**TRIBHUVAN UNIVERSITY**

**INSTITUTE OF ENGINEERING**

ADVANCED COLLEGE OF ENGINEERING AND MANAGEMENT

DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING

KALANKI, KATHMANDU



**A Minor Project Proposal Defense Report On**

**“Evaluating LSTM and Transformer Models for Cross-Regional Air Quality Forecasting”**

[CT 654]

**Submitted By:**

Aditya Pandit ACE078BCT006

Ayush Man Shrestha ACE078BCT018

Bibek Acharya ACE078BCT019

Bibek Bhakta Shrestha ACE078BCT020

A Minor Project Proposal report submitted to the Department of Electronics and Computer Engineering in the partial fulfillment of the requirements for degree of Bachelor of Engineering in Computer Engineering

Kathmandu, Nepal

25 December, 2024

## 

# ACKNOWLEDGEMENT

We take this opportunity to express our deepest and sincere gratitude to our Academic Project Coordinator **Er. Laxmi Prasad Bhatt**,Department of Electronics and Computer Engineering for his insightful advice, motivating suggestions, invaluable guidance, help and support in this project selection and also for his constant encouragement and advice throughout our journey till the date.

We express our deep gratitude to **Er. Prem Chandra Roy**, Head of Department of Electronics and Computer Engineering, **Er. Dhiraj Pyakurel**,Deputy Head, Department of Electronics and Computer Engineering for their regular support, co-operation, and coordination.

We also express our deep and sincere gratitude to **Er. Ukesh Thapa**, Department of Electronics and Computer Engineering for invaluable guidance, help and support in this project.

The in-time facilities provided by the department are also equally acknowledgeable.

We would like to convey our thanks to the teaching and non-teaching staff of the Department of Electronics & Communication and Computer Engineering, ACEM for their invaluable help and support hitherto. We are also grateful to all our classmates for their help, encouragement and invaluable suggestions.

Finally, yet more importantly, we would like to express our deep appreciation to our

grandparents, parents, siblings for their perpetual support and encouragement.

Aditya Pandit ACE078BCT006

Ayush Man Shrestha ACE078BCT018

Bibek Acharya ACE078BCT019

Bibek Bhakta Shrestha ACE078BCT020

## 

# Table of Contents

[ACKNOWLEDGEMENT ii](#_Toc186010202)

[Table of Contents iii](#_Toc186010203)

[List of Tables v](#_Toc186010204)

[List of Figures vi](#_Toc186010205)

[List of Abbreviations/Acronyms vii](#_Toc186010206)

[CHAPTER 1 1](#_Toc186010207)

[INTRODUCTION 1](#_Toc186010208)

[1.1 Background 2](#_Toc186010209)

[1.2 Motivation 2](#_Toc186010210)

[1.3 Statement of the Problem 2](#_Toc186010211)

[1.4 Project objective 3](#_Toc186010212)

[1.5 Significance of the study 3](#_Toc186010213)

[CHAPTER 2 4](#_Toc186010214)

[LITERATURE REVIEW 4](#_Toc186010215)

[CHAPTER 3 7](#_Toc186010216)

[REQUIREMENT ANALYSIS 7](#_Toc186010217)

[3.1 Software Requirements 7](#_Toc186010218)

[3.2 Functional Requirements 7](#_Toc186010219)

[3.3 Non-Functional Requirements 8](#_Toc186010220)

[3.4 Feasibility Study 8](#_Toc186010221)

[CHAPTER 4 9](#_Toc186010222)

[SYSTEM DESIGN AND ARCHITECTURE 9](#_Toc186010223)

[4.1 Block Diagram 9](#_Toc186010224)

[4.2 Flowchart 10](#_Toc186010225)

[4.3 Use Case Diagram 11](#_Toc186010226)

[4.4 DFD 12](#_Toc186010227)

[CHAPTER 5 13](#_Toc186010228)

[METHODOLOGY 13](#_Toc186010229)

[5.1 Data Collection 13](#_Toc186010230)

[Details of Collected Data 13](#_Toc186010231)

[5.2 Data Preprocessing 13](#_Toc186010232)

[Steps : 13](#_Toc186010233)

[5.3 Model Development 14](#_Toc186010234)

[Algorithms Used: 14](#_Toc186010235)

[5.4 Model Evaluation 14](#_Toc186010236)

[5.5 Comparative Analysis 14](#_Toc186010237)

[5.6 Deployment 15](#_Toc186010238)

[Deployment Process: 15](#_Toc186010239)

[5.6 System Workflow 15](#_Toc186010240)

[5.7 Tools and Technologies Used 15](#_Toc186010241)

[5.8 Algorithm 16](#_Toc186010242)

[5.8.1 LSTM Model 16](#_Toc186010243)

[5.8.2 Transformer Model 17](#_Toc186010244)

[5.9 Development Model 18](#_Toc186010245)

[5.9.2 Agile Model 18](#_Toc186010246)

[CHAPTER 6 19](#_Toc186010247)

[EXPECTED OUTPUT 19](#_Toc186010248)

[CHAPTER 7 20](#_Toc186010249)

[TIME SCHEDULE 20](#_Toc186010250)

[REFERENCES 21](#_Toc186010251)

# List of Tables

**Title** **Page**

[Table 2.1 : Summary of the Literature Review 6](#_Toc186009667)

# List of Figures

**Title Page**

[Figure 4.1 : Block Diagram of the System 9](#_Toc186009684)

[Figure 4.2 : Flowchart of the System 10](#_Toc186009685)

[Figure 4.3 : Use Case Diagram 11](#_Toc186009686)

[Figure 4.4 : Level 0 DFD 12](#_Toc186009687)

