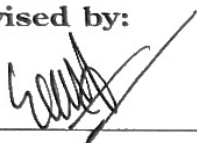
Dr. Md. Manowarul Islam
Professor
Department of Computer Science and Engineering
Jagannath University

External Examiner

DECLARATION

I hereby declare that this project has been done by me under the supervision of **Md. Sazzadur Ahamed, Assistant Professor**, Department of Computer Science and Engineering, Daffodil International University. I also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:



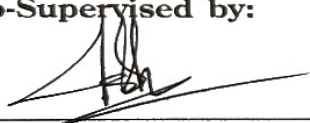
Md. Sazzadur Ahamed

Assistant Professor

Department of Computer Science and Engineering

Daffodil International University

Co-Supervised by:



Mr. Golam Rabbany

Lecturer

Department of Computer Science and Engineering

Daffodil International University

Submitted by:

Rihan Jamil Rahat

Rihan Jamil Rahat

Student ID: 213-15-4517

Department of Computer Science and Engineering

Daffodil International University

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. I am deeply grateful to everyone who has assisted me in one way or another.

First, I express my heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for me to complete the **Final Year Design Project(FYDP)** successfully.

I am grateful and wish our profound indebtedness to **Md. Sazzadur Ahamed, Assistant Professor**, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of my supervisor in the field of **Machine Learning** to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

I would like to express my heartfelt gratitude to the Head of the Department of Computer Science and Engineering, for his kind help in finishing my project and also to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

I would like to thank my entire course-mates at Daffodil International University, who took part in this discussion while completing the coursework.

Finally, I must acknowledge with due respect the constant support and patience of my parents.

ABSTRACT

In recent years, Dhaka has experienced rapid urbanization, which has intensified sources of air contamination such as vehicular emissions, industrial activity, and unregulated construction. These factors have contributed to a noticeable deterioration in air quality. For this causes people faces health problems such as asthma, lung disease, and heart disease, especially among children and old people. In this work, I tried to create a device that could observe and also predict air quality in live. For collecting data, I used IoT-based sensors that measure PM2.5, PM10, CO and some other pollutants. After collecting the data, I experimented with several traditional and deep learning models, including regression-based, probabilistic, distance-based, and recurrent neural network approaches. These models are used for predicting air quality levels and to see which model gives better results. From the comparison, I found that using a mix of models can predict more accurately than using only one model. The system can also give an early warning when pollution is very high. I believe this research will be helpful for Dhaka city planners and authorities so that they can take actions to reduce pollution.

Table of Contents

Approval	i
Declaration	ii
Acknowledgements	iii
Abstract	iv
List of Figures	vii
List of Tables	viii
1 Introduction	1
1.1 Introduction.....	1-2
1.2 Motivation.....	2
1.3 Objectives.....	2-3
1.4 Methodology.....	3
1.5 Project Outcome.....	3
1.6 Organized of the Report.....	3-4
2 Background	5
2.1 Introduction.....	5
2.2 Literature Review.....	5-14
2.2.1 Similar Applications.....	14-15
2.3 Gap Analysis.....	15
2.4 Summary.....	16
3 Research Methodology	17
3.1 Methodology.....	17
3.1.1 Overview.....	17
3.1.2 Proposed Methodology.....	17-19
3.1.3 Functional and Nonfunctional Requirements.....	19-20
3.1.4 Data Flow Diagram.....	20-21
3.1.5 Dataset Overview.....	21-22
3.1.6 UI Design.....	22-23
3.2 Detailed Methodology and Design.....	24
3.2.1 Alternate Solutions Considered.....	24
3.2.2 System Design.....	25
3.2.3 Circuit Diagram.....	26
3.3 Project Plan.....	26-27

3.4	Task Allocation.....	27-28
3.5	Summary.....	28
4	Implementation and Results	29
4.1	Environment Setup.....	29
4.1.1	System Setup Diagram.....	30
4.2	Testing and Evaluation.....	30-33
4.3	Results and Discussion.....	34
4.4	Summary.....	35
5	Engineering Standards and Design Challenges	36
5.1	Compliance with the Standards.....	36
5.1.1	Software Standards.....	36
5.1.2	Hardware Standards.....	36
5.1.3	Communication Standards.....	37
5.2	Impact on Society, Environment and Sustainability.....	37
5.2.1	Impact on Life.....	37
5.2.2	Impact on Society & Environment.....	37
5.2.3	Ethical Aspects.....	37
5.2.4	Sustainability Plan.....	37-38
5.3	Project Management and Financial Analysis.....	38
5.4	Complex Engineering Problem.....	39
5.4.1	Complex Problem Solving.....	39-40
5.4.2	Engineering Activities.....	41
5.5	Summary.....	41-42
6	Conclusion	43
6.1	Summary.....	43-44
6.2	Limitation.....	44
6.3	Future Work.....	44-45
	References	46-48

List of Figures

3.1	Methodology Pipeline Diagram.....	18
3.2	Data Flow diagram.....	20
3.3	Data Collection Page of Air Quality Monitoring (Temperature, Humidity, CO).....	22
3.4	Data Collection Page of Air Quality Monitoring (NH3, PM2.5, PM10).....	23
3.5	Result Dashboard of air quality Prediction.....	23
3.6	Circuit Diagram of This System.....	26
4.1	Context Diagram of This System.....	30
4.2	Feature Correlation Heatmap.....	31
4.3	Confusion Matrix of LR.....	32
4.4	Confusion Matrix of KNN.....	32
4.5	Confusion Matrix of NB.....	32
4.6	Confusion Matrix of SVM.....	32
4.7	Confusion Matrix of LSTM.....	33
4.8	Actual vs Prediction Air Quality.....	34

List of Tables

2.1	Summary of Literature Reviewed.....	11-14
2.2	Feature Comparison with Existing Systems.....	15
3.1	Sample Dataset of Collected Air Quality Data.....	21
3.2	Summarye of Dataset's Attributes.....	22
3.3	Research Plan and Expected Outcomes.....	27
3.4	Task Allocation Table.....	27
4.1	Model Performance Metrics Table.....	32
5.1	Cost and Financial Analysis.....	38
5.2	Mapping with Complex Engineering Problem.....	39
5.3	Mapping with knowledge Profile.....	40
5.4	Mapping with Complex Engineering Activities.....	41

Chapter 1

Introduction

This chapter provides an overview of the research work. It begins with the background and problem statement of air pollution in Dhaka City, followed by the motivation behind choosing this research topic. The chapter also describes the objectives, methodology, and expected outcomes of the project. Finally, the organization of the full report is presented.

1.1 Introduction

Air pollution is one of the biggest problems in Dhaka City. Rapid urban growth, heavy traffic, industrial factories, and construction work have made the air dirty. Pollutants like PM_{2.5}, PM₁₀, carbon monoxide (CO), and smoke are common in the air, especially near busy roads and industrial areas. For example, during rush hours in Motijheel or Ashulia, vehicle emissions increase PM_{2.5} levels sharply. Similarly, in winter, smoke from brick kilns in Keraniganj raises air pollution in surrounding areas.

Air pollution has serious effects on human health. People can suffer from asthma, bronchitis, lung problems, and even heart diseases. Children and elderly people are at higher risk. Poor air quality also affects the environment, damaging plants, reducing visibility, and increasing climate-related problems.

Currently, air monitoring systems in Dhaka are limited. Most stations are fixed in certain locations and cannot provide continuous real-time updates. Because of this, people do not always know the current air quality, and authorities may not respond quickly when pollution levels are high.

Modern technology can play an important role in solving Dhaka's air pollution problem. With the help of IoT-based air sensors and machine learning models, it is possible to collect pollution data in real time and make accurate predictions about future air quality. Machine learning can study the data, find patterns, and give forecasts that help us understand when pollution will rise or fall. Deep learning models like LSTM can also capture time-based changes, which makes the prediction more reliable.

This research uses different machine learning models to analyze air quality data collected by IoT sensors. The system can predict pollution levels and classify air quality, helping citizens know when the air is unsafe. Machine learning models like Random Forest, Decision Tree, and Support Vector Machine (SVM) will study the data, find patterns, and forecast air pollution trends. By using these models, the project aims to make air quality monitoring more accurate and useful for daily life and policy planning.

Therefore, this research focuses on building a system that combines IoT devices for real-time data collection and different machine learning models for accurate forecasting. This will not only allow people to know about the current air quality but also help the government and policymakers plan better actions to protect public health and the environment.

1.2 Motivation

The main reason behind choosing this topic is the urgent need to reduce the harmful effects of air pollution in Dhaka. Every day, millions of people in this city breathe in polluted air without knowing how dangerous it is. Many times, we hear about Dhaka being ranked as one of the worst cities in terms of air quality. This reality motivated me to find a solution that could help people live healthier lives.

On a personal level, I feel connected to this issue because I live in Dhaka and face the same challenges as other residents. Breathing fresh air is a basic human right, but in many parts of this city, it feels like a luxury. If a real-time air quality monitoring and forecasting system can be built, it will not only increase awareness but also save lives by providing early warnings during high-pollution periods.

From an academic point of view, I was motivated by the chance to apply machine learning and IoT technologies to a real-world problem. Air pollution is a complex issue, and it requires advanced tools to analyze and predict patterns. By working on this project, I can improve my skills in data analysis, predictive modeling, and system design. These are important skills for my future career, but at the same time, the project will contribute to society in a meaningful way.

Finally, the motivation also comes from the hope that this research can guide urban planners, health organizations, and policymakers. If they have access to accurate air quality predictions, they can make better decisions, such as controlling traffic, planning industrial zones, or issuing health advisories. In short, my motivation is both social (helping people live in a cleaner environment) and technical (gaining knowledge and experience by solving a real-world challenge).

1.3 Objectives

The main Objectives of my study are:

- i. My first goal is to build a sensor-based framework that continuously tracks and monitoring air pollutants in Dhaka.

- ii. Another aim is to apply machine learning methods to predicts future air pollution levels with acceptable accuracy.

1.4 Methodology

First, I define the air pollution problem in Dhaka city. Then, I use IoT devices to collect information about air, weather, and traffic. After that, I clean the data and get it ready for the ML model. I have divided the data into two parts. First is training data and another is testing data. To analyze air pollution, I have experimented with some machine learning techniques, like Logistic Regression, Naive Bayes, Support Vector Machine, KNN, and a deep learning model called Long Short-Term Memory (LSTM) to predict air quality. After all, this model tells us if the air is Good, Moderate, or Poor. This helps us see the air quality at any time.

1.5 Project Outcome

The main outcomes of this project are:

- A functional IoT-based system for live monitoring of air pollution across Dhaka City.
- Accurate predictions of air pollution levels using various machine learning techniques, like Logistic Regression, Naive Bayes, Support Vector Machine, KNN, and a deep learning model called Long Short-Term Memory.
- Live air quality classification, which can alert citizens when pollution reaches unsafe levels.
- Comparison of machine learning models, showing which approaches perform best for predicting air pollution.
- Practical recommendations for city planners, health organizations, and environmental experts to take timely actions against pollution.
- A scalable monitoring framework that can be adapted and applied to other cities facing similar air quality challenges.

1.6 Organization of the Report

The structure of the report progresses from context and background to methodology, implementation, results, and concludes with challenges, standards, and future directions. Each discusses a key element of the final year design project and collectively presents a comprehensive study on live air quality observation in Dhaka with IoT and machine learning techniques.

- **Chapter 1: Introduction**

This chapter provides an overview of the study, highlighting the background, problem statement, motivation, specific objectives, methodology, expected outcomes, and the overall organization of the report. It sets the foundation for understanding the scope and significance of the research.

- **Chapter 2: Background**

This chapter gives an in-depth background on air pollution in Dhaka, a summary of existing literature, related applications, relevant studies, and a gap identification. It establishes the need for a multiple, IoT-enabled predictive model by identifying what existing studies lack and how this project addresses those gaps.

- **Chapter 3: Research Methodology**

This chapter discusses the technical approach of the project. It includes the requirement analysis, system design, proposed methodology, functional and non-functional requirements, data flow diagrams, UI designs, and the project planning strategy. It also explains the rationale behind choosing specific models and techniques.

- **Chapter 4: Implementation and Results**

This chapter presents the environment setup, system implementation, and the results obtained through testing and evaluation. It includes a comparative performance analysis of various machine learning and deep learning techniques, along with a discussion of findings and observations.

- **Chapter 5: Engineering Standards and Design Challenges**

This section explores the standards followed during the development of the system, including hardware, software, and communication protocols. It also discusses the project's societal and environmental impacts, sustainability considerations, ethical aspects, financial analysis, and the project's alignment with complex engineering problem-solving frameworks.

- **Chapter 6: Conclusion**

The final chapter summarizes the entire project, highlighting key achievements, limitations encountered during the development process, and possible directions for future work to enhance the system's functionality and scalability.

