API ENABLED AQI PREDICTION USING DEEP LEARNING

by

SANSKAR SHARMA
(Roll No.: B200038CS)

ANURAG SINGH
(Roll No.: B200058CS)

PROJECT REPORT

submitted for partial fulfillment for Degree in

Bachelor of Technology
in
Computer Science and Engineering

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

NATIONAL INSTITUTE OF TECHNOLOGY SIKKIM

May, 2024

NATIONAL INSTITUTE OF TECHNOLOGY SIKKIM
Ravangla Campus
Barfung Block, Ravangla Sub Division, South Sikkim-737139
(AN INSTITUTE OF NATIONAL IMPORTANCE, GOVT. OF INDIA)

Certificate by Supervisor

This is to certify that the project report entitled "API Enabled AQI Prediction using Deep Learning" is being submitted by **Sanskar Sharma** (Roll No. B200038CS) and **Anurag Singh** (Roll No. B200058CS), students in the Department of Computer Science and Engineering, National Institute of Technology Sikkim, for the award of the degree of Bachelor of Technology (B.Tech). It is an original work carried out by them under my supervision and guidance. The results embodied in this report have not been submitted to any other University or Institute for the award of any degree or diploma.

Dr. Pankaj Kumar Keserwani
(Assistant Professor)
Department of CSE
NIT Sikkim

NATIONAL INSTITUTE OF TECHNOLOGY SIKKIM
Ravangla Campus
Barfung Block, Ravangla Sub Division, South Sikkim-737139
(AN INSTITUTE OF NATIONAL IMPORTANCE, GOVT. OF INDIA)

Certificate by Students

We hereby declare that the work presented in the report entitled "API Enabled AQI Prediction using Deep Learning" is a bonafide record of the work done by us under the supervision of **Dr. Pankaj Kumar Keserwani**, Department of Computer Science and Engineering, National Institue of Technology Sikkim and that no part thereof has been presented for the award of any other degree.

- We have followed the guidelines provided by the institute in writing the report.
- The report does not contain any classified information.
- Whenever, we have used materials from other sources, we have given due credits to those by citing them in the text of the report and giving the details in the references.
- Whenever, we have quoted written materials from other sources, we have put them under quotation marks and given due credits to the sources by citing them in the text of the report and giving their details in the references.

Dated: May 2024
Place: NIT Sikkim

Sanskar Sharma
Roll No.: B200038CS

Anurag Singh
Roll No.: B200058CS

Acknowledgments

We would like to express our profound gratitude to my dedicated and insightful supervisor, **Dr. Pankaj Kumar Keserwani**, Assistant Professor, Department of Computer Science and Engineering, National Institute of Technology Sikkim, for their unwavering guidance and support throughout this process. Their invaluable expertise and keen insights were instrumental in the development and advancement of this work. Furthermore, the invaluable skills and knowledge gained during this period will be of significant benefit to our future academic and professional pursuits. We are grateful to all the professors in the Computer Science and Engineering department at the National Institute of Technology, Sikkim, for teaching us the fundamentals of computer science. Their knowledge was very helpful to us when we were doing our research and making progress.

Dated: May 2024
Place: NIT Sikkim

Sanskar Sharma
Roll No.: B200038CS

Anurag Singh
Roll No.: B200058CS

Contents

List of Figures vi

List of Tables vii

List of Acronyms viii

Notations ix

Abstract x

1 Introduction 1
1.1 Background 2
1.2 Motivations 3
1.3 Problem Statement 3
1.4 Objectives 3
1.5 Chapter Summary 4

2 Literature Review 6

3 Methods and Materials 12
3.1 Proposed Framework 12
3.1.1 API 13
3.1.2 Dataset Collection 14
3.2 Data Visualisation 15
3.2.1 AQI vs Temperature 16
3.2.2 AQI vs Humidity 16
3.2.3 PM2.5 vs Temperature 16
3.2.4 PM2.5 vs Humidity 17
3.3 Data Preprocessing 17
3.4 Feature Selection 18
3.4.1 Why Feature Selection 19

3.4.2 Feature Importance Score 19
3.5 Evaluation Parameters 20
3.6 Proposed Model 21
3.6.1 Long Short-Term Memory(LSTM) 21
3.6.2 Advantages of LSTM 22
3.6.3 Model Architecture 22
3.7 Mobile App Integration 22

4 Result and Analysis **24**
4.1 Experimental Setup 24
4.2 Model Evaluation 24
4.2.1 Evaluation Metrics 25
4.3 Comparison of Proposed Architecture with Different Approaches and Models 26
4.4 Result of Proposed Architecture 26

5 Conclusions and Future Works **27**
5.1 Conclusions 27
5.2 Future Scope 27

Bibliography **29**

List of Figures

3.1 Proposed Framework 13
3.2 Comparison of AQI vs Temperature for North and South Sikkim 16
3.3 Comparison of AQI vs Humidity for North and South Sikkim 16
3.4 Comparison of AQI vs PM2.5 for North and South Sikkim 16
3.5 Comparison of PM2.5 vs Humidity for North and South Sikkim 17
3.6 Correlation Matrix of AQI and Meterological Parameters 19
3.7 Architecture of LSTM Model 21
3.8 Mobile App Integration with Sequential LSTM 23

4.1 Scatter plot of Actual vs Predicted AQI 25

List of Tables

1.1 Table to Indian AQI Range 2

2.1 Literature Survey 8
2.2 Literature Survey 9
2.3 Literature Survey 10
2.4 Literature Survey 11

3.1 Location Information 14
3.2 Air Quality Indices 14
3.3 Timestamps 14
3.4 Meteorological Parameters 15
3.5 Normalized Data Format 18
3.6 Feature Importance Scores for North and South Sikkim using Random Forest model 20
3.7 Sequential LSTM Model Architecture 22

4.1 Comparitive Analysis with other approaches 26

List of Acronyms

API	Application Programming Interface
AQI	Air Quality Index
BLSTM	Bidirectional Long Short-Term Memory
CC	Correlation Coefficient
CNN	Convolutional Neural Network
CO	Carbon Monoxide
CPCB	Central Pollution Control Board
DL	Deep Learning
DLSTM	Deep Long Short-Term Memory
DSN-PS	Deep Sparse Network with Pairwise Selection
GBDT	Gradient Boosting Decision Trees
GRU	Gated Recurrent Unit
GWO	Grey Wolf Optimizer
IMD	India Meteorological Department
IOT	Internet Of Things
IQR	Interquartile Range
LSTM	Long Short-Term Memory
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MBGD	Mini-Batch Gradient Descent
ML	Machine Learning
MLP	Multilayer Perceptron
mRMR	minimum Redundancy Maximum Relevance
MSE	Mean Squared Error
NSE	Nash-Sutcliffe Efficiency
PBIAS	Percentage Bias
PM2.5	Fine particles with diameters of 2.5 micrometers or smaller
PM10	Fine particles with diameters of 10 micrometers or smaller
PSO	Particle Swarm Optimization
RDD	Resilient Distributed Datasets
ResNet	Residual Network
RMSE	Root Mean Square Error
RMSLE	Root Mean Squared Logarithmic Error
RNN	Recurrent Neural Network
SGD	Stochastic Gradient Descent
S-BLSTM	Stacked Bidirectional Long Short-Term Memory
S-LSTM	Stacked Long Short-Term Memory
SPO	Swarm Particle Optimization
SVR	Support Vector Regression
XGB	Extreme Gradient Boosting

Notations

β	It denotes the momentum and controls the impact of past gradients on the weight updates.
x	It denotes an attribute in a dataset
μ	It denotes the mean of the feature x
σ	It is the standard deviation of the feature x
O	It denotes the outliers
$Q1$	It denotes the first quartile
$Q3$	It denotes the third quartile
R^2	Coefficient of determination

Abstract

Air pollution poses significant health risks, necessitating the development of forecasting mechanisms for early interventions by authorities. Machine Learning, especially Deep Learning models, has gained importance in air quality prediction. In today's context, monitoring and predicting air quality, particularly in developing nations like India, are important. Modern predictive technologies based on machine learning techniques have proven to be efficient tools in studying these contemporary hazards.