[Figure 4.5 : Level 1 DFD 12](#_Toc186009688)

[Figure 5.1 : System Workflow 15](#_Toc186009689)

[Figure 5.2 : Flowchart of LSTM Model 16](#_Toc186009690)

[Figure 5.3 : Flowchart of Transformer Model 17](#_Toc186009691)

[Figure 5.4 : Agile Model 18](#_Toc186009692)

[Figure 7.1 : Gantt Chart 20](#_Toc186009693)

# List of Abbreviations/Acronyms

AQI Air Quality Index

CNN Convolutional Neural Network

CO Carbon monoxide

LSTM Long Short-Term Memory

NLP Natural Language Processing

NO2 Nitrogen dioxide

O3 Ozone

PM Particulate Matter

SO2 Sulfur dioxide

SVM Support Vector Machine

WAQI World Air Quality Index

WHO World Health Organization

## 

# CHAPTER 1

# INTRODUCTION

Air pollution is contamination of the indoor or outdoor environment by any chemical, physical or biological agent that modifies the natural characteristics of the atmosphere. [11] Aside from harming ecosystems, air pollution also has negative effects on human health, including premature death, skin rashes, lung infections, respiratory tract infections, pneumonia, lung cancer, and heart failure (Zhang et al., 2024). Among other, Particulate Matter 2.5 (PM2.5), Particulate Matter 10 (PM10), [Carbon dioxide](https://www.sciencedirect.com/topics/chemical-engineering/carbon-dioxide) (CO2), [Carbon monoxide](https://www.sciencedirect.com/topics/earth-and-planetary-sciences/carbon-monoxide) (CO), [Sulfur Oxides](https://www.sciencedirect.com/topics/earth-and-planetary-sciences/sulfur-oxide) (SOX), [Nitrogen Oxides](https://www.sciencedirect.com/topics/chemical-engineering/nitrogen-oxides) (NOX), Ozone (O3), and Ammonia (NH3) are key contributors to AQI. Particulate Matter is a complex mixture with components having diverse chemical and physical characteristics. PM2.5, i.e. particles with an aerodynamic diameter equal to or less than 2.5 μm, and PM10, i.e. particles with an aerodynamic diameter of equal to or less than 10 μm. However, other commonly measured air pollutants such as ozone (O3), nitrogen dioxide (NO2), sulfur dioxide (SO2) and carbon monoxide (CO) are also of concern, as are other components of air pollution. Considering the health implications of air pollutants, it is imperative to constantly monitor the concentration levels in order not to exceed the specified threshold. Thus, measures to check this include the use of the air quality index (AQI). [12] The air quality index (AQI) is a quantitative air quality assessment tool that provides a standard measurement framework and the quality of air that is safe for human beings. Air pollution is a global issue and while some developed countries have devised methods to check the pollutants’ concentration, developing countries struggle to monitor air pollutants, let alone establish ground monitoring stations in multiple locations. Fortunately, some developing countries such as South Africa have established 133 monitoring stations to monitor the quality of air across the country and also in some industrial hubs of the country. Research has shown that environmental monitoring stations collect large volumes and varieties of data, which need to be analyzed to understand the nonlinear nature of air pollutants using robust computational approaches. Approaches that have been used to analyze air pollutants include numerical models, statistical approaches, and machine learning techniques. Several machine learning approaches, including random forest (RF), decision tree (DT), and support vector machine (SVM), have been used for air quality predictions. As mentioned earlier, these machine learning approaches have their limitations. Moreover, deep learning approaches include long short-term memory (LSTM) and a convolutional neural network (CNN). The relationship between deep learning and machine learning can be considered in two aspects. Firstly, deep learning is a subset of machine learning. Secondly, deep learning models are algorithms using neural networks with multiple layers(Agbehadji & Obagbuwa, 2024). A transformer model is a type of [deep learning](https://www.ibm.com/topics/deep-learning) model that have quickly become fundamental in [natural language processing](https://www.ibm.com/topics/natural-language-processing) (NLP), and have been applied to a wide range of tasks in machine learning and artificial intelligence.

## 1.1 Background

Air pollution has become one of the most pressing environmental and health concerns worldwide. Rapid urbanization, industrialization, and increased vehicular emissions have significantly deteriorated air quality, particularly in densely populated regions. [2] Poor air quality not only contributes to environmental degradation but also poses severe health risks, including respiratory diseases, cardiovascular problems, and premature mortality (Ravindiran et al., 2023). Air quality is generally measured using the Air Quality Index (AQI), which aggregates data from various pollutants such as PM2.5, PM10, NO2, CO, SO2, and O3. According to the World Health Organization (WHO), exposure to fine particulate matter (PM2.5) is one of the leading causes of air pollution-related deaths, with millions of lives affected annually. Regions like South Asia have been identified as hotspots for critical air pollution levels. While air quality monitoring has been established in many regions, real-time predictionand forecasting of air quality remain underdeveloped. Predictive models can provide early warnings, allowing authorities and individuals to take preventive measures. While use of models is essential the use of effective model is important too. The development of proper predictive tool by comparative analysis between them is essential.

## 1.2 Motivation

Air pollution is becoming a major hazard to human health and ecosystems. Despite the enormous advances in air quality monitoring systems, the capacity to forecast air quality over different locations is an essential requirement. The increasing need for proactive measures to lessen the effects of air pollution through prompt and precise forecasting led to the creation of this initiative. Present-day prediction algorithms frequently ignore pollution that transcends regional borders and concentrate on isolated areas. In this study, innovative machine learning models like Transformer and LSTM are used to investigate novel approaches for cross-regional air quality forecasting. The objective is to create more precise forecasting tools that will improve our understanding of regional variations in air quality.

