Regional Clustering, Time Series Forecasting, and Satellite-to-Ground Data Mapping for Environmental Analysis in Bangladesh Using Explainable AI

by

Thufa Anwar Tamanna
ID: 21241080

Shah Newaz Khan Joy
ID: 21201591

Zayed Masum
ID: 22101461

Fariha Binte Abdullah
ID: 23241050

S M Kafi Anam
ID: 24241280

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
BRAC University
February 6, 2026

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at BRAC University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Students' Full Names & Signatures:

S M Kafi Anam
24241280

Thufa Anwar Tamanna
21241080

Shah Newaz Khan Joy
21201591

Zayed Masum
22101461

Fariha Binte Abdullah
23241050

Approval

The thesis titled **"Regional Clustering, Time Series Forecasting, and Satellite-to-Ground Data Mapping for Environmental Analysis in Bangladesh Using Explainable AI"** submitted by

1. Thufa Anwar Tamanna (ID: 21241080)
2. Shah Newaz Khan Joy (ID: 21201591)
3. Zayed Masum (ID: 22101461)
4. Fariha Binte Abdullah (ID: 23241050)
5. S M Kafi Anam (ID: 24241280)

of Fall 2025 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science in Fall 2025.

Examining Committee:

Supervisor:
(Member)

Dr. Chowdhury Mofizur Rahman
Professor, Dept. of CSE, BRAC University

Co Supervisor:
(Member)

Dewan Ziaul Karim
Senior Lecturer, Dept. of CSE, BRAC University

Program Coordinator:
(Member)

Dr. Md. Golam Rabiul Alam
Professor, Dept. of CSE, BRAC University

Head of Department:
(Chair)

Dr. Sadia Hamid Kazi
Associate Professor and Chairperson
Dept. of CSE, BRAC University

Abstract

Air pollution and changes in the weather make it hard to keep an eye on the environment and protect public health, especially in areas where there aren't many ground-based sensors or they are spread out unevenly. Ground measurements give accurate local readings, but they fail to encompass much of ground data.

Spatial coverage limits analysis on a regional scale. Satellite observations provide extensive spatial data but lack local accuracy. This thesis suggests a complete multimodal machine learning framework that combines satellite-derived spatial features, ground-based environmental measurements, regional clustering, and time-series forecasting to better predict and understand air quality and important weather variables.

The suggested framework uses gradient-boosted regression models for multimodal learning and pre-trained convolutional neural networks (ResNet-18, ResNet-50, and DenseNet-121) to get features from satellite images. Time-series forecasting is used to find patterns over time and changes in environmental factors that happen at different times of the year. Unsupervised regional clustering is used to find groups of districts that have similar pollution and weather patterns. The framework is assessed against various air pollutants (SO, NO_2, O_3, $PM_{2.5}$, PM_{10}) and meteorological factors (solar radiation, relative humidity, and rainfall) utilizing standard performance metrics, such as RMSE, MAE, and R^2. The experimental results demonstrate the effectiveness of multi-modal fusion is highly dependent on the target and region. For variables that were affected, like sulfur dioxide and rainfall, explained variance went up a lot and prediction error went down a lot. On the other hand, O_3 and solar radiation got a little better. Again, NO_2, particulate matter, and relative humidity were best modeled using only ground-based data. This suggests that there are many localized emission sources and processes that happened close to the ground. To improve the transparency of the model, explainable artificial intelligence (XAI) methods were added through global feature importance analysis with XGBoost. This gave us a better idea of what environmental factors were most important for making predictions over time. The results underscore the significance of pollutant- and region-specific modeling approaches and illustrate that multimodal learning can enhance environmental prediction when informed by the physical and statistical attributes of the target variables.

Keywords: Machine Learning, Climate Change, Environmental monitoring, Satellite Imagery, Ground-Based information, Regional groupings, Time Series Forecasting, Bangladesh

Contents

Chapter 1
Introduction

1.1 Background

The 21st century is characterized by challenges never experienced before in the world, and at the top of the list are environmental degradation and climate change. Over the past few years, environmental surveillance and research have become even more essential due to fast urbanization, global warming, and poor air quality, particularly in developing countries in the world, such as Bangladesh. Being one of the most densely populated countries in the world and producing a delta on the river, Bangladesh is particularly susceptible to both types of risks that global warming poses: local pollution concentrations and the impact of climate change as a whole [1].

The megacity of the developing world that has epitomized the strains of the environment is Dhaka, the capital of Bangladesh. It is regularly listed among the most polluted cities in the World, and the quality of the air is often very harmful [2].The combination of ground station and space-derived measurements of the environmental parameters can be very instrumental in drawing sustainable city planning and policymaking. The leading pollutants of the air are emissions of brick kilns, auto-exhaust, industrial effluent, and transboundary air pollution, and the biggest concern in the area of the health of the people is Particulate Matter ($PM_{2.5}$). In addition to the air, the surface water bodies in Dhaka, including the rivers 'Buriganga', 'Turag', 'Shitalakshya' are extremely polluted through direct discharge of the untreated effluents of the industries and municipal sewage, which have turned biologically dead in many parts. [3].

New breakthroughs in machine learning (ML) and deep learning (DL) have developed potential resources of environmental data analysis. However, as useful as they are, traditional environmental monitoring approaches typically have problems of spatial coverage, temporal resolution, and the capacity to detect the interaction of complex environmental processes. It can be possible by using clustering techniques to find similar regions with similar environmental conditions and time series forecasting to draw out the tentative trends to follow in the future and satellite-to-ground data mapping to intensify large-scale environmental surveillance.

The access to satellite imagery provided on site, such as Google Earth Engine has provided an opportunity to capture environmental characteristics in a large spatial extent and high temporal spatial resolution. Nevertheless, the interpretation of such models is not trivial, and that is where Explainable AI (XAI) comes into the picture and works to enable interpretation and actionability of complex models using stakeholders. The purpose of this thesis is to combine the methods of classical and deep learning and study the environmental data in Bangladesh, focusing on Dhaka and other big cities. We have the opportunity to use the ground-based data and the satellite view to build an inclusive model which can be used in guiding the evaluation and forecasting of the environment coupled by the interpretability enabled by XAI.

Machine learning, especially deep learning and explainable artificial intelligence (XAI), provides unprecedented opportunities both to analyze environmental data comprehensively and make solid predictions on it.

1.2 Motivation

The proposed research will be inspired by the need to have a more generalized, scalable and interpretable monitoring and analysis system of the environment in Bangladesh. The present paradigm which is based on limited ground data is not sufficient to guide the complex environmental crisis that is cast upon the urban centers of this country. Besides, though some studies have applied machine learning and deep learning, it is always important to select the best machine learning algorithm so as to train the deep learning model. Some papers have analyzed the environmental components but none of them have tackled region-wise clustering in managing the environment locally nor have they forecast the future of environmental conditions at a city level or developed interpretable mappings between satellite images and ground information.

It is not easy to translate satellite observations to imagery owing to the widespread spatial coverage therefore real-life applicable environmental conditions. This satellite analyzing mismatch is the basis of the contradiction between the satellite and terrestrial analyses of the region. Also,the use of remote sensing is limited by ground truth measurements information.The development of sensing in environmental monitoring in Bangladesh would be largely enhanced by integrating satellite sensing with the monitoring procedure.

In Bangladesh, there is a provision of high-fidelity temporal data that gives spatial coverage by ground stations. Satellite data, on the other hand, has a huge spatial scale that can afford to cover large areas and may be translated into meaningful and ground-level indicators of the environment through intricate modeling. Such a study is driven by the necessity to combine the two data inputs in a synergetic way to produce a comprehensive environmental image, not only, as wide-ranging as it is, but also precise in resolution.

Classical ML models provide good solid baselines, whereas deep learning models (like LSTMs on such as time series and CNNs on imagery have been found to work towards management of complex and non-linear behavior which is a characteristic of environmental information. The research under proposal is caused by the opportunity to compare these methods in the organized manner and determine the most effective instruments when it comes to the specific conditions of Bangladesh.

Moreover, a regional clustering related to environmental characteristics can guide specific policy interventions and resource directions. Different parts of Bangladesh may have various environmental characteristics depending on aspects like industrial works, population, closeness to water, and how the land is used. The knowledge of such regional agglomerates can be used to create specific environmental management practices, taking into consideration the local services and requirements. The time series forecasting module offers the urgent requirement to be able to predict environmental knowledge in order to enable a proactive environmental management.

Additionally, there is the fact that most of the current models are what we call black boxes that give little or no information on how they come up with their decisions. Such models cannot be readily applied in environmental policy formation as a result of this lack of interpretability. In addition to developing predicting models, we are interested in making their outcomes reasonable and predictive through Explainable AI (XAI), in order to develop a correct value to the policymakers, environmentalists, etc. and the urban planners in Bangladesh. XAI can provide details of the most significant environmental parameters in different parts and times which would be submitted to allow more effectual interventions on the environment based on an increased degree of specificity.

1.3 Problem Statement

Although the environment information on the ground and satellite images can be obtained very easily it has been observed that there is a vacuum in the efficient integration of these data systems in a bid to provide accurate, interpretable and usable environmental information in Bangladesh. The current strategies tend not to deliver spatial clustering analysis that can give regional differences and credible time series forecasts that would make future environmental conditions and development of fit-to-purpose mapping techniques The ones that would transform satellite pictures to practical information on-ground.

Besides, the prediction of the deep learning models is not transparent which complicates it too. Such gaps are aimed to be filled through this thesis with the purpose of developing a single view that entails clustering, prediction and mapping of image to data with the assistance of explainable AI (XAI) to deliver deep learning structures transparency.

1.4 Objective

The main objective of the given thesis is to formulate, test, and analyze an in-depth data-driven model based upon both traditional and deep learning models that allow a multi-dimensional assessment of environmental situations in Bangladesh.To achieve this aim, the following specific objectives are set:

Clustering of Regions: To group the areas of Bangladesh, in particular cities, on the basis of multi-dimensional ground data on the environment to identify comparable zones of the environment. Use classical K-means clustering, in order to identify the optimal regional combinations. Complete confirmation of clustering results through knowledge of the environment area and measurements.

Time Series Forecasting: Model and implement detailed time series forecasting of environment parameters using classical (SARIMA) and deep learning (LSTM). Gather performance of forecasting in terms of short, medium and long-term predictions. **Satellite-to-Ground Data Mapping:** At first, retrieval and processing of accessible satellite pictures(e.g. The entire Bangladesh covered with Sentinel-2, Sentinel-5P, MODIS) is taken through Google Earth Engine. Then, training a DL model (e.g., CNN, vision transformer) that connects satellite imagery data with ground sensor data . Develop a transferable model that can be applied in prediction of ground-level environmental parameters of future satellite imagery.

Explainable AI Integration: To apply the state of the art XAI techniques (e.g., SHAP, LIME, Integrated GRAD) on the model, chosen on the basis to find out what characteristics determine the prediction of future environmental components and which visual indications were represented in the satellite pictures that are linked, in order to provide meaningful data to policy makers.

Comprehensive Evaluation and Validation: Critically validate all models established based on proper statistical measures and domain-level assessment measures. Compare the functioning of both the classical and deep learning system. Evaluate the feasibility or the use of the system developed through a case study and through the feedback of the other stakeholders.

1.5 Methodology in Brief

The study follows a thorough mixed-methodology whereby classical machine learning, deep learning, and explainable AI methods are combined to handle the four key elements of analysis. The methodology aims at utilizing the synergetic mixture of the strengths of various methods of analysis so that the outcomes are interpretable and practically applicable.

Data Collection and Preprocessing

Ground Data Collection: From DoE Bangladesh, the ground based environment data was collected. A set of parameters were measured: air quality ($PM_{2.5}$, PM_{10}, NO_2, SO_2, O_3), weather conditions (temperature, humidity, precipitation, solar sun radiation), along with land use parameters. The informations were obtained from government environmental observation stations(DoE), research centers, and citizen science sources.

Satellite Data Acquisition: We got satellite images from Google Earth Engine, Landsat, Sentinel-2, and MODIS. We used these images to get information like vegetation indices (NDVI, EVI), land surface temperature, aerosol optical depth, and maps showing how land is used. After that, these datasets were aligned in space and time with ground-based estimation points.

Data Preprocessing: Adequate normalization and rigorous data cleaning methods were performed. The missing data is handled using imputation methods. The time-based spatial alignment of the satellite and the ground data will be performed using interpolation techniques.

Modeling and Analysis

Regional Clustering Analysis: K-means clustering was used. Feature engineering was performed to identify key environmental variables. Cluster validation was conducted using silhouette analysis.

Time Series Forecasting: Classical model SARIMA was employed. Deep learning model LSTM was also used.

Mapping (Satellite to Ground Data): Deep learning approach CNN will be utilized for image embedding. Different pretrained models will be implemented.

Explainability and Interpretation (XAI)

Gradient-based visual explanations like (GRAD Cam/ IG) was used for interpreting deep learning time series forecast model.

Validation and Evaluation

Clustering: Silhouette Score.

Forecasting & Mapping: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Coefficient of Determination (R^2).

Tools and Technologies

Using libraries like `pandas`,`scikit-learn`, `TensorFlow/PyTorch`, `GeoPandas`, `Captum`, and the Google Earth Engine (GEE) API, the whole workflow was done in Python.

1.6 Scopes and Challenges

1.6.1 Scope of the Study

Geographical Scope

Bangladesh is the subject of research here, and the focus of the analysis is on megacities.The satellite-to-ground data mapping part will cover all of the areas of each city to make sure that the model works in all situations and has full spatial coverage.

Temporal Scope

The study analyzes environmental data of several years to identify seasonal patterns, long-term trends, and significant events. The time periods was highly depended on the data availability, which included at least 40 months of historical data to help train and test the model.

Environmental Parameters

The study will concentrate on essential environmental indicators, encompassing air quality (particulate matter and gaseous pollutants), meteorological variables (temperature, humidity, rainfall), and land use characteristics. The choice of parameters will depend on how easy it is to get the data, how important it is to the environment, and how important it is to policy.

Methodological Scope

This study combines traditional and advanced learning methodologies.Explainable AI (XAI) techniques will be used to make sure that deep learning models are easy to understand and useful in real life.

Stakeholder Scope

The study seeks to benefit a diverse array of stakeholders, encompassing environmental scientists, policymakers, urban planners, and environmental management agencies. The explainable AI part is to make sure that all stakeholders, even those who aren't technical, can understand and use complex analytical results. It also encourages cost-effective environmental monitoring using satellite images, which can be used in addition to or instead of traditional installations on the ground.

1.6.2 Challenges

Data Quality and Availability

Ensuring data quality was a big challenge for us due to the incomplete records for the essential environmental parameters. Challenges included missing values, sensor drift, and inconsistencies in data reported by ground stations. Although strict preprocessing steps were applied, they may not completely eliminate data quality issues.

Spatio-Temporal Data Alignment

Aligning sparsely distributed, gridded, and time-integrated satellite data with point- and grid-based ground measurements is complex and error-prone. Matching satellite image timestamps with ground readings may be affected by time gaps, cloud cover, or missing records.

Model Generalization

It is hard to make sure that the ground mapping of the satellite is useful when it comes to the new and unseen data. It is the tuning between making sure that the model is accurate and has the ability to interpret it.

Computational Resources

Training deep learning models, especially Convolutional Neural Networks (CNNs), on large-scale satellite images can consume a lot of computer power. This research required access to high performance computing resources like GPUs and cloud platforms.

Validation of XAI

Statistically validating the accuracy of the XAI explanations remains a significant research challenge. This study will focus on qualitative validation by comparing XAI outputs to well-known domain knowledge.

Despite these challenges, the study is expected to make significant contributions to environmental monitoring and forecasting in Bangladesh. It promotes both traditional and modern analytical practices in environmental science. Emphasis on precision and layman term explanation will ensure that the results will be valuable to most developing countries for policy making alongside Bangladesh.