This study investigates factors influencing the Air Quality Index (AQI), including CO (Carbon Monoxide), NO_2 (Nitrogen Dioxide), O_3 (Ozone), PM10 (Particulate Matter), PM2.5 (Particulate Matter), and SO_2 (Sulfur Dioxide) over a one-year period (2022-2023). It covers all districts of Sikkim state and major cities like Delhi, Mumbai, Kolkata, and Chennai. The investigation encompasses 16 attributes, such as Temperature, Relative Humidity, Dew Point, Precipitation, Rainfall, Snowfall, Mean Sea Level Pressure, Surface Pressure, Cloud Cover, etc.

Utilizing Deep Learning techniques on our dataset enables the prediction of AQI variations across Indian cities. The integration of Deep Learning in forecasting air quality variations across diverse geographical regions offers scalable and robust analysis. This approach enhances prediction accuracy for time-series data.

This study aims to provide actionable insights, empowering active measures in addressing air quality concerns and promoting healthier living environments for communities across globe.

Chapter 1

Introduction

Air pollution is one of the most serious environmental problems that threaten human health and the economy. Regional transports, acid storage, and increasing greenhouse gases reveal the effects of air pollution reaching global dimensions. Pollutants from traffic, transportation, industry, and heating are the main causes of air pollution [1]. Air pollution has emerged as a critical global challenge, affecting nations regardless of their developmental status. Particularly in urban areas of developing countries, the escalating pace of industrialization and the surging vehicular count have led to a substantial release of harmful air pollutants. The outcome of this pollution are profound, ranging from minor allergic responses such as throat, eye, and nose irritations to severe health conditions like bronchitis, heart diseases, pneumonia, and aggravated asthma. According to a survey, due to air pollution 50,000 to 100,000 premature deaths per year occur in the U.S. alone. Whereas in EU, number reaches to 300,000 and over 3,000,000 worldwide [2,4].

Air pollution forecasting is very useful for informing about the pollution level that will allow policy-makers to adopt measures for reducing its impact. For example, traffic restrictions could be imposed with the goal of avoiding high pollution episodes. Usually, the Air Quality Index (AQI) is used to indicate the air pollution level. AQI is a piece-wise linear function of the following pollutant concentrations: Ozone (O_3), particulate matters (PM2.5, PM10), carbon monoxide (CO), sulphur dioxide (SO_2) and nitrogen dioxide (NO_2). Nevertheless, there does not exist a global standard and different countries and regions have their own AQI indexes, based on their own air quality standards [2].

Table 1.1: Table to Indian AQI Range

AQI Range	Category	Meaning
0-50	Good	The air is fresh and free from toxins.
51-100	Satisfactory	Acceptable air quality for healthy adults.
101-150	Moderate	Inhaling such air can cause slight difficulty in breathing.
151-200	Poor	This could be typically problematic for children.
201-300	Very Poor	Exposure to air can cause chronic morbidities.
301-500	Severe	Life under threat

Machine Learning (ML) techniques have emerged as the predominant approach for forecasting air quality. Over the course of the 21st century, a wealth of literature has surfaced, featuring numerous works dedicated to implementing diverse models. These models aim to achieve optimal accuracy in forecasting the Air Quality Index (AQI) or predicting concentrations of specific pollutants [3]. Extensive research endeavors have been directed towards employing various ML algorithms for predicting both AQI and the concentration levels of individual pollutants linked to air quality, considering the interplay with meteorological factors.Utilizing an API(Application Programming Interface)-enabled AQI prediction through Deep Learning, considering meteorological factors, could be a game-changer in enhancing AQI prediction accuracy [4].

1.1 Background

Air pollution has become a pervasive concern, transcending boundaries and impacting human health and economies globally. Its sources range from vehicular emissions to industrial activities, contributing to a complex web of pollutants that affect both developed and developing nations. Rapid urbanization and industrial growth in many developing countries, notably in their urban centers, have intensified this issue, leading to a surge in detrimental air pollutants. The health repercussions, ranging from mild irritations to severe respiratory illnesses and premature deaths, underscore the urgency of addressing air quality concerns. Statistical data worldwide, such as the staggering numbers of premature deaths attributed to air pollution, highlights the critical need for proactive measures to mitigate its impact [5].

In recent years, Machine Learning (ML) techniques have emerged as indispensable tools in forecasting air quality. This paradigm shift has resulted in a vast array of research dedicated to deploying diverse ML models for accurate predictions of the Air Quality Index (AQI) and concentrations of specific pollutants. These models integrate meteorological factors, recognizing their crucial interplay with air quality [6]. The evolution of ML algorithms for AQI prediction showcases a growing body of work striving for optimal precision in gauging air quality variations. Leveraging an API-enabled AQI prediction through Deep Learning, incorporating meteorological elements, holds the promise of significantly elevating the accuracy of AQI predictions, marking a potential paradigm shift in this domain [7, 8].

1.2 Motivations

Advancing air quality prediction methods forms the core motivation of this study. The integration of Deep Learning emerges as a pivotal approach, aiming to reshape existing models by fostering collaborative and location-specific training. Unlike conventional methods, Deep Learning acknowledges the diverse pollutant characteristics and sources across regions, addressing a critical challenge previously overlooked. This approach bridges the gap, ensuring predictive models encapsulate the unique attributes of each locale, thus enhancing accuracy and applicability [5, 8].
Moreover, the imperative for real-time model updates underpins the motivation for employing Deep Learning. Safeguarding sensitive environmental data in an era dominated by evolving data security concerns becomes paramount. Deep Learning not only enables efficient collaboration in model training but also ensures adaptive and responsive predictive models through real-time data integration [4].

1.3 Problem Statement

The research endeavors to revolutionize Air Quality Index (AQI) prediction through the integration of Deep Learning and Application Programming Interfaces (APIs) [9]. Addressing critical challenges in air quality forecasting, including the need for location-specific adaptability and seamless real-time data integration, the study seeks innovative solutions. Leveraging the power of Deep Learning, this research aims to develop an API-enabled system for predicting AQI, ensuring both accuracy and location specificity in air quality assessment across diverse geographical regions.

1.4 Objectives

- **Harness Deep Learning:** Implement Deep Learning techniques to enable location-specific model training, addressing the diversity of pollutant characteristics across regions.

- **Develop an API-Enabled System:** Create an API-enabled platform for Air Quality Index (AQI) prediction, ensuring accuracy and privacy in air quality assessment across diverse geographical regions.

- **Facilitate Real-Time Data Integration:** Enable efficient and real-time updates of predictive models, ensuring adaptability to changing air quality dynamics across various locations.

- **Enhance Predictive Accuracy:** Develop robust and generalizable air quality prediction models by aggregating knowledge from diverse locations through Deep Learning, thereby improving accuracy in forecasting AQI variations.

- **Mobile App Integration:** Integration with a mobile application to allow users to access timely air quality predictions.

These objectives collectively aim to revolutionize the approach to air quality prediction, leveraging innovative methodologies to ensure accuracy while addressing the challenges inherent in forecasting air quality variations across diverse regions[10].

1.5 Chapter Summary

This chapter delineates the critical facets of air quality prediction, emphasizing the pivotal role of innovative methodologies in addressing challenges and advancing predictive accuracy. The chapter navigates through the motivations, objectives, and methodologies underpinning the study.