## 1.3 Statement of the Problem

Predicting air quality is a crucial technique for controlling the negative impacts of air pollution. Most conventional models often fail to capture the complex nature of patterns of air pollution, especially when those patterns are regionally distributed. When pollution sources like weather patterns and geographical interactions differ greatly, this shortcoming becomes apparent, which results in forecasting becoming very inconsistent and unreliable. The problem is that there aren't many trustworthy methods that can reliably and precisely generalize across borders. The scalability and applicability of conventional approaches are limited because they typically rely on training data that is unique to a particular region. To get around this problem, modern deep learning models like Transformers and LSTMs have also not been sufficiently compared. This study was conducted in an effort to bridge the gap between local and global air quality forecast models. Numerous studies have demonstrated that forecasting abilities can be significantly enhanced by the use of complex temporal models. This study aims to provide a thorough evaluation of these models and evidence-based insights into their efficacy and limitations in cross-regional forecasting.

## 1.4 Project objective

To evaluate the efficiency of LSTM and Transformer model across different regions for forecasting AQI.

## 1.5 Significance of the study

The goal of this project is to use machine learning techniques to create an efficient system for predicting the air quality index (AQI). This study can assist policymakers in making data-driven decisions on the implementation of mitigation measures by providing accurate and timely insights into trends in air pollution. It can aid in the planning of preventative actions by health organizations and raise public awareness of the health risks linked to air quality. The findings are particularly relevant to Nepal, where rapid urbanization and industrialization have exacerbated air pollution problems. By adapting the solution to the local situation, this project also promotes sustainable development initiatives and enhances the quality of life in the community by fostering a healthier environment.

# CHAPTER 2

# LITERATURE REVIEW

[1] The research by Liu et al. (2019) explores the application of machine learning algorithms to predict air quality metrics, specifically the Air Quality Index (AQI) in Beijing and nitrogen oxides (NOₓ) concentrations in an Italian city. Using Support Vector Regression (SVR) and Random Forest Regression (RFR), the models achieved high accuracy, with the SVR model excelling in AQI prediction (RMSE of 7.666, R² of 0.9776, and correlation coefficient of 0.9887) and the RFR model performing better in NOₓ prediction (RMSE of 83.6716, R² of 0.8401, and correlation coefficient of 0.9180). These results highlight the potential of machine learning to provide precise air quality predictions, supporting environmental monitoring and public health initiatives.

[2] The study by Ravindiran et al. (2023) applied machine learning models to predict the Air Quality Index (AQI) in Visakhapatnam, India, utilizing data on 12 pollutants and 10 meteorological parameters from July 2017 to September 2022. Among the models tested—LightGBM, Random Forest, Catboost, Adaboost, and XGBoost—the Catboost model demonstrated superior performance, achieving an R² of 0.9998, a mean absolute error (MAE) of 0.60, a mean square error (MSE) of 0.58, and a root mean square error (RMSE) of 0.76. In contrast, the Adaboost model was the least effective, with an R² of 0.9753. These findings suggest that machine learning, particularly the Catboost algorithm, offers a promising approach for accurately predicting urban air quality, which is crucial for environmental monitoring and public health planning.

[4] The study by Bhatta and Yang (2023) introduces a machine learning approach to reconstruct historical hourly PM₂.₅ concentrations in the Kathmandu Valley from 1980 to the present. Utilizing the Extreme Gradient Boosting (XGBoost) model, trained on PM₂.₅ data from the U.S. Embassy in Phora Durbar and meteorological inputs from NASA's MERRA-2 reanalysis data, the model achieved a 10-fold cross-validation score of approximately 83.4%, an R² of ~84%, an RMSE of ~15.82 µg/m³, and an MAE of ~10.27 µg/m³. Cross-validation with unseen data from 2018 to 2020 yielded R² scores between 56% and 67%. The reconstructed data revealed that MERRA-2 underestimates PM₂.₅ levels in the region and confirmed higher concentrations during the dry pre- and post-monsoon seasons compared to the monsoon period. Additionally, a strong inverse relationship between PM₂.₅ concentrations and humidity was observed. Notably, none of the years met the World Health Organization's annual mean air quality standards, underscoring persistent air quality challenges in the Kathmandu Valley.

[5] The paper "Deep Learning-Based PM 2.5 Long Time-Series Prediction by Fusing Multisource Data—A Case Study of Beijing" introduces a deep learning model to predict PM 2.5 concentrations in Beijing over long periods (48 hours to 30 days) by combining multisource data, including meteorological and air quality information. Using the Informer model, the study achieves superior prediction accuracy compared to traditional models like LSTM and attention-LSTM. The proposed model improves forecasting performance, offering valuable insights for long-term air quality management and policy decisions in polluted cities like Beijing.

[7] The study by Mogollón-Sotelo et al. (2021) developed a Support Vector Machine (SVM) model to forecast ground-level PM₂.₅ concentrations in Bogotá, a densely populated city characterized by complex terrain. Focusing on days with high Air Quality Index (AQI) values, the model was trained using data from an air quality monitoring network and employed a radial basis function kernel. The model's performance was evaluated using statistical metrics, yielding a root mean square error (RMSE) of 9.302 μg/m³, mean bias of 1.405 μg/m³, index of agreement of 0.732, and a correlation coefficient of 0.654. These results indicate that the SVM model can accurately predict short-term PM₂.₅ concentrations in urban areas with complex topography, demonstrating its potential applicability for air quality forecasting in other cities.