Chapter 2

Literature Review

2.1 Preliminaries

Bangladesh is a fast developing country, thus environmental control here requires ecological research and integration of advanced data related skills. Deep Learning(DL) and Machine Learning(ML) therefore emerges as highly efficient methods to analyze air and water quality(Reichstein et al., 2019). Large scale spatiotemporal data provided via satellite remote sensing can be accessible by platforms like Google Earth Engine. However, the results often lack interpretability if used alone(Gorelick et al., 2017). Again, using time series forecasting models (like ARIMA and LSTM) and traditional clustering techniques (like K-means and hierarchical clustering) have been extensively usable in environmental research, however their incorporation with Explainable AI (XAI) is still not well recognized in developing nations (Samek et al., 2021). XAI approaches (e.g. SHAP, LIME, Grad-CAM) enables boosting model transparency which are vital for environmental management for the policymakers (Arrieta et al., 2020). Hence, combining satellite-ground data, predictive modeling along with, XAI, this study extends on these frameworks to address Bangladesh's environmental issues while filling in interpretability and scalability deficiencies.

2.2 Review of Existing Research

A research conducted by Jianbang Du [4] suggested a way to use a decade's worth of hourly data on air pollution, weather, and traffic statistics from the Texas Commission on Environmental Quality (TCEQ) and Houston TranStar to predict long-term and hourly levels of air pollution in cities. The pollutants included in the data were ground-level O_3, $PM_{2.5}$, NO_2, and NO_x. For NOx concentrations, the NRMSE was 2.16% and the smallest angle of prediction error was 0.108°. The study evaluated seven machine learning regression algorithms, among which the Random Forest and XGBoost models performed best at predicting different pollutants. The suggested method made very accurate hourly forecasts for the whole year. Similarly, the research conducted by Pugliese et al. [5] research provided an overview on how Machine Learning (ML) and Deep Learning (DL) techniques can be used for hazard susceptibility mappings. The research investigated forecasts concerning air pollution, urban heat islands, floods, and landslides, highlighting the escalating significance of machine learning and deep learning methodologies owing to the enhanced accessibility of environmental data and advancements in computational power.The authors systematically reviewed 1,454 articles published between 2020 and 2023,selected using strict inclusion criteria from the Scopus and Web of Science databases. The review identified ML, DL algorithms and usage of ML and DL models such as Random Forest (RF), Light Gradient Boosting Machine (LGBM), Extreme Gradient Boosting (XGB), Support Vector Machines (SVM), and Artificial Neural Networks(ANN) as high-performing techniques for hazard susceptibility modeling. Specifically, tree based models like RF and XGB were noted for their superior performance in air pollution modeling. RF, SVM, and ANN were also frequently applied in mapping susceptibility to urban heat islands, floods, and landslides. Although ML,DL models exhibited strong predictive performance across broad geographic domains, the study highlighted concerns about the spatial consistency of susceptibility maps generated by different models, even when their accuracy metrics were similar.

The authors discussed about some of the major problems with threir research, such as not having enough spatial validation and not dealing with data imbalance issues well enough, especially in studies of flood and landslide susceptibility.

A prominent study by Berrocal et al. [6] addressed the critical challenge of ambient air pollution exposure assessment by systematically comparing statistical and machine learning (ML) methods for creating national daily maps of PM2.5 concentrations. The goal was to create a standard for how well models could predict space in areas that weren't watched and how fast they can run. The authors used a number of different methods, including Inverse Distance Weighting, Universal Kriging, Bayesian Downscaler models, Random Forest, Support Vector Regression, and Neural Networks. Their results showed that Universal Kriging and the Bayesian Downscaler worked well. The UK model used CMAQ had the smallest RMSE of 3.08 μg/m^3. The best machine learning method, RF, had an RMSE of 3.41 μg/m^3.

The study also underscored the importance of spatial dependence modeling and contributed to open science by making both the data and analytical code publicly available via GitHub. However, it did not introduce any new modeling frameworks or advanced techniques in their study.

In another study, Eman Said Al Abri [7] conducted a study utilizing advanced machine learning techniques to enhance the prediction and forecasting of atmospheric ozone levels. The research emphasized the advantages of ensemble learning models compared to conventional techniques such as Artificial Neural Networks (ANN) and Support Vector Machines (SVM), which have been prevalently utilized in this domain.The study utilized two datasets that were from Sohar University, Oman and from the Department for Environment, Food, and Rural Affairs (DEFRA) in the UK. These datasets were able to predict ozone concentrations in both space and time series. The research incorporated ensemble models, specifically Multilayer Perceptron (MLP) and SMOreg, with ensemble techniques such as Bagging, Random Forest (RF), Stacking, and Voting, executed via the WEKA toolkit. The Bagged Random Forest model provided best result with a Correlation Coefficient (CC) of 0.92 and a Mean Absolute Error (MAE) of 7.052. It beat all the other single classifiers.Also, meta-learning techniques like Bagging, Stacking, and Voting were very good at making predictions, with an average CC of 0.91 for each. The study further showed that the multivariate time series forecasting methods performed much better than univariate models, and ensemble methods was better than single-base methods. The comparative insights shows the model performance and validated the efficacy of ensemble learning for O^3 prediction, its primary emphasis was on predictive accuracy. It did not incorporated image processing or spatial object detection, thereby allowing for future research development in those areas.

The work of Cheng et al. [8] performed a spatial machine learning evaluation focusing on air pollution exposure prediction and model interpretability by introducing a novel variable importance measure. Their paper suggested a *leave-one-out quantile contrast*, an efficient method to evaluate model performance across different quantiles. This approach provides a single, meaningful measure of variable importance that yields deeper insight into spatial ML models compared to traditional accuracy metrics.

The technique was utilized in two sophisticated modeling approaches: Universal Kriging with Partial Least Squares (UK PLS) and Spatial Random Forest with Partial Least Squares (SpatRF-PL). We tested these models on two high-quality datasets: the number of ultrafine particles in Seattle and the national PM 2.5 sulfur concentration levels.

The method was applied to two advanced modeling techniques: Universal Kriging with Partial Least Squares (UK-PLS) and Spatial Random Forest with Partial Least Squares (SpatRF-PL). These models were tested on two high-quality datasets—ultrafine particle counts in Seattle and national $PM_{2.5}$ sulfur concentration levels. The results demonstrated high predictive performance, with the UK-PLS and SpatRF models achieving R^2 values of 0.81 and 0.78 respectively for ultrafine particles in Seattle, and R^2 values between 0.89 and 0.90 for national sulfur concentration prediction.

Cheng et al. illustrated that despite similar predictive accuracies, the interpretation of variable importance varied significantly between models. Their novel metric revealed subtle but critical differences in how models captured spatial and covariate relationships—differences that would remain hidden using conventional performance metrics like R^2 alone.

However, the study did not address the application of their framework to imbalanced datasets or incorporate more recent deep learning methods such as transformer-based architectures or object detection approaches. This limitation restricts the broader applicability of their method in complex, multi-modal spatial prediction scenarios.

Research by Alam et al. [9] proposed a deep learning-based approach for accurately predicting vehicle CO_2 emissions, integrating eXplainable Artificial Intelligence (XAI) to address the longstanding challenge of model interpretability in emission estimation. The paper introduced a highly efficient and computationally lightweight multilayer perceptron architecture named *CarbonMLP*, designed specifically for high-accuracy emission prediction.

The model utilized SHAP (SHapley Additive exPlanations) to interpret the influence of various vehicle characteristics on CO_2 emissions, providing transparency and facilitating practical insights relevant to regulatory and policy applications. The study addressed key limitations of conventional emission estimation models, which often fail to adapt to real-world, dynamic driving conditions while maintaining interpretability.

To evaluate model performance, several metrics were employed: Coefficient of Determination (R^2), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). CarbonMLP outperformed traditional machine learning algorithms and other deep learning models such as LSTM and BiLSTM, achieving exceptional results: R^2 of 0.9938, MSE of 0.0002, RMSE of 0.0142, and MAPE of 2.59%.

Beyond model accuracy, the study emphasized the value of SHAP in providing interpretable and actionable insights, contributing to the practical deployment of such models in sustainable environmental strategies. However, a notable limitation was the use of singlemodality input data, with no integration of multi-source or spatiotemporal data fusion techniques. This restricts the model's applicability across broader geographic contexts and more complex environmental datasets.

Similarly, the research conducted by Quynh et al. [10] focused on enhancing the accuracy of air quality predictions for $PM_{2.5}$ and PM_{10} concentrations using hybrid deep learning models. The study constructed and compared multi-stage forecasting systems incorporating Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), and encoder-decoder architectures to address the limitations of short-range air quality predictions resulting from the dynamic nature of meteorological conditions.

The research utilized air quality monitoring data from Hanoi, Vietnam, covering the period from January 2018 to May 2022. The models were evaluated based on Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2). Among the tested architectures, the encoder-based LSTM model demonstrated the highest performance, achieving an MAE of 2.77, RMSE of 3.348, and R^2 of 0.991 for $PM_{2.5}$ prediction, and an MAE of 1.850, RMSE of 2.209, and R^2 of 0.985 for PM_{10} prediction.

Despite the strong forecasting performance, the study did not incorporate explainable AI (XAI) techniques or spatial mapping strategies, limiting the interpretability and applicability of its findings in policy-related and regional environmental planning contexts. Nevertheless, the paper provides valuable insights into improving time series forecasting models and emphasizes the importance of advanced feature engineering in air quality prediction systems.

A prominent systematic review by Houdou et al. [11] focused on evaluating interpretable machine learning (IML) approaches for forecasting and predicting air pollution. The study categorized various interpretability methods, including SHAP (SHapley Additive exPlanations), Partial Dependence Plots (PDP), and LIME (Local Interpretable Model-agnostic Explanations), to assess how effectively and transparently air quality forecasting models employed these tools.

The review synthesized results from 56 high-quality articles selected from an initial pool of 5,396 articles retrieved from PubMed, Scopus, and Web of Science databases. The authors emphasized the critical importance of interpretability in environmental models, particularly in regions with sparse air quality monitoring infrastructure. They also highlighted a disproportionate concentration of studies conducted in China and the United States, leaving other regions underrepresented.

Among the findings, SHAP-based models achieved an R^2 of 0.991 in $PM_{2.5}$ prediction and reduced NO_2 prediction errors by 22% compared to traditional black-box models. PDP analysis revealed nonlinear relationships between $PM_{2.5}$ concentrations and aerosol product features, further emphasizing the value of interpretability tools.

Despite its thorough examination of IML techniques, the review did not explore spatial mapping or object detection models, which could complement interpretability with geographic forecasting precision. Moreover, it identified significant gaps in current research, such as the lack of time-aware interpretability tools—only 9% of studies addressed temporal forecasting in conjunction with model interpretability.

In another study, Mesin et al. [12] explored the use of nonlinear adaptive filtering—particularly artificial neural networks (ANNs)—for forecasting air pollution concentrations in urban settings. The study introduced an adaptive ANN trained using both historical and real-time data to improve the accuracy of time series predictions. The authors compared adaptive versus non-adaptive models and highlighted the importance of addressing non-stationary and nonlinear behaviors in environmental time series.

Their best-performing model achieved a Root Mean Squared Error (RMSE) of 10.42 μg/m^3 and an R^2 value of 0.86 on the test set, significantly outperforming the baseline persistence model, which reported an RMSE of 19.85 μg/m^3 and an R^2 of 0.43. These results provided strong evidence that adaptive ANNs can deliver superior predictive performance compared to static models and conventional statistical techniques.

However, the study did not explore adaptivity in broader spatial dimensions or include complex atmospheric reactants such as volatile organic compounds (VOCs). This omission limits the model's applicability for multi-regional or cross-location forecasting tasks. Nonetheless, the study made important contributions by demonstrating the advantages of adaptive modeling approaches in forecasting air pollution under highly dynamic environmental conditions.

In the paper by Li *et al.*, titled "Spatiotemporal Interpolation Methods for the Application of Estimating Population Exposure to Fine Particulate Matter in the Contiguous U.S. and a Real-Time Web Application," the authors aim to develop efficient spatiotemporal interpolation techniques to estimate population exposure to $PM_{2.5}$ concentrations across the United States. In particular, the research compares shape function-based (SF) technique with indirect distance weighting (IDW) to achieve the primary research gap of finding few real-time and scalable interpolation systems to deal with large environmental data and exposure diagrams. The authors base their study on 146,125 daily $PM_{2.5}$ data points provided at 955 monitoring stations in 2009 and combine it with census data of the intervening population to design a real-time web app with the ability to visualize pollution exposure and measure the model performance with the help of 10-fold cross-validation. Their findings show that SF-based approaches are always better than IDW in all possible error measures, and the RMSE is 5.97 μg/m^3, MAE is 3.18 μg/m^3, and R^2_{CV} is 0.56, which is much better than the other baseline, namely persistence (RMSE: 19.85 μg/m^3; R^2_{CV}: 0.43). Though the study has effectively demonstrated the advantage of spatiotemporal interpolation in national-scale exposure assessment, there is an issue in boundary effects that needs to be covered, and the assessment of other pollutants, such as VOCs, is seldom possible. The paper will also provide a valuable approach to spatial analysis that relates to our project, especially within regional pollution exposure mapping in Bangladesh using explainable AI frameworks.

A study by Chakraborty et al. [13] explored the use of Explainable Artificial Intelligence (XAI) techniques in air quality prediction to improve model transparency and pollutant interpretability. The authors utilized SHAP and LIME methods combined with machine learning algorithms such as Random Forest, K-Nearest Neighbors (KNN), and XGBoost, applied to air quality data from several cities in India. Among the models tested, XGBoost achieved the highest accuracy of 83%, along with superior precision, recall, and F1-score compared to other algorithms.

Although the study successfully identified the most impactful pollutants, PM_{10} and $PM_{2.5}$, it did not extend its analysis to time-aware forecasting models or incorporate temporal dependencies. Consequently, the approach is limited in its applicability to dynamic, real-time air pollution monitoring and forecasting.

A recent study developed machine learning models to estimate hourly concentrations of $PM_{2.5}$, PM_{10}, and NO_2 at two locations in Stuttgart and tested their generalizability in

Karlsruhe, aiming to replace costly physical monitoring stations with virtual ones [14]. The methodology employed five regression models: Ridge Regression, Support Vector Regression (SVR), Random Forest, Extra Trees, and XGBoost, trained across four scenarios with progressively increasing feature sets, including temporal, meteorological, traffic, and pollutant data.

Ensemble methods, especially in the most feature-rich scenario, demonstrated the best performance in terms of R^2 and error metrics. However, the study's reliance on pollutant data reported by existing stations limits its generalizability to regions with well-developed monitoring infrastructure. Additionally, it lacks forecasting capabilities, deep learning techniques, and interpretability analyses such as SHAP, restricting the understanding of feature impacts. The transferability was evaluated only in the Karlsruhe region, further limiting its broader applicability.

This research is relevant to our project as it highlights both the potential and limitations of machine learning for virtual air quality monitoring. It underscores the need for more advanced models (e.g., deep learning), incorporation of satellite data, real-time deployment frameworks, and explainability tools as crucial steps toward reliable environmental monitoring and forecasting, especially in regions like Bangladesh.

A study by Kumar Singh et al. [15] predicted the Air Quality Index (AQI) for Ahmedabad in 2015 using three machine learning models: Random Forest, Linear Regression, and Gaussian Naive Bayes. The data preprocessing techniques included mean imputation for missing values, feature selection via a correlation heatmap, and normalization. Among the models, Random Forest achieved the highest accuracy of 91.77%, outperforming Linear Regression (85.84%) and Gaussian Naive Bayes (79.21%), establishing it as a stable baseline for AQI prediction in the city.