Chapter 2

Background

This chapter provides the necessary background and previous research studies related to air pollution monitoring and prediction in Dhaka City. It explains the importance of monitoring air pollution, reviews previous studies and similar applications, identifies gaps in existing research, and highlights the contribution of the proposed system.

2.1 Introduction

Air pollution has become one of the most critical environmental issues in Dhaka City. Rapid urbanization, heavy traffic, industrial activities, and construction work release harmful pollutants into the air, including PM_{2.5}, PM₁₀, carbon monoxide (CO), Ammonia (NH₃). These pollutants can cause serious health problems, such as asthma, bronchitis, cardiovascular diseases, and even neurological disorders. Children, elderly people, and individuals with pre-existing health conditions are particularly vulnerable.

Understanding air pollution requires collecting accurate data from multiple sources. Traditional monitoring systems in Dhaka are limited to a few fixed stations, which cannot provide real-time information or cover the whole city. Modern approaches, such as IoT-based sensors and machine learning models, allow researchers to collect data continuously, analyze patterns, and predict future pollution levels. Using these technologies, it is possible to classify air quality, issue early warnings, and help city authorities and citizens take preventive actions.

This chapter also includes a review of previous research on air pollution in Dhaka and other cities. The literature shows how various sensors, data analysis methods, and machine learning algorithms have been applied to monitor and forecast air quality. By understanding past research, this study identifies gaps and motivates the development of a system that can provide more accurate and real-time predictions for Dhaka City.

2.2 Literature Review

Air pollution in Dhaka City has become a critical public health issue due to rapid urbanization, traffic congestion, and industrial activities. Over the past decade, numerous studies have focused on monitoring, analyzing, and predicting air quality using various technologies and machine learning approaches. The following literature highlights key contributions in this area.

Alam et al. (2018) conducted a comprehensive study on air pollutants in different locations of Dhaka City. They measured levels of PM_{2.5} , PM₁₀ , CO, NO_x, SO₂, H₂S, and other pollutants using real-time gas analyzers and dust samplers. Their findings revealed that most pollutants exceeded safe limits and were strongly correlated with health issues such as respiratory diseases, cardiovascular problems, and skin conditions. This study emphasized the importance of continuous monitoring and statistical analysis to understand pollution patterns [1].

Chowdhury et al. (2020) explored the use of data mining and machine learning models for forecasting air quality in Dhaka. They applied regression and deep learning models, including LSTM networks, to predict the Air Quality Index (AQI) and classify air quality levels. Their results demonstrated that machine learning models, particularly LSTM, could provide accurate forecasts and help identify periods of high pollution, assisting urban planners in taking timely measures [2] .

Dasgupta et al. (2020) examined global technologies such as satellite imagery and traffic data to monitor local air pollution in Dhaka. By comparing satellite-based PM_{2.5} and NO₂ measurements with ground station data, they found that these global datasets can effectively complement local monitoring networks. Their study suggested that integrating satellite, traffic, and sensor data could improve spatial coverage and provide more actionable insights for air quality management [3] .

Hossain et al. (2020) proposed a hybrid deep learning model combining GRU and LSTM layers to predict AQI in Dhaka and Chattogram. The model captured temporal changes in air quality and outperformed individual GRU or LSTM models in terms of prediction accuracy. This research highlighted the potential of advanced neural networks for forecasting urban air pollution [4] .

Islam et al. (2024) developed an IoT-based air pollution monitoring framework for Dhaka City, collecting real-time data on SO₂, O₃, NO₂, PM_{2.5} , PM₁₀ , and CO. They applied several machine learning algorithms, including Decision Tree, Random Forest, and Gradient Boosting, to classify air quality levels. Random Forest achieved the highest accuracy in predicting pollution categories. This study demonstrated the effectiveness of combining IoT sensing with machine learning for precise and scalable air quality monitoring [5] .

Lisun-UI-Islam et al. (2023) focused on hourly AQI prediction using deep learning time-series models such as LSTM, CNN, RNN, and GRU. They highlighted the importance of feature selection, showing that models trained on relevant pollutants and meteorological variables achieved better prediction accuracy. LSTM performed particularly well,

confirming its suitability for forecasting urban air quality [6] .

Hussain et al. (2021) applied the Decision Tree algorithm to predict categorical air quality levels in Dhaka and other major cities of Bangladesh. Using historical AQI and climate data, the model achieved over 96% accuracy, demonstrating that simple supervised learning methods can effectively forecast air pollution and support timely public health interventions [7] .

Rahman et al. (2019) analyzed spatial gradients and temporal trends of air pollution in Dhaka. Their study showed that pollutant concentrations varied significantly across the city and over time, emphasizing hotspots of high pollution and the potential health impacts. This research underlined the need for targeted monitoring and interventions in critical urban areas [8] .

Islam (2025) reviewed the application of AI and machine learning for air pollution management in Dhaka. The study examined various deep learning architectures, including spatiotemporal RNNs and CNNs, for forecasting AQI and assessing pollutant sources. It also highlighted explainable AI methods for interpreting predictions. The work emphasized integrating IoT, satellite data, and advanced ML models for adaptive and data-driven air quality management [9] .

Saha et al. (2021) investigated the effects of COVID-19 lockdowns on air pollution in Dhaka using an IoT-based sensor network. Their study revealed a significant reduction in CO and NO₂ levels during the lockdown, demonstrating how reduced traffic and industrial activity improved air quality. This research showcased the ability of real-time IoT monitoring to capture dynamic changes in urban pollution [10] .

Rahman et al. (2024) developed AirNet, a predictive machine learning system for real-time air quality forecasting. The study used large-scale air quality datasets and applied multiple machine learning classifiers, including Random Forest, Decision Tree, Logistic Regression, SVM, KNN, and Naive Bayes. Their results showed that Random Forest and Decision Tree models provided the highest accuracy. The research also implemented a web-based interface to display real-time predictions, emphasizing the importance of combining data preprocessing, model selection, and visualization for effective air quality monitoring [11].

Alam et al. (2024) proposed an IoT-based environmental monitoring system suitable for developing regions. Using Arduino microcontrollers and multiple sensors to measure parameters such as CO₂, CO, TVOC, humidity, temperature, and particulate matter, the system transmitted data to a Firebase cloud platform. The study demonstrated that low-cost IoT devices could reliably collect and display real-time environmental data, providing an affordable solution for local authorities and raising public awareness about pollution

levels [12].

Ferdous and Ali (2005) analyzed traffic-related air pollution in Dhaka City using Gaussian plume and Caline4 dispersion models. They collected traffic count data and pollutant measurements near road intersections to validate the models. Their findings indicated that Gaussian plume models and Caline4 could predict CO and other traffic-related pollutants with reasonable accuracy, highlighting the potential of modeling approaches for urban air pollution assessment and traffic planning [13].

Khandker et al. (2023) conducted a comprehensive literature review on air pollution in Bangladesh, examining both ambient and household sources. They identified brick kilns, traffic emissions, open burning, and industrial activities as major contributors to PM_{2.5} and other pollutants. The study emphasized the associated health risks, including respiratory and cardiovascular diseases, and highlighted the need for continuous monitoring and policy interventions to reduce pollution levels [14].

Yi et al. (2015) surveyed wireless sensor network (WSN)-based air pollution monitoring systems, reviewing various architectures, sensor types, communication protocols, and data aggregation techniques. They found that while WSNs are promising for distributed monitoring, challenges remain in sensor calibration, power management, and data reliability. The study suggested integrating low-cost sensors with centralized analytics and periodic calibration for accurate and scalable air quality monitoring [15].

Fattah et al. (2022) investigated multi-sensor IoT systems for air pollution monitoring, using semiconductor gas sensors combined with neural networks and fuzzy logic for data fusion. The methodology involved calibration, preprocessing, and short field campaigns to validate performance. Their findings revealed that multi-sensor fusion improves classification accuracy of pollutants, reduces measurement drift, and enables deployable portable monitoring systems suitable for urban environments [16].

Okokpujie et al. (2018) developed a smart air pollution monitoring system using Arduino-based sensors and cloud integration through ThingSpeak. The system measured CO, NO_x, and particulate matter and displayed real-time data on LCD and online dashboards. Results indicated that low-cost IoT devices could reliably monitor air quality and provide public visibility for early warnings and awareness, demonstrating the feasibility of citizen-focused pollution monitoring systems [17].

Mandal et al. (2024) proposed a medium-term AQI prediction model using Bidirectional GRU neural networks. They collected daily air quality data from multiple locations in Bangladesh and compared BiGRU with GRU, LSTM, and CNN models. Their study found that BiGRU outperformed other architectures in prediction accuracy, showing the

importance of capturing bidirectional temporal patterns in air quality forecasting for effective planning and public health measures [18].

Osman et al. (2022) reviewed real-time air quality monitoring systems that combine IoT devices and predictive analytics. The review examined multiple ML techniques, including regression, ANN, and SVM, and highlighted challenges in sensor calibration, data transmission, and real-time prediction. The study concluded that integrating IoT with machine learning models can significantly improve prediction accuracy and enable timely decisions for environmental management [19].

Islam et al. (2024) conducted a narrative review of air pollution and its health impacts in Bangladesh. They summarized evidence on high PM_{2.5} exposure, respiratory and cardiovascular diseases, and the effectiveness of existing policy interventions. Their findings stressed the urgent need for widespread monitoring networks and improved regulatory measures to reduce exposure and protect public health [20].

Begum et al. (2013) investigated the presence of fine particulate matter (PM_{2.5}) and its sources in Bangladesh, with a particular focus on urban environments. Using receptor modeling and chemical characterization methods, the study identified biomass burning, brick kilns, and vehicular emissions as the major contributors to air pollution in Dhaka. Their findings highlight the alarming rise of particulate pollution in densely populated cities and its strong association with health risks. The methodology combined field sampling with elemental analysis, providing a scientific basis for pollution source apportionment. This research emphasized the urgent need for effective policy interventions to reduce particulate matter exposure in urban areas [21].

Idrees et al. (2020) conducted a comprehensive review of low-cost air pollution monitoring systems, with particular emphasis on the enabling technologies and communication protocols that make such systems feasible. The study explored the application of Internet of Things (IoT) devices, wireless sensor networks, and cloud-based platforms in real-time air quality monitoring. Their review demonstrated that low-cost sensors can provide scalable, community-level monitoring, although challenges remain regarding calibration, accuracy, and data reliability. The paper concluded that integrating these technologies with advanced data processing methods, including machine learning, can significantly improve air pollution monitoring and prediction systems [22].

Dhingra et al. (2019) introduced the IoT-Mobair system, a mobile air pollution monitoring platform that integrates portable sensor nodes with GPS, cloud computing, and a mobile application. Their methodology involved designing compact sensor units mounted on vehicles, enabling real-time data collection of particulate and gaseous pollutants. The

collected data was transmitted to the cloud and visualized through an Android app and web dashboard. The study highlighted that mobile IoT sensing significantly improved spatial coverage, detected local pollution hotspots missed by stationary monitors, and could support safe-route recommendations for urban users. This research demonstrated that combining IoT sensors with cloud analytics offers a scalable, cost-effective complement to fixed monitoring stations [23].

Sakib et al. (2024) developed an IoT-based toxic gas monitoring system integrated with machine learning models for Air Quality Index (AQI) prediction in Dhaka. Their prototype used Arduino Uno and NodeMCU with MQ-series sensors (MQ-4, MQ-135, MQ-136) to detect CH₄, CO/SO₂, and H₂S in real time, transmitting data to ThingSpeak for cloud storage. On the modeling side, they compared Logistic Regression, Discriminant Analysis, Decision Tree, K-Nearest Neighbors, and SVM using preprocessed AQI datasets. The findings revealed that the Decision Tree algorithm achieved superior performance with RI ≈ 0.9999 , outperforming other classifiers. The system effectively distinguished between clean and polluted conditions and showed how low-cost IoT sensors combined with machine learning can deliver accurate AQI predictions for real-time urban monitoring [24].

Islam et al. (2022) developed a low-cost IoT system to collect air component data for machine learning experiments. Using sensors like MQ-series, DHT11, and particulate sensors connected to Arduino/NodeMCU, the system logged air quality data to the cloud. The project demonstrated that low-cost IoT sensors can produce useful datasets for AQI prediction and that simple machine learning models, such as LSTM, can learn temporal pollution patterns effectively [25].

Ahmed et al. (2024) proposed a hybrid deep-learning model (CLSTM-BiGRU) integrating CNN, LSTM, and bidirectional GRU layers with satellite-derived hydro-climatological variables to forecast AQI. Using optimization techniques like Grey Wolf Optimization, their model significantly outperformed standard benchmark models, showing that combining remote-sensing data with hybrid deep learning can provide accurate AQI predictions in regions with sparse ground sensors [26].