[10] The study by Wang et al. (2021) introduces a novel air quality prediction model named CT-LSTM, which integrates a chi-square test (CT) with a long short-term memory (LSTM) network to forecast the Air Quality Index (AQI) in Shijiazhuang, China. The CT component identifies significant factors influencing air quality, while the LSTM network models temporal dependencies in the data. Trained on hourly air quality and meteorological data from January 1, 2017, to December 31, 2018, and evaluated on data from January 1, 2019, to December 31, 2019, the CT-LSTM model achieved an accuracy of 93.7% in predicting daily AQI levels, outperforming other methods such as Support Vector Regression (SVR), Multi-Layer Perceptron (MLP), Backpropagation (BP) neural network, and Simple Recurrent Neural Network (RNN). Additionally, the CT-LSTM model demonstrated superior performance in terms of Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), indicating its effectiveness in accurately forecasting air quality levels.

Table 2.1 : Summary of the Literature Review

|  |  |  |  |
| --- | --- | --- | --- |
| Date published | Title | Author | Key finding |
| 8 February 2023 | Deep Learning-Based PM2.5 Long Time-Series Prediction by Fusing Multisource Data—A Case Study of Beijing | Zhao, Y  Zhang, L  Wang, X  Zhang, R  Li, Y  Zhou, Y  Li, J | The paper finds that the Informer model, using multisource data and Spearmancorrelation for feature selection, outperforms traditional models like LSTM in long-term PM 2.5 prediction, improving air quality forecasting in cities like Beijing. |
| 2019 | To identify and analyze the sources of PM2.5 and PM10 pollution in Kathmandu, Nepal | Bhattrai et al | The study identified vehicular emissions, construction dust, and agricultural burning as the main sources of PM₂.₅ pollution in Kathmandu and called for measures like cleaner transportation and better construction management to reduce pollution. |
| 2021 | Evolution of neural network to deep learning in prediction of air, water pollution and its Indian context | B P Nandi, G Singh, A Jain, D K Tayal | deep learning models, especially CNNs and RNNs, significantly improve the accuracy of predicting PM2.5 concentrations compared to traditional methods. By integrating multiple data sources, such as air quality, weather, and environmental factors, these models provide more reliable predictions. |
| 25 June 2023 | Reconstructing PM2.5 Data Record for the Kathmandu Valley Using a Machine Learning Model. | Surendra Bhatta and Yuekui Yang | machine learning models can effectively reconstruct missing PM2.5 data records for regions with sparse monitoring, such as the Kathmandu Valley |
| 2014 | Dynamics of PM2.5 concentrations in Kathmandu Valley, Nepal | Ramesh Aryal,  Narayan Prakash,  Ram Krishna Sharma and  Hari Prasad Gautam | PM2.5 levels in Kathmandu Valley are highest during the winter, influenced by temperature inversions, low wind speeds, and the valley’s topography, which traps pollutants. |
| 29 September 2019 | Air Quality Index and Air Pollutant Concentration Prediction Based on Support Vector Regression and Random Forest Regression | Zhao Wei, Wang Bing, Xu Xian, Liu Zhi, and Zhou Xiaoping. | machine learning models, specifically support vector regression (SVR) and random forest regression (RFR), can effectively predict air quality levels and pollutant |

# CHAPTER 3

# REQUIREMENT ANALYSIS

## 3.1 Software Requirements

The development and implementation require appropriate tools and platforms.

**Development Tools:**

**Programming Languages:** Python-preferred for ML  
**IDE:** PyCharm, VS Code, or Jupyter Notebook.

**Machine Learning Libraries:**

**Data Processing:** Pandas, NumPy.  
**Visualization:** Matplotlib, Seaborn, Plotly.  
**Modeling:** Scikit-learn, TensorFlow, Keras, or PyTorch.  
 **Database:** SQLite or PostgreSQL for storing historical data.  
**Version Control:** Git/GitHub for version management and collaboration.  
**Deployment Tools:** Flask/Django for web app development (optional). Docker for containerization. Heroku or AWS for cloud deployment.

## 3.2 Functional Requirements

Functional requirements describe what the system will do to meet the project goals.

**Functionalities:**

**Data Collection:** Collect historical air pollutant and weather data. Store the historical data in the database.

**Data Preprocessing:** Clean the missing data. Normalize data for modeling.

**Forecasting:** Forecast air quality metrics using machine learning algorithms.

**Visualization:** Show the trend and forecast in an interactive dashboard or visual charts.

**Model Evaluation:** Evaluate model performance using metrics like Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

## 3.3 Non-Functional Requirements

Non-functional requirements define the system’s quality attributes.

**Performance:** Ensure low latency (real-time if possible).   
**Security:** Secure sensitive data (e.g., user information). Encrypt data transmissions over the internet  
**Usability:** User-friendly interface for visualizing air quality trends and forecast.  
**Reliability:** System should consistently produce accurate forecast.

## 3.4 Feasibility Study

**Technical Feasibility:**

**Strengths:** Availability of open-source machine learning libraries and free datasets. Scalability using cloud platforms (AWS, Google Cloud).  
**Challenges:** Handling missing or inconsistent data.

**Economic Feasibility:**

**Costs:** Free or low-cost air quality data. Free access to Python libraries, platforms for datasets like OpenAQ, AQICN, Kaggle.

**Operational Feasibility:**

The project is viable and solves a real-world problem; for example, air quality monitoring and its health effects. Easy-to-use dashboards and alerts ensure that the solution will be adopted by the end-users-for example, individuals and policymakers.

**Time Feasibility:**

Assuming consistent effort, a basic prototype can be developed within a few months.  
Advanced features, such as deployment or complex dashboards, may take more time.