However, the study's generalizability is limited as it focused on a single city and a one-year timeframe. It did not address advanced temporal modeling, deeper feature engineering, uncertainty quantification, or considerations for real-time deployment. Furthermore, only accuracy was reported, whereas metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are critical for comprehensive evaluation.

This research is relevant to our project as it demonstrates the foundational use of machine learning for AQI forecasting, while highlighting the need for multi-city datasets, spatial and temporal features, advanced models like LSTM or transformers, and explainable AI methods to enhance robustness and interpretability for environmental monitoring in Bangladesh.

The study by Rahman et al. [16] developed AirNet, a real-time web-based system for classifying Air Quality Index (AQI) categories such as Good, Moderate, and Hazardous, leveraging machine learning models to provide city-specific air quality alerts. Using a Kaggle dataset with 23,463 records from different cities, seven classification algorithms were applied, including Random Forest, Decision Tree, K-Nearest Neighbor (KNN), Linear SVC, SVM, Logistic Regression, and Multinomial Naive Bayes. Preprocessing involved handling missing data, proper encoding of AQI categories, and feature selection through correlation analysis. The implementation utilized Python's scikit-learn and a Django web interface.

Random Forest and Decision Tree achieved perfect accuracy of 100%, outperforming other models such as KNN (99%) and SVM (93%). However, such high accuracy likely indicates overfitting. The study lacked forecasting capabilities, deep learning architectures, integration of meteorological and spatial features, k-fold cross-validation, and handling of class imbalance. Additionally, model interpretability methods like SHAP or LIME were not incorporated, and the web app relied on static CSV data instead of real-time sensor inputs.

Despite these limitations, the study offers significant potential for AQI classification with an accessible web interface. Future work could improve the system by incorporating LSTM or GRU forecasting models, weather and spatial data fusion, data imbalance treatment, robust cross-validation, and explainability techniques to enable practical deployment.

In a recent study, Matara et al. [C. Matara, S. Osano, A. O. Yusuf, and E. O. Aketch, ”Prediction of Vehicle-induced Air Pollution based on Advanced Machine Learning Models”, Eng. Technol. Appl. Sci. Res., vol. 14, no. 1, pp. 12837–12843, Feb. 2024.] (https://doi.org/10.48084/etasr.6678) applied advanced machine learning models—including Extra Trees (ET), Random Forest (RF), XGBoost, Gradient Boosting, and LightGBM—to predict PM2.5 concentrations along the JKIA-Westlands Expressway in Kenya. The research tried to improve the accuracy and explainability of vehicle-induced air pollution through hyperparameter optimization, Bayesian optimization, and Interpretability SHAP. The highest performance is identified as that of the ET model, which gave an MAE of 1.48, resulting R^2 of 0.76 on the training data and showed high robustness on testing data with an MAE of 1.69 and R^2 of 0.71, showing high predictive accuracy. Linear Regression, on the other hand, gave bad results with MAE of 3.57 and R^2 of 0.064. Even though the study utilized explainability methods such as the SHAP, it reported that integrated multi-source real-world data on a combination of traffic, weather, and air quality were often limited, which limited the predictive capacity. Although this work is a good example of a methodological approach to ensemble ML hyperparameter optimization and interpretability in air pollution prediction, it also indicates pitfalls of data integration that would require more thorough discrimination to be used in modelling.

Sunder et al. [17] constructed a forecast of the spatial PM2.5 over Singapore using different machine learning models, such as Random Forest (RF), Gradient Boosting (GB), and Tree-based Pipeline Optimization (TP). The study assimilated seven years (2014–2021) of data at five monitoring stations, combining satellite-retrieved MODIS Aerosol Optical Depth (AOD), relative humidity, wind speed, and temperature. Performance criteria were assessed using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and coefficient of determination in this study.

Predictive performance was excellent for the RF model (high R^2 with low error), and global quality was also outstanding for the TP model, reaching its best Global Performance Index of 7.4 at selected times (e.g., August 2016). The work focused on explainable AI techniques by using SHAP to increase the interpretability of models and investigate how weather features influence PM_2. 5 concentrations. The paper discovered some shortages, such as the spatial prediction aspects not being emphasized enough, the risk of overprediction caused by outside factors not being adequately discussed, and no processing for peak pollutant intake. These shortcomings also reveal possible ways to improve the interpretability of spatiotemporal air quality models.

A study by Stirnberg et al. [18] analyzed PM_{10} concentration variability using explainable machine learning on seven years of data from the SIRTA atmospheric observatory near Paris. The study utilized meteorological parameters, including wind speed, temperature, mixed layer height, and solar radiation, to elucidate daily variation in PM_{10}. The PM_{10} variable had an R^2 R^2 of approximately 0.58, indicating moderate explanatory power; the model's predictions correlated better with black carbon and organic matter than with nitrates, which showed greater variance.

However, the study acknowledged limitations such as the inability to translate findings into practical applications, inadequate modeling of interaction effects, and the complexity of meteorological conditions influencing pollution events.

A notable review by Talaie, Kamyab, and Razmfarsa [19] has shed light on the utilization ofenvironmental engineering in integrating artificial intelligence (AI) to enhance sustainability. The paper presents the potential of machine learning, neural network and natural language processing for simplifiedresource management, environment monitoring and decision making. Leveraging various sources of data, such as air quality sensor networks, satellite imagery and traffic information, as well as historical weather records, the paper illustrates AI's ability toincrease prediction accuracy and streamline environmental activities. While it has long been shown that air pollution, water demand forecast were improved using AI or ML technique, the authors highlight obstacles such as absence of freely available data in developing countries; high energy consumption of AI systems; andrisks in terms of misuse of unethical AI.The paper makes a call for continued improvement of design of energy-efficient AI models that enable collective intelligence without the enormous energy costs associated with today'splatforms, and development of datasets at scale that can produce more informed environmental outcomes to meet New Zealand sustainability objectives.

A recent study by Islam et al. [20]were devotedto the estimation of ground-based PM2. 5 based concentrations in Dhaka through ahybrid method of best subset regression and machine learning models. Themodel used Random tree, Additive regression, REPTree and a Random Subspace for improved prediction. The best model performed wasthe Random Subspace model which attained AUC of 0.94 and RMSE 34.50 and it was applied on Darus Salam site. However, to our best knowledge, the deep learning models andreal-time prediction ability have not been investigated yet, which could potentially provide better performance in terms of the predictive power and generalization for dynamically changing air pollution conditions.

A dataset study by Rahman et al. [16] supplied hyper-local measurements ofPM2. 5, PM10 and CO in Dhaka'sExport Processing Zone (DEPZ) from 2019 to 2023. The authors obtained RMSE values of 20.31 and 3.18 by employing ARIMA models to predict the concentrations of PM2.5 and CO, respectively, indicating moderate accuracy in pollutant concentration prediction. Seasonalityand pollution spikes after lockdown were well captured by this study, although its application to a single location together with the absence of spatial variability analysis, incorporation of meteorological conditions and explainability of forecasting model components limited it's overall ability.

Ahmed et al. [21]compared the USEPA single-pollutantmethod based global AQI and locally developed PCA-based multi-pollutant AQI in DNCC and DSCC areas. The local AQI combines data from eight satellites (Sentinel-5P/TROPOMI) withground sensors to enhance spatial resolution near industrial areas and road networks. On the other hand, the world AQI overem- phasized thePM2. 5 and mischaracterized several areas of industrialactivity. The local AQI more effectively reflectedcollaborative effects of pollutants and their spatial distribution. However, this study suffered from poor on-the-ground measurements anda narrow time span without seasonal variability.

Hameed et al. [22] have proposed also a multimodal deep learning approach that combines real-time air quality sensorsand CCTV-based traffic analytics – with the application of YOLOv8 for vehicle counting in Dalat City, Vietnam. Sensor failures werehandled using particle swarm optimization (PSO)-based sensor fusion and ConvLSTM models were able to successfully forecast AQI at short-term, medium-term and long-term timescales outperforming existing ARIMA-based methods. Asignificant association between traffic density and PM2. 5 surges during rush hours. However, the sampling period without seasonal change was only five months and the model was not interpretable enough to beused for policy-making. Other technical limitation was the poornight vehicle detection and a low spatial resolution (small number of sensors).

Kalajdjieski et al. [23] proposed a deep learning approach utilizingmulti-modal data (static camera images and weather data) to predict air pollution in North Macedonia's Skopje. They presented a novel model based on Inception CNN with ConditionalGANs for dealing with the class imbalance and reported 88% training and 76% testing accuracy in binary categorization of pollution. As the model works well on binary case, there are some challenges to handle fine-grained 6-class AQI classificationbecause of pollution levels visually resembling each other. This method also takes advantage of available camera infrastructure, rather than requiring dense sensor coverage. However, itis spatially limited since providing single-location predictions without interpretability frameworks such as SHAP and Grad-CAM, predicting capability and integration with the satellite imagery data.

2.3 Summary of Key Findings

This study demonstrates that machine learning (ML) and deep learning (DL) models have huge potential to predict and capture the true picture of the extent of air pollution and identify the underlying mechanisms. Several key results were discussed in the literature andanalysis of models. All ensemble method include RF and XGB were of the first order with combination of deep learning models LSTM, encoder-decoder performedbetter in time series predictions. The accuracy improvement was consistently realized by integrating all of the meteorological, traffic andsatellite data, using the feature engineering, multi-source data fusion. Another developing approach in the latest works is using explainable AI (XAI) methods like SHAPand LIME, which have improved our understanding of how the pollution impacts. The results indicate that the further implementation of highly-sophisticated machine learning, deep learning, and XAI architectures might create a considerable improvement in modern air quality prediction systems, and its implications include the inference that such methods are essential to the environmental monitoring market, the urban planning process, and the decision making of public health. Research efforts can lead to the enhancement of spatiotemporal generalizability, real-time pipelines, and explainability to facilitate more practical deployments in third-world countries such as Bangladesh.

Chapter 3

Project Specifications and Impact Assessment

3.1 Final Specifications and Requirements

The finalized specification is based on three major functional domains such as regional clustering, time-series forecasting, and satellite to ground data mapping, while integrating explainable artificial intelligence (XAI).

In order to ensure data acquisition and integration, both ground-station CSV data (air-quality + meteorological indicators) and multiband satellite GeoTIFF imagery (NO_2, SO_2, CO, O_3, rainfall, temperature, humidity, solar radiation, vegetation index) must be ingested. While the spatial resolution $\approx$ 1.1 km per pixel; temporal coverage taken from Jan 2022 – May 2025.

To preprocess data and accurate the imputations, the study employed MICE with Random Forest estimators (800 trees, 25 iterations $\times$ 5 seeds) excluding temporal identifiers such as district, year, month during imputation to prevent data leakage. Again, validation for completeness was done via RMSE and MAE comparison against Bayesian Ridge / Extra Trees.

In order to extract multimodal features, satellite images were resized to 224$\times$224 px and normalized. Pretrained ResNet-18, ResNet-50, DenseNet-121 CNNs were utilized for spatial feature extraction and tabular features were standardized by StandardScaler. Again, for fusion CNN embeddings were concatenated with the tabular matrix after principal component analysis. RMSE, MAE, and R^2 were used for evaluation and validation. For each components the best performing CNN pretrained model was used for embedding. Classical model SARIMA and deep learning model LSTM was implemented. XAI was implemented on LSTM to help intepretability.

In order to maintain scalability and efficiency, 5GB GPU training support was needed so the large size of nationwide air pollution dataset could be handled. The dataset used was not robust, since most component had on average 15% missing data. Codebase was written in Python 3.10 using PyTorch.

3.2 Societal Impact

The developed multimodal environmental intelligence framework provides wide societal advantages in Bangladesh, such as accurate prediction of pollutants ($PM_{2.5}$, SO_2, O_3 etc.)

allows early warnings that help citizens reduce exposure and supports governmental health advisories. Transparent XAI outputs explain why certain regions cluster as high-risk, enabling policymakers to design targeted interventions, e.g., regulating brick-kiln zones or vehicular corridors. Again, regional clustering highlights environment-sensitive zones guiding future zoning, industrial permitting, and green-belt allocation. This study also serves as an open methodological blueprint for Bangladeshi universities and agencies to replicate AI-driven environmental analytics.Taken together, the system also contributes to UN SDGs 11(Sustainable Cities) and 13 (Climate Action) by providing accessible, explainable environmental intelligence that allows users to take informed action.

3.3 Environmental Impact

The project itself produces a minimal ecological footprint and yields long-term environmental benefits that enables data driven pollution mitigation, reducing industrial emissions through informed control and supports sustainable resource management by forecasting rainfall or temperature trends relevant to agriculture and disaster preparedness. Promotes satellite-based virtual sensing, decreasing dependency on additional physical stations and thus material waste. There are other environmental costs as training CNN/XGBoost models requires electricity, somewhat lessened via GPU time scheduling andcloud carbon-offset policies. Big GeoTIFF archivesneed space on disk, which is minimized with compression and spatio-temporal decimation.

The cost in terms of computational resources is minimal compared to the environmentalgain in accuracy of policy, emission reduction and prevention.

3.4 Ethical Issues

Ethical concerns were borne in mind throughout this study, sothat data usage, analysis and reporting adhered to present research norms. The dataused in this research, i.e., ground-station measurements and satellite images were collected directly from the open governmental sites and open scientific sources such as Depart- ment of Environment (Bangladesh) and Google Earth Engine. Nohuman subjects or sensitive personal information was involved; therefore, no institutional ethics approval was required. Datamanagement was in accordance with the standards of transparency, fairness and reproducibility. All pre-processing (ie imputation) and modeling were recorded with transparent, reproducible computationalenvironments (Google colab + Jupyter notebook). Multiple imputationof missing values by multivariate statistical methods (MICE with Random Forest) was performed only for analytical completeness, without influencing the results.

Intellectual property and software licensing were also respected. Open-source libraries such as scikit-learn, Captum and PyTorch were used for this research under their respective permissive licenses. Satellite imagery was acquired through APIs in compliance with Earth Engine's terms of service and proper attribution was provided wherever data or methods were reused or adapted.

Finally, the research maintained environmental and social responsibility by interpreting model predictions in a policy-neutral manner and by explicitly stating the uncertainty and limitations of the results. The findings are intended for scientific and developmental use only, not for commercial or discriminatory applications.

3.5 Risk Management

Throughout the thesis, risk management was adopted to guarantee validity, stability and on-time completion. Incomplete or inconsistent data, model overfitting, computing constraints and scheduling delays were the main concerns associated with this thesis. These concerns were addressed by the use of GPU-based platforms like Google Colab to maximize performance, regularization and cross-validation during model training and thorough imputation using MICE with Random Forest. Weekly progress reviews helped in the early detection and resolution of problems. In order to avoid data loss and integration difficulties, wekept the backup sets of data and code modules updated. During theprogress of 17 weeks, such a posture ensured reliability, reproducibility and transparent implementation.

3.6 Economic Impact

From an economic standpoint, the integrated framework provides a cost-effective, high-value solution as the satellite-driven monitoring cuts expenses of installing and maintaining numerous ground sensors ($\approx$ 60–70% cost reduction). Automated imputation and forecasting lessen manual analytical labor. Early detection of pollution hotspots saves public health expenditure by preventing hospitalizations. Helps allocate limited environmental control budgets toward most affected districts. Industries can plan emission-control upgrades economically by analyzing localized pollution data. Reduced environmental degradation indirectly enhances productivity, tourism, and human capital.

This study aims to help in the research and innovation economy that creates potential for local AI-environmental startups, increasing technological self-reliance. Open-data frameworks may attract international collaborations and funding. Reduced environmental degradation indirectly enhances productivity, tourism, and human capital. Hence evaluation of the cumulative economic benefit exceeds the initial computational investment, reinforcing national sustainability goals.