Overall, the reviewed literature indicates that combining IoT-based sensing, machine learning, and deep learning approaches can provide accurate and timely predictions of air pollution in Dhaka. Decision Tree, Random Forest, and LSTM models are particularly effective, while satellite and traffic data can complement local measurements. These studies highlight the importance of real-time monitoring, predictive modeling, and data-driven strategies to mitigate the adverse effects of air pollution on public health and the environment.

Table 2.1: Summary of Literature Reviewed.

Author(s)	Year	Study Focus	Methodology	Key Findings
Alam et al.	2018	Health impacts of air pollutants in Dhaka	Field surveys with pollutant measurement and statistical evaluation	Reported high levels of pollution in Dhaka; strongly linked with respiratory and heart-related illnesses.
Chowdhury et al.	2020	Data mining for air pollution analysis	Machine learning regression models and LSTM for AQI forecasting	LSTM provided better prediction accuracy; using temperature data further enhanced results.
Dasgupta et al.	2020	Local monitoring of air pollution	Integration of satellite imagery and traffic data with ground-level sensors	Global datasets strengthened local monitoring and widened spatial coverage.
Hossain et al.	2020	Air Quality Index prediction	Hybrid GRU–LSTM model applied for AQI prediction	The hybrid approach surpassed individual GRU/LSTM models, yielding more precise AQI estimation.
Islam et al.	2024	IoT & ML for air pollution monitoring	IoT sensor readings analyzed with RF, DT, and Gradient Boosting	Random Forest performed best in classification, confirming IoT+ML as a reliable solution.
Lisun-Ul-Islam et al.	2023	Hourly AQI prediction in Dhaka	Time-series deep learning (LSTM, CNN, RNN, GRU)	LSTM gave the highest accuracy, and applying feature selection improved forecasting quality.
Hussain et al.	2021	Decision tree for AQI prediction	Decision Tree using historical AQI and climate datasets	Model achieved more than 96% accuracy in categorizing air quality.
Rahman et al.	2019	Spatial & temporal trends in Dhaka air pollution	Statistical evaluation of spatial and temporal pollution data	Identified pollution-prone zones and associated them with significant health threats.

Islam et al.	2025	Comprehensive air pollution analysis	Review of AI/ML techniques with IoT and satellite data	Combining ML, IoT, and satellite improves AQ predictions and allows explainable AI analysis.
Saha et al.	2021	Monitoring air quality during COVID-19	IoT-based sensor data before and during lockdown	Recorded significant pollution reduction during lockdown; IoT devices tracked changes effectively.
Rahman et al.	2024	Predictive ML for AQI forecasting	ML classifiers (RF, DT, LR, SVM, KNN, NB) with web dashboard	Random Forest and Decision Tree models produced the highest accuracy; real-time visualization was effective.
Alam et al.	2024	IoT-based environmental observation	Arduino with CO ₂ , CO, TVOC, humidity, temperature, PM sensors; Firebase cloud storage	Low-cost IoT system successfully collected real-time data, raising awareness in developing regions.
Ferdous & Ali	2005	Traffic-related pollution modeling	Gaussian plume and Caline4 dispersion modeling with traffic inputs	Models provided reasonable predictions for CO and traffic pollution, aiding urban planning.
Khandker et al.	2023	Air pollution sources & impacts	Literature survey of ambient and household pollution	Found brick kilns, transport, and industries as primary sources; stressed policy action for health safety.
Yi et al.	2015	Wireless sensor network monitoring	Review of WSN structures, sensor types, protocols, and aggregation methods	WSNs show promise but require better calibration and energy efficiency; analytics integration enhances accuracy.
Fattah et al.	2022	Multi-sensor IoT systems	Neural networks with fuzzy logic for multi-sensor fusion; field validation	Fusion approach improved accuracy, minimized drift, and supported portable monitoring devices.

Okokpujie et al.	2018	Smart air pollution monitoring	Arduino-based sensors (CO, NOx, PM) integrated with ThingSpeak cloud	Provided cost-effective real-time monitoring and early warnings through online dashboards.
Mandal et al.	2024	Medium-term AQI prediction	BiGRU vs GRU, LSTM, CNN models on daily AQ data	BiGRU captured bidirectional dependencies better, delivering higher accuracy than others.
Osman et al.	2022	Live air quality monitoring	Review of IoT systems integrated with ML (ANN, SVM, regression)	IoT + ML improved prediction accuracy and supported timely environmental management.
Islam et al.	2024	Health impacts of air pollution in Bangladesh	Narrative review of studies and policies	High PM _{2.5} levels connected to lung and heart diseases; recommended broader monitoring and regulation.
Begum et al.	2013	PM _{2.5} source investigation in urban areas	Receptor modeling, chemical analysis, and field sampling	Found biomass burning, kilns, and vehicles as key contributors; emphasized public health and policy needs.
Idrees et al.	2020	Low-cost air pollution monitoring	Literature on IoT, WSN, and cloud platforms	Low-cost sensors support community monitoring, but reliability issues require calibration; ML enhances performance.
Dhingra et al.	2019	Mobile air pollution monitoring	Portable sensing nodes with GPS, cloud backend, and mobile interface	Improved spatial mapping and detection of hotspots; scalable and economical for city use.

Sakib et al.	2024	Toxic gas monitoring with ML for AQI	Arduino/NodeMCU with MQ sensors, ThingSpeak, ML classifiers	Decision Tree gave near-perfect performance (RI \approx 0.9999); proved low-cost sensors with ML can predict AQI reliably.
Islam et al.	2022	Low-cost IoT data collection	MQ-series sensors, DHT11, particulate sensors with Arduino/NodeMCU logging	Generated useful AQI datasets; LSTM effectively modeled temporal trends.
Ahmed et al.	2024	Hybrid deep learning for AQI forecasting	CNN, LSTM, BiGRU combined with satellite climate variables, optimized with Grey Wolf	Hybrid deep learning with remote sensing enhanced AQI forecasts in sensor-limited regions, surpassing benchmarks.

2.2.1 Similar Applications

Several applications and systems have been developed to monitor air quality and predict pollution levels using sensors and machine learning. Some notable examples include:

1. **AirVisual (IQAir):** A mobile and web application that collects real-time data from thousands of air-quality stations worldwide. It provides AQI predictions, historical trends, and pollution alerts. Methodologically, it integrates sensor data with weather and satellite information to forecast air quality.
2. **Plume Labs – Air Report App:** This app provides hourly air-quality forecasts and health recommendations. It uses predictive modeling to classify air quality levels and issue alerts for specific pollutants like PM_{2.5} and NO₂.
3. **BreezoMeter:** A web and mobile platform offering hyper-local air-quality data. It uses real-time sensor readings and machine learning algorithms to estimate pollution exposure in specific locations.
4. **AQICN (Air Quality Index China Network):** This global web-based system tracks AQI data from multiple cities, including Dhaka. It uses real-time sensor data to forecast pollution levels and provide visual trends on maps.

5. **Dhaka Air Quality Apps (Local IoT Initiatives):** Some projects in Dhaka have implemented IoT sensors with simple ML algorithms to monitor PM2.5, PM10, CO, and NO₂ levels. These applications provide real-time air-quality alerts and allow city residents to check pollution levels on mobile or web dashboards.

2.3 Gap Analysis

From the literature and existing applications, the following gaps were observed:

- Most existing systems in Dhaka monitor only a few pollutants (usually PM2.5 and PM10), ignoring gases like CO, NH₃, SO₂, Temperature, and Humidity.
- Many applications do not provide predictive forecasts; they mainly report real-time measurements.
- Some ML-based solutions use only a single model, which may limit prediction accuracy.
- Existing apps often lack classification of air quality severity with early warning notifications.
- Integration of IoT sensors with multiple machine learning models for better accuracy is rarely implemented.

Table 2.2: Feature Comparison with Existing Systems

Features	AirVisual	Plume Labs	BreezoMeter	AQICN	Local Dhaka IoT Apps	Proposed System
Monitor multiple pollutants (PM1.0, PM2.5, PM10, CO, NH ₃)	Yes	Yes	Yes	Yes	Partial	Yes
Live air quality reporting	Yes	Yes	Yes	Yes	Yes	Yes
Predictive forecasting	Yes	Yes	Yes	Partial	No	Yes
Severity classification & alerts	Yes	Partial	Yes	Partial	No	Yes
Multiple ML models for prediction	No	No	No	No	No	Yes
Mobile & web accessibility	Yes	Yes	Yes	Yes	Partial	Yes

2.4 Summary

This chapter provided the necessary background and reviewed previous studies on air pollution monitoring in Dhaka and other regions. It discussed the methodologies and findings of different research papers, identified similar applications, and highlighted the gaps in current solutions. The review shows that while some real-time monitoring systems exist, they often cover limited pollutants and rely on single models, which may reduce prediction accuracy. There is a strong need for an integrated system that combines multiple machine learning models, IoT-based real-time sensors, and predictive analytics to provide accurate forecasts and early warnings for all major pollutants. Addressing these gaps will not only improve public awareness but also support policymakers in taking timely actions to reduce pollution. This sets a clear foundation for the development of the proposed system in this research.

Chapter 3

Research Methodology

This chapter explains the overall research process and the steps taken to design and develop the proposed air pollution monitoring and prediction system. It describes the adopted methodology, requirements, data flow, system design, and planning of tasks. The goal is to ensure that the system is built in a structured and efficient way.

3.1 Methodology

3.1.1 Overview

This study implemented the project through an organized series of steps. Since the work is focused on monitoring air quality using IoT devices and machine learning models, the method combines both device setup and data analysis.

First, the project started by finding out the main problem and reading some related works. Then, IoT devices were applied to collect real data like air parameters, and weather. After collecting, the data was cleaned and set in order so that it could be used easily. Finally, the data was split into two parts: one part was used to train the ML model and the other part was used to test it. Finally, the trained model was applied to monitor air quality and classify it into simple categories such as Good, Moderate, and Poor.

The basic steps are:

- i. IoT Device Setup
- ii. Data Collection
- iii. Preprocessing Data
- iv. Split Data
- v. Select and Train Model
- vi. Machine Learning Model
- vii. Air Quality Monitoring (Good, Moderate or Poor)

3.1.2 Proposed Methodology

This project is designed around a clear step-by-step plan to create a smart air quality observation system using IoT devices. The idea is to collect air data through sensors, process it carefully, and then use machine learning to predict the air quality level, whether it's good, moderate, or poor.

First, the problem is understood and relevant background information is gathered. Then, IoT sensors are set up to gather the necessary data. Once the data is collected, it's cleaned and split into parts for training and testing the machine learning techniques. Then machine learning models tell whether it's actually working. Finally, the system is used for live monitoring and prediction of air quality.

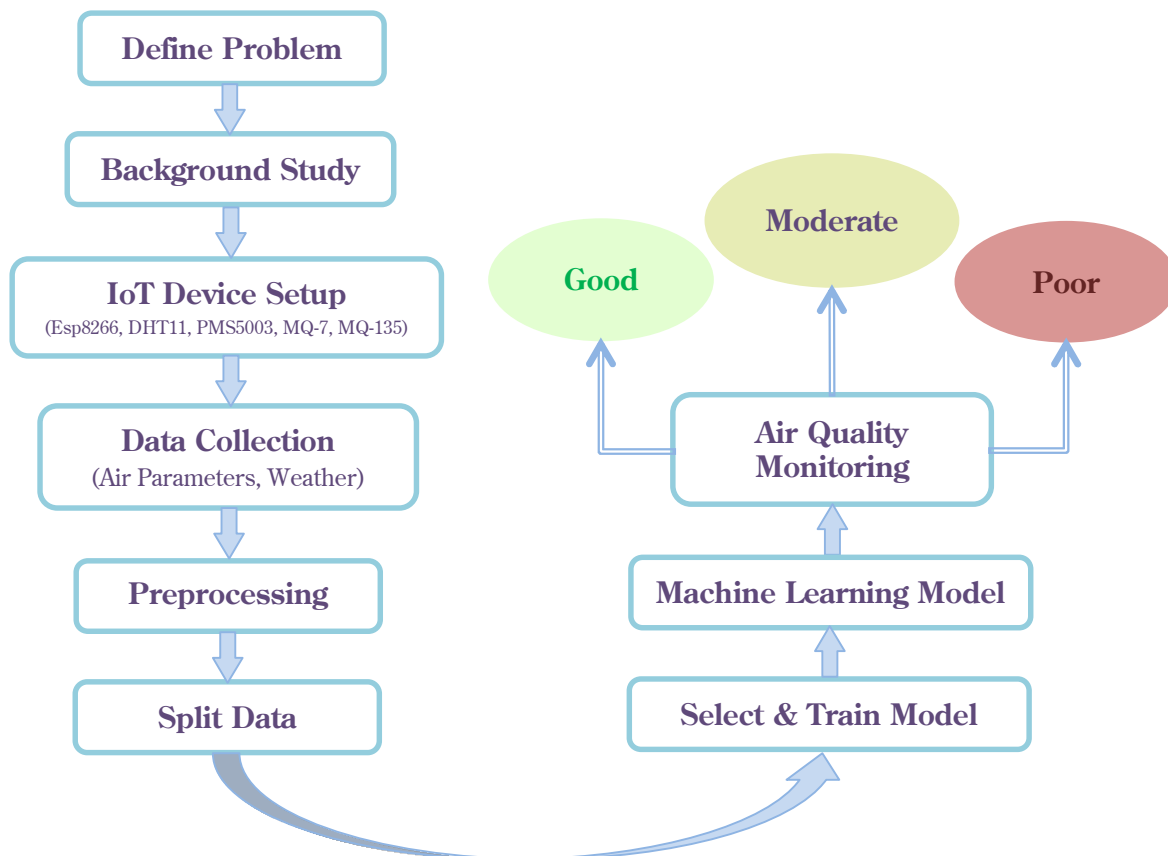


Figure 3.1: Methodology Pipeline Diagram

- i. **Define Problem:** The first step of this study is to identify the real-world issue of air pollution in Dhaka city. This study aims to design a system that can observe environmental data in real time and predict air quality levels to provide useful insights for people and policymakers.
- ii. **Background Study:** A background study was conducted to review previous works on air pollution monitoring systems. Existing methods, IoT device capabilities, and machine learning techniques were analyzed to understand the strengths and limitations. This study helped to shape the design of the proposed system.
- iii. **IoT Device Setup:** IoT-based sensors are used for real-time monitoring. Devices are configured to capture multiple environmental factors such as temperature,

humidity, CO, NH₃, PM1.0, PM2.5, and PM10. These devices continuously collect data and send it to the system for processing.