Motivated by the pressing need to revolutionize air quality prediction, the study explores the integration of Deep Learning, an approach fostering collaborative and location-specific model training. Contrary to conventional methods, this approach acknowledges the diverse pollutant characteristics and sources across regions, aiming to enhance predictive accuracy by encapsulating the unique attributes of each locale. The imperatives of real-time model updates stand as foundational motivations for employing Deep Learning.

The objectives encapsulate the study's essence, highlighting the pursuit of harnessing Deep Learning, developing an API-enabled system for AQI prediction, and ensuring privacy preservation. Facilitating real-time data integration and enhancing predictive accuracy by aggregating knowledge from diverse locations remain focal objectives, collectively aiming to transform air quality prediction methodologies[11].

Through an complex balance of motivations and objectives, the study endeavors to propel the understanding and forecasting of air quality variations across diverse geographical regions, promising advancements in accuracy.

Chapter 2

Literature Review

In recent years, numerous advancements have been made in the prediction of Air Quality Index (AQI) utilizing various artificial intelligence and machine learning techniques. The increasing awareness of air pollution's adverse effects on public health and the environment has driven researchers to explore innovative methods for accurate AQI prediction. This chapter reviews the existing literature on AQI, focusing on the evolution from traditional statistical methods to advanced deep learning models in order to enhance prediction accuracy.

In the year 2019, Krishan et al. [5] proposed a novel approach to air quality modeling using Long Short-Term Memory (LSTM) networks over the National Capital Territory of Delhi (NCT-Delhi), India1. Their study focused on predicting concentrations of O_3, PM2.5, NO_x, and CO, utilizing factors such as vehicular emissions, meteorological conditions, traffic data, and pollutant levels.

In the year 2019, Samaher et al. [1] proposed an intelligent predictor for air pollutants concentrations using deep learning techniques with a recurrent neural network (RNN). They optimized the network structure using a particle swarm optimization (PSO) algorithm, resulting in the smart air quality prediction model (SAQPM).

In the year 2020, Zhang et al. [12] proposed a deep learning and image-based model for air quality estimation, which utilizes a residual neural network to detect air quality from images collected by mobile devices. They designed a self-supervision module, improved the air quality level detection model, and enhanced recognition accuracy.

In the year 2019, Samaher et al. [10] proposed an intelligent predictor for air pollutants concentrations using deep learning techniques with a recurrent neural network (RNN). They optimized the network structure using a particle swarm optimization (PSO) algorithm, resulting in the smart air quality prediction model (SAQPM).

In the year 2023, Pan et al. [13] proposed a comprehensive study on ozone pollution prediction using machine learning methods, evaluating nineteen different models to find the most efficient and accurate one.

In the year 2019, Hocine et al. [14] proposed an Internet of Things (IoT)-based system for air quality monitoring, highlighting the necessity for such systems due to the significant health impacts of air pollution. They emphasized the advantages of IoT systems in providing efficient and economical solutions compared to traditional methods.

In the year 2020, Kumar et al. [3] proposed a comprehensive study on air pollution prediction using machine learning techniques, focusing on the case of Indian cities. They investigated six years of air pollution data from 23 Indian cities, preprocessed the dataset, and selected key features through correlation analysis. Their exploratory data analysis revealed hidden patterns and identified pollutants directly affecting the air quality index.

In the year 2022, Wang et al. [15] proposed a novel air pollutant prediction and early warning framework that innovatively combines feature extraction techniques, feature selection methods, and intelligent optimization algorithms. They introduced a two-stage deep learning hybrid framework optimized by the gray wolf optimization algorithm, which significantly outperformed other comparative models in terms of prediction accuracy, warning accuracy, and prediction stability.

In the year 2020, Rahimpour et al. [4] proposed a novel approach to predict the Air Quality Index (AQI) for urban environments using machine learning techniques. They introduced hybrid single decomposition (HSD) and hybrid two-phase decomposition (HTPD) models.

In the year 2021, Zhang et al. [9] proposed a hybrid deep learning model VMD-BiLSTM for PM2.5 air quality forecasting. This model combines variational mode decomposition (VMD) and bidirectional long short-term memory network (BiLSTM) to predict PM2.5 changes in cities in China.

In the year 2020, Liu et al. [16] proposed a spatial multi-resolution data-driven AQI forecasting model based on real-time decomposition, which utilizes spatial correlation analysis to identify auxiliary sites relevant to the target site for AQI forecasting.

In the year 2021, Yeo et al. [7] proposed a deep learning model integrating a convolutional neural network (CNN) and a gated recurrent unit (GRU) to accurately predict PM2.5 concentrations in Seoul, South Korea. Utilizing data from various meteorological and air quality stations, they trained the model with historical data from 2015 to 2017 to predict PM2.5 levels for 2018.

In the year 2021, Jiang et al. [6] proposed an insightful study on the use of controlled release urea (CRU) in rice production, highlighting its benefits for both yield improvement and environmental protection. Their research, grounded in global meta-analysis and machine learning techniques, demonstrated that CRU application can enhance rice yield, nitrogen use efficiency, and net benefits while simultaneously reducing emissions of N_2O and CH_4, global warming potential, nitrogen leaching, and NH_3 volatilization.

In the year 2022, Cheng et al. [17] proposed a novel stacked ResNet-LSTM model to enhance prediction accuracy for PM2.5 concentration level forecast. They demonstrated that the model outperforms other models such as boosting algorithms or general recurrent neural networks. Additionally, to address the issue of missing air quality data in certain areas, they introduced a correlation alignment (CORAL) method for prediction by aligning the second-order statistics between source and target areas.

In the year 2023, Yonar et al. [11] proposed an approach to model air pollution by integrating the Adaptive Network Fuzzy Inference System (ANFIS) with metaheuristic algorithms such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and Differential Evolution (DE). They utilized various air pollution parameters like PM2.5, PM10, SO_2, O_3, NO_2, CO, and meteorological data to predict daily air pollution in Istanbul

This section provides an overview of significant research and advancements in the field of air quality prediction.

Ref.	**YOP**	**Dataset Used**	**Model Used**	**Result**	**Hyperparameters Used**
[5]	2019	Safdarjung Meteorology Monitoring Station	LSTM	NSE: 0.86 to 0.94, CC: 0.93 to 0.97	Length of input sequence or the batch size, learning, rate and epochs, statistical parameters (NSE, CC, PBIAS, and RMSE)
[1]	2019	Resilient Distributed Datasets (RDDs) such as Spark	RNN	SMAPE values are : PM2.5 : 13.61,PM10 : 12.25, NO_2 : 11.83, O_3 : 4.52	Parameter values (weight, bias, number of hidden layers, number of nodes in each hidden layer, and activation function)
[12]	2020	High-quality outdoor air quality data set (NWNU-AQI)	DL and image-based model	SMAPE for AQC-Net :- MSE :0.3426, RMSE : 0.5854, MAE : 0.2919	Optimized by SGD, momentum (β) = 0.9, mini-batch size is 32, with an initial learning rate of 10^{-2}

Table 2.1: Literature Survey

Ref.	YOP	Dataset Used	Model Used	Result	Hyperparameters Used
[10]	2020	China National Environmental Monitoring Centre Anhui, China	TLS-BLSTM	RMSE Values :- PM2.5 :12.1., NO_2:10.32, O_3: 11.27	BLSTM units with Nf (16-256), 200 epochs, MBGD, L2 reg (l=0.01, rd=0.1), ReLU, batch 128. Optimal: Stacked BLSTM, Nh=7, Nf=128
[13]	2023	KAUST	ANN, SVR, RF, XGBoost	RMSE: 3.032 MAE: 1.979 MAPE: 0.213 R^2 : 0.921	For linear models, various norms were used, For SVR, multiple kernels were tested, MLP configurations included different layer shapes
[14]	2019	IoT based system	Epi-demiological data	Measures outdoor CO, CO_2 concentration	LSTM model optimization encompassed a thorough search across learning rates (0.01 - 1), epochs (1 - 100), and batch sizes (50 - 1000). Fine-tuning, guided by Nash - Sutcliffe efficiency (NSE) values, was performed with a consistent learning rate of 0.01.
[3]	2023	Central Pollution Control Board (CPCB)	Gaussian Naive Bayes, SVM, XG-Boost	With SMOTE-SVM-MAE:0.153, RMSE:2.098, R^2 : 0.512 XGBoost-MAE:0.174, RMSE:1.008, R^2 : 0.325	The standard performance parameters are MAE, RMSE, Root Mean Squared Logarithmic Error (RMSLE), and coefficient of determination, i.e. R^2