Chapter 4

Methodology

Design decisions for our proposed study are driven by four core objectives: (1) robust imputation of missing station data using multivariate techniques, (2) accurate satellite-to-ground mapping for spatial prediction using a multimodal approach to improve data quality and approximate missing values, (3) comparative analysis between different models, and (4) reliable time-series forecasting and regional clustering with interpretable outputs. Constraints include heterogeneous data quality (numerous missingness in the existing dataset), limited station density, computational resources, satellite image findings, and the need for model interpretability for policy use.

4.1 Design Process or Methodology

Overview

The design and methodology of this study include statistical data processing, multimodal machine learning, and geospatial analytics to unify satellite-based environmental indicators with ground-based measurements in Bangladesh. The goal is to create a streaming time-series framework for temporal prediction, local clustering, andmapping of environmental phenomena in support of sustainable decision making via explainable AI (XAI).

Two main sources of data were used in this research: (1) effective ground station observations from theMinistry of Environment, Forest and Climate change, Bangladesh over January 2022–May 2025 (2) satellite imagery data through Google Earth Engine(Sentinel 5P, CHIRPS, MODIS, ERA5). The ground-based environmental data wereoriginally in PDF and have been converted to CSV after extensive filtering/pre-processing and combined to create a cohesive dataset representation for air quality and meteorological parameters. In parallel, satellite data were spatially aligned with district-level administrative boundaries from GADM (Level-2). Each image was projected to EPSG:4326 coordinates, sampled bilinearly at an approximate spatial resolution of 1.1 km per pixel, and stored in GeoTIFF format. The imagery contained nine environmental bands stacked on top of one another: SO_2, NO_2, CO, O_3, rainfall, vegetation, air temperature, relative humidity, and surface solar radiation. To facilitate the missing and incomplete ground measurements, Multivariate Imputation by Chained Equations (MICE) [24]with Random Forests [25] as a prediction estimator was used. There were chosento avoid the intricate nonlinear dependence of ambient variables (e.g. co-pollutant mixes, temperature, humidity, and rainfall), with strong seasonal patterns that violate assumptions of linear imputation used in previouswork. The Random Forest imputation employed800 decision trees per iteration, with five random seeds and 25 iterations per seed. Toensure stability, median was used for aggregation. the arithmetic mean is used All irrelevant identifiers (e.g. Station ID, year, month) have been dropped to prevent data leakage during the point-wise imputation approach applied as an initialphase. Pearson correlation and mutual information The reduction in pairs being positively correlated indicates thatmost pairs do not have linear relationships, while this is contradicted by the strong relationship between negatively/positively correlated pairs from mutual information analysis. The comparison withBayesian ridge regressor and Extra Trees regressor revealed a lower RMSE/MAE for the tree-based models which further confirms the linearity of the dataset. Following imputation, tabular features were rescaled intostandard units by applying StandardScaler. Model- ing framework The modeling pipeline was based on multimodal fusion: satellite images were resized to 224×224pixels and fed into pre-trained convolutional neural network (CNN) [26] —ResNet-18, ResNet-50 [27], and DenseNet-121 [28]—, given spatial embedding. They were concatenated with the scaled tabular feature matrix to create a combinedmultimodal representation that preserved local (station level) and global (satellite) characteristics.

The concatenated features were subsequently used to train an XGBoost [24] regressor with 700 estimators, maximum depth 8, learning rate $\alpha = 0.05$, and L2 regularization to mitigate overfitting. A 90/10 train–test split was applied. Performance was evaluated using RMSE, MAE, and R^2. Comparative analysis between the hybrid CNN+XGBoost model and a tabular-only XGBoost baseline highlighted the effectiveness of the multimodal approach, especially for pollutants (e.g., SO_2, O_3) exhibiting strong spatial dependence. The best fit pretrained model was later used for each components image embedding to create the fusion dataset. PCA was implemented to reduce embeddings to 32 dimensions with highest variance.

Regional pattern discovery was applied on the multimodal dataset. Instead of clustering for a single time frame, each district was represented by a feature vector made from the entire timeline. K means clustering was implemented and silhouette score was used to determine the optimal number of clusters. For each components the mean and standard deviation was calculated for that specific PCA embedding for all months. All district vectors were scaled using standard scaler. K means was implemented independently with (2, 3 and 4) as number of clusters. This enabled the identification of similar environmentally homogeneous zones based on various meteorological and pollution factors. For forecasting the multimodal dataset was restructured at the district and station levels. Monthly environmental indicators were used to capture seasonal and long term relations. Forecasting was performed independently. Choronological split was enforced to avoid data leakage. AdamW optizer with gradient clipping was used to stablize training. Accuracy was measured using RMSE, MAE and R^2. Components like SO_2 along with humidity and solar sun radiation showed predictable performance.

Explainable AI techniques were applie on the trained LSTM model to increase transparency. Integrated Gradients (IG) [25] was applied using Captum framework. For each components IG measured the feature contributions which could help policy makers to understand which components are responsible for the irregularity of the other. This increases trust and usability for policymakers and researchers, supporting data-driven environmental planning in Bangladesh with a transparent, reliable AI system.

4.2 Preliminary Design / Model Specification

4.2.1 Overview

The developed model adopts a multimodal data fusion approach that integrates station-based grounddata with satellite assemblages to enhance prediction accuracy and fill in missing or incomplete measurements. The architecture uses classical machine learning on structured tabular data (treebased regressors) and deep learning on images (pre-trained CNNs as feature extractors). These data shows precise ground data and broad relative satellite data.

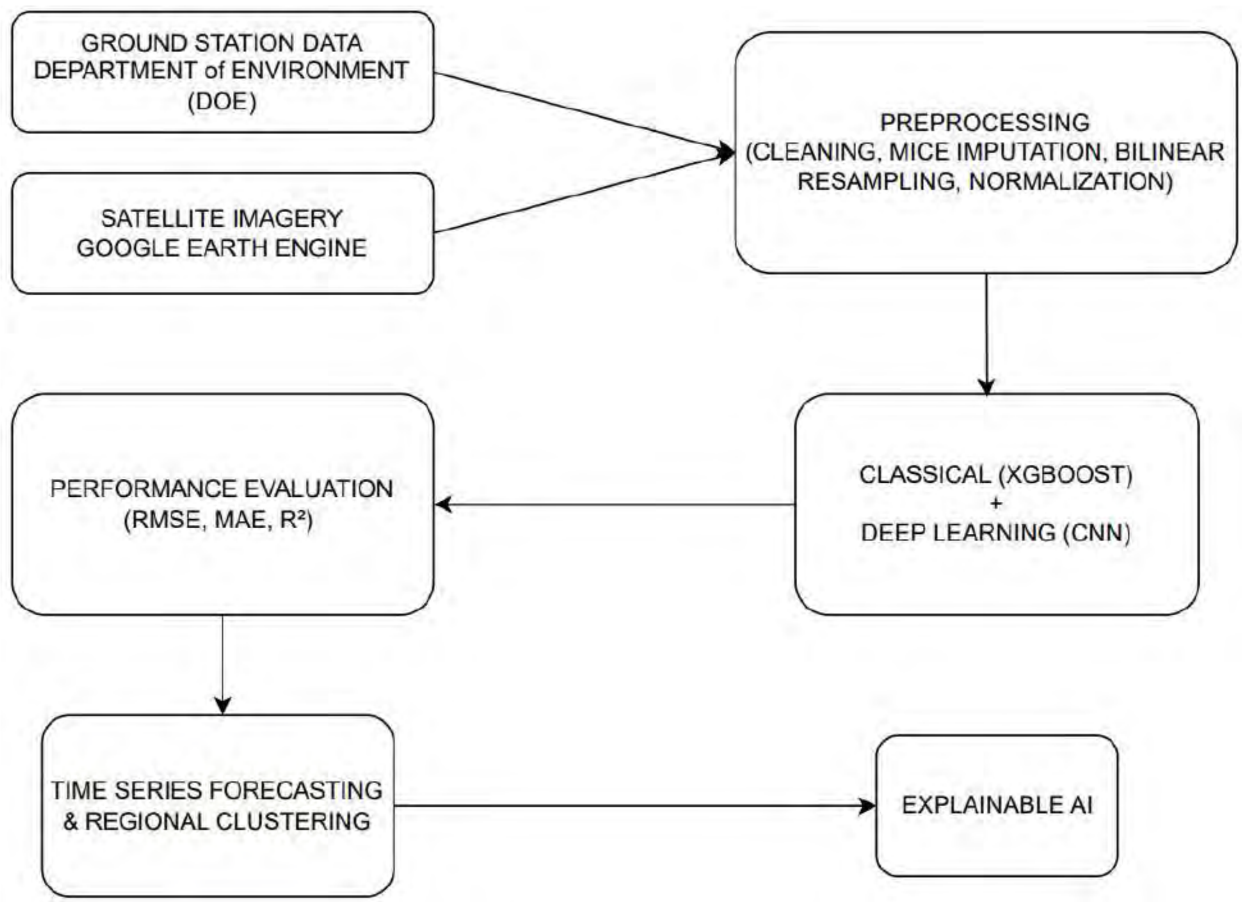

Figure 4.1: Proposed Workflow.

4.2.2 Data Preparation and Imputation

The baseline data indicated the MAR and MNAR patterns of missingness values, requiring robust imputation. Random Forest was used as MICE model to simulate variables until each missing data was imputed sequentially. Random Forests were chosen for their ability to separate non linear data with high precision and without strict distributional assumptions. Each forest contained 800 trees. 5 seeds were used. Each seed had 25 iterations. Arithmetic mean was used for initial fills. Median fills were used for subsequent fills. Features such as district, year, and month were excluded to avoid leakage. Resulting imputations achieved substantially lower RMSE and MAE with Random Forest/Extra Trees than Bayesian Ridge, confirming nonlinear dependencies. Imputation using Random Forest had significantly lower RMSE and MAE compared to Bayesian Ridge. This solidifies the nonlinear dependency analysis.

year	month	station	SO2	NO2	CO	O3	PM2.5	PM10	Solar radia	Relative h	Ambient temper	Rainfall
2022	7	Mymensingh	1.0	6.2	1.1	89.5	37.9	112.2	202.6	93.9	28.7	nan
2022	5	Narayanganj	3.5	7.25	nan	13.8	45.1	59.6	nan	64.1	28.1	0.29
2023	11	Narshingdhi	1.7	7.6	0.6	4.7	49.0	148.6	328.8	48.6	24.4	0.02
2023	12	Rangpur	3.0	nan	1.3	37.5	177.1	247.7	270.3	83.1	19.3	0.01
2022	6	Dhaka	1.7	2.45	0.9	9.2	32.3	69.6	119.6	82.4	30.75	5.86
2023	4	Comilla	2.4	6.3	0.9	28.5	116.5	151.5	313.6	79.4	28.7	0.01
2022	3	Mymensingh	4.4	15.0	0.8	19.1	132.7	252.7	294.1	66.7	26.0	0.08
2022	10	Rajshahi	5.0	5.3	0.9	12.3	73.8	116.6	60.2	73.2	32.8	nan

Figure 4.2: Raw Dataset Sample.

year	month	district	so2	no2	co	o3	PM2_5	PM10	ssrd_J_m2	rh_2m_pct	t2m_C	rain_mm
2022	7	Mymensingh	1.0	6.2	1.1	89.5	37.9	112.2	202.6	93.9	28.7	1.13
2022	5	Narayanganj	3.5	7.25	1.369	13.8	45.1	59.6	255.93	64.1	28.1	0.29
2023	11	Narshingdhi	1.7	7.6	0.6	4.7	49.0	148.6	328.8	48.6	24.4	0.02
2023	12	Rangpur	3.0	5.974	1.3	37.5	177.1	247.7	270.3	83.1	19.3	0.01
2022	6	Dhaka	1.7	2.45	0.9	9.2	32.3	69.6	119.6	82.4	30.75	5.86
2023	4	Comilla	2.4	6.3	0.9	28.5	116.5	151.5	313.6	79.4	28.7	0.01
2022	3	Mymensingh	4.4	15.0	0.8	19.1	132.7	252.7	294.1	66.7	26.0	0.08
2022	10	Rajshahi	5.0	5.3	0.9	12.3	73.8	116.6	60.2	73.2	32.8	0.964

Figure 4.3: Imputed Dataset Sample.

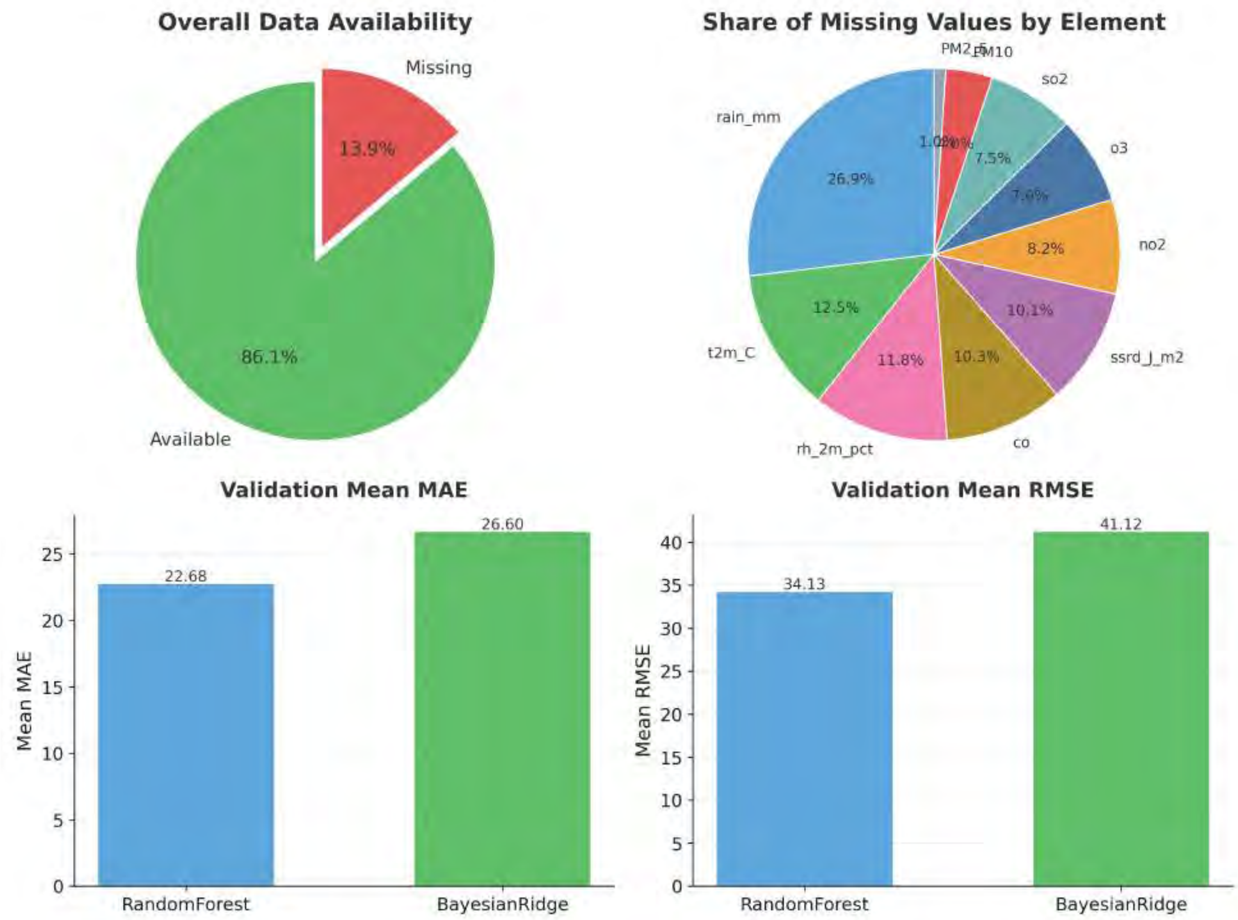

Figure 4.4: Raw Dataset Details.