# CHAPTER 4

# SYSTEM DESIGN AND ARCHITECTURE

## 4.1 Block Diagram

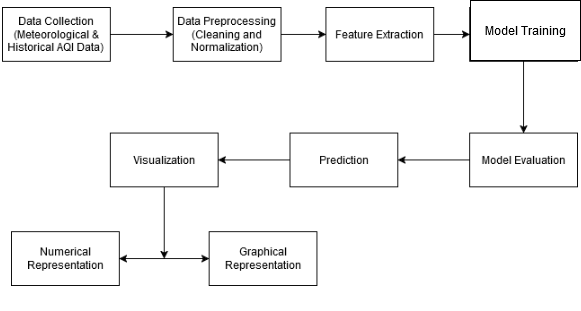


Figure 4.1 : Block Diagram of the System

## 4.2 Flowchart

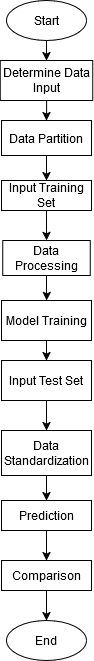


Figure 4.2 : Flowchart of the System

## 4.3 Use Case Diagram

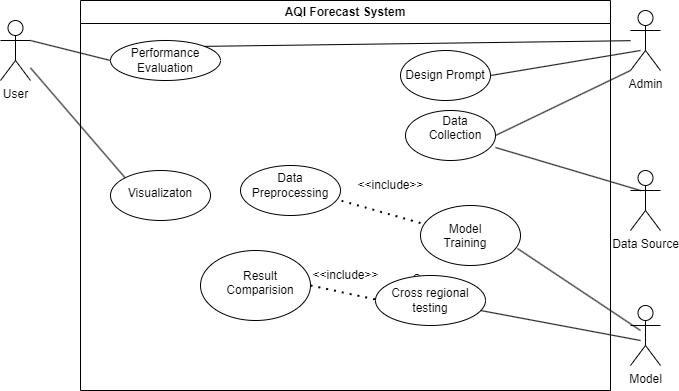


Figure 4.3 : Use Case Diagram

## 4.4 DFD

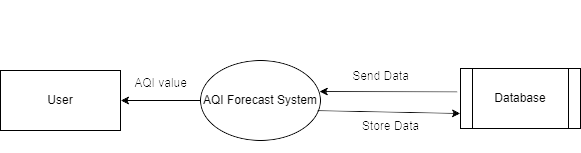


Figure 4.4 : Level 0 DFD

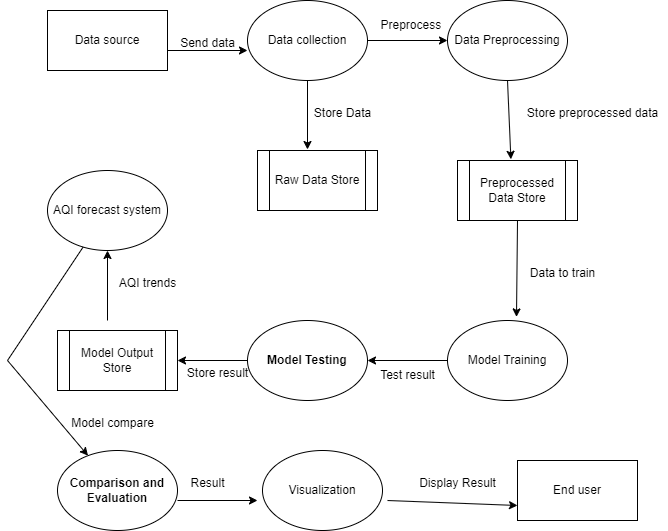


Figure 4.5 : Level 1 DFD

# CHAPTER 5

# METHODOLOGY

This chapter shows the step-by-step approach used to achieve the objectives of this project.

## ****5.1 Data Collection****

Air quality data was collected from reliable sources such as:

* Public air quality (e.g., OpenAQ, AQICN).

### ****Details of Collected Data****

The dataset includes the following parameters:

* **Pollutants:** PM2.5, O3.
* **Weather Conditions:** Temperature, Humidity.

## ****5.2 Data Preprocessing****

### Steps :

1. **Handling Missing Data:**
   * Missing values were filled using **mean or interpolation** techniques.
   * Rows with critical missing values were removed.
2. **Data Normalization/Scaling:**
   * Min-Max Scaling was applied to scale features between [0, 1] for better model convergence.
3. **Outlier Detection and Removal:**
   * Outliers were detected using **Z-score analysis** and handled appropriately.
4. **Feature Selection:**
   * Correlation matrices and domain knowledge were used to select significant features for model training.
5. **Time Series Formatting:**
   * The dataset was structured into time-based formats suitable for sequence forecasting.

**Tools Used:** Python Libraries – Pandas, NumPy, Scikit-learn.

## ****5.3 Model Development****

The model was developed using **Machine Learning (ML) and Deep Learning** techniques.

### ****Algorithms Used:****

#### ****LSTM Model:****

* 1. Use sequential input data for temporal dependencies.
  2. Define architecture:
     + Input layer for time-series data.
     + One or more LSTM layers with dropout for regularization.
     + Fully connected dense layers for output.
  3. Train with sequences containing data from multiple regions.

#### ****Transformer Model:****

* 1. Adapt the Transformer architecture for time-series forecasting.
  2. Key components:
     + Positional encodings to capture sequential order.
     + Attention mechanism to weigh the importance of different time steps and regions.
     + Encoder-decoder setup, or a simpler encoder-only variant.
  3. Incorporate spatial embeddings to model cross-regional relationships.