- iv. Data Collection (Air Parameters, Weather):** Data from IoT sensors such as Temperature, Humidity, CO, NH₃, PM1.0, PM2.5, and PM10 were collected and stored in a structured format. These parameters were chosen because they directly influence the Air Quality Index (AQI) and provide meaningful insights into pollution levels.
- v. Data Preprocessing:** Raw data can contain errors, empty fields, or random fluctuations. Before using it for prediction, I correct these problems and organize the data properly. Doing this improves the accuracy of the models.
- vi. Split Data:** The processed dataset is divided into a training and testing data set. The training data set helps the model learn, and the testing data set shows how well it performs on new data, keeping the results reliable. This guarantees fair testing and avoids overfitting.
- vii. Select and Train Model:** Different models like Logistic Regression, Naive Bayes, Support Vector Machine, KNN, and a deep learning model Long Short-Term Memory (LSTM) are chosen and trained using our dataset. They learn the connection between air quality and various environmental factors.
- viii. Machine Learning Model:** The trained machine learning models explore the dataset and predict air quality values. Their accuracy and reliability are tested using the testing dataset, and the most accurate model is chosen for deployment.
- ix. Air Quality Monitoring (Good, Moderate, Poor):** Finally, this system provides real-time classification of air condition into Good, Moderate, and Poor categories. This helps people to stay informed about air pollution levels and enables policymakers to take preventive actions when the air quality becomes poor.

This method makes sure the system is useful, correct, and can grow if needed. It uses IoT devices to gather data and machine learning models to make predictions. This way, the research builds a dependable way to watch and forecast air quality in cities.

3.1.3 Functional and Nonfunctional Requirements

Functional Requirements

- The system must be capable to capture air quality data from IoT devices in live.

- It should store the collected information, clean it, and prepare it for analysis.
- The solution needs to predict the air condition and categorize the pollution severity.
- Whenever pollution levels rise, the system should provide alerts.
- A user-friendly dashboard should be available to show both current conditions and predictions.

Non-Functional Requirements

- The prediction accuracy should remain high (90% or above).
- The system should respond quickly, ideally within two seconds.
- It must be flexible enough to handle information from multiple areas.
- Continuous 24/7 operation with minimal downtime is expected.

3.1.4 Data Flow Diagram

The data flow diagram (DFD) of the proposed air quality monitoring system shows how data moves through different components to provide accurate air quality predictions.

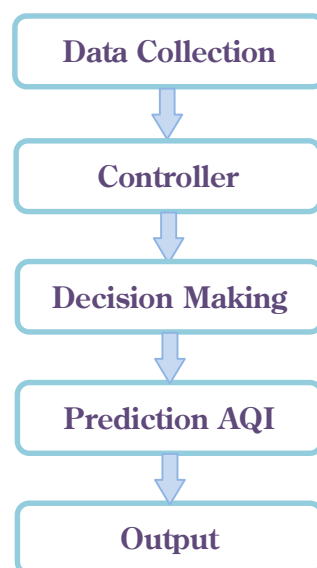


Figure 3.2: Data Flow Diagram

- i. **Data Collection:** Environmental information such as temperature, humidity, CO, NH₃, PM1.0, PM2.5, PM10, and traffic data is collected continuously from IoT sensors.
- ii. **Controller:** The controller receives and manages the incoming sensor data, ensuring it is organized and sent to the appropriate processing modules.
- iii. **Decision Making Unit:** Machine learning Techniques analyze the data to detect

trends and classify pollution levels. This module processes the data using machine learning models to identify types and observed tendencies in air quality. It helps the system decide the current pollution level and predict future conditions.

- iv. **Prediction (AQI):** Using the processed data, the trained machine learning technique predicts the Air Quality Index, based on the collected and processed data.
- v. **Output:** The system shows the final air quality results as categories: Good, Moderate, or Poor. So that citizens and authorities can take appropriate actions.

3.1.5 Dataset Overview

I collected a total of 10,264 records using IoT sensors. The dataset includes values of different air quality parameters such as temperature, humidity, carbon monoxide, ammonia, and dust. These readings were taken continuously to capture real environmental changes.

Before using the dataset, I cleaned it by handling missing values, removing unusual data points, and scaling all attributes to the same range. This step was important for me to make sure the data was balanced and ready for model training. A subset of the dataset is displayed in Table 3.1 to demonstrate the structure and content of the collected data.

Table 3.1: Sample Dataset of Collected Air Quality Data

Temperature (°C)	Humidity (%)	CO (ppm)	NH3 (ppm)	PM1.0 (µg/m3)	PM2.5 (µg/m3)	PM10 (µg/m3)	Air Quality
31.6	78	4.32	0.32	45	47	55	Moderate
30.1	66	4.12	0.34	38	49	56	Moderate
31.3	75	2.48	2.76	75	108	117	Poor
31.9	81	17.48	0.96	71	90	95	Poor
29.5	68	4.32	0.96	23	30	36	Good
29.1	67	3.83	0.96	25	30	33	Good
31.3	75	6.64	2.49	50	59	62	Moderate
30.6	87	18.38	4.01	73	102	112	Poor

This data was collected using IoT sensors to measure air quality. Each entry contains different environmental readings. Temperature shows current air temperature, how warm or cool the air is, and Humidity indicates the moisture level in the air. Carbon Monoxide (CO) measures the level of CO gas, while Ammonia (NH₃) shows how much ammonia is present. PM1.0, PM2.5, and PM10 represent the density of tiny particles in the air of different sizes, which help to understand pollution levels. Based on these readings, the air conditions is labeled as Good,

Moderate, or Poor.

Table 3.2: Summary of Dataset's Attributes

Attribute	Unit	Description
Temperature	°C	Measures the air temperature.
Humidity	%	Measures the moisture level in the air.
CO	ppm	Carbon monoxide concentration.
NH ₃	ppm	Ammonia gas concentration.
PM1.0	(µg/m ³)	Particles with a diameter of 1.0 µm or smaller.
PM2.5	(µg/m ³)	Particles whose size does not exceed 2.5 µm.
PM10	(µg/m ³)	Particulate with a maximum diameter of 10 µm.
Air Quality	Good/Moderate/Poor	Categorizes overall air quality.

3.1.6 UI Design

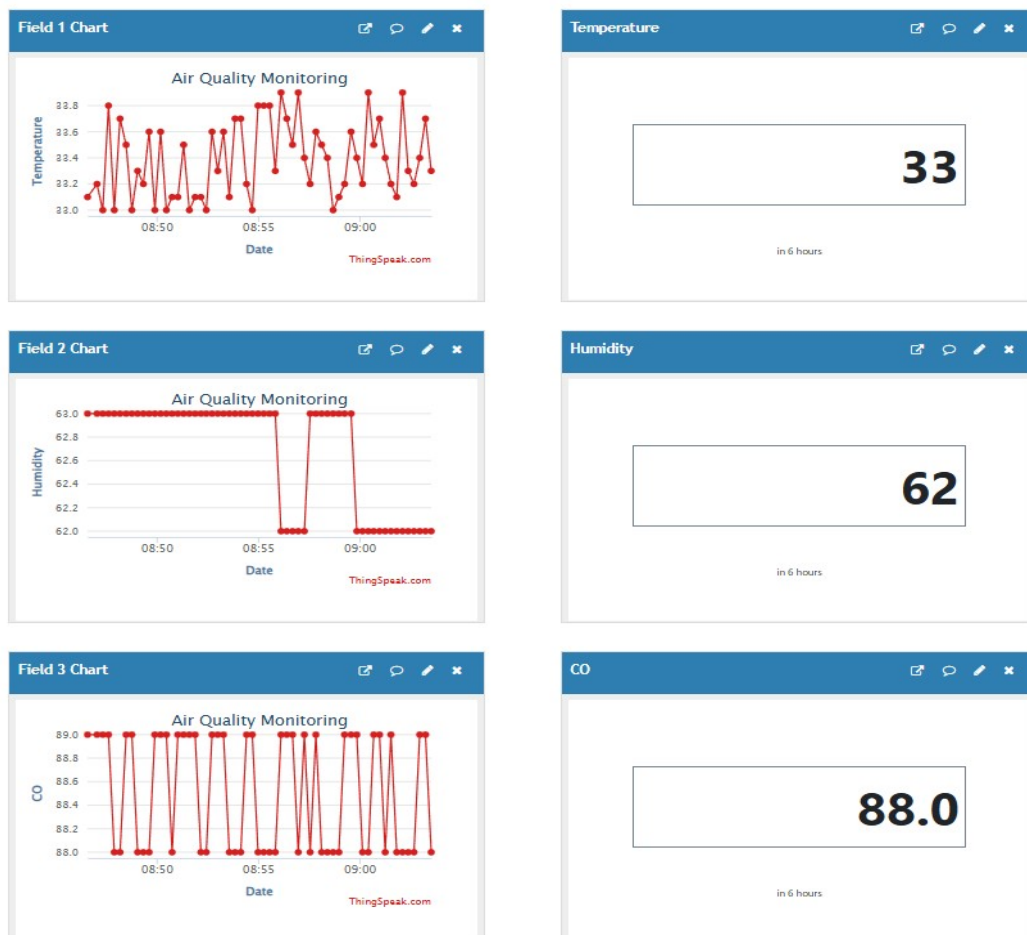


Figure 3.3: Data Collection page of air quality monitoring (Temperature, Humidity, CO).



Figure 3.4: Data Collection page of air quality monitoring (NH3, PM2.5, PM10).

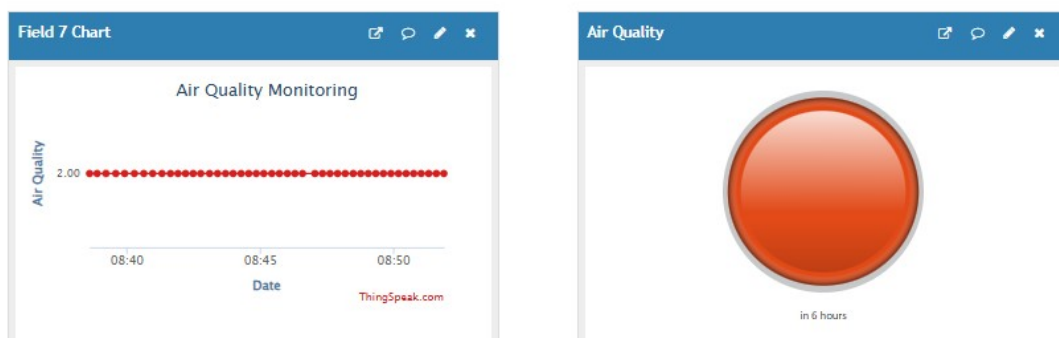


Figure 3.5: Result Dashboard of air quality Prediction.

In this designed user interface enables users to input the required data and obtain prediction results. The outcomes are displayed not only in textual form but also through graphical visualizations, ensuring better understanding and clarity for the users.