4.2.3 Image Data and Preprocessing

Satellite imagery spanning January 2022 to May 2025 was obtained from GEE, projected to EPSG:4326 using bilinear resampling at ∼1.1 km/pixel. Each GeoTIFF contained nine bands (NO_2, SO_2, CO, O_3, rainfall, vegetation index, air temperature, relative humidity, surface solar radiation). All images were resized to 224×224 pixels and normalized using Z score normalization to ensure compatibility with pretrained CNN backbones.

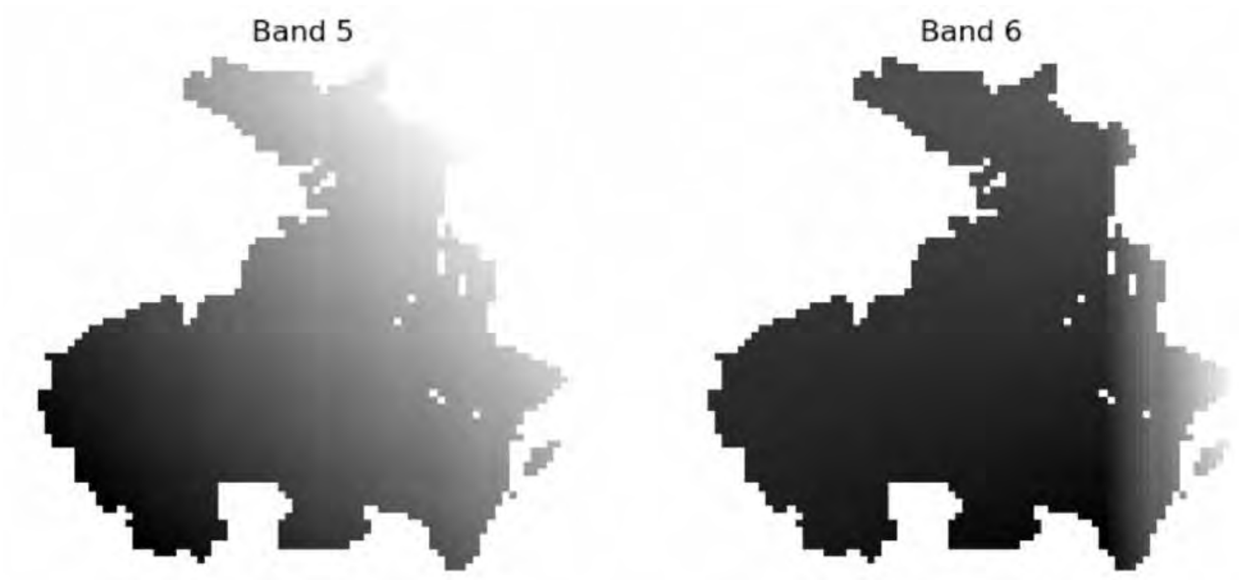

Figure 4.5: Satellite Imagery Sample.

dist_Rajshahi	dist_Rangpur	dist_Sylhet	month.1	no2	co	o3	PM2_5	pm10	ssrd_J_m2	rh_2m_pct	t2m_C	rain_mm	emb_1	emb_2	emb_3	emb_4	emb_5	emb_6	emb_7	emb_8	emb_9
True	False	False	-0.3272	-0.0314	-0.6041	2.0531	-0.8411	-0.4701	0.4244	-0.8259	1.2704	-0.1574	0.0002	0.0015	0.0096	0.0011	0.1280	0.0677	0.0003	0.0030	0.0297
False	False	False	-0.9092	-0.2912	-0.3531	-0.2146	0.9015	0.7686	0.2608	0.2113	-0.0081	-0.3005	0.0003	0.0010	0.0091	0.0008	0.1036	0.0589	0.0003	0.0019	0.0464
False	False	False	0.5458	0.0057	1.0278	-0.1015	-1.2395	-1.0643	-0.2521	0.7883	1.9285	0.1007	0.0003	0.0004	0.0032	0.0032	0.1069	0.6346	0.0006	0.0039	0.1297
False	False	False	0.8367	-0.6499	0.7052	5.0356	-1.1731	-1.1147	-0.8348	-0.8382	0.9508	-0.2316	0.0001	0.0027	0.0028	0.0012	0.1195	0.1205	0.0001	0.0014	0.1040
False	False	True	0.2548	0.0640	0.0988	-0.7443	-1.0583	-0.9324	-0.0898	-0.3489	0.5454	-0.2357	0.0003	0.0012	0.0101	0.0007	0.1021	0.0445	0.0004	0.0036	0.0536
False	True	False	-1.4911	-0.6870	-0.5539	-0.4717	1.5242	1.0815	0.2044	1.0706	1.7003	-0.3331	0.0002	0.0013	0.0071	0.0008	0.1012	0.0712	0.0002	0.0028	0.0509
False	False	False	-0.0362	-0.5880	-0.3531	-0.3329	-0.9344	-1.1276	-0.2143	-1.4458	0.6499	-0.3715	0.0003	0.0004	0.0032	0.0032	0.1069	0.6346	0.0006	0.0039	0.1297
False	False	False	1.1277	-0.5633	-0.8050	-0.2815	-0.3978	-0.4877	0.0437	1.2486	0.2175	-0.4143	0.0005	0.0012	0.0090	0.0007	0.1301	0.0471	0.0001	0.0020	0.0367

Figure 4.6: Processed Dataset Sample (Fusion+Encoding+Scaling).

4.2.4 Model Architecture

The multimodal fusion model includes (i) image feature extraction, (ii) tabular data processing, and (iii) feature fusion followed by regression.

- **Image feature extraction:** ResNet-18, ResNet-50, and DenseNet-121 were used as fixed feature extractors. ResNet uses residual (skip) connections. This helps models to create deeper connections without the fear of vanishing gradients; ResNet-50 employs bottleneck blocks for expressiveness. DenseNet-121 densely connects layers. This enables feature reuse.

- **Tabular processing:** Imputed ground features were standardized to zero mean and unit variance (`StandardScaler`). Each record represents meteorological and pollutant readings per district and month. One hot encoding was used for districts ensuring the model understands spatial difference. Cyclic encoding was applied on monthly values to preserve their circular nature.

- **Fusion & regression:** CNN embeddings were concatenated with tabular features to create a unified representation, then fed to an XGBoost regressor (700 estimators, max_depth 8, $\alpha = 0.05$, L2 regularization). Data were split into 90% train and 10% test.

- **Fusion Dataset & Principal Component Analysis :** After analysis the multimodal fusion dataset was created by concatenating the ground based dataset and image embedding vectors. For image embedding the best performing pretrained models were used for each components. In general Resnet 18 performed the best. Each embeddings were PCA into 32 dimensions to mitigate curse of dimensionality.

- **Regional Clustering :** K Means (k=2, 3, 4, n_init = 10, random state = 42, Standard Scaling) clustering was used to group similar districts into clusters based on components.

- **Time Series Forecasting & XAI :** Classical model SARIMA was applied on the dataset, however since the dataset expands for a short amount of months, SARIMA shows unstable behavior. In addition to that strong non linear relations between meteorological component makes SARIMA unusable. Switching to deep learning model LSTM showed greater improvement. As such LSTM was implemented and integrated gradients (IG) was used to properly understand the feature importance dictating the forecasting. Later Bi directional variant of LSTM named Bi-LSTM was implemented to check whether incorporating forward and backward contexts as input could improve performance or not.

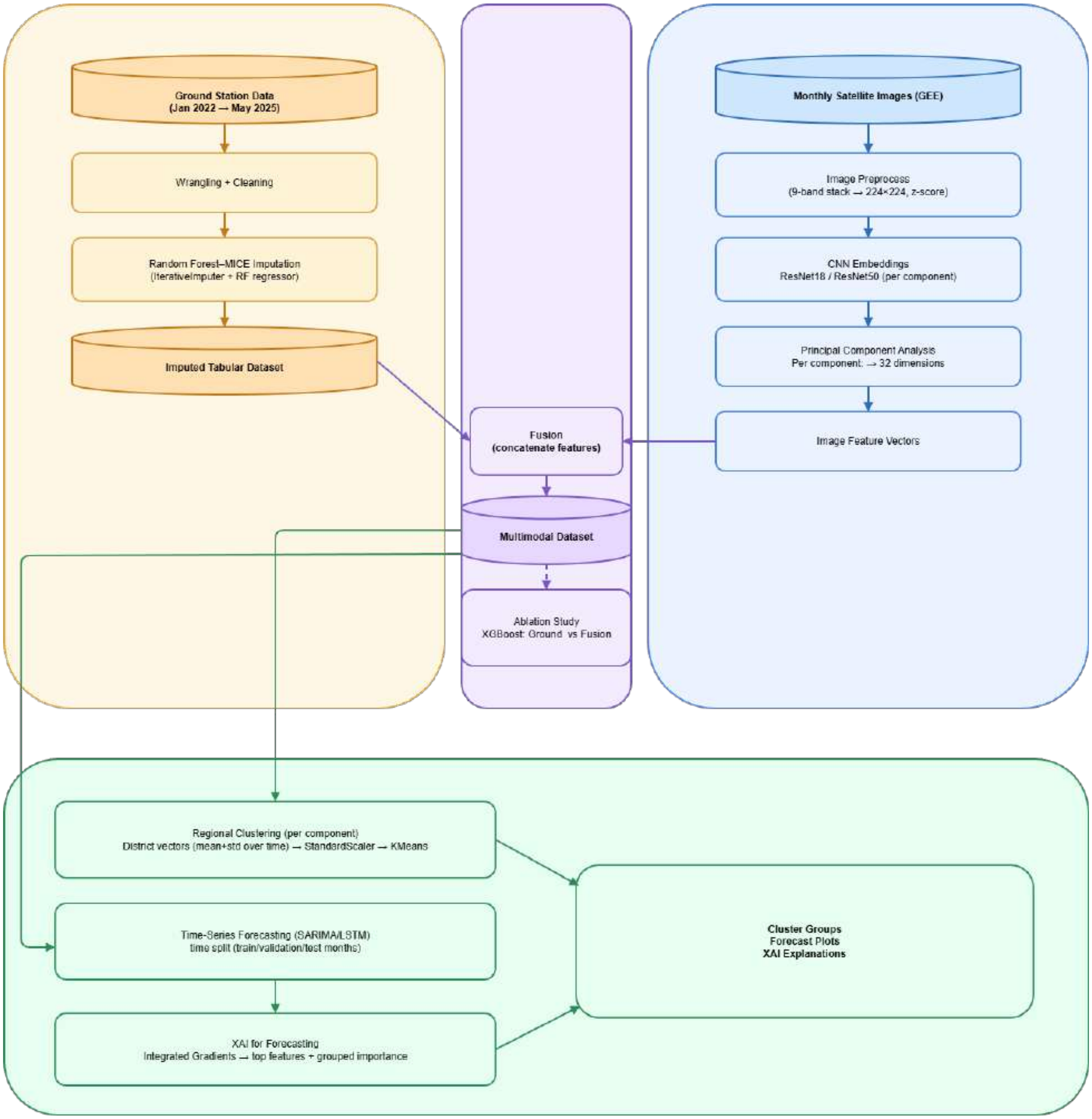

Figure 4.7: Model Pipeline.

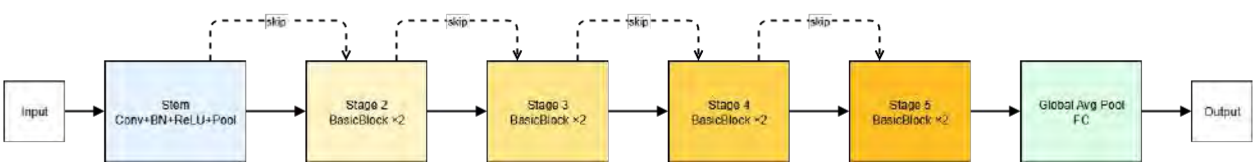

Figure 4.8: ResNet-18 Architecture.

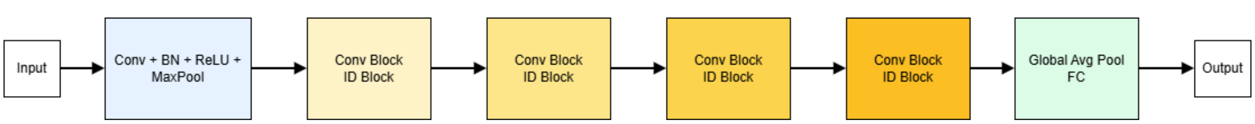

Figure 4.9: ResNet-50 Architecture.

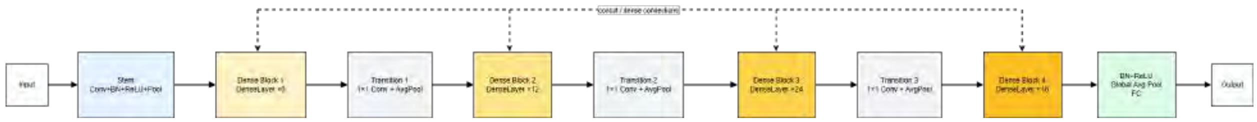

Figure 4.10: DenseNet Architecture.

Activation Function,Normalization and Its Significance on Model Accuracy

In the proposed CNN pretrained models, **Rectified Linear Unit (ReLU)** was used as the activation function for the deep learning components. ReLU introduces non-linearity to the data. Non linearity enables the model to learn complex patterns while avoiding the vanishing gradient problem common in other activation functions such as sigmoid [26].

The ReLU function is defined as:

$$f(x) = \max(0, x)$$

where x is the input to the neuron. Its derivative is simple:

$$f'(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases}$$

This simplicity makes ReLU computationally efficient and suitable for deep architectures.

For feature normalization, two approaches were used: **Z-score normalization** for satellite images and **Standard Scaling** for ground station data. Both ensure features are centered and scaled to similar ranges, improving gradient-based optimization and model convergence.

Z-score Normalization:

$$x' = \frac{x - \mu}{\sigma}$$

where μ is the mean and σ is the standard deviation of the feature. This transformation produces zero-mean and unit-variance data, stabilizing model training.

Standard Scaling:

$$x_{\text{scaled}} = \frac{x - \mu_{\text{train}}}{\sigma_{\text{train}}}$$

The parameters μ_{train} and σ_{train} are computed only from the training dataset to prevent data leakage. This approach aligns with the implementation of `StandardScaler` in `scikit-learn`.

Chapter 5

Performance Evaluation and Analysis

5.1 Performance Evaluation

This section discusses the performance of the introduced environmental analysis framework with respect to the results achieved by our ML and DL pipelines. The model is a combination of spatial land use features determined from satellite data and on-site ground level based environmental measurements for prediction of various air quality parameters across the territory of Bangladesh. Model quality is evaluated by several standard regression metrics: RMSE, MAE, R^2 based on test-set checks.

5.1.1 Evaluation Setup and Data Overview

In order to quantitatively evaluate the performance of the predictive models, three standard regression metrics were used: **Root Mean Square Error (RMSE)**, **Mean Absolute Error (MAE)**, and the **Coefficient of Determination (R^2)** [27].

1. **Root Mean Square Error (RMSE):**

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

RMSE measures the standard deviation of prediction errors. It penalizes larger deviations more heavily due to the squared term. A lower RMSE indicates higher predictive accuracy and better model calibration. RMSE is sensitive to outliers as a result emphasizes larger errors.

2. **Mean Absolute Error (MAE):**

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

MAE calculates the average magnitude of prediction errors without considering their direction. Unlike RMSE, it provides an intuitive interpretation of the average absolute deviation in the same units as the target variable.

3. Coefficient of Determination (R^2):

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

where $\bar{y}$ represents the mean of the observed data. The R^2 score expresses how well the predicted values approximate the actual distribution. A value of $R^2 = 1$ indicates the prediction was perfect, whereas $R^2 = 0$ indicates that the model predicted like mean value.