Table 2.2: Literature Survey

Ref.	YOP	Dataset Used	Model Used	Result	Hyperparameters Used
[15]	2022	Xingtai Dataset,Anyang Dataset,Beijing Dataset	RNN, mRMR, GWO-LSTM	Xingtai Dataset-MAE: 0.0237, RMSE: 0.0281, MAPE: 5.2616 Anyang Dataset: MAE: 0.0216 RMSE: 0.0244 MAPE: 4.2446 Beijing Dataset MAE: 0.0235 RMSE: 0.0279 MAPE: 5.0212 Accuracy - Xingtai: 87Anyang: 90Beijing: 93 MAPE: 5.2616	Number of iterations, gray wolves ,Hidden layer neurons, Batch size, Dropout rate
[8]	2022	Air pollution sensors	IoT-based system integrating microcontrollers	System effectively monitors and transmits air quality data	Not applicable (NA)
[4]	2020	AQI data of Orumiyeh city	Hybrid Single Decomposition (HSD) model, Hybrid Two-Phase Decomposition (HTPD) model	HSD Models: Similar accuracy, HTPD Model: Highest accuracy - Training Phase: $R^2 = 0.98$, RMSE = 4.13, MAE = 2.99	Not specified
[9]	2021	PM2.5 data from four observation stations in China	Combination of VMD and BiLSTM	MAE of 3.562, RMSE of 5.232, MAPE of 5.427, $R^2 = 0.992$, and accuracy of 77.391%	Not explicitly mentioned

Table 2.3: Literature Survey

Ref.	YOP	Dataset Used	Model Used	Result	Hyperparameters Used
[16]	2021	AQI data from several sites including Zhangjiakou, Cangzhou, and Nantong.	ORELM, ORELM-LR, ORELM-HR, MREM	MAE: 8.81, MAPE: 26.44%, RMSE: 13.94	Stochastic Gradient Descent, Particle Swarm Optimization (PSO)
[7]	2021	26 city monitoring stations in China	WPD, Stacked Auto-Encoder, Outlier Robust Extreme Learning Machine, Multi-Objective Wolf Colony Algorithm	model demonstrated excellent forecasting performance	Number of iterations, population size, search agents
[2]	2018	Registry of Motor Traffic, AirMAC	Emission standards.	Aims to reduce SO_2 and PM pollutants	Not applicable
[6]	2021	global meta-analysis	RF, MLPR	Random Forest performed better than MLPR	Not available
[17]	2022	PM2.5 concentrations and meteorological data from Beijing	Stacked ResNet-LSTM and CORAL model	RMSE:40.679, MAE:23.746 ,R^2 : 0.804	Not available
[11]	2022	Air Quality Open Data	ANFIS ,trained with GA, PSO, DE	Better than classical ANFIS	Membership functions, GA/PSO/DE parameters

Table 2.4: Literature Survey

Chapter 3

Methods and Materials

This section outlines the methodology employed in the research and details the materials used during the investigative process. The study leverages an extensive dataset that encompasses meteorological and air quality parameters. These datasets are sourced in real-time from prominent APIs, notably the India Meteorological Department (IMD), the Central Pollution Control Board (CPCB) and Open-Meteo Web Server.

By adopting a dual-sourced strategy, the research aims to establish a comprehensive understanding of the intricate relationship between meteorological conditions and air quality indices across diverse geographical locations. The inclusion of API data enriches the dataset, providing a nuanced perspective on environmental dynamics. The subsequent sections delve into the specific methods applied and the materials utilized for a thorough exploration of the research objectives [8, 14].

3.1 Proposed Framework

The methodology commences by sourcing raw data through the Open-Meteo API [18], forming the foundational dataset. This data undergoes a structured preprocessing phase comprising three pivotal steps. Initially, missing values within the dataset are handled through mean imputation, replacing absent values with the mean of the entire dataset to ensure data integrity. Following this, outlier elimination takes place utilizing the Z-score method, effectively identifying and removing anomalies that could skew the dataset. Subsequently, normalization is performed via min-max normalization, standardizing the data to a uniform scale, thereby preparing it for further processing [4].

Upon completion of preprocessing, a feature selection phase ensues, leveraging a correlation-based approach. This strategy evaluates the correlation values among features and the target attribute, identifying pertinent features essential for the predictive model's evaluation. The processed data is then partitioned into training and testing subsets. The training data is fed into the Deep Learning (DL) model for learning, while the testing data is utilized for validating the DL model's predictive capabilities. Ultimately, the model generates air quality index (AQI) predictions, subsequently assessed for performance using evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-Squared to gauge its accuracy and effectiveness in forecasting air quality. The Flowchart for the proposed System Model is below in Figure 3.13.

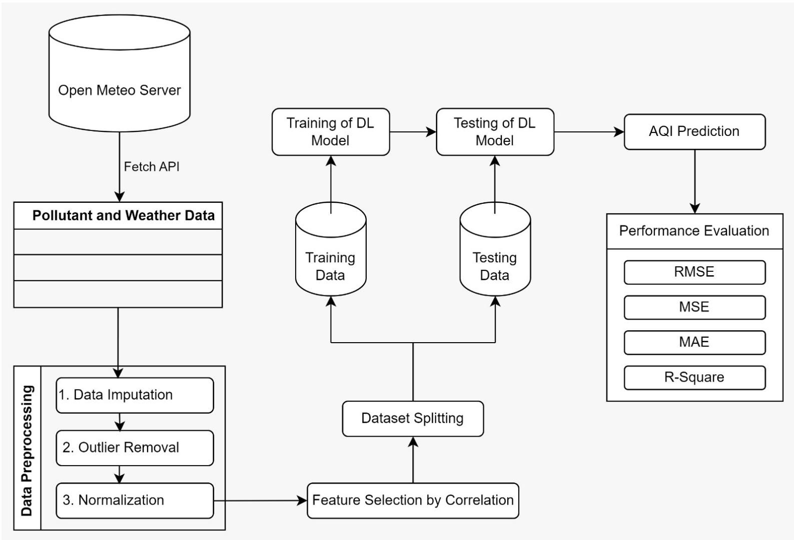


Figure 3.1: Proposed Framework

3.1.1 API

Application Programming Interfaces (APIs) are critical in enabling communication between different software systems [19]. In this project, APIs are used to fetch real-time AQI and meteorological data from various sources. These APIs facilitate seamless integration between the mobile application and the cloud-based LSTM model. The API endpoints are designed to receive input data, process it through the trained model, and return the predicted AQI values to the mobile application.

The API used in this project adheres to RESTful principles, ensuring stateless communication and scalability. Each API call is authenticated and authorized to prevent unauthorized access and ensure data security. The response times of the API are optimized to provide real-time feedback to users, enhancing the application's usability[19, 20].