3.2 Detailed Methodology and Design

This detailed methodology section explains the detailed methodology and design of the proposed IoT-based air pollution observation setup using machine learning techniques to analyze and predict air quality. Here, alternate solutions that were considered are discussed, the reasons for selecting the proposed solution are explained, and the overall system design is described. Additionally, the role of diagrams such as circuit diagram and context diagram in representing the system is outlined.

3.2.1 Alternate Solutions Considered

Before designing the final system, I considered several alternative approaches:

1. Manual Monitoring:

- Collecting air quality data using conventional stations and manual recording.
- **Problem:** This method is expensive, slow, and cannot provide real-time updates.

2. Single Sensor IoT System:

- Using only one sensor (for example, PM2.5) for prediction.
- **Problem:** Single sensors cannot capture all environmental factors, so the predictions may not be accurate or reliable.

3. Third-Party APIs for Air Data:

- Using existing online APIs to collect air quality data and make predictions.
- **Problem:** Data is generalized and may not be accurate for specific locations in Dhaka. It also depends on internet availability.

4. Proposed Solution (Multi-sensor IoT device + ML models):

- Multiple sensors measure pollutants such as PM1.0, PM2.5, PM10, CO, NH₃, as well as temperature, humidity, and traffic data.
- Data is processed and analyzed with ML techniques like LR, KNN, NB, SVM, and LSTM.
- **Advantage:** Provides accurate, real-time, localized predictions and alerts.

I chose the multi-sensor IoT system with machine learning because it ensures accuracy, scalability, and real-time monitoring. It takes into account multiple environmental factors and predicts air quality effectively, which other alternatives could not achieve.

3.2.2 System Design

This system is structured to efficiently collect, process, and analyze environmental data. The design includes the following parts:

1. IoT Sensor Setup:

- I configured multiple IoT sensors to monitor air pollutants (CO, NH₃, PM1.0, PM2.5, PM10), temperature, and humidity.
- Sensors send data to the controller (ESP8266) in real time.

2. Data Collection:

- I collected a dataset combining air quality, weather, and traffic information.
- Data is stored in a well-organized format for preprocessing and analysis.

3. Data Preprocessing:

- The raw dataset contains missing values, noise, or inconsistencies. I cleaned the data, normalized it, and handled missing or extreme values to make it perfect for machine learning.

4. Data Splitting:

- I divided the dataset into training and testing sets. Training data is used to teach the machine learning models, while testing data evaluates their performance.

5. Model Selection and Training:

- I experimented with multiple machine learning techniques, including Logistic Regression, KNN, Naive Bayes, SVM, and LSTM.
- Models were trained using the preprocessed data to learn the patterns between environmental conditions and air quality levels.

6. Air Quality Prediction:

- The trained models predict air quality and classify it as Good, Moderate, or Poor.
- This prediction allows citizens to be informed and policymakers to take early actions during pollution episodes.

3.2.3 Circuit Diagram

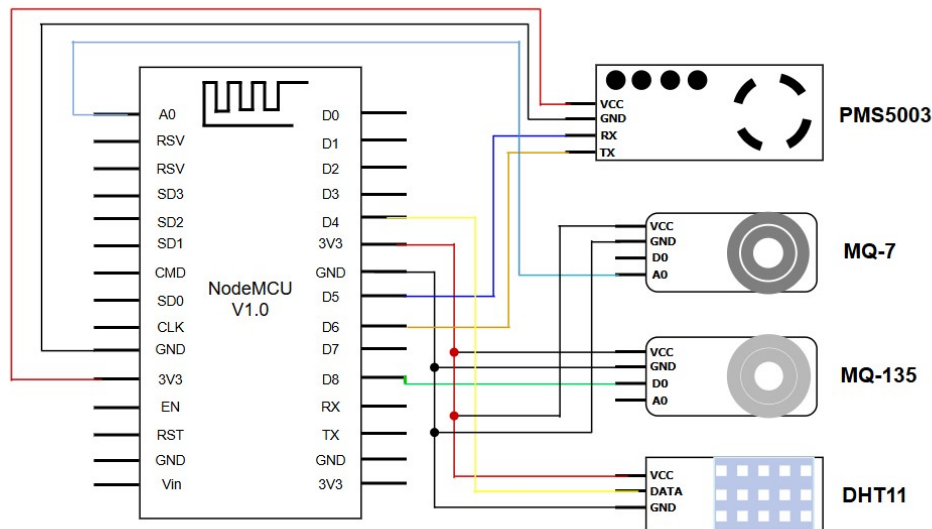


Figure 3.6: Circuit Diagram of This System

In this project, we connect different sensors to a NodeMCU (ESP8266) to check air quality. We use four sensors. The DHT11 sensor measures temperature and humidity. The MQ-135 sensor detects bad gases like ammonia (NH₃). The MQ-7 sensor measures carbon monoxide (CO). The PMS5003 sensor checks small dust particles like PM1.0, PM2.5, and PM10. The sensors get power from the NodeMCU using the 3.3V or 5V pins and GND. They collect data and send information to the NodeMCU using pins like D1, D2 and A0. The PMS5003 sensor sends data through the TX and RX pins. This way, the NodeMCU can read the air quality right away. We can show the data on a screen or send it to a server to check the air from anywhere.

3.3 Project Plan

In this part, I present a clear plan for my project using a table. The table shows each stage of the research process in order, making it easy to see how the work was carried out step by step. Each task is explained, including what I did and why I did it. Also I included the expected outcomes for each task so that it is clear how it contributes to the progress of the project. Using this organized plan helped me keep track of my work, spot any delays, and maintain a steady pace throughout the research.

Table 3.3: Research Plan and Expected Outcomes

No.	Task	Actions Taken	Expected Outcome
1	Requirement Analysis	I identified the objectives, scope, and requirements of my project to clearly understand what needed to be done.	A clear roadmap of my project goals and system requirements.
2	Literature Review	I reviewed previous research papers and studies on IoT-based air monitoring systems and machine learning methods.	Understanding of current solutions, gaps, and best practices.
3	IoT Sensor Integration	I set up and connected multiple air quality sensors (DHT11, PMS5003, MQ-135, MQ-7) to my controller for real-time data collection.	A fully functional IoT system ready to collect environmental data.
4	Data Collection & Preprocessing	I collected environmental data including air quality, weather, and traffic, and cleaned the data to remove inconsistencies.	A clean and structured dataset ready for training machine learning models.
5	Model Selection and Training	I selected machine learning techniques (LR, NB, KNN, SVM & LSTM) and trained them using my prepared dataset.	Trained models capable of predicting air quality accurately.
6	Dashboard Development	I developed a user friendly dashboard to display live air quality, predictions, and alerts.	Interactive system for users to monitor air quality and take action.
7	Final Testing and Evaluation	I tested the entire system using live data to evaluate the efficiency and correctness of my models.	Verified and reliable system ready for real-world use.
8	Report Preparation	I compiled all my research, methodology, results, and conclusions into this report.	Complete project documentation ready for submission.

3.4 Task Allocation

Table 3.4: Task Allocation Table

Tasks	Weeks																		
	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46	48
Data collection phase	█	█	█	█	█														
Preprocess all the data						█	█	█	█	█	█								
Model training											█	█	█	█					
Create a demo application.																█	█	█	█

This table presents the timeline of the main tasks in my project, covering weeks 12 to 48. It shows when each task was carried out and helps to understand how the work was planned to keep the project running smoothly.

- **Data Collection Phase:** During this period, I collected environmental data, including air quality, weather, and traffic information, using IoT sensors.
- **Preprocessing Phase:** I cleaned and prepared the collected data, removed inconsistencies, and formatted it to make it ready for model training and testing.
- **Model Training Phase:** I trained different machine learning techniques, including LR, NB, KNN, SVM and LSTM, using the preprocessed data.
- **Demo Application:** Finally, I developed a demo dashboard to display real-time air quality and predictions, showing how the system works.

3.5 Summary

This chapter presented a detailed explanation of the research methodology and the design of the proposed IoT-based air quality monitoring and prediction system. It described the overall workflow, starting from problem definition, background study, IoT sensor setup, data collection, preprocessing, and splitting of data, to model selection, training, and real-time air quality monitoring. Alternate solutions were discussed, and the reasons for choosing a multi-sensor IoT system with machine learning models were explained. The chapter also included the system design, functional and non-functional requirements, data flow, and the role of diagrams such as circuit and context diagrams. Additionally, a detailed project plan and task allocation table were presented to show the timeline and sequence of activities. Overall, this chapter ensures that the methodology is systematic, practical, and scalable, forming a solid foundation for the development and implementation of the air pollution monitoring system in Dhaka City.

Chapter 4

Implementation and Results

This chapter describes how the air quality monitoring system was built and tested. It covers setting up the environment, checking how well the machine learning models work, comparing the results, and discussing what was found. The purpose is to show how the system predicts air quality and how accurate the models are.

4.1 Environment Setup

The system was built and tested using the necessary hardware and software tools.

➤ **Hardware Environment:**

- **Microcontroller:** ESP8266
- **Sensors:** DHT11, MQ-135, MQ-7, PMS 5003, Flame sensor.
- **Actuators:** Buzzer, LEDs
- **Connectivity:** Wi-Fi module (ESP8266) for IoT-based integration
- **Power Supply:** 5V regulated power adapter

➤ **Software Environment:**

- **Programming Language:** C/C++ (Arduino IDE), Python for run ML models.
- **Libraries:** ESP8266WiFi.h, WiFiClientSecure.h, DHT.h, SoftwareSerial.h
- **Database/Storage:** Google Sheets (via API/IoT integration)
- **Tools:** Arduino IDE for code uploading, Serial Monitor for debugging

I completed my work on Google Colab because it is free and provides access to GPU/TPU, which significantly speeds up the process. Used Python 3.10 for coding. Mostly used scikit-learn, pandas, numpy and matplotlib. These were used for data cleaning, training and making graphs.

Before model training, I fixed missing values, normalized the data and changed text columns to numbers. I used supervised learning methods. Also split the data in a way (stratified) so that all classes are properly represented in both training and test sets.

4.1.1 System Setup Diagram

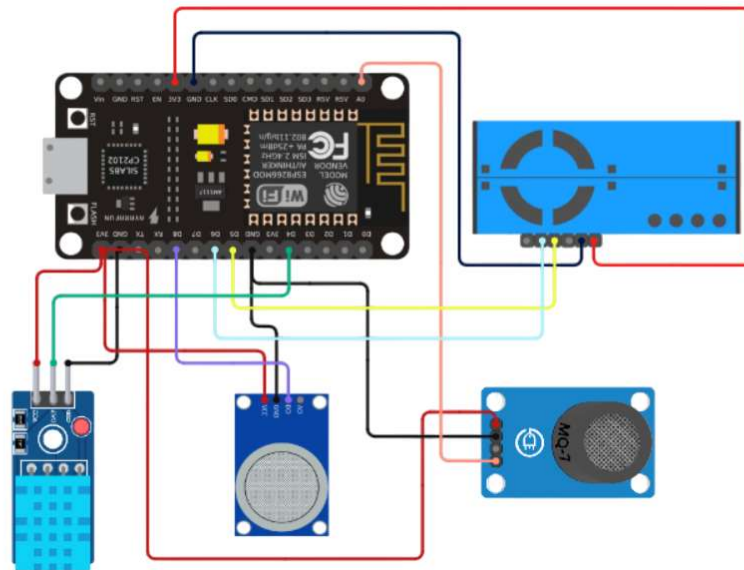


Figure 4.1: Context Diagram of This System.

In this project, multiple sensors are connected to a NodeMCU (ESP8266) to monitor air quality. The DHT11 sensor records temperature and humidity, the MQ-135 identifies harmful gases such as NH_3 , the MQ-7 measures CO levels, and the PMS5003 detects particulate matter (PM1.0, PM2.5, and PM10). These sensors receive power through the NodeMCU and transmit data via specific pins, including digital and analog connections. The PMS5003 communicates using TX/RX pins. Once collected, the data is either displayed directly or transferred to a server, allowing users to observe air quality remotely.

4.2 Testing and Evaluation

In this research, I worked with some machine learning models. So this are Logistic Regression, K-Nearest Neighbors, Naive Bayes, Support Vector Machine, and Long Short-Term Memory. The goal was to estimate air quality in three groups – Good, Moderate, and Poor. The dataset I collected and used for machine learning techniques there was 10,264 rows and 8 columns, where the features included both environmental and traffic factors such as temperature, humidity, carbon monoxide, NH_3 , PM1.0, PM2.5, and PM10.

Before starting the training, I created a heatmap to check the relationship between different features. This helped me to see which parameters were more useful for predicting the air quality.