4. Silhouette Score (for K-Means Clustering):

$$\text{Silhouette} = \frac{1}{n}\sum_{i=1}^{n}\frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

where $a(i)$ denotes the average intra-cluster distance of sample i, defined as the mean distance between i and all other data points within the same cluster, and $b(i)$ represents the minimum average inter-cluster distance between sample i and all points in the nearest neighboring cluster.

The Silhouette Score evaluates the quality of clustering by simultaneously measuring cluster compactness and separation. The score ranges from -1 to 1, where values close to 1 indicate clusters that are furthest away from points of other clusters and closest to points of its own cluster, values near 0 suggest overlapping clusters, and negative values implies miss clustering.

Together, these metrics provide a balanced assessment of model reliability.

Atmospheric condition and pollutant data (except for weather data) on the ground level were from government monitoring station, while satellite data could were obtained using Google Earth Engine and spatially matched to district-level observations. Spatial features were derived from satellite images with pre-trained CNN models (including ResNet-18, ResNet-50 and DenseNet-121) as fixed feature extractors to represent high-level spatial patterns without retraining.

5.1.2 Quantitative Model Performance Results

Model performance was evaluated on held-out test data using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). The key quantitative results for each air pollutant and meteorological variable are summarized below.

Satellite to Ground Data Mapping

Sulfur Dioxide (SO_2, ppb) Observation: Multimodal fusion achieved significant improvements on SO_2 predictions against the base-line based on the fixed stations measurements. Addition of spatial features derived from satellite led to a major uplift in R^2 value, this reflects the increased role played by distance and regional dispersion.

Table 5.1: Performance comparison of different models

Model	**RMSE**	**MAE**	R^2
Ground Data	6.94	3.63	-0.03
XGBoost + ResNet18	5.12	2.90	0.44
XGBoost + ResNet50	5.08	2.99	0.45
XGBoost + DenseNet121	5.54	3.07	0.34

Table 5.2: Performance comparison of different models

Model	**RMSE**	**MAE**	R^2
Ground Data	11.78	4.10	0.25
Fusion models	12.22–12.38	–	0.18–0.20

Nitrogen Dioxide (NO_2, ppb) Observation: For NO_2, it is worth noticing that multimodal fusion did not overcome the performance from ground station based baseline. This indicates that NO_2 variation in the dataset is more dominated by local emissions rather than large-scale spatial features observed with satellite.

Table 5.3: Performance comparison of different models

Model	**RMSE**	**MAE**	R^2
Ground Data	15.69	9.90	0.57
XGBoost + ResNet50	15.03	9.34	0.61

Ozone (O_3, ppb) Observation: Some performance boost was achieved for O_3 with spatial features based on ResNet50. This suggests ozone creation and distribution are influenced by regional air quality conditions and spatial land use.

PM2.5 (μg/m³)

Table 5.4: Performance comparison of different models

Model	RMSE	MAE	R^2
Ground Data	20.63	13.93	0.88
Fusion models	21.07–21.19	–	0.87

Observation: For PM2.5, the ground station data model had a slightly better performance than the fusion models. This indicates that the signal of PM2 ground-based measurements and meteorological covariates are already captured by most of the predictors for PM2.5, with marginal augmentation from satellite-derived spatio-features.

Table 5.5: Performance comparison of different models

Model	RMSE	MAE	R^2
Ground Data	27.83	20.86	0.88
Fusion models	28.77–29.80	–	0.86–0.87

PM10 (μg/m³) Observation: Similar to PM2.5, PM10 predictions gained from multi-modal fusion, suggesting that local ground-based effects and near-surface emission processes prevail.

Table 5.6: Performance comparison of different models

Model	RMSE	MAE	R^2
Ground Data	96.60	71.47	-0.01
XGBoost + ResNet18	88.03	65.58	0.16
XGBoost + ResNet50	97.29	72.18	-0.01
XGBoost + DenseNet121	99.43	73.91	-0.04

Solar Radiation (J/m²) Observation: Multimodal fusion meaningfully improved solar radiation predictions only when relying on ResNet18-based spatial features. The increase in R^2 also indicated that shallow spatial dimensions can effectively capture cloud- and surface-related features, while deeper CNN layers failed to add any value in the current configuration.

Relative Humidity (%)

Table 5.7: Performance comparison of different models

Model	**RMSE**	**MAE**	$\boldsymbol{R^2}$
Ground Data	8.28	6.38	0.57
Fusion models	8.90–9.10	–	0.48–0.50

Observation: The fusion model performed consistently below ground station based models for relative humidity. This suggests that moisture variations are already well represented by surface meteorological data set, with a weak contribution from satellite retrievals of the spatial characteristics.

Table 5.8: Performance comparison of different models

Model	**RMSE**	**MAE**	$\boldsymbol{R^2}$
Ground Data	3.06	1.72	0.82
XGBoost + ResNet18	2.03	1.18	0.92
XGBoost + ResNet50	2.47	1.44	0.88
XGBoost + DenseNet121	2.25	1.33	0.90

Rainfall (mm/hr) Observation: Multimodal fusion also brought great betterment to rain prediction. Lowering RMSE by more than 60% and increasing R^2 by the same amount suggest that rainfall relies heavily on spatial context extracted from satellite imagery (i.e. cloud structure, atmospheric patterns).

Table 5.9: Relative performance change of best multimodal fusion model compared to ground-station baseline, sorted by ΔR^2

Variable	Best Fusion Model	R^2 Ground	R^2 Fusion	ΔR^2
SO_2 (ppb)	XGBoost + ResNet50	-0.03	0.45	+0.48
Solar Radiation (J/m^2)	XGBoost + ResNet18	-0.01	0.16	+0.17
O_3 (ppb)	XGBoost + ResNet50	0.57	0.61	+0.04
PM_{10} ($\mu g/m^3$)	Fusion (average)	0.88	0.86	-0.02
$PM_{2.5}$ ($\mu g/m^3$)	Fusion (average)	0.88	0.87	-0.01
NO_2 (ppb)	Fusion (average)	0.25	0.19	-0.06

Regional Clustering

Table 5.10: Silhouette scores for K-Means regional clustering across environmental components. Maximum score per component is shown in bold.

Component	**K=2**	**K=3**	**K=4**
SO_2	**0.0707**	0.0667	0.0478
NO_2	**0.0701**	0.0658	0.0472
CO	**0.0799**	0.0779	0.0794
O_3	**0.0728**	0.0429	0.0487
Rainfall	**0.0767**	0.0632	0.0739
Temperature	0.0652	**0.0695**	0.0509
Relative Humidity	0.0660	**0.0704**	0.0641
Solar Radiation	0.0664	**0.0790**	0.0637

The figure below shows all the regions that show similar behaviors for each components. Clustered using K-Means clustering algorithm.

Figure 5.1: Regional Clustering.

Time Series Forecasting

Table 5.11: Notable SARIMAX forecasting results across environmental targets. Only targets with meaningful or comparatively stable performance are reported.

Target	**Train (mo)**	**Val (mo)**	**Test (mo)**	**RMSE**	R^2
Relative Humidity	29	3	3	14.4200	**0.5177**
SO_2 (ppb)	29	3	3	9.9863	-0.8066

Table 5.12: LSTM forecasting performance across environmental targets (sorted by R^2 in descending order).

Target	**Train (mo)**	**Val (mo)**	**Test (mo)**	**RMSE**	**MAE**	R^2
Relative Humidity$_{\mathrm{2m}}$	29	6	3	12.6023	10.4968	0.6317
SO_2 (ppb)	18	1	1	3.1088	2.5160	0.3262
Solar Radiation (J/m^2)	24	3	6	136.2426	77.0781	0.2843
NO_2 (ppb)	29	3	3	8.5742	4.2185	-0.0092
CO	18	1	6	4.1023	1.7644	-0.0708
Rainfall (mm/hr)	24	1	6	8.8747	4.4540	-0.1361
Temperature (C)	29	3	3	7.6157	7.0951	-0.4005
O_3 (ppb)	29	1	1	11.4219	8.2716	-0.4231

Table 5.13: R^2 comparison of LSTM and BiLSTM forecasting performance across environmental targets.

Target	**LSTM** R^2	**BiLSTM** R^2	$\boldsymbol{\Delta R^2}$
SO_2	0.2893	0.5299	+0.2406
NO_2	0.5287	-0.0244	-0.5531
CO	-0.0902	-0.0655	+0.0247
O_3	-0.2809	-0.4185	-0.1376
Rainfall	-0.1174	-0.0972	+0.0202
Temperature	-0.0536	0.1409	+0.1945
RH_{2m}	0.5318	0.5353	+0.0034
SSRD	0.3315	0.4610	+0.1295

Explainable AI on LSTM

Following graphs show how Integrated Gradients can be used to find the most significant features to predict each components.

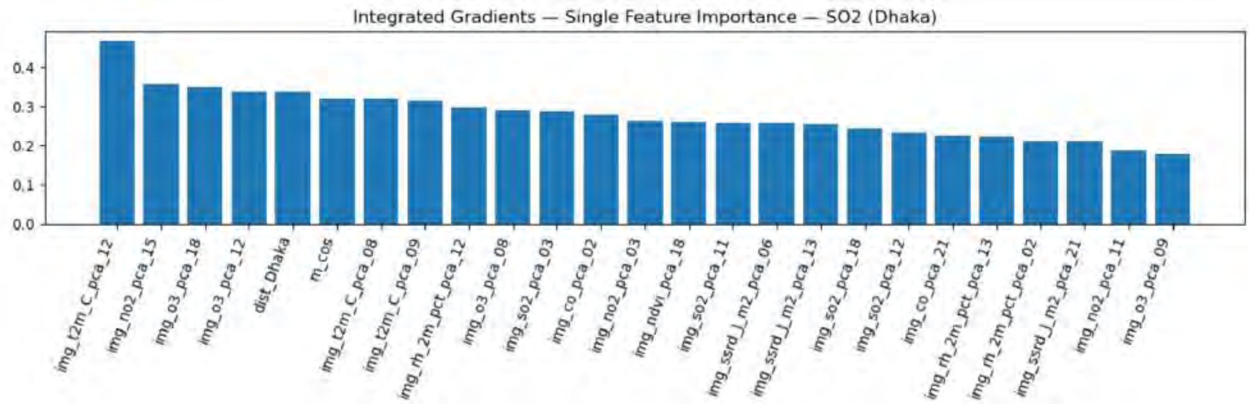

Figure 5.2: XAI So2(Dhaka).

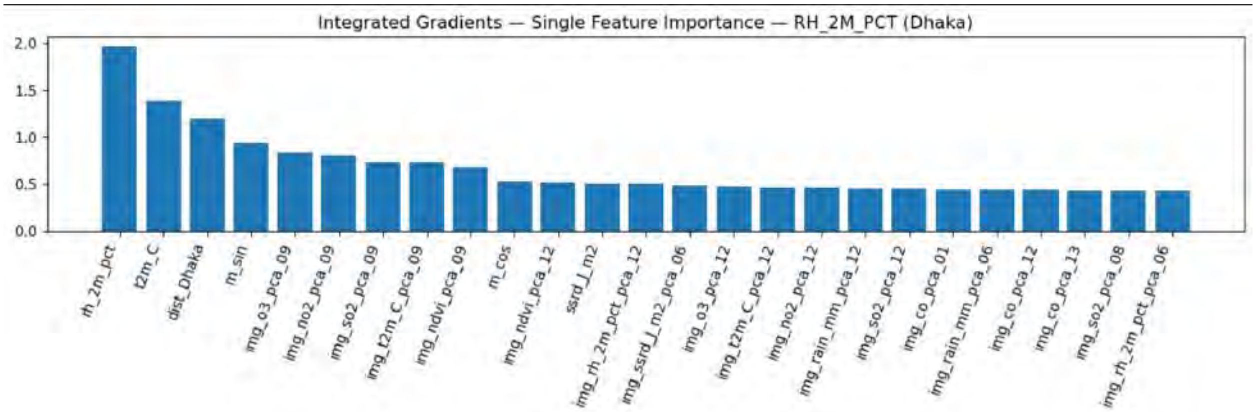

Figure 5.3: Relative Humidity(Dhaka).

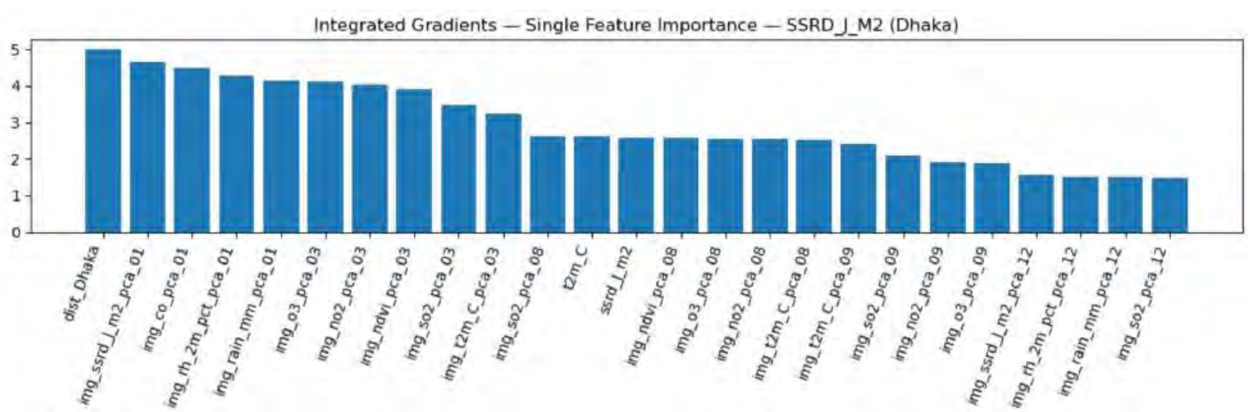

Figure 5.4: Solar Sun Radiation(Dhaka).

Overall Summary

In general, the quantitative results indicate that the effectiveness of multimodal fusion is quite target-dependent. Substantially decreased prediction errors and improved explained variances were demonstrated for SO_2 and rainfall, which reflected spatial feature fusion is of great advantage. O_3 and solar radiation indicated "moderate" or architecture-dependent enhancements, indicative of the partial contribution of regional atmospheric dynamics. In contrast, NO_2, PM2.5, PM10 and relative humidity were most effectively estimated based only on ground station data, indicating predominance of local emission sources and near-ground meteorological parameters. Results suggest that multimodal approaches improve predictive performance only for spatially patterned variables, while ground-based measurements are adequate for targets governed by local dynamics.

5.2 Analysis of Design Solutions

This paper evaluates design solutions adopted for the proposed environmental monitoring framework. The comparison considers architectural decisions, data integration approaches, model selection rules and its effect on predictive power for several pollutants. Instead of focusing on the numerical performance, this section systematically assesses each proposed decision with respect to different nature of environmental data.

5.2.1 Effectiveness of Multimodal Data Integration

Multimodal fusion is often the most effective when there exists strong spatial dependence in our target variable. One-size-fits-all combination approach would be too complex for pollutants, which are controlled by local dynamics.

In this work, a major design choice was to combine spatial features derived from satellite with on-ground environmental data. This multimodal approach was inspired by the assumption that satellite imagery can describe large-scale spatial patterns—such as regional pollution transport, land use type, and atmospheric condition—that ground sensor measurements cannot fully present.

The findings suggest this design choice is selectively successful. For pollutants that have regional dispersal and atmospheric transport characteristics like SO_2 and O_3, including satellite features brought significant enhancements in performances.The large increase in R^2 in SO_2 implied that the spatial context were very important to predict its concentration.