## ****5.4 Model Evaluation****

The developed models are evaluated using standard regression metrics:

* **Mean Absolute Error (MAE)**
* **Root Mean Squared Error (RMSE)**
* **R-squared (R²)**

## 5.5 Comparative Analysis

Compare the performance of LSTM and Transformer models on:

* + Predictive accuracy across regions.
  + Ability to handle long-term temporal dependencies.
  + Computational efficiency.
  + Robustness to data from regions with varying pollution levels.

## ****5.6 Deployment****

**Model Selection:** Choose the model with better overall performance for real-time forecasting.

Once the model is optimized, it is deployed as a web-based application for real-time air quality forecasting.

### ****Deployment Process:****

1. **Backend Framework:** Flask (Python) to serve the model.
2. **Frontend Development:** React.js for the user interface.
3. **Hosting Platform:** AWS or Heroku for cloud deployment.

## 

## ****5.6 System Workflow****

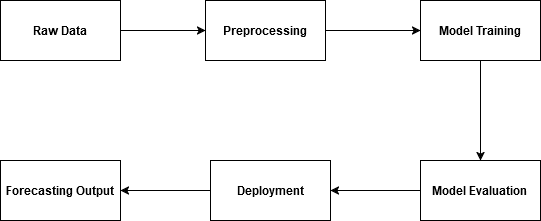


Figure 5.1 : System Workflow

## ****5.7 Tools and Technologies Used****

* **Programming Language:** Python
* **Libraries:** Pandas, NumPy, Scikit-learn, TensorFlow/Keras
* **Deployment Tools:** Flask, AWS, React.js

**Data Visualization Tools:** Matplotlib, Seaborn

## 5.8 Algorithm

### 5.8.1 LSTM Model

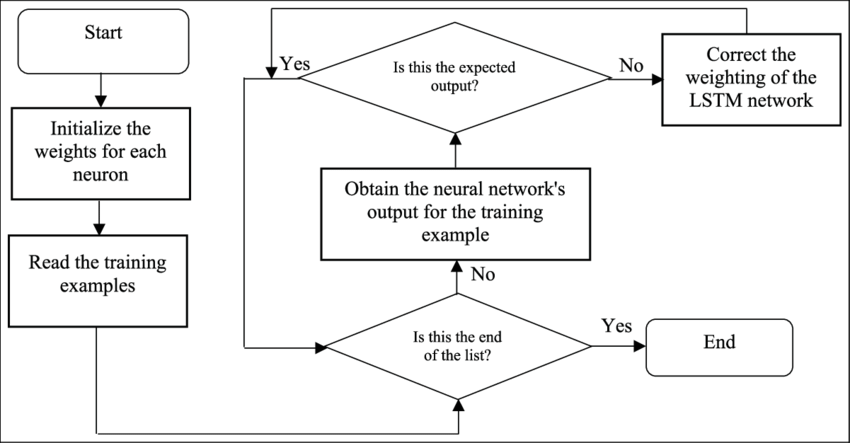
****

Figure 5.2 : Flowchart of LSTM Model

Source : https://www.researchgate.net/figure/Flowchart-for-LSTM-training\_fig3\_328881984