3.1.2 Dataset Collection

The dataset for this research is extensive, capturing both meteorological and air quality parameters. It is enriched with real-time data from APIs, including sources like the India Meteorological Department (IMD) [21] and the Central Pollution Control Board (CPCB) [22]. This dual-sourced approach ensures a robust data-set, offering valuable insights into the relationship between meteorological conditions and air quality indices across diverse locations. The inclusion of API data enhances the dataset's depth, contributing to a nuanced understanding of environmental dynamics[23]. The attributes include:

Feature	**Description**
city_name	Name of the city
country_code	Country code
lat	Latitude
lon	Longitude
state_code	State code
timezone	Timezone

Table 3.1: Location Information

Feature	**Description**
aqi	Air Quality Index
co	Carbon Monoxide concentration
no2	Nitrogen Dioxide concentration
o3	Ozone concentration
pm10	Particulate Matter (PM10)
pm25	Particulate Matter (PM2.5)
so2	Sulfur Dioxide concentration

Table 3.2: Air Quality Indices

Feature	**Description**
datetime	Local time
timestamp_local	Local timestamp
timestamp_utc	UTC timestamp
ts	General timestamp

Table 3.3: Timestamps

Feature	Description
temperature_2m	Temperature at 2 meters above ground
relativehumidity_2m	Relative Humidity at 2 meters above ground
dewpoint_2m	Dew point at 2 meters above ground
apparent_temperature	Apparent Temperature
precipitation	Precipitation
rain	Rainfall
snowfall	Snowfall
pressure_msl	Mean Sea Level Pressure
surface_pressure	Surface Pressure
cloudcover	Cloud Cover
cloudcover_low	Low Cloud Cover
cloudcover_mid	Mid-level Cloud Cover
cloudcover_high	High Cloud Cover
et0_fao_evapotranspiration	Reference Evapotranspiration (FAO Penman-Monteith)
windspeed_10m	Wind Speed at 10 meters above ground
windspeed_100m	Wind Speed at 100 meters above ground
winddirection_10m	Wind Direction at 10 meters above ground
winddirection_100m	Wind Direction at 100 meters above ground
windgusts_10m	Wind Gusts at 10 meters above ground

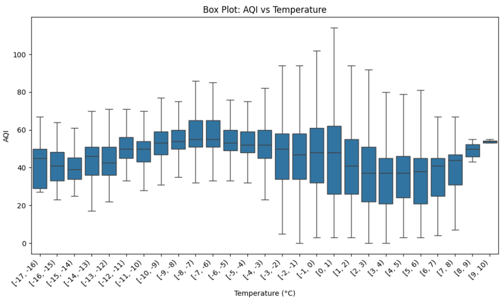
Table 3.4: Meteorological Parameters

3.2 Data Visualisation

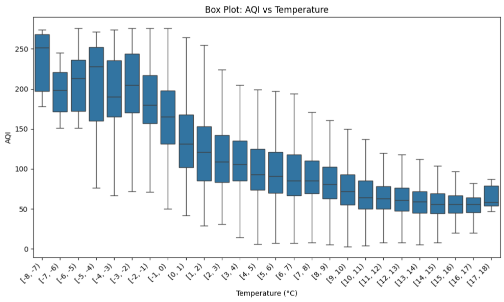
Research employs data visualization techniques to objectively analyze the correlation between Air Quality Index (AQI) and meteorological parameters[16].

Utilizing datasets from India Meteorological Department, Open -Meteo and CPCB [18, 21, 22], our visualizations provide clear insights into the relationship between temperature, humidity, and AQI variations across Sikkim. These findings contribute to a comprehensive understanding of air quality dynamics for informed decision-making in environmental management.

3.2.1 AQI vs Temperature



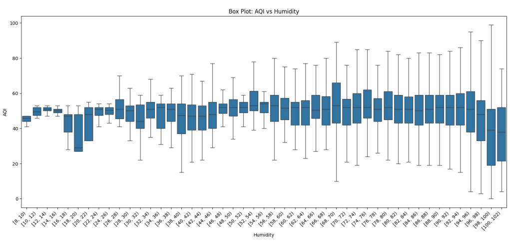
((a)) North Sikkim



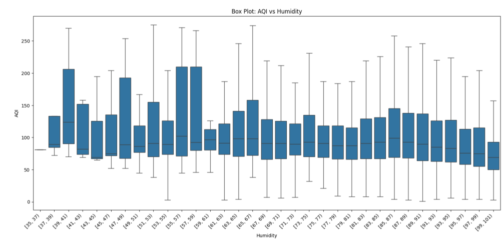
((b)) South Sikkim

Figure 3.2: Comparison of AQI vs Temperature for North and South Sikkim

3.2.2 AQI vs Humidity



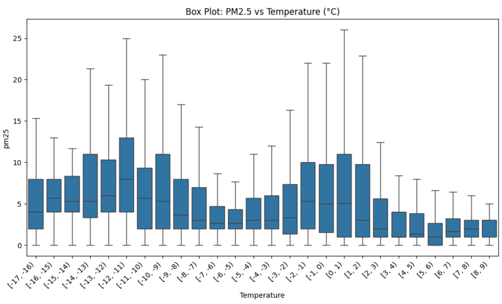
((a)) North Sikkim



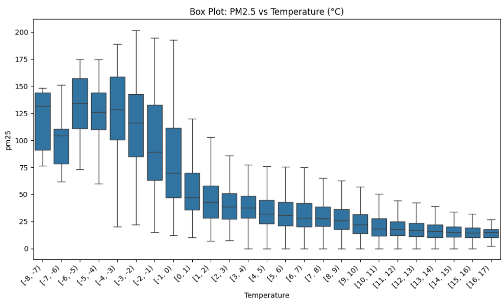
((b)) South Sikkim

Figure 3.3: Comparison of AQI vs Humidity for North and South Sikkim

3.2.3 PM2.5 vs Temperature



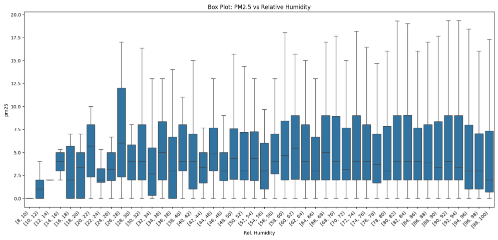
((a)) North Sikkim



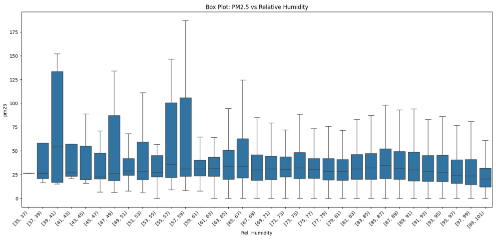
((b)) South Sikkim

Figure 3.4: Comparison of AQI vs PM2.5 for North and South Sikkim

3.2.4 PM2.5 vs Humidity



((a)) North Sikkim



((b)) South Sikkim

Figure 3.5: Comparison of PM2.5 vs Humidity for North and South Sikkim

3.3 Data Preprocessing

Data pre-processing is a critical step in ensuring the quality and reliability of datasets, providing a solid foundation for meaningful analyses and accurate interpretations. The following essential steps are undertaken in this pivotal phase:

- **Data Imputation:** Addressing missing values through strategic techniques, such as mean or median imputation, to create a complete and informative dataset.
- **Outlier Removal:** Enhancing dataset reliability by identifying and mitigating the impact of anomalies. Outliers O can be detected using statistical methods like the interquartile range (IQR). Data points outside the range $Q1 - 1.5 \times \text{IQR}$ to $Q3 + 1.5 \times \text{IQR}$ are considered outliers and removed.
- **Normalization:** Standardizing data scales using Min-Max scaling. For a feature x in a dataset, Min-Max scaling is represented as:

$$\text{Normalized}(x) = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Alternatively, Z-score normalization can be used:

$$\text{Normalized}(x) = \frac{x - \mu}{\sigma}$$

where μ is the mean and σ is the standard deviation of the feature x.