On the other hand, for pollutants such as PM2.5, PM10 and NO_2 the multimodal approach did not perform better than the Ground station based data baseline. This implies that these pollutants may be more sensitive to local sources, episodic variability (on shorter time scales), and near-surface properties well-represented by ground-based measurements. Therefore, the added spatial information from satellite imagery was of little additional predictive value for these targets.

5.2.2 Choice of CNN-Based Feature Extractors

Deeper CNNs are useful only when the spatial heterogeneity is a significant factor to the prediction learning. Models should therefore match the complexity to the pollutant behaviour, rather than uniformly over-applied. We adopted the fixed feature extraction on 'off-the-shelf' pre-trained convolutional neural networks (ResNet-18, ResNet-50 and DenseNet-121) to make use of learned spatial representations without sacrificing end-to-end training's computations.

Among the tested network configurations, ResNet-50 demonstrated good performance for spatially dependent pollutants like SO_2 and O_3. Its deeper structure can extract more expressive spatial features that may be responsible for outperforming shallower architectures in some cases.Although DenseNet-121 was successful in learning fine-grained patterns, the performance of its variants was not always superior to ResNet ones, and this is likely because extracted features become redundant when used with tabular data.

Nevertheless, for pollutants with less important spatial influence, the CNN architecture selection did not affect significantly the performance. This points to a law of diminishing (integer) returns for increasingly complex CNN if spatial factors are not the predominant reason behind scoring.

5.2.3 Evaluation of XGBoost as the Fusion Model

For ground station based data and the fusion experiments, we have used XGBoost as the main regression model, because it performs well on structured data, is capable of modeling nonlinear relationships and is robust to feature interactions. XGBoost is a powerful fusion backbone, however the performance greatly depends on input features relevance and quality. When integrating heterogeneous data sources, feature selection or dimensionality reduction may be applied.

The conclusions are that XGBoost is an appropriate modeling tool in the domain of environmental modelling, especially when facing a complex and non-linear data structure. Its performance on PM2.5 and PM10 based on ground elements was particularly pronounced and can be characterized as strong, obtaining a high R^2 even without any spatial extension.

However, the performance improvements of XGBoost over high-dimensional CNN features varies for different pollutants. Although it well utilized spatial features for SO_2 and O_3, the higher dimension of features did not bring along performance improvement for all the targets. This exposes a tension between expressiveness of the model and relevancy of the features.

XGBoost is a strong feature combination backbone that depends on relevance and quality of the input features. In case of heterogeneous data, feature selection or dimensionality reduction could be required when the data sources are being combined.

5.2.4 Effectiveness of Regional Clustering

Regional clustering was employed to identify districts exhibiting similar long-term environmental behavior, thereby enabling spatial generalization beyond point-wise regression. For each component, a single district-level representation was constructed by aggregating observations across all months and years into summary statistics (mean and standard deviation). These summaries combined the tabular component values, when available, with the corresponding satellite-derived PCA embedding features. The resulting district-level vectors were standardized and clustered using the KMeans algorithm, yielding groups that can be interpreted as distinct environmental regimes—that is, districts characterized by comparable pollution–meteorology patterns.

To avoid an arbitrary choice of the number of clusters, cluster validity was evaluated using the silhouette score across a range of candidate values of k, and the value that best balanced within-cluster cohesion and between-cluster separation was selected. This approach ensures that the clustering captures persistent spatial structure rather than short-term temporal variability, and it is particularly well suited for variables that exhibit coherent regional behavior.

5.2.5 Evaluation of Time-Series Forecasting

Time series forecasting used fixed lookback window of 12 months such that any training can train on at least a full season of data. Chronological splitting was used to ensure that no future data was leaked during preprocessing. Train-validation-test splitting was variable for each components.

Table 5.14: Statistical Comparison of Forecasting Models

Target	**LSTM**		**BiLSTM**		**SARIMA**		**BiLSTM − LSTM**	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	ΔRMSE	ΔR^2
SO_2	3.19	0.29	**2.60**	**0.53**	9.99	−0.81	−0.60	+0.24
NO_2	**5.86**	**0.53**	8.64	−0.02	145.96	−291.45	+2.78	−0.55
CO	4.14	−0.09	**4.09**	**−0.07**	$\gg 10^6$	$\ll 0$	−0.05	+0.02
O_3	**10.84**	**−0.28**	11.40	−0.42	513.46	−2799.33	+0.57	−0.14
Rainfall	8.80	−0.12	**8.72**	**−0.10**	427.98	−2402.41	−0.08	+0.02
Temperature	5.58	−0.05	**5.04**	**0.14**	10.84	−1.84	−0.54	+0.19
Humidity	14.21	0.53	**14.16**	**0.54**	14.42	0.52	−0.05	+0.00
SSRD	138.17	0.33	**124.07**	**0.46**	942.02	−39.62	−14.10	+0.13

Across all components performance of SARIMA was overshadowed by LSTM and Bi directional LSTM as the data used was highly nonlinear. Forecasting performance varied across components, with stronger results for variables exhibiting smooth seasonal structure, notably SO_2, RH_{2m}, and SSRD.

5.2.6 Explainability Analysis of LSTM Forecasting Using Integrated Gradients

To improve interpretability, Integrated Gradients (IG) was applied to the LSTM forecasting model. IG attributes predict sensitivity to input features. It does that by integrating gradients from a baseline to the observed input. It is ideal for neural time-series models.

Usage of IG significantly increased the readability of the model and ensured that any policy makers can target important features for every components.

5.2.7 Impact of Temporal Variability and Data Distribution

Environmental records are inherently temporal, with pronounced seasonal behavior and long-term trends interspersed with sudden jumps associated with weather or human activities. Although the present architecture is mainly based on spatial and cross-sectional learning, one limitation is temporal variation. Due to the lack of explicit temporal modeling, the framework is not capable of fully accounting for seasonal and long-term dynamics. Just including time-aware features or sequences of such could improve robustness and generalization.

The performance degradation we saw for some pollutants could be due, at least in part, to the distributional change and seasonal effect. Modeling on unrolled time windows and without straightforward temporal encoding may have difficulty generalizing across seasonal regimes (eg, for pollutants very sensitive to meteorology).

5.2.8 Computational Trade-offs and Scalability

Computational efficiency is another crucial design factor. Training time and data dimensionality are greatly increased when CNN-extracted features are included. This expense may not be justified for targets where ground station-based data models already perform well, but it is justified for pollutants where fusion produces significant gains.

This emphasizes the necessity of adaptive design strategies, in which multimodal models are not applied consistently to all pollutants but rather selectively based on cost-benefit analysis.

5.2.9 Overall Assessment of Design Choices

In general, it can be seen that the design choices made in this framework are closely related to the physical properties of environmental data. It also combines satellite with ground observation and ensemble learning model, making it an elastic air quality analysis platform.

Design-solutions and their effects are presented as evidence that the proposed model is technically feasible and applicable in practice. While multi-modal fusion can enhance the prediction in certain scenarios. These findings indicate a clear path for future improvement such as feature selection, and selective multimodal employment.

5.3 Statistical Analysis

The predictive performance of the suggested models is statistically analyzed in this section using quantitative error metrics derived from the experimental findings. The analysis focuses on error magnitude, variance behavior, and comparative performance across models.

5.3.1 Analysis of Comparative Error Magnitude

RMSE, MAE, and R^2 values calculated on held-out test datasets were used to statistically compare the ground only dataset and multimodal fusion models.

The ground data had poor explanatory power for SO_2, with an RMSE of 6.935 ppb and a R^2 of 0.0285. The multimodal fusion models, on the other hand, significantly decreased prediction error. With an RMSE of 5.077 ppb and an R^2 of 0.45, the XGBoost + ResNet50 model reduced RMSE by about 27% and significantly increased explained variance.

The baseline for O_3 was RMSE 15.6937 ppb and R^2 0.5718, whereas XGBoost + ResNet50 improved R^2 to 0.6075 and decreased RMSE to 15.0251 ppb. The consistent improvement across RMSE/MAE and R^2 indicates improved fit rather than isolated metric fluctuation, despite the moderate absolute RMSE reduction ($\approx 4.3\%$).

Fusion model did not reduce error for particulate pollutants. For PM2.5, R^2 dropped from 0.8781 to 0.8714–0.8729 while baseline RMSE 20.6277 μg/m^3 increased under fusion (e.g., 21.0677–21.1947 μg/m^3). For PM10, R^2 decreased from 0.8794 to 0.8617–0.8711 and baseline RMSE 27.8327 μg/m^3 increased under fusion (28.7664–29.7983 μg/m^3). These trends show that, in the present configuration, satellite-derived spatial embeddings did not lower statistical error for particulate matter.

For relative humidity, the baseline RMSE was 8.2790 with R^2 0.5703, and all fusion variants degraded performance RMSE 8.8997–9.0971, R^2 0.4812–0.5035. This suggests that humidity in this dataset is already well captured by ground or auxiliary variables and is not improved by the extracted spatial embeddings.

The baseline RMSE for solar radiation was 96.5968 with R^2 0.0073. Fusion performance depended strongly on the CNN backbone. XGB + ResNet18 improved RMSE to 88.0347 and increased R^2 to 0.1634, while ResNet50 and DenseNet121 remained near or below baseline (R^2 0.0130 and 0.0401, respectively). This indicates that spatial features can add value for solar radiation, but the gain is architecture-dependent.

For Rainfall the baseline RMSE was 3.0614 with R^2 0.8195, while fusion models improved substantially. XGB + ResNet18 achieved RMSE 2.0287 and R^2 0.9207, with ResNet50 and DenseNet121 also improving relative to baseline (R^2 0.8839 and 0.9046). This indicates that rainfall benefits strongly from spatial context (e.g., cloud/atmospheric texture proxies captured in satellite imagery).

For regional clustering K=2 and k=3 showed best silhouette scores. SO_2 had best k=2 silhouette score of 0.0707, NO_2 had best k=2 silhouette score of 0.0701, CO had best k=2 silhouette score of 0.0799 O_3 had best k=2 silhouette score of 0.0728 and Rainfall had best k=2 silhouette score of 0.0767. Whereas Temperature had best k=3 silhouette score of 0.0695, Relative Humidity had best k=3 silhouette score of 0.0704 and Solar Radiation had best k=3 silhouette score of 0.0790.

For time series forecasting SARIMAX had only one meaningful results Relative Humidity had 0.5177 R^2 value. LSTM showed few good results Relative Humidity had 0.6317 R^2 value, SO_2 had 0 .3262 R^2 value and Solar Radiation had 0.2843 R^2 value meaning that they can be safety forecasted with out current dataset.

XAI implementation using Integrated Gradients showed how much each features are contributing for predicting each components. For sample of Dhaka city while predicting SO_2 images of temperature and NO_2 had the most impact. Whereas for relative humidity of Dhaka city ground value of temperature, the month and district name had most impact. Similarly for solar radiation the district name, images of solar radiation and image of CO had most impact.

5.3.2 Bias and Variance Behavior

Bias–variance characteristics are inferred from comparative generalization trends rather than explicit decomposition. For SO_2 and rainfall, the substantial reductions in test RMSE and simultaneous increases in R^2 (SO_2: $0.0285 \rightarrow 0.4488$, rainfall: $0.8195 \rightarrow 0.9207$) suggest improved generalization and reduced error dispersion. For O_3, the smaller but consistent metric improvements also indicate better fit.

For NO_2, PM2.5, and PM10, fusion increased RMSE and reduced R^2, suggesting that additional spatial embeddings introduced variance (or noise) without adding commensurate explanatory signal.

For solar radiation, variance behavior appears model-dependent: ResNet18 improves generalization, while deeper backbones do not, implying that the usefulness of spatial embeddings is sensitive to representation choice. For relative humidity, the consistent degradation across fusion variants indicates that the added spatial features likely increase variance without improving signal.

5.3.3 Pollutant-Specific Statistical Characteristics

Across targets, distinct statistical patterns emerge:

Strong positive fusion effect: SO_2, rainfall (large RMSE reductions and large R^2 gains).

Moderate positive fusion effect: O_3 (modest RMSE reduction with clear R^2 increase).

Architecture-dependent effect: solar radiation (improves with ResNet18; not with ResNet50/DenseNet).

No benefit / negative effect: NO_2, PM2.5, PM10, relative humidity (baseline remains superior).

These trends confirm that multimodal gains depend on whether the target contains a strong spatial component that is captured by the satellite-derived embeddings.

5.3.4 Paired Model Comparison and Significance Interpretation

Although formal hypothesis testing (e.g., paired t-tests or bootstrapped confidence intervals) was not explicitly conducted, paired comparisons of test-set RMSE and MAE between baseline and fusion models provide indicative evidence of performance differences.

For SO_2, O_3, and rainfall, improvements appear consistently across RMSE/MAE and R^2, supporting the interpretation of robust and practically meaningful performance gains. For solar radiation, improvements occur only for the ResNet18 fusion model, implying that gains are plausible but not uniform across architectures. For NO_2, PM2.5, PM10, and relative humidity, metric trends are inconsistent or negative, indicating no statistically meaningful advantage under the current modeling setup.

From a statistical interpretation perspective, only the improvements observed for SO_2—and to a lesser extent O_3—can be considered robust and practically significant.

5.3.5 Reliability and Uncertainty Considerations

The overall statistical evidence suggests that the fusion framework improves reliability for targets with strong spatial structure (SO_2, rainfall, and partially O_3), but does not improve and can worsen performance for locally dominated or highly variable targets NO_2 and particulate matter and for relative humidity. These results emphasize that uncertainty-aware deployment should be pollutant/variable-specific rather than uniform across all targets.

5.4 Comparisons and Relationships

This section examines comparative relationships between models, data modalities, pollutants, and spatial characteristics. The comparisons are based on test-set RMSE, MAE, and R^2 values reported in Section 5.1 and statistically analyzed in Section 5.3.

5.4.1 Comparison Between Ground-Based and Multimodal Models

A central comparison in this study is between the ground data XGBoost models and the multimodal fusion models that integrate satellite features. The comparative results can be summarized as follows:

Table 5.15: Effect of Multimodal Fusion on Prediction Performance

Target	**Ground Only**		**Fusion Model**		**RMSE Trend**	**R^2 Trend**
	RMSE	R^2	RMSE	R^2		
SO_2 (ppb)	6.94	−0.0285	**5.08**	**0.4488**	↓ Significant	↑ Large
O_3 (ppb)	15.69	0.5718	**15.03**	**0.6075**	↓ Moderate	↑ Moderate
NO_2 (ppb)	**11.78**	**0.25**	>12.0	<0.20	↑ Degradation	↓ Degradation
PM2.5 (μg/m^3)	**20.63**	–	∼21.10	–	↑ Slight	–
PM10 (μg/m^3)	**27.83**	–	28.77–29.80	–	↑ Degradation	–
Solar Radiation	–	0.0073	–	**0.1634** (ResNet18)	Architecture-dependent	↑ Selective

Ground data remains superior for NO_2, PM2.5, PM10, and relative humidity, where fusion reduces R^2 and increases RMSE

This comparison demonstrates that multimodal fusion provides clear benefits only for select pollutants and does not universally improve prediction accuracy.

5.4.2 Relationship Between Pollutant Characteristics and Model Effectiveness

A strong relationship is observed between pollutant behavior and the effectiveness of multimodal modeling. Pollutants such as SO_2 and O_3 show improved performance when satellite features are included.

In contrast, PM2.5, PM10, and NO_2 are primarily driven by local ground sources. For these pollutants, ground-based measurements already capture most of the predictive signal, and the addition of satellite-derived features introduces little to no new information.

This relationship confirms that the success of multimodal learning is dependent more on pollutants change than by model complexity.

5.4.3 Comparison Across CNN Feature Extractors

Comparisons among ResNet-18, ResNet-50, and DenseNet-121 reveal that deeper CNN architectures generally perform better for geography influenced pollutants. ResNet-50 consistently achieved the lowest RMSE and highest R^2 values for SO_2 and O_3 among the tested architectures.