### 5.8.2 Transformer Model

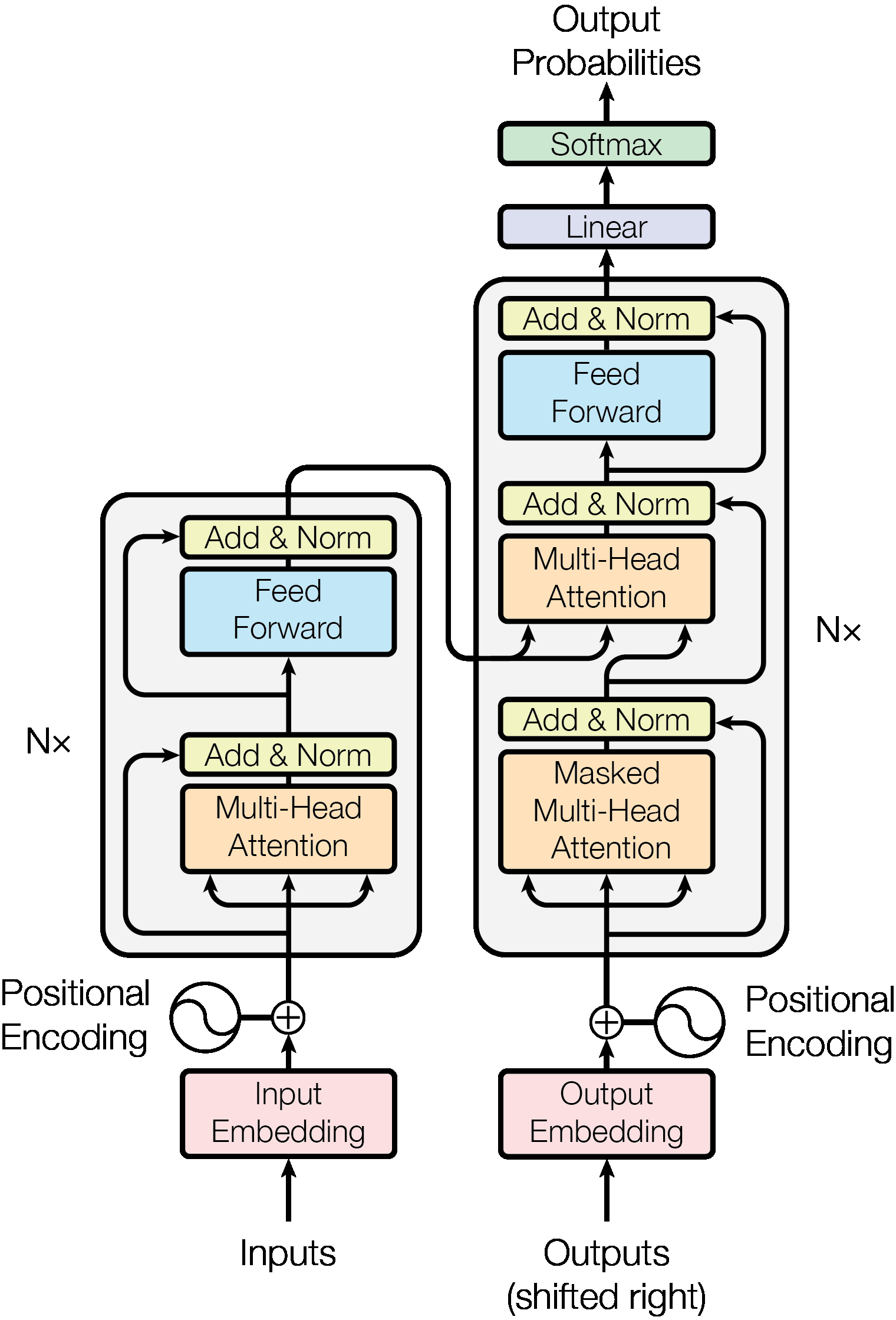
****

Figure 5.3 : Flowchart of Transformer Model

Source : https://production-media.paperswithcode.com/methods/new\_ModalNet-21.jpg

## 5.9 Development Model

### 5.9.2 Agile Model

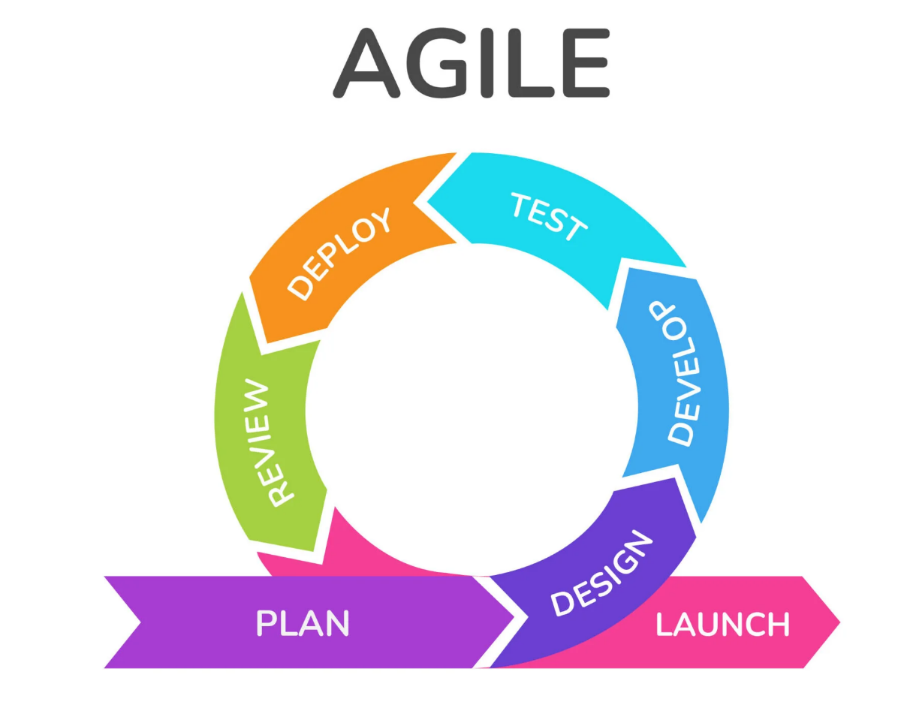
****

Figure 5.4 : Agile Model

Source : <https://www.orientsoftware.com/blog/v-model-vs-agile/>

1. **Requirements:** In this phase, we must define the requirements. We should explain business opportunities and plan the time and effort needed to build the project. Based on this information, we can evaluate technical and economic feasibility.
2. **Design:** After identifying the project, we work with stakeholders to define requirements. Then, we can use UML diagram to show the work of new features and how it will apply to your existing system.
3. **Construction/ Iteration:** When the team defines the requirements, the work begins. Designers and developers start working on their project, which aims to deploy a working product. The product will undergo various stages of improvement, so it includes simple, minimal functionality.
4. **Testing:** In this phase, the Quality Assurance team examines the product's performance and looks for the bug.
5. **Deployment:** In this phase, the team issues a product for the user's work environment.
6. **Feedback:** After releasing the product, the last step is feedback. In this, the team receives feedback about the product and works through the feedback.

# CHAPTER 6

# EXPECTED OUTPUT

The objective of this study is to assess Transformer and LSTM models for air quality forecasting using several measures, including as R2, MAE, and RMSE. Insights into cross-regional forecasting, tools for visual depiction, and best practices for model construction will also be beneficial. The project also aims to provide a comprehensive project report and a reusable codebase for data preprocessing, model training, and evaluation. The results are intended to show that sophisticated deep learning models for cross-regional air quality forecasting are both feasible and efficient.