AQI	CO	NO_2	O_3	PM_{10}	$PM_{2.5}$	SO_2
0.305	0.199	0.052	0.205	0.193	0.329	0.102
0.325	0.499	0.462	0.825	0.368	0.102	0.199
0.303	0.178	0.047	0.303	0.178	0.047	0.367
0.527	0.156	0.042	0.527	0.156	0.042	0.846
0.497	0.135	0.037	0.497	0.135	0.037	0.234

Table 3.5: Normalized Data Format

By meticulously implementing these pre-processing steps, we optimize the dataset's integrity, mitigating potential biases and outliers, thus enhancing the robustness of our research results and outcomes.

3.4 Feature Selection

Feature selection is pivotal in refining models for predicting Air Quality Index (AQI) based on indices and meteorological data. The objective is to identify key features influencing AQI prediction, enhancing model efficiency and interpretability[4]. Relevant features include CO, NO_2, O_3, PM10, PM2.5, SO_2, and meteorological parameters. The process involves examining correlations and statistical significance, strategically retaining features crucial for accurate predictions. This methodology not only improves model accuracy but also deepens understanding of intricate relationships between air quality indices and meteorological parameters.

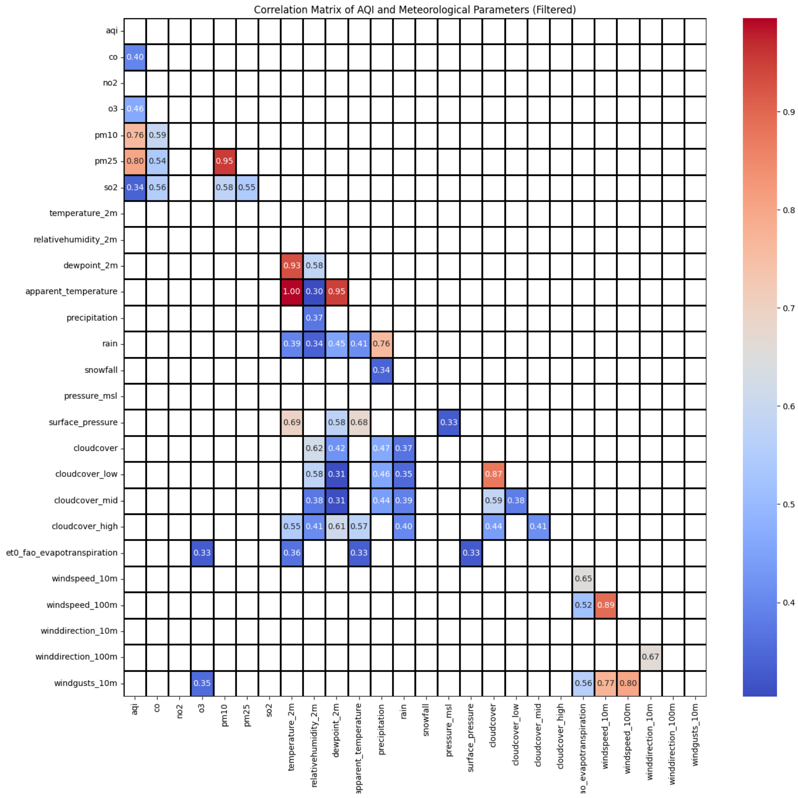


Figure 3.6: Correlation Matrix of AQI and Meterological Parameters

3.4.1 Why Feature Selection

Feature selection is imperative in the context of predicting Air Quality Index (AQI) based on indices and meteorological data. By strategically choosing relevant features such as CO, NO_2, O_3, PM10, PM2.5, SO_2, and key meteorological parameters, we streamline the machine learning model. This not only improves model interpretability but also mitigates the curse of dimensionality, leading to enhanced model efficiency and generalization. By focusing on essential variables, we aim to uncover the intricate relationships between air quality indices and meteorological factors, facilitating a more accurate and insightful research outcome[13].

3.4.2 Feature Importance Score

In the realm of predicting Air Quality Index (AQI) based on indices and meteorological data, feature importance scores play a pivotal role. These scores quantify the contribution of each variable, such as CO, NO_2, O_3, PM10, PM2.5, SO_2, and key meteorological parameters, towards the predictive performance. Utilizing techniques like Random Forest or Gradient Boosting, we obtain feature importance scores, guiding us in selecting the most influential variables for our predictive model. This meticulous approach ensures that our model captures the essential relationships, leading to a more robust and accurate prediction of AQI.

Feature	Importance
dewpoint_2m	0.190760
evapotranspiration	0.085652
winddirection_100m	0.076235
surface_pressure	0.072456
apparent_temperature	0.068297
pressure_msl	0.060347
winddirection_10m	0.056761
cloudcover_mid	0.050835
windgusts_10m	0.047888
windspeed_100m	0.045710
temperature_2m	0.040836
windspeed_10m	0.040754
cloudcover_high	0.038808
cloudcover_low	0.037241
relativehumidity_2m	0.034210
cloudcover	0.025479
precipitation	0.011978
snowfall	0.011037
rain	0.004717

((a)) North Sikkim

Feature	Importance
apparent_temperature	0.338034
dewpoint_2m	0.126070
winddirection_100m	0.088331
pressure_msl	0.049219
winddirection_10m	0.049044
temperature_2m	0.047387
surface_pressure	0.038936
cloudcover_mid	0.032912
windgusts_10m	0.029869
cloudcover_high	0.028563
relativehumidity_2m	0.026315
windspeed_100m	0.025302
cloudcover_low	0.023254
evapotranspiration	0.019466
windspeed_10m	0.019457
cloudcover	0.017666
precipitation	0.016889
rain	0.013315
snowfall	0.009972

((b)) South Sikkim

Table 3.6: Feature Importance Scores for North and South Sikkim using Random Forest model

3.5 Evaluation Parameters

- **MSE:** Calculates the average squared difference between predicted and actual values

  MSE = $\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

- RMSE = Measures the average magnitude of the difference between predicted ($\hat{y}_i$) and actual values (y_i) after taking the square root of the mean squared error (MSE)

  RMSE = $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$

- MAE = Represents the average absolute difference between predicted and actual values, providing a less sensitive metric to outliers compared to MSE.

$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$

- $\mathbf{R}^2$ (Coefficient of Determination): R-squared measures the proportion of variance in the dependent variable explained by the independent variable(s) in a regression model.
  $$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$
  - In the above formulas:
  - y_i: Actual value for observation i
  - $\hat{y}_i$: Predicted value for observation i
  - $\bar{y}$: Mean of the actual values
  - n: Number of observations

3.6 Proposed Model

In our proposed work, we have made a Sequential LSTM model in order to deal with time series data efficiently.

3.6.1 Long Short-Term Memory(LSTM)

The LSTM model is a type of recurrent neural network (RNN) capable of learning long-term patterns. LSTMs are designed to avoid long-term dependency problems, which are a common issue in traditional RNNs[5, 17]. They can maintain information over extended periods, making them suitable for time-series forecasting tasks like AQI prediction.

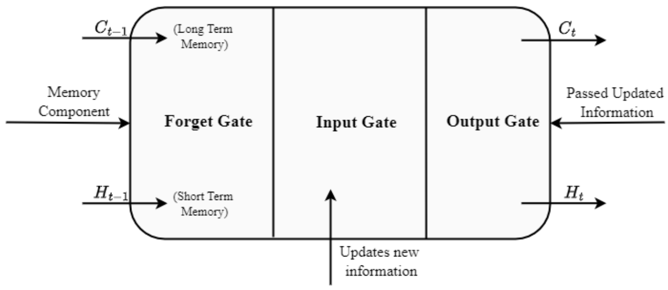


Figure 3.7: Architecture of LSTM Model

- **Forget Gate:** Decides what past information to keep.