However, for pollutants where geographic influence is weak, performance differences between CNN architectures were minimal. This indicates that increasing architectural depth does not yield benefits unless geographic features are strongly correlated with the target pollutant.

5.4.4 Relationship Between Error Metrics and Model Interpretation

A consistent relationship was observed between RMSE, MAE, and R^2 across experiments. For SO_2 and O_3, reductions in RMSE were accompanied by increases in R^2, indicating genuine improvements in model fit.

For PM2.5, PM10, and NO_2, small changes in RMSE or MAE were often associated with drop of R^2.

5.4.5 Spatial and Regional Performance Relationships

Across targets, performance is affected by regional unevenness and station density. Districts with richer ground measurement support tend to yield more stable results for ground models, particularly for PM2.5/PM10 and relative humidity. Where satellite image is informative (SO_2, rainfall), image-derived features can improve generalization. However, dense ground rich data are still more valuable.

5.4.6 Consolidated Comparative Summary

Multimodal fusion is beneficial for SO_2, rainfall, moderately effective for O_3. On the other hand for PM2.5, PM10, and NO_2 its kind of ineffective.

Pollutant physical behavior strongly determines whether satellite image based features improve prediction or not. Deeper models out perform Classical statistical model for forecasting. Ground-based data remain critical for better accuracy and reliability.

5.5 Final Adjustment

5.5.1 Final Model Adjustment via Target Normalization

Although LSTM demonstrated stable forecasting performance for some components, additional refinement was applied as a final adjustment step to improve training stability and predictive accuracy. Specifically, target-wise normalization was introduced prior to LSTM training.

Environmental variables in this study (e.g., SO_2, NO_2, rainfall, surface solar radiation) exhibit heterogeneous numerical scales and variance structures. The target variable for each forecasting was standardized using StandardScaler computed on the training set:

$$y' = \frac{y - \mu_{\text{train}}}{\sigma_{\text{train}}}$$

where μ_{train} and σ_{train} denote the mean and standard deviation of the training targets, respectively.

The LSTM model was trained to predict the normalized target values. Afterwards inverse transformation was applied to all validation and test predictions before evaluation. Table 5.16 summarizes the comparative results before and after applying target normalization.

Table 5.16: Impact of Target Normalization on LSTM Forecasting Performance

Target	**Before Normalization**			**After Normalization**		
	RMSE	MAE	R^2	RMSE	MAE	R^2
SO_2	3.1088	2.5160	0.3262	**2.5562**	**1.9380**	**0.5445**
NO_2	8.5742	4.2185	-0.0092	**7.3031**	**4.1642**	**-3.8642**
CO	4.1023	1.7644	-0.0708	**4.0899**	1.7724	-0.0643
O_3	11.4219	8.2716	-0.4231	**11.2110**	**7.7748**	**-0.3711**
Rainfall (mm)	8.8747	4.4540	-0.1361	**8.7044**	**4.2729**	**-0.0929**
T2M (°C)	7.6157	7.0951	-0.4005	**5.4589**	**3.6850**	**-0.0087**
RH (%)	12.6023	10.4968	0.6317	**12.5922**	**9.0326**	**0.6322**
SSRD (J/m^2)	136.2426	77.0781	0.2843	**131.6851**	92.6032	**0.3928**

5.6 Discussion

The experimental results confirm that multimodal fusion has a pollutant dependent effect, instead of providing universally better performance. Pollutants with more pronounced spatial and atmospheric transport signatures (SO_2, O_3) demonstrated evident advantages of satellite-derived predictors. In contrast, particulate matter (PM2.5 and PM10) along with NO_2 were more accurately modeled using only ground-based data.

These results emphasize the importance of pollutant-specific modeling strategies and the fact that multimodal approaches should be selectively used depending on the particle physical and chemical properties.

In general, the performance assessment supported that combining satellite images and ground-based observations can contribute to considerable improvements in predictive capability of certain air pollutants. Despite the increased computation and lack of consistent improvement over ground only models, multimodal fusion resulted in marked improvements for several critical gaseous pollutants. These findings confirm the efficacy of the proposed framework.

Chapter 6
Conclusion

A multimodal environmental monitoring framework is introduced by this thesis. The framework integrates satellite derived features with ground based measurements. Both classical and deep learning techniques were incorporated. The primary target was to evaluate the justification of this merge between two different types of data for prediction of air quality and meteorological variables. The results of the experiment indicate that multimodal fusion's effectiveness is strongly pollutant dependent. Significant gains in performance were noted for variables and pollutants with strong atmospheric and geographic structure. Specifically, the inclusion of satellite images resulted in significant increases in explained variance and significant decreases in prediction error for rainfall and sulfur dioxide (SO). Satellite imagery also provided moderate or conditional benefits for solar radiation and ozone (O), indicating a partial reliance on local atmospheric conditions and cloud–surface interactions. In contrast, nitrogen dioxide (NO), particulate matter (PM2.5 and PM10), and relative humidity were best modeled using ground station based data alone. For these variables, multimodal fusion did not improve and in some cases slightly degraded—predictive performance. This outcome suggests that these targets are primarily governed by close to ground emission sources that are already well captured by ground-based measurements. These findings were further supported by the statistical analysis, which demonstrated that increases in R^2 and consistent decreases in RMSE and MAE accompanied improvements in SO, rainfall, and O. On the other hand, fusion models showed higher variance for variables driven by ground data without commensurate increases in explanatory power. These results demonstrate that improved predictive performance is not always a result of more complex models. All things considered, this study shows that when used carefully and in accordance with the physical properties of the target variable, multimodal learning can significantly improve environmental prediction. The suggested framework emphasizes the significance of pollutant- and variable-specific model pipelines while offering a flexible framework for integrating data sources.

6.1 Key Contributions

The main contributions of this thesis can be summarized as follows:

1. **Multimodal Environmental Prediction Framework:** Creation of a multimodal environmental forecasting framework which combines satellite data and ground-based data based on CNN-driven feature extraction and XGBoost regression.

2. **Comprehensive Multivariable Evaluation:** Thorough analysis of a range of air pollutants and weather conditions, such as SO_2, NO_2, O_3, $PM_{2.5}$, PM_{10}, solar radiation, relative humidity, and rainfall.

3. **Target-Dependent Multimodal Effectiveness:** Evidence of target-specific effects of fusion of multimodes, indicating significant performance improvement on spatially affected variables and only small changes in locally-favored variables.

4. **Statistically Grounded Performance Analysis:** Statistically based analysis of model performance, with a focus on error behavior, trends of variance and predictive reliability, as opposed to the use of aggregate measures of accuracy.

5. **Guidelines for Adaptive Model Deployment:** Hands-on lessons regarding the use of adaptive models, concerning the situations when the multimodal approach can be justified, and the situations when a plain ground-based model can be used.

6.2 Limitations

This study has a number of limitations even though it has made contributions. First, the models mainly concentrate on cross-sectional learning and they do not explicitly include temporal sequence modeling that can constrain robustness to strong seasonal variability as well as long term tendencies. Second, CNN-based image features were not fine-tuned but through pre-trained models, so they could be limited to capturing specific patterns of the atmosphere associated with the domain. Third, the formal statistical significance testing and the quantification of uncertainty was not fully exploited which interferes with the ability to provide a probabilistic interpretation of predictions. Lastly, performance is vulnerable to the availability of data and the density of the monitoring of the region of interest, especially in the case of particulate matter and humidity.

6.3 Future Work

Several directions for future research emerge from this work:

1. **Temporal Modeling:** Time aware models like GRU or graph models might improve performance for seasonal forecasting.

2. **Feature Optimization and Selection:** The variance caused by high-dimensional spatial embeddings could be reduced by applying features selection or attention mechanism and enhance fusion efficiency.

3. **Uncertainty Quantification:** Probabilistic modeling, or the use of ensemble based uncertainty estimation, could be part of future work to give confidence ranges to point predictions.

4. **Adaptive Deployment Strategies:** Selective approach is identified depending on the component, a pollutant aware model pipeline may be useful and more practical in the long-term.

5. **Expanded Data Sources:** Spatial representations may be enhanced by the addition of more satellite products, finer-resolution images, or mobile sensing data, which may enhance generalization.

6.4 Final Remarks

To sum up, it can be concluded that multimodal machine learning has a good potential of providing environmental monitoring in case it is used wisely. The proposed framework will provide significant performance benefits with a necessary complexity level by basing model design on the physical and statistical characteristics of environmental variables. The experience of this work can be added to the expanding amount of literature on data-driven environmental models and serve as a viable base in the deeds of future developments in air quality forecasts and environmental decision support.

Bibliography

[1] World Bank, "Bangladesh environmental report," 2023, accessed: 2025-06-16. [Online]. Available: https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099510110202254063/p1767570f316560ae08d1b0e29aef0a3f14

[2] IQAir, "World air quality report," 2023, accessed: 2025-06-16. [Online]. Available: https://www.iqair.com/world-air-quality-report

[3] The Daily Star, "79 rivers dead or on deathbed," 2023, accessed: 2025-06-16. [Online]. Available: https://www.thedailystar.net/news/bangladesh/news/79-rivers-dead-or-deathbed-3852851

[4] J. Du, "The temporal and frequent pattern mining analysis and machine learning forecasting on mobile sourced urban air pollutants," Ph.D. dissertation, Texas Southern University, 2021, accessed: 2026-01-25. [Online]. Available: https://core.ac.uk/reader/542922313

[5] J. Pugliese, A. Folini, D. Carrion, and M. A. Brovelli, "Hazard susceptibility mapping with machine and deep learning: A literature review," *Remote Sensing*, vol. 16, no. 18, pp. 3374–3374, Sep. 2024, accessed: 2026-01-25. [Online]. Available: https://doi.org/10.3390/rs16183374

[6] V. J. Berrocal *et al.*, "A comparison of statistical and machine learning methods for creating national daily maps of ambient pm2.5 concentration," *arXiv preprint arXiv:1904.08931*, 2019, accessed: 2026-01-25. [Online]. Available: https://arxiv.org/pdf/1904.08931

[7] E. S. A. Abri, "Modelling atmospheric ozone concentration using machine learning algorithms," Ph.D. dissertation, Loughborough University, 2020, accessed: 2026-01-25. [Online]. Available: https://core.ac.uk/reader/288368404

[8] S. Cheng, M. N. Blanco, L. Sheppard, A. Szpiro, and A. Shojaie, "Variable importance measure for spatial machine learning models with application to air pollution exposure prediction," Department of Biostatistics, University of Washington and Department of Environmental & Occupational Health Sciences, University of Washington, Tech. Rep., 2023, accessed: 2026-01-25. [Online]. Available: https://arxiv.org/abs/2406.01982

[9] G. M. I. Alam *et al.*, "Deep learning model based prediction of vehicle co emissions with explainable ai integration for sustainable environment," *Scientific*

Reports, vol. 15, no. 1, Jan. 2025, accessed: 2026-01-25. [Online]. Available: https://doi.org/10.1038/s41598-025-87233-y

[10] T. P. T. Quynh, T. N. Viet, H. D. Thi, and K. H. Manh, "Enhancing air quality prediction accuracy using hybrid deep learning," *International Journal of Environmental Science and Development*, vol. 14, no. 2, pp. 155–159, 2023, accessed: 2026-01-25. [Online]. Available: https://doi.org/10.18178/ijesd.2023.14.2.1428

[11] A. Houdou *et al.*, "Interpretable machine learning approaches for forecasting and predicting air pollution: A systematic review," *Aerosol and Air Quality Research*, vol. 24, no. 1, p. 230151, 2024, accessed: 2026-01-25. [Online]. Available: https://doi.org/10.4209/aaqr.230151

[12] L. Mesin, F. Orione, and E. Pasero, "Nonlinear adaptive filtering to forecast air pollution," in *Adaptive Filtering Applications*. InTech, 2011, accessed: 2026-01-25. [Online]. Available: https://doi.org/10.5772/18191

[13] S. Chakraborty, B. Misra, and N. Dey, "Explainable artificial intelligence (xai) for air quality assessment," https://www.researchgate.net/publication/378110823, 2024, accessed: 2025-06-19.

[14] A. Samad, S. Garuda, U. Vogt, and B. Yang, "Air pollution prediction using machine learning techniques – an approach to replace existing monitoring stations with virtual monitoring stations," *Atmospheric Environment*, vol. 310, p. 119987, 2023, accessed: 2026-01-25. [Online]. Available: https://doi.org/10.1016/j.atmosenv.2023.119987

[15] R. K. Singh, S. Raghav, T. Maini, M. K. Singh, and M. Arquam, "Air quality prediction using machine learning," Proc. Advancements in Electronics & Communication Engineering 2022, Jul. 14, 2022, 2022, accessed: 2025-06-19. [Online]. Available: https://ssrn.com/abstract=4157651

[16] M. M. Rahman, E. H. Nayeem, S. Ahmed *et al.*, "Airnet: Predictive machine learning model for air quality forecasting using web interface," *Environmental Systems Research*, vol. 13, p. 44, 2024, accessed: 2025-06-19. [Online]. Available: https://environmentalsystemsresearch.springeropen.com/articles/10.1186/s40068-024-00378-z

[17] M. S. S. Sunder, V. A. Tikkiwal, A. Kumar, and B. Tyagi, "Unveiling the transparency of prediction models for spatial pm2.5 over singapore: Comparison of different machine learning approaches with explainable artificial intelligence," *AI*, vol. 4, no. 4, pp. 787–811, 2023, accessed: 2025-06-19. [Online]. Available: https://doi.org/10.3390/ai4040040

[18] R. Stirnberg, J. Cermak, S. Kotthaus, M. Haeffelin, H. Andersen, J. Fuchs, M. Kim, J.-E. Petit, and O. Favez, "Meteorology-driven variability of air pollution (pm1) revealed with explainable machine learning," *Atmospheric Chemistry and Physics*, vol. 21, pp. 3919–3948, 2021, accessed: 2025-06-19. [Online]. Available: https://doi.org/10.5194/acp-21-3919-2021

[19] A. Talaie, H. Kamyab, and A. Razmfarsa, "Advancing environmental engineering: The role of artificial intelligence in sustainable solutions - a short review," *Journal of Environmental Treatment Techniques*, vol. 12, no. 3, pp. 14–18, 2024, accessed: 2025-06-19. [Online]. Available: https://digitalcommons.odu.edu/management_fac_pubs/73

[20] A. R. M. T. Islam, M. A. Awadh, J. Mallick, S. C. Pal, R. Chakraborty, M. A. Fattah, B. Ghose, M. K. A. Kakoli, M. A. Islam, H. R. Naqvi, M. Bilal, and A. Elbeltagi, "Estimating ground-level pm2.5 using subset regression model and machine learning algorithms in asian megacity, dhaka, bangladesh," *Air Quality, Atmosphere & Health*, vol. 16, pp. 1117–1139, 2023.

[21] A. Ahmed *et al.*, "Comparison between local and global methods to develop aqi in representing the spatial pattern of air quality of dhaka city," *Dhaka University Journal of Earth and Environmental Sciences*, vol. 11, no. 1, pp. 131–149, 2023.

[22] S. Hameed *et al.*, "Deep learning-based multimodal urban air quality prediction and traffic analytics," *Scientific Reports*, vol. 13, no. 1, p. 22181, 2023.

[23] J. Kalajdjieski *et al.*, "Air pollution prediction with multi-modal data and deep neural networks," *Remote Sensing*, vol. 12, no. 24, p. 4142, 2020.

[24] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2016, pp. 785–794.

[25] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," *arXiv preprint arXiv:1703.01365*, 2017.

[26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems (NeurIPS)*, 2012, pp. 1097–1105.

[27] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance," *Climate Research*, vol. 30, no. 1, pp. 79–82, 2005.