# CHAPTER 7

# TIME SCHEDULE

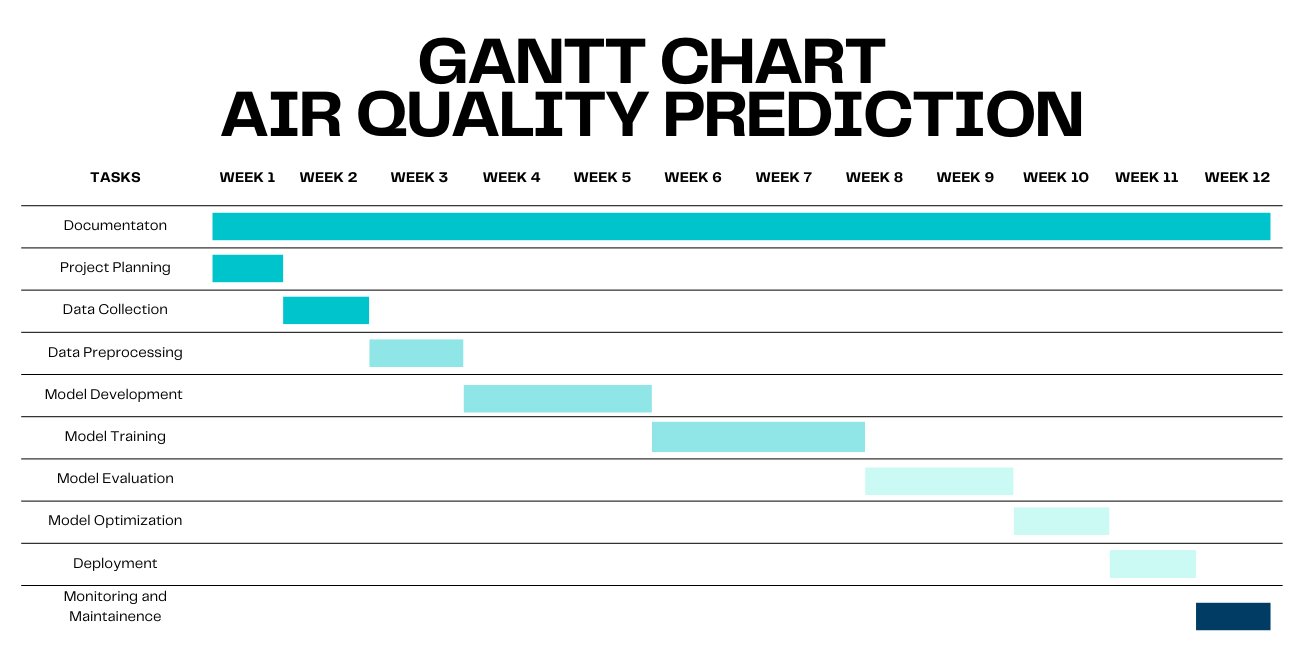


Figure 7.1 : Gantt Chart

# REFERENCES

[1] Liu H, Li Q, Yu D, Gu Y. Air Quality Index and Air Pollutant Concentration Prediction Based on Machine Learning Algorithms. Applied Sciences. 2019; 9(19):4069. <https://doi.org/10.3390/app9194069>

[2] Ravindiran, G., Hayder, G., Kanagarathinam, K., Alagumalai, A., & Sonne, C. (2023). Air quality prediction by machine learning models: A predictive study on the indian coastal city of Visakhapatnam. Chemosphere, 338, 139518. <https://doi.org/10.1016/J.CHEMOSPHERE.2023.139518>

[3] El-Habil, B. Y., & Abu-Naser, S. S. (2022). Global climate prediction using deep learning. Journal of Theoretical and Applied Information Technology, 100(24), 4824-4838.

[4] Bhatta, Surendra & Yang, Yuekui. (2023). Reconstructing PM2.5 Data Record for the Kathmandu Valley Using a Machine Learning Model. Atmosphere. 14. 1073. 10.3390/atmos14071073

[5] Niu, M., Zhang, Y., & Ren, Z. (2023). Deep Learning-Based PM2.5 Long Time-Series Prediction by Fusing Multisource Data—A Case Study of Beijing. Atmosphere, 14(2), 340. <https://doi.org/10.3390/atmos14020340>

[6] Shahriar SA, Kayes I, Hasan K, Hasan M, Islam R, Awang NR, Hamzah Z, Rak AE, Salam MA. Potential of ARIMA-ANN, ARIMA-SVM, DT and CatBoost for Atmospheric PM2.5 Forecasting in Bangladesh. Atmosphere. 2021; 12(1):100. <https://doi.org/10.3390/atmos12010100>

[7] Mogollón-Sotelo, Caroline & Casallas García, Alejandro & Vidal, Sergio & Celis Mayorga, Nathalia & Ferro, Camilo & Belalcazar, Luis. (2021). A support vector machine model to forecast ground-level PM 2.5 in a highly populated city with a complex terrain. Air Quality Atmosphere & Health. 13. 10.1007/s11869-020-00945-0.

[8] Rui, F.; Han, G.; Kun, L. Analysis and accurate prediction of ambient PM2.5 in China using Multi-layer Perceptron. Atmos. Environ. **2020**, 232, 117534.

[9] Zhang, Jiaxuan & Li, Shunyong. (2022). Air quality index forecast in Beijing based on CNN-LSTM multi-model. Chemosphere. 308. 136180. 10.1016/j.chemosphere.2022.136180.

[10] Wang, J., Li, J., Wang, X. et al. Air quality prediction using CT-LSTM. Neural Comput & Applic 33, 4779–4792 (2021). https://doi.org/10.1007/s00521-020-05535-w

[11] Zhang, Z., Zhang, S., Chen, C., & Yuan, J. (2024). A systematic survey of air quality prediction based on deep learning. Alexandria Engineering Journal, 93, 128–141. <https://doi.org/10.1016/J.AEJ.2024.03.031>

[12] Agbehadji, I. E., & Obagbuwa, I. C. (2024). Systematic Review of Machine Learning and Deep Learning Techniques for Spatiotemporal Air Quality Prediction. Atmosphere, 15(11). <https://doi.org/10.3390/ATMOS15111352>