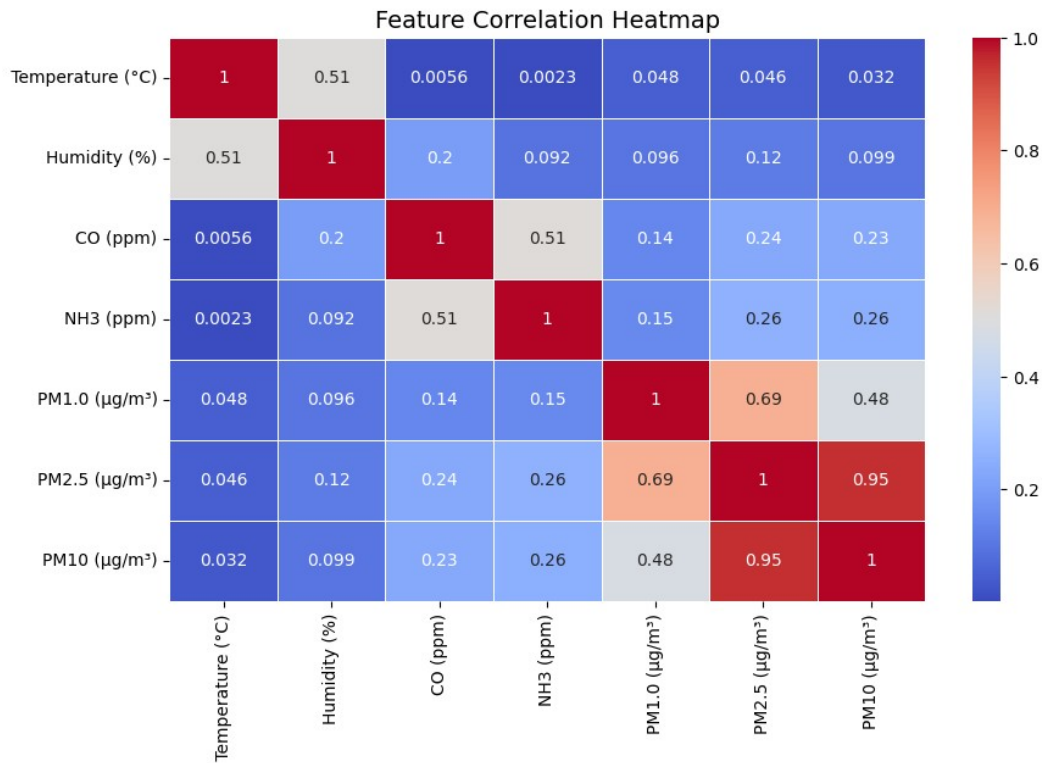


Figure 4.2: Feature Correlation Heatmap

The figure shows a heatmap of the dataset features. It shows a clear figure of how the factors like temperature, humidity, CO, NH₃, PM1.0, PM2.5 and PM10 are related to each other. From the heatmap, I noticed that PM2.5 and PM10 are closely linked, and they also connect with gases such as CO and NH₃. This means these factors have a strong role in deciding air quality. By checking these relations, I could understand which features are more useful for predicting the air quality levels (Good, Moderate, Poor). It also gave me an idea of how to use machine learning models more effectively.

➤ **Steps I Followed:**

1. **Data Preparation:** I cleaned the dataset, then fixed missing values, after then normalized the data, and removed outliers to ensure the dataset was suitable for machine learning.
2. **Train-Test Split:** I divided the dataset into two parts once is training set and another is testing sets to train the models and evaluate their performance fairly.
3. **Model Training:** I trained each model using the training dataset.
4. **Prediction and Metrics:** I used the testing dataset to predict air quality levels and calculated Accuracy, Precision, Recall, F1-Score, and Support for each model.

Table 4.1: Model Performance Metrics Table

Model	Accuracy	Precision (Avg)	Recall (Avg)	F1-Score (Avg)
Logistic Regression	98%	0.87	0.90	0.89
Naive Bayes	96 %	0.86	0.93	0.89
SVM	97%	0.85	0.85	0.85
KNN	98%	0.91	0.88	0.98
LSTM	98%	0.95	0.97	0.96

➤ **Confusion Matrices:**

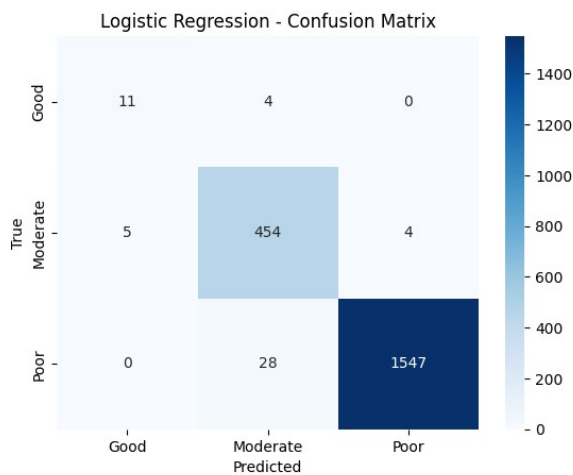


Figure 4.3 : Confusion Matrix of LR

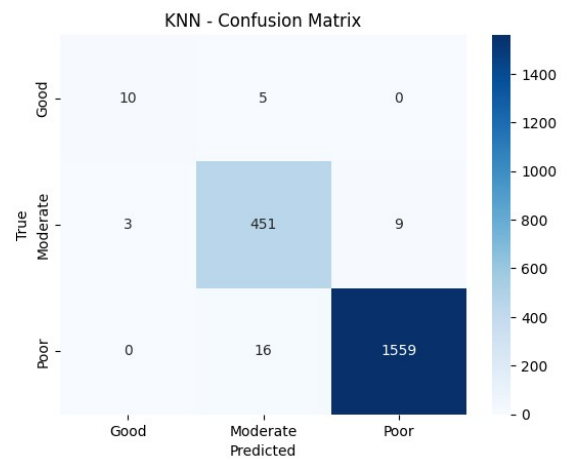


Figure 4.4: Confusion Matrix of KNN

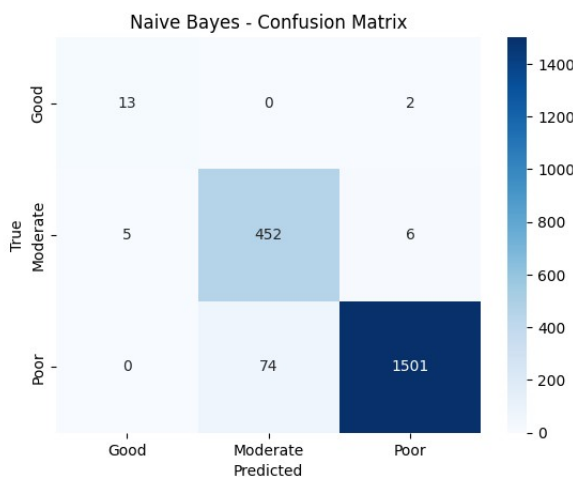


Figure 4.5: Confusion Matrix of NB

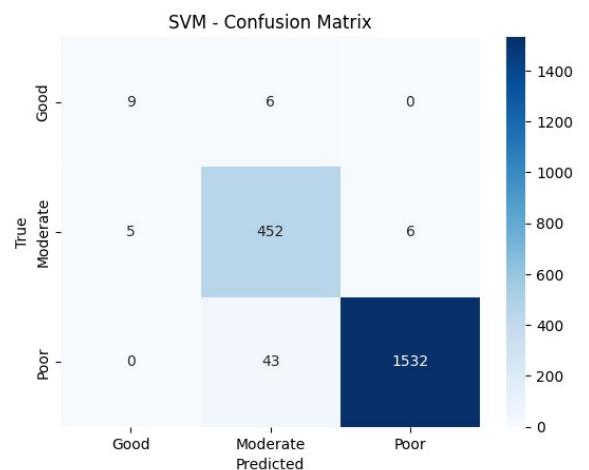


Figure 4.6: Confusion Matrix of SVM

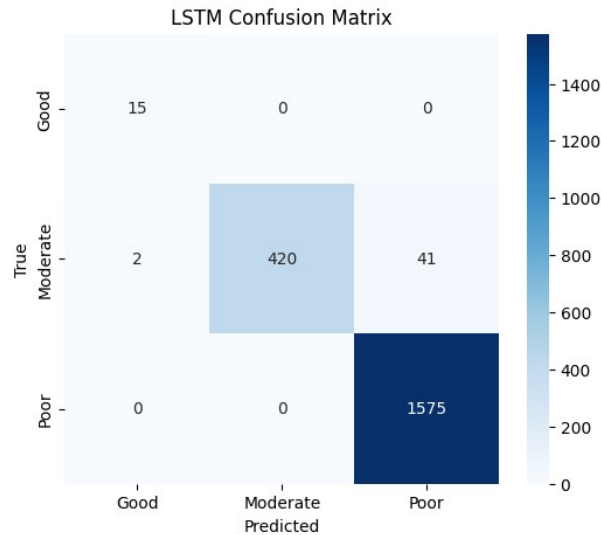


Figure 4.7: Confusion Matrix of LSTM

I created confusion matrix charts for all five models to check and visualize how accurately each model classified the air quality categories. The matrices clearly show that LSTM achieved near-perfect classification, especially for the Moderate and Poor categories, capturing almost all instances correctly. Logistic Regression (LR) and SVM performed well overall but showed minor misclassifications in the Good category. Naive Bayes (NB) worked strongly for Moderate and Poor, but it was a little weaker for Good. While KNN demonstrated the lowest accuracy, with more misclassifications across all categories. These confusion matrices provide a clear visual understanding of each model's strong and weak point in predicting air quality.

➤ **Comparative Analysis:**

- Among the models, the deep learning technique (LSTM) provided the highest accuracy (~98%), with balanced precision and recall, making it the highly suitable for this dataset. Its capability to detect sequential patterns in sensor data makes it more suitable for live air quality prediction.
- Logistic Regression also performed very well but had minor misclassification in the Good category.
- Naive Bayes showed slightly lower accuracy (96%) but still produced reliable predictions for Moderate and Poor levels.
- SVM performed reasonably but was less accurate for the Good category.
- KNN had the lowest performance due to sensitivity to noise and high-dimensional features.

Overall, the multi-sensor IoT setup combined with LSTM provides the most accurate and reliable air quality predictions.

4.3 Results and Discussion

For my research, I used a dataset of 10,264 records. The data included temperature, humidity, CO, NH₃, PM1.0, PM2.5, and PM10. I divided the air quality into three parts: Good, Moderate, and Poor. At first, I checked the data with a heatmap. From this, I saw that PM2.5 and PM10 had the most effect on air quality. Then I cleaned the dataset, split it, and trained five machine learning techniques. These were LR, KNN, Naive Bayes, SVM, and LSTM. I trained and tested them and then compared their performance using accuracy, precision, recall, F1-score, and confusion matrices. Among these models, LSTM gave the best results with 98% accuracy and showed strong performance in all three categories. Logistic Regression also gave 98% accuracy but was slightly weaker in predicting the “Good” category. Naive Bayes and SVM worked at a moderate level, while KNN was less effective because it is sensitive to noise and large datasets. The results show that using IoT sensor data with the LSTM model can predict air quality very accurately. This kind of system can be useful for both citizens and policymakers to take quick steps when pollution levels increase. I also created an actual vs predicted chart, which showed that the model’s predictions are very close to the real measurements, proving the method works well.

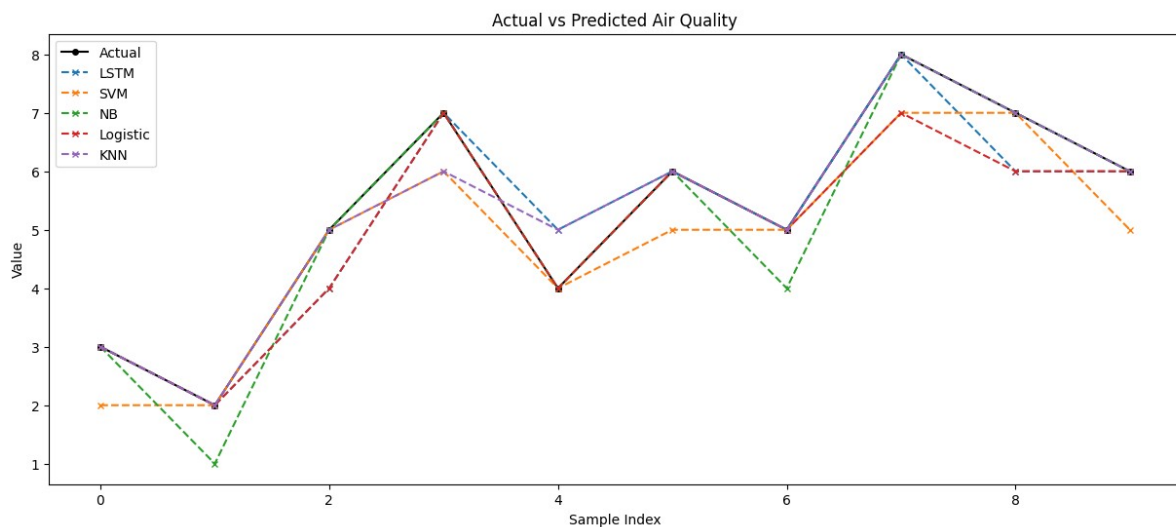


Figure 4.8 : Actual vs Prediction Air Quality

This chart visualizes the comparison between actual air quality measurements and the predicted categories from the all-performing models. It highlights how accurately the system can classify air quality as Good, Moderate, or Poor in real time.

