ABSTRACT

Air pollution is a ubiquitous issue in the contemporary world with PM2.5 being one of the most dangerous pollutants because it can penetrate deep into the blood and lungs. Its measurement and forecast is important for environmental policy, public health, and early warning systems. This project emphasizes the monitoring of air quality through deep learning models for forecasting PM2.5 levels based on past air quality and weather conditions. The goal is to find the most precise and reliable model for real-time air pollution forecasting.

The project utilizes two time series datasets: one of the world capital cities' weather and air quality, and another of air quality measurement from Delhi. The datasets contain different features like temperature, humidity, wind speed, and concentration of pollutants. Preprocessing involved working with missing data, using scaling mechanisms like MinMaxScaler, StandardScaler, and RobustScaler, and feature extraction of temporal attributes like day, month, and year from date columns. Time series framing was done through sliding window methods, and positional encoding was specifically utilized in models based on Transformers. Models used are LSTM, CNN, Bi-LSTM, Vanilla Transformer, and hybrid models like CNN-Transformer and LSTM-Transformer.

The performance of the models can be assessed by using metrics like Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared (R^2) score. Visual aids such as loss over epochs graphs, actual vs. predicted value plots, and scatter plots were employed to analyze model behaviour and detect overfitting or underfitting. Among all the models that had been experimented with, the hybrid LSTM-Transformer model was the most promising one, capitalizing on the sequential learning capability of LSTM and the attention mechanism of Transformers to efficiently detect temporal patterns and long-term dependencies within the data.

This project proves that deep learning architectures, particularly hybrid models, can improve the accuracy of PM2.5 predictions quite extensively and contribute to air quality monitoring activities. These models can also be extended further for real-world deployment and inclusion in smart city systems. The project was deployed with Python and TensorFlow within the Jupyter Notebook system.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1. INTRODUCTION

This chapter presents the subject matter of air quality monitoring and describes the context, motivation, goals, and organization of the project. It explains the contemporary relevance of measuring air pollutants such as PM2.5, gives a cursory description of current methodologies, and identifies the unique approach employed by this project. It also explains why the results of this project are significant and the organization of the report.

1.2. Introduction to the Area of Work

Air pollution is a key global environmental issue, particularly in urban and industrial regions. Air quality monitoring has gained prominence to evaluate environmental situations, increase public awareness, and make policy decisions. Fine particulate matter, or PM2.5, is a key pollutant that has significant influences on human health, especially respiratory and cardiovascular functions. PM2.5 prediction from past data can assist the authorities in taking prompt measures to counteract its influence.

1.3. Present Day Scenario

Most cities in the world today consistently surpass the safe PM2.5 thresholds set by environmental authorities. Owing to advances in sensing devices and the extent of available data, vast amounts of air quality data are now available. Still, conventional statistical approaches are not usually equipped to deal with the intricate temporal relationships present in such data. Deep learning models provide an attractive alternative by learning from past data to make precise projections of future pollutant concentrations.

1.4. Motivation to Do the Project Work

While a number of studies have investigated air quality forecasting with traditional machine learning techniques or single deep models such as LSTM, they tend not to generalize across a variety of datasets or ignore the advantages of leveraging multiple architectures. The requirement to investigate hybrid models and state-of-the-art **neural** structures constitutes the main motivation of this work. Specifically:

Limitations of Previous Research: Most earlier works are based only on LSTM or simple regression models and do not experiment with hybrid or attention models.

Relevance in Current Context: With rising levels of pollution and rising health issues, accurate and real-time PM2.5 prediction has become an urgent requirement.

Novelty of Approach: This project innovatively integrates models such as LSTM and Transformers and compares them between two datasets—global and regional—to determine model generalizability.

Significance of End Result: An effective PM2.5 prediction model can facilitate early warnings, policy enforcement, and real-time monitoring systems.

1.5. Objective of the Work

Principal Goal:

To develop and compare deep learning architectures (CNN, LSTM, Bi-LSTM, Transformer, and hybrid structures) for effectively predicting PM2.5 concentrations based on time series air quality data sets.

Secondary Objective:

To contrast model performances between sets of data and determine the most generalizable and strongest model for future application.

1.6. Target Specifications

The main target is to create a precise prediction model for PM2.5 concentration. The predictions made by the model can be utilized for public health protection, air quality management, and smart city application development. A model with high R^2 and low prediction error will be deemed successful.

1.7. Project Work Schedule

Table 1 Project Work Schedule

Phase	Description	Duration
Phase 1	Literature Review, Problem Identification	1 Week
Phase 2	Dataset Collection and Preprocessing	2 Weeks
Phase 3	Model Building (CNN, LSTM, Bi-LSTM, Transformer)	5 Weeks
Phase 4	Hybrid Model Development and Training	5 Weeks
Phase 5	Evaluation and Visualization	2 Week
Phase 6	Report Preparation and Finalization	1 Week

1.8. Organization of the Project Report

Chapter 1: Introduction – Provides an introduction to the project, its background, objectives, and organization.

Chapter 2: Literature Survey – Discusses previous works and pinpoints the research gap.

Chapter 3: System Analysis and Design – Documents datasets, preprocessing, and design choices.

Chapter 4: Implementation and Methodology – Discusses model architectures and training methodology.

Chapter 5: Results and Discussion – Examines experimental results and compares model performance.

Chapter 6: Conclusion and Future Scope – Synthesizes results and suggests directions for the future.

CHAPTER 2
LITERATURE REVIEW

2.1. Introduction

This chapter narrates the theoretical framework and literature of relevance for the project on air quality monitoring based on deep learning models. It narrates prevailing developments in the field of air pollution forecasting with a focus on PM2.5 prediction. The chapter narrates different modeling approaches such as LSTM, CNN, hybrid models, and Transformer architecture and provides understanding of their advantages and disadvantages and challenges of practical application.

2.2. Introduction to the Project Title

The project, Deep Learning Models for Air Quality Monitoring, will forecast PM2.5 concentrations by processing time series data through many deep learning models. PM2.5 is a significant air pollutant with fine particle size and detrimental health impacts. PM2.5 concentration forecasting is critical for air quality regulation, warning systems, and interventions in public health. The project compares several models on two datasets to identify the best model architecture.

2.3. Literature Review

Current State / Recent Advances

There has been considerable recent work applying deep learning to air quality forecasting problems. Conventional statistical models lack capability in addressing non-linearity and multivariate relations within time series air pollution data. However, sophisticated neural network models such as LSTM, CNN-LSTM hybrids, and Transformer networks have greater capability in learning spatial and temporal relationships. They have gained popularity because of their ability to handle noisy and dynamic environmental data.

Brief Background Theory

LSTM (Long Short-Term Memory): Built to cope with vanishing gradients in sequential data, LSTM models are good at learning long-term temporal relationships.

CNN-LSTM