- **Input Gate:** Controls new information flow into the cell state.
- **Cell State:** The memory unit that carries information through time steps.
- **Output Gate:** Determines what information from the cell state affects the output.

3.6.2 Advantages of LSTM

The choice of LSTM for AQI prediction is driven by its ability to handle sequential data and capture temporal patterns effectively. Air quality data is inherently sequential, with each reading influenced by previous measurements and meteorological conditions[5]. LSTM networks can learn these dependencies and provide accurate predictions by maintaining long-term context.

3.6.3 Model Architecture

Layer (type)	**Output Shape**	**Parameters**
LSTM (LSTM)	(None, 24, 64)	20,736
Dropout (Dropout)	(None, 24, 64)	0
LSTM 1 (LSTM)	(None, 24, 64)	33,024
Dropout 1 (Dropout)	(None, 24, 64)	0
LSTM 2 (LSTM)	(None, 32)	12,416
Dropout 2 (Dropout)	(None, 32)	0
Dense (Dense)	(None, 32)	1,056
Dropout 3 (Dropout)	(None, 32)	0
Dense 1 (Dense)	(None, 1)	33

Table 3.7: Sequential LSTM Model Architecture

- Total parameters: 201,797
- Trainable parameters: 67,265
- Non-trainable parameters: 0
- Optimizer parameters: 134,532

3.7 Mobile App Integration

The app collects AQI and weather data for a user-selected region and city through APIs. The working is as follows:

- **Data Collection:** Users select their desired region and city within the app. The app then retrieves real-time AQI and weather data from various reliable sources using APIs. This includes data on pollutants and meteorological data.

- **Data Processing and Transmission:** The collected data is pre-processed within the app to ensure compatibility and accuracy. Once processed, this combined data is sent to the Sequential LSTM model hosted on the cloud.

- **Prediction Retrieval:** The cloud-hosted Sequential LSTM model processes the input data to predict future AQI levels. The model's architecture, designed for time-series forecasting, efficiently handles the sequence of data points to generate accurate predictions.

- **User Display:** The predicted AQI values are sent back to the app, which then displays it to the user through a user-friendly interface.

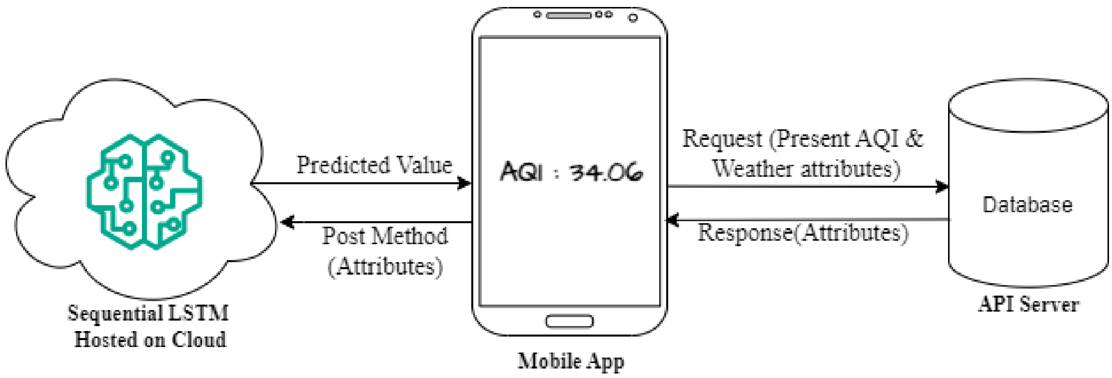


Figure 3.8: Mobile App Integration with Sequential LSTM

Chapter 4

Result and Analysis

4.1 Experimental Setup

In this, dataset from the Sikkim region was utilized. which includes Air Quality Index (AQI) and meteorological data. This dataset spanned a duration of two years using various APIs [18,21,22] providing hourly data entries. Consequently, each day comprises 24 data entries.

The dataset was preprocessed and divided into training and testing sets in a 7:3 ratio. The training set was used to train a sequential model, and the testing set was used to evaluate the model's performance.

4.2 Model Evaluation

- **Actual vs Predicted Graph:** Below scatter plot shows Actual vs Predicted AQI using Sequential LSTM Model.

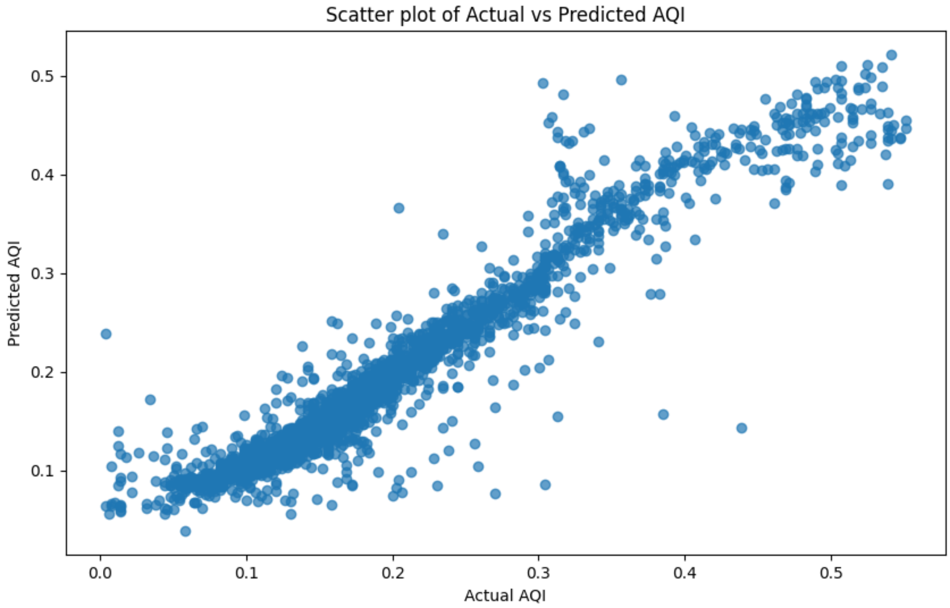


Figure 4.1: Scatter plot of Actual vs Predicted AQI

4.2.1 Evaluation Metrics

The Sequential model underwent evaluation using RMSE, MSE, MAE, and R-squared metrics, yielding the following results:

- **RMSE Score:** 0.03082560357256409
- **MSE Score:** 0.00950217835612876
- **Mean Absolute Error:** 0.018605553604013166
- **R^2 :** 0.9188031327046687

The Root Mean Squared Error (RMSE) score, indicating the average magnitude of errors between predicted and actual values, was found to be 0.0308. The Mean Squared Error (MSE) score, measuring the average of squared errors, stood at 0.0095. Additionally, the Mean Absolute Error (MAE) was calculated to be 0.0186, representing the average absolute difference between predicted and actual values. Notably, the coefficient of determination (R-squared) was 0.9188, indicating that approximately 91.88% of the variance in the dependent variable is explained by the independent variable(s) in our regression model. Overall, these metrics suggest that our Sequential model demonstrates strong predictive performance and effectively captures the relationship between the variables under consideration.

4.3 Comparison of Proposed Architecture with Different Approaches and Models

Model Name	MSE	RMSE	MAE	R^2
Decision Tree	NA	38.852	28.159	71.3
Random Forest	NA	27.243	18.678	85.9
SVR	NA	25.302	16.853	47.826
MLP	NA	25.349	16.853	48.696
ANFIS	1.3	0.117	NA	74.7
KNN	NA	4.023	0.834	45.3
GNB	NA	3.487	0.564	38.2
SVM	NA	3.803	NA	62.3
XGBoost	NA	1.027	NA	83.4
Sequential LSTM	0.9502	3.08256	1.8605	91.8803

Table 4.1: Comparitive Analysis with other approaches

The table above presents a comparative analysis of various models, highlighting the performance metrics of each. Our proposed architecture, the Sequential LSTM, demonstrates superior performance across multiple evaluation metrics, including the lowest RMSE and MSE scores, and the highest R^2 value. This underscores its effectiveness and accuracy in AQI prediction when compared to traditional and contemporary models.

4.4 Result of Proposed Architecture

The proposed Sequential LSTM architecture achieved remarkable results in AQI prediction, showcasing its robustness and precision. The model recorded an exceptionally low RMSE score of 0.0308, indicating minimal deviation between predicted and actual values. Additionally, it achieved an MSE score of 0.0095, reflecting its high accuracy. The Mean Absolute Error (MAE) stood at 0.0186, further demonstrating the model's reliability. Most notably, the model attained an R^2 value of 0.9188, illustrating a high degree of correlation between the predicted and actual AQI values. These metrics collectively highlight the superior performance of our proposed architecture.

Chapter 5

Conclusions and Future Works

5.1 Conclusions

In conclusion, the utilization of sequential LSTM (Long Short-Term Memory) models for Air Quality Index (AQI) prediction signifies a significant advancement in environmental research and public health management. Through this study, it becomes apparent that LSTM networks, with their ability to capture long-term dependencies and temporal patterns in sequential data, offer a robust framework for accurate AQI forecasting.

Integrating the LSTM-based AQI prediction model into a mobile app is a transformative step in empowering users to access vital air quality information conveniently and efficiently. With this app, users can gain valuable insights into the projected AQI levels of specific regions, enabling them to plan activities, travel routes, and outdoor engagements with greater awareness and consideration for their health and well-being.

5.2 Future Scope

The present research lays a solid foundation for future investigations and advancements in the domain of air quality prediction and environmental monitoring. Several promising avenues for future work emerge from the current study:

- **Integration of Additional Data Sources:** Incorporating data from diverse sources, such as satellite imagery, social media, and urban planning data, can provide a more comprehensive understanding of the factors influencing air quality. This expanded dataset could enhance the accuracy and granularity of predictive models.

- **Ensemble Models and Hybrid Approaches:** Exploring the potential of ensemble models or hybrid approaches that combine the strengths of different machine learning techniques could lead to enhanced predictive performance. This could involve integrating deep learning architectures with traditional machine learning algorithms.

- **Integration of IoT:** Use of IoT devices, like affordable air quality sensors, offers an oppurtunity to improve air quality monitoring. The IoT devices can be placed widely in cities and countryside areas, providing real-time and detailed data on the environment which can improve predictive models, making forecasts more accurate and timely.

References

[1] S. Al-Janabi, M. Mohammad, and A. Al-Sultan, "A new method for prediction of air pollution based on intelligent computation," *Soft Computing*, vol. 24, no. 1, pp. 661–680, 2020.

[2] C. WEERATHUNGHE and R. BALASUBRAMANIAN, "Assessment and abatement measures for vehicular air pollution in colombo, sri lanka," in *Sustainability Matters: Asia's Green Challenges*. World Scientific, 2013, pp. 35–71.

[3] K. Kumar and B. Pande, "Air pollution prediction with machine learning: a case study of indian cities," *International Journal of Environmental Science and Technology*, vol. 20, no. 5, pp. 5333–5348, 2023.

[4] A. Rahimpour, J. Amanollahi, and C. G. Tzanis, "Air quality data series estimation based on machine learning approaches for urban environments," *Air Quality, Atmosphere & Health*, vol. 14, pp. 191–201, 2021.

[5] M. Krishan, S. Jha, J. Das, A. Singh, M. K. Goyal, and C. Sekar, "Air quality modelling using long short-term memory (lstm) over nct-delhi, india," *Air Quality, Atmosphere & Health*, vol. 12, pp. 899–908, 2019.

[6] Z. Jiang, S. Yang, X. Chen, Q. Pang, Y. Xu, S. Qi, W. Yu, and H. Dai, "Controlled release urea improves rice production and reduces environmental pollution: A research based on meta-analysis and machine learning," *Environmental Science and Pollution Research*, vol. 29, pp. 3587–3599, 2022.

[7] I. Yeo, Y. Choi, Y. Lops, and A. Sayeed, "Efficient pm2. 5 forecasting using geographical correlation based on integrated deep learning algorithms," *Neural Computing and Applications*, vol. 33, no. 22, pp. 15 073–15 089, 2021.

[8] S. Malleswari and T. K. Mohana, "Air pollution monitoring system using iot devices," *Materials Today: Proceedings*, vol. 51, pp. 1147–1150, 2022.

[9] Z. Zhang, Y. Zeng, and K. Yan, "A hybrid deep learning technology for pm 2.5 air quality forecasting," *Environmental Science and Pollution Research*, vol. 28, pp. 39 409–39 422, 2021.

[10] J. Ma, Z. Li, J. C. Cheng, Y. Ding, C. Lin, and Z. Xu, "Air quality prediction at new stations using spatially transferred bi-directional long short-term memory network," *Science of The Total Environment*, vol. 705, p. 135771, 2020.

[11] A. Yonar and H. Yonar, "Modeling air pollution by integrating anfis and metaheuristic algorithms," *Modeling Earth Systems and Environment*, vol. 9, no. 2, pp. 1621–1631, 2023.

[12] Q. Zhang, F. Fu, and R. Tian, "A deep learning and image-based model for air quality estimation," *Science of The Total Environment*, vol. 724, p. 138178, 2020.

[13] Q. Pan, F. Harrou, and Y. Sun, "A comparison of machine learning methods for ozone pollution prediction," *Journal of Big Data*, vol. 10, no. 1, p. 63, 2023.

[14] H. Mokrani, R. Lounas, M. T. Bennai, D. E. Salhi, and R. Djerbi, "Air quality monitoring using iot: A survey," in *2019 IEEE International Conference on Smart Internet of Things (SmartIoT)*. IEEE, 2019, pp. 127–134.

[15] J. Wang, W. Xu, J. Dong, and Y. Zhang, "Two-stage deep learning hybrid framework based on multi-factor multi-scale and intelligent optimization for air pollutant prediction and early warning," *Stochastic Environmental Research and Risk Assessment*, vol. 36, no. 10, pp. 3417–3437, 2022.

[16] H. Liu and R. Yang, "A spatial multi-resolution multi-objective data-driven ensemble model for multi-step air quality index forecasting based on real-time decomposition," *Computers in Industry*, vol. 125, p. 103387, 2021.

[17] X. Cheng, W. Zhang, A. Wenzel, and J. Chen, "Stacked resnet-lstm and coral model for multi-site air quality prediction," *Neural Computing and Applications*, vol. 34, no. 16, pp. 13 849–13 866, 2022.

[18] O. Meteo, "Air quality data," https://open-meteo.com/, 2024, accessed: 2023-08-01.

[19] S. Sachdeva, R. Kaur, Kimmi, H. Singh, K. Aggarwal, and S. Kharb, "Meteorological aqi and pollutants concentration-based aqi predictor," *International Journal of Environmental Science and Technology*, vol. 21, no. 5, pp. 4979–4996, 2024.

[20] A. Z. Ikromovna, “Programming environments for creating mobile applications on the android operating system,” *American Journal of Public Diplomacy and International Studies (2993-2157)*, vol. 1, no. 10, pp. 305–309, 2023.

[21] I. M. Department, “Weather data,” https://mausam.imd.gov.in/, 2024, accessed: 2023-08-20.

[22] C. P. C. Board, “Air quality data,” https://cpcb.nic.in/, 2024, accessed: 2023-07-26.

[23] M. T. Udristioiu, Y. E. Mghouchi, and H. Yildizhan, “Prediction, modelling, and forecasting of pm and aqi using hybrid machine learning,” *Journal of Cleaner Production*, vol. 421, p. 138496, 2023.