4.4 Summary

In this chapter, the implementation of the IoT-based air quality monitoring system and its evaluation using machine learning models were presented. The system setup involved multiple sensors connected to a microcontroller (ESP8266) to collect environmental data such as temperature, humidity, CO, NH₃, PM1.0, PM2.5, and PM10. The collected dataset of 10,264 records was preprocessed, normalized, and split into training and testing sets for model development. Five models: Logistic Regression, KNN, Naive Bayes, SVM, and LSTM were trained and evaluated using metrics including accuracy, precision, recall, F1-score, and confusion matrices. Among these, LSTM emerged as the most accurate and reliable model for predicting air quality levels, followed closely by Logistic Regression, while KNN showed the least performance. Heatmaps and confusion matrices helped visualize feature importance and classification performance, and the actual vs predicted air quality chart confirmed the model's effectiveness. Overall, this chapter demonstrates that the combination of a multi-sensor IoT system with LSTM provides precise and real-time air quality monitoring, supporting informed decision-making and timely action against pollution.

Chapter 5

Engineering Standards and Design Challenges

In this chapter, I explained the engineering standards I followed while designing and implementing my IoT-based air quality observation system. I also describe the challenges I faced during the design, the effects on society and the environment, ethical considerations, my plan for sustainability, project management and budget analysis, and the complex engineering problems I solved throughout the project.

5.1 Compliance with the Standards

Compliance with engineering standards is crucial to ensure system reliability, safety, and interoperability. Both software and hardware components, as well as communication protocols, were selected and evaluated based on established standards. Alternatives were considered for each category, and the rationale for the final selection is described below.

5.1.1 Software Standards

In this research, I followed the software quality guideline ISO/IEC 25010, which defines a comprehensive model for software product quality evaluation (ISO/IEC, 2011). It talks about things like how fast the software works, if it keeps running without problems, and if it is easy to fix or use. I did the coding in Arduino IDE and Python 3.10. These worked well with my sensors and also for machine learning stuff. I also thought about using MATLAB and R. MATLAB is good for analyzing data but it costs a lot and does not work well with live sensor data. R is good for statistics but it cannot handle real-time data from devices. That's why I went with Python and Arduino IDE. They are free, easy to use, and have lots of online help. They also work well with sensors and the cloud, which made things much easier for my project.

5.1.2 Hardware Standards

For hardware, this system followed standard rules for microcontrollers and sensors. The ESP8266 microcontroller supports Wi-Fi using the IEEE 802.11 standard, which helps send data reliably. Sensors like DHT11, MQ-135, MQ-7, and PMS5003 work safely with 3.3V, following the manufacturers' instructions. Other options like ESP32 were also considered. Raspberry Pi is more powerful but uses more energy and is bigger, while ESP32 has dual-core processing and Bluetooth but is more complex to use. I chose ESP8266 because it uses little power, has enough processing speed, is small, and is affordable.

5.1.3 Communication Standards

For sending data, Wi-Fi (802.11b/g/n) was used to connect the IoT devices to the cloud and dashboards. Other options like LoRa and Zigbee were considered. LoRa is good for long distances and low power use but sends data slowly. Zigbee works well for short distances but needs extra setup. Wi-Fi was chosen because it sends data fast, is widely available, and easily connects with cloud services and mobile apps.

5.2 Impact on Society, Environment and Sustainability

The proposed air quality monitoring system has a beneficial effect on on peoples, society, and the environment, while being ethical responsible and sustainable.

5.2.1 Impact on Life

The system provides live monitoring and prediction of air condition index. This enables individuals to make better decisions about outdoor activities and their safety. By alerting people about pollution levels (Good, Moderate or Poor), it reduces health risks caused by exposure to harmful pollutants such as CO, NH₃, and particulate matter. This contributes to improved public health and better quality of life.

5.2.2 Impact on Society & Environment

This system helps people take early steps to reduce air pollution by focusing on issues like traffic control, industrial smoke, and proper city planning. The collected data can guide policymakers in making rules for the environment and in managing the city better. Regular monitoring also makes the public more conscious about pollution and its harmful effects, which finally encourages everyone to work for a cleaner and healthier environment.

5.2.3 Ethical Aspects

I have ensured that ethical practices are followed while developing the system. The data used is limited only to environmental factors, and no personal details are gathered or shared. The purpose of the system is to give clear, fair, and correct information to everyone. It has been developed with respect for privacy and responsibility in the use of data.

5.2.4 Sustainability Plan

This project has been planned in such a way that it can run for a long time without creating extra cost. The devices used in the system need very little electricity, so they save energy. If any part gets damaged or becomes old, it can be changed easily without rebuilding the whole system. Because the system is connected to cloud service, it can be checked and controlled from anywhere, which reduces the need for manual checking. The machine learning part can

also be trained again with new data, so the system will still work properly even if the environment changes in the future.

5.3 Project Management and Financial Analysis

This part shows the estimated cost needed to set up the proposed IoT-based air quality observation system. It includes the cost of hardware, software, and other materials, along with an alternative budget plan and rationale for choices.

Table 5.1: Cost and Financial Analysis Table

Category	Item/Description	Qty.	Unit Price (BDT)	Total Cost (BDT)
Hardware	ESP8266 Wi-Fi Module	2	490	980
	DHT 11 Sensor (Temp., Humidity)	1	180	180
	MQ-135 Gas Sensor (NH3 detection)	1	200	200
	MQ-7 Gas Sensor (CO detection)	1	200	200
	PMS5003 Sensor (PM2.5, PM10)	1	3300	3300
	Flame Sensor	1	100	100
	Power Supply	1	850	850
	Breadboard	1	120	120
	Jumper wires	50	3	150
	Led	3	10	30
	Cable	1	140	140
Software	Arduino IDE	-	0	0
	ThingSpeak	-	0	0
	Google Sheet	-	0	0
	Google Colab Pro (Python, ML Tools)	1 month	850	850
Installation	Sensor Installation & Calibration	1 time	150	150
Maintenance	Sensor & Hardware Maintenance	-	0	0
Miscellaneous	Internet/Data Charges	4 months	350	1400
Field Survey	Visiting data collection sources	15 visit	100	1500
Documentation	Documentation and report printing	5	350	1750
Total Cost				11,900

5.4 Complex Engineering Problem

The Air Quality Monitoring System project involved tackling multiple complex engineering problems that required advanced knowledge of sensors, IoT integration, and machine learning. The project demanded careful data handling, model selection, and system-level design to provide accurate air quality predictions. Below we detail how the project maps into complex problem-solving and engineering activity frameworks.

5.4.1 Complex Problem Solving

The project satisfies the key characteristics of complex engineering problems as defined in engineering standards.

Table 5.2: Mapping with Complex Engineering Problem.

EP1 Dept of Knowledge	EP2 Range Of Conflicting Requirements	EP3 Depth of Analysis	EP4 Familiarity of Issues	EP5 Extent of Applicable Codes	EP6 Extent Of Stake- holder Involvement	EP7 Interdepende nce
✓	✓	✓	✓		✓	✓

Justifications:

- EP1 (Depth of Knowledge): Applied advanced knowledge of IoT systems, multi-sensor integration, and machine learning algorithms (Logistic Regression, KNN, Naive Bayes, SVM, LSTM) to design, train, and evaluate predictive models.
- EP2 (Range of Conflicting Requirements): Balanced requirements such as real-time prediction, hardware limitations, sensor accuracy, and power constraints.
- EP3 (Depth of Analysis): Conducted systematic experimentation with multiple models, hyperparameters, and dataset partitions to identify the most accurate model for air condition classification.
- EP4 (Familiarity of Issues): Addressed everyday challenges in sensor noise, missing data, and environmental variability affecting air quality prediction.
- EP6 (Extent of Stakeholder Involvement): Coordinated design decisions to ensure usability for both citizens and policymakers.

- EP7 (Interdependence): Integrated sensor data collection, preprocessing, model prediction, and visualization into a single coherent system.

Mapping with Knowledge Profile for EP1

This table designed to map the overall problem and EP1 to the Knowledge Profile.

Table 5.3: Mapping with knowledge Profile.

K1	K2	K3	K4	K5	K6	K7	K8
Natural Science	Mathematics	Engineering Fundamentals	Specialist Knowledge	Engineering Design	Engineering Practice	Comprehension	Research Literature
	✓	✓	✓	✓	✓		✓

Justification for Each Knowledge Profile Element:

- K2 (Mathematics): Applied statistical and numerical analysis for predictive modeling, including probability theory for Naive Bayes, distance metrics for KNN, optimization in Logistic Regression and SVM, and matrix operations for LSTM. Mathematical concepts were essential for model evaluation (accuracy, precision, recall, F1-score) and understanding data patterns.
- K3 (Engineering Fundamentals): Applied knowledge of sensor operation, signal processing, and fundamental machine learning principles.
- K4 (Specialist Knowledge): Expertise in IoT communication, environmental monitoring, and predictive modeling informed the selection of suitable sensors and algorithms.
- K5 (Engineering Design): Designed an end-to-end system architecture, integrating data acquisition, preprocessing, model training, prediction, and visualization following sound engineering principles.
- K6 (Engineering Practice): Implemented practical hardware-software integration using ESP8266, sensors, and cloud-based storage while ensuring reliability and scalability.
- K8 (Research Literature): Engaged with recent studies and methods in air quality prediction and IoT-enabled environmental monitoring to guide model selection and system design.

5.4.2 Engineering Activities

Table 5.4: Mapping with Complex Engineering Activities.

EA1 Range of re- sources	EA2 Level of Interaction	EA3 Innovation	EA4 Consequences for society and environment	EA5 Familiarity
✓	✓	✓	✓	✓

Justifications:

- EA1 (Range of Resources): Combined multiple sensors, microcontrollers, power supplies, cloud storage, and machine learning frameworks for model development.
- EA2 (Level of Interaction): Resolved challenges arising from interactions between noisy sensor data, multi-model performance, and real-time prediction requirements.
- EA3 (Innovation): Applied creative integration of LSTM and traditional ML models to improve prediction accuracy and reliability.
- EA4 (Consequences for Society and Environment): Enabled citizens and policymakers to make timely decisions during pollution events, contributing to public health and environmental safety.
- EA5 (Familiarity): Extended beyond prior experience by combining IoT data acquisition, machine learning model deployment, and real-time visualization in a single automated system.

5.5 Summary

In this chapter 5, I discussed the key engineering standards, design difficulties, and problem-solving steps involved in building the IoT-based air quality monitoring system. I carefully chose software, hardware and communication tools to make sure the system works reliable, efficient and can easily connect with other devices. Different options were considered and the final choices were made based on cost, performance, and easy of use. This system is designed to benefit society and the environment, so that people can understand the levels of pollution, make healthy and good decisions with take sustainable practices. Ethical issues like keeping data private and being transparent about its use, were also taken into account. To make the

system last longer, energy-saving and affordable components were used, which can also be upgraded in the future. A financial analysis outlined the budget, including alternative choices and the reasoning behind cost-effective decisions. Finally, the complex engineering challenges were linked to relevant knowledge areas and engineering activities, showing how multidisciplinary knowledge, creative solutions, and practical engineering approaches were combined to build a reliable, accurate, and socially useful air quality monitoring system.

Chapter 6

Conclusion

This chapter provides a comprehensive overview of my IoT-based air quality monitoring project. It summarizes the work I carried out, highlights the limitations encountered during implementation, and suggests future improvements to make the system more robust and effective. The chapter aims to give a complete understanding of the research process, the outcomes achieved, and the potential for further development.

6.1 Summary

In this research, I tried to make a system that could check and predict air quality in real time. I used IoT devices and a few simple sensors. These sensors measured temperature, humidity, carbon monoxide (CO), ammonia (NH₃), and tiny particles like PM1.0, PM2.5, and PM10. So my goal was to find out if the air quality was Good, Moderate or Poor. I collected 10,264 data for the research. Before using the data, I had to clean it. Then I noticed some values were missing, so I filled them in. Some numbers were very different from the rest, so I removed them. I also made sure all numbers were on the same scale. Cleaning the data took time, but I found it was really important. Without cleaning, the models didn't work as well.

I trained and tested five machine learning techniques: Logistic Regression, K-Nearest Neighbors, Naive Bayes, Support Vector Machine, and LSTM. I wanted to see which model would predict the air quality most accurately. I looked at accuracy, precision, recall, and F1-score. I also used heatmaps to see how the sensor readings were related. Confusion matrices and charts helped me compare the models. From my tests, LSTM worked the best. It gave around 98% accuracy. It did well for all air quality categories. Logistic Regression was okay, but it sometimes made mistakes for the Good category. KNN, Naive Bayes, and SVM were average and less reliable. I noticed LSTM handled patterns in the data better, and it seemed more stable for real-time predictions.

The system can send data in real time using Wi-Fi. It also saves the data in the cloud. This means anyone can check the air quality anytime. People can decide whether it's safe to go outside. Authorities can also use the data to track pollution and take action. While working on this, I noticed some sensors reacted faster than others. For example, PM2.5 readings changed quicker than PM10. That was interesting because it showed which sensors are more sensitive. I also realized that real-time monitoring is not just about sensors. It needs proper data handling and a good model.

Finally, I found that combining IoT devices with machine learning creates a simple but effective system. It is easy to use, gives quick results and also helps people and authorities stay aware of pollution. I think this kind of system can be very useful in everyday life, especially in cities with high air pollution.

6.2 Limitation

Even though my system works well, I faced some problems while building and testing it. These problems mostly came from sensors, internet, and hardware.

Limitations I faced:

- **Sensor readings sometimes change:** I noticed that the sensors don't always give the same readings, especially when the weather changes.
- **Data from only a few places:** I collected data from several places, so the system may not work well in other areas unless retrained from all places in the entire city.
- **Need for stable internet:** My system depends on Wi-Fi to update data in live. If the connection is slow or goes off, updates are delayed.
- **Measures only a few pollutants:** I could only include some types of pollution, so the system doesn't show the full air quality.
- **Hardware limitations:** The ESP8266 and sensors have limited power and range, so I couldn't make the system handle bigger tasks or more complex calculations.

6.3 Future Work

In the future, I plan to improve this system in several ways to make air quality observation more accurate and useful. Some of the things I want to work on are:

- **Add More Sensors:** I will include sensors for other pollutants like NO₂, SO₂, and O₃ so that the system can observe air quality more completely.
- **Collect More Data:** My aim to gather data from different areas to make the machine learning model's predictions more reliable.
- **Better Communication:** I want to use technologies like LoRa or NB-IoT to increase the coverage area and reduce dependence on Wi-Fi.

- **Mobile Apps and Alerts:** I plan to develop a mobile app that can alert people immediately when air quality drops.
- **Improve Prediction Models:** I will try hybrid or advanced machine learning models to make predictions more accurate.
- **Long-Term Analysis:** I want to analyze long-term data to identify trends and help policymakers make better environmental decisions.

References

- [1] Z. M. Alam, E. Armin, M. M. Haque, J. Halsey, and M. A. Qayum, "Air pollutants and their possible health effects at different locations in Dhaka city," *J. Curr. Chem. Pharm. Sci.*, vol. 8, no. 1, p. 111, Apr. 2018, doi: 10.21767/2277-2871.1000111.
- [2] A. Chowdhury, M. Shihab Uddin, M. Rashad Tanjim, F. Noor, and R. M. Rahman, "Application of Data Mining Techniques on Air Pollution of Dhaka City," in *Proc. 2020 IEEE 10th Int. Conf. on Intelligent Systems (IS)*, Varna, Bulgaria, Aug. 2020, doi: 10.1109/IS48319.2020.9200125.
- [3] S. Dasgupta, M. Khaliquzzaman, and D. Wheeler, *Global Technology for Local Monitoring of Air Pollution in Dhaka*, World Bank Policy Research Working Paper 9429, Oct. 2020.
- [4] E. Hossain, M. A. U. Shariff, M. S. Hossain, and K. Andersson, "A Novel Deep Learning Approach to Predict Air Quality Index," in *Advances in Intelligent Systems and Computing*, Springer, Dec. 2020, pp. 367–381, doi: 10.1007/978-981-33-4673-4_29.
- [5] M. M. Islam, S. U. P. Shakil, N. M. Nayan, M. A. Kashem, and J. Uddin, "Air Pollution Monitoring Using IoT and Machine Learning in the Perspective of Bangladesh," *Ann. Emerg. Technol. Comput. (AETiC)*, vol. 8, no. 3, pp. 17–33, Jul. 2024, doi: 10.33166/AETiC.2024.03.002.
- [6] M. Lisun-Ul-Islam, M. R. H. Rahat, S. Esha, A. Faiyaz, and A. I. Jony, "Hourly Air Quality Prediction in Dhaka City using Time Series Forecasting Techniques: Deep Learning Perspectives," *Tuijin Jishu/Journal of Propulsion Technology*, vol. 44, no. 5, pp. 568–579, Dec. 2023, doi: 10.52783/tjjpt.v44.i5.2518.
- [7] M. Hussain, S. Afrin, A. Irin, and S. K. Park, "Applying Decision Tree Algorithm for Air Quality Prediction in Bangladesh," in *Proc. 2021 5th Int. Conf. on Electrical Information & Communication Technology (EICT)*, Khulna, Bangladesh, Dec. 2021, doi: 10.1109/EICT54103.2021.9733443.
- [8] M. M. Rahman, S. Mahamud, and G. D. Thurston, "Recent spatial gradients and time trends in Dhaka, Bangladesh, air pollution and their human health implications," *J. Air Waste Manag. Assoc.*, vol. 69, no. 4, pp. 478–501, 2019, doi: 10.1080/10962247.2018.1548388.
- [9] F. A. Samiul Islam, "A Comprehensive Analysis of Air Pollution in Dhaka City, Bangladesh, and the Application of Artificial Intelligence and Machine Learning for Enhanced Management

and Forecasting,” *Int. J. Appl. Nat. Sci.* (IJANS), Bluemark Publishers, 2025 (rev. version on ResearchGate), doi: 10.61424/ijans.v3.i1.303.

[10] R. Saha, S. N. M. Azizul Hoque, M. M. R. Manu, and A. Hoque, “Monitoring air quality of Dhaka using IoT: Effects of COVID-19,” in *Proc. 2nd Int. Conf. on Robotics, Electrical and Signal Processing Techniques (ICREST)*, Dhaka, Jan. 2021, pp. 715–721, doi: 10.1109/ICREST51555.2021.9331026.

[11] Rahman, M. M., Nayeem, M. E. H., Ahmed, M. S., Tanha, K. A., Sakib, M. S. A., Uddin, K. M. M., & Babu, H. M. H., “AirNet: predictive machine learning model for air quality forecasting using web interface,” *Environmental Systems Research*, vol. 13, Article no. 44, 2024. doi: 10.1186/s40068-024-00378-z.

[12] Alam, M., Islam, M. M., Nayan, N. M., & Uddin, J., “An IoT Based Real-Time Environmental Monitoring System for Developing Areas,” *Journal of Advanced Research in Applied Sciences and Engineering Technology*, vol. 52, no. 1, pp. 106-121, Oct. 2024, doi: 10.37934/araset.52.1.106121.

[13] Ferdous, M. R., & Ali, M. A., “Air Quality Modeling for Predicting Traffic Pollution in Dhaka City,” *UAP Journal of Civil and Environmental Engineering*, vol. 1, no. 1, pp. 27- 33, 2005.

[14] Khandker, S., Mohiuddin, A. S. M., Ahmad, S. A., McGushin, A., & Abelsohn, A., “Air Pollution in Bangladesh and Its Consequences,” *Research Square*, posted May 16, 2023. doi: 10.21203/rs.3.rs- 1184779/v2.

[15] Yi, W. Y., Lo, K. M., Mak, T., Leung, K. S., Leung, Y., & Meng, M. L., “A Survey of Wireless Sensor Network Based Air Pollution Monitoring Systems,” *Sensors*, vol. 15, no. 12, pp. 31392- 31427, Dec. 2015. doi: 10.3390/s151229859.

[16] G. Fattah, J. Mabrouki, F. Ghrissi, M. Azrou, and Y. Abrouki, "Multi-sensor system and Internet of Things (IoT) technologies for air pollution monitoring," in *Futuristic Research Trends and Applications of Internet of Things*, CRC Press, 2022, pp. 16. DOI: 10.1201/9781003244714-5.

[17] K. Okokpujie, E. Noma-Osaghae, O. Modupe, S. John, and O. Oluwatosin, "A Smart Air Pollution Monitoring System," *Int. J. Civ. Eng. Technol.*, vol. 9, no. 9, pp. 799–809, Sep. 2018.

[18] A.K. Mandal and R. Sen, "Medium-Term AQI Prediction in Selected Areas of Bangladesh Based on Bidirectional GRU Network Model," *SN Computer Science*, vol. 5, no. 627, pp. 1–10, Jun. 2024. doi: 10.1007/s42979-024-02971-6.

- [19] N. Osman, M. F. Jamlos, F. Dzaharudin, A. R. Khan, Y. K. Yeow, and K. A. Khairi, "Real-Time and Predictive Analytics of Air Quality with IoT System: A Review," in *Recent Trends in Mechatronics Towards Industry 4.0*, Springer, 2021, pp. 107–116, doi: 10.1007/978-981-33-4597-3_11.
- [20] S. Islam, I. Tarik, M. K. Islam, M. H. Kabir, S. M. Sajib, and A. Bhadra, "Air Pollution and Health Hazards: A Narrative Review from the Bangladesh Perspective," *Bangabandhu Sheikh Mujib Medical College Journal*, vol. 3, no. 2, pp. 101–106, Sep. 2024. doi: 10.3329/bsmmcj.v3i2.75908.
- [21] B. A. Begum, P. K. Hopke, and A. Markwitz, "Air Pollution by Fine Particulate Matter in Bangladesh," *Atmospheric Pollution Research*, vol. 4, no. 1, pp. 75-86, Jan. 2013, doi: 10.5094/APR.2013.008.
- [22] Z. Idrees and L. Zheng, "Low-cost air pollution monitoring systems: A review of protocols and enabling technologies," *J. Ind. Inf. Integr.*, vol. 17, pp. 100123, Mar. 2020, doi: 10.1016/j.jii.2019.100123.
- [23] S. Dhingra, R. B. Madda, A. H. Gandomi, R. Patan, and M. Daneshmand, "Internet of Things Mobile–Air Pollution Monitoring System (IoT-Mobair)," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 5577–5584, Jun. 2019, doi: 10.1109/JIOT.2019.2903821.
- [24] M. Sakib, A. S. Mahmud, F. T. Labiba, M. U. Khalil, and M. A. Hossain, "Real-Time IoT-Based Toxic Gas Monitoring and Comparative Analysis of Machine Learning Techniques for Air Quality Index Prediction in Dhaka," in *Proc. 2024 Int. Conf. on Innovations in Science, Engineering and Technology (ICISSET)*, Chattogram, Bangladesh, Feb. 2024, pp. 1–6, doi: 10.1109/ICISSET62123.2024.10939534.
- [25] T. Islam, F. Rabbi, R. Ahmed, M. M. Rahman, and M. Ahmed, "IoT Based Air Components Collection for Machine Learning Reinforcement," B.Sc. thesis, Dept. of Computer Science & Engineering, BRAC University, Dhaka, Bangladesh, 2022.
- [26] A. A. M. Ahmed, S. J. J. Jui, E. Sharma, M. H. Ahmed, N. Raj, and A. Bose, "An advanced deep learning predictive model for air quality index forecasting with remote satellite-derived hydro-climatological variables," *Sci. Total Environ.*, vol. 906, p. 167234, Jan. 2024, doi: 10.1016/j.scitotenv.2023.167234.

ORIGINALITY REPORT

17%	15%	12%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	3%
2	downloads.hindawi.com Internet Source	1%
3	Sivakumaran AR, Cuddapah Anitha, Manjula Arunraj, Ebinezer M D J, Venkatesh Babu S, Gokila S. "Mitigating Air Pollution Risks with Deep Learning: A Quantum-Optimized Approach for Nitrogen Dioxide Prediction in Los Angeles", Journal of Machine and Computing, 2025 Publication	1%
4	Anik Alvi, Minhajuddin Ahmed, S. N. M. Azizul Hoque. "Consequences of Lockdown Caused by COVID-19 Outbreak on the Quality of Air in Dhaka", 2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI), 2021 Publication	<1%
5	Nahin Mostofa Niloy, Md. Morshedul Haque, Shafi M Tareq. "Characterization of dissolved organic matter at urban and industrial rainwater of Bangladesh by fluorescence spectroscopy and EEM-PARAFAC modeling", Environmental Challenges, 2021 Publication	<1%
6	mjsat.com.my Internet Source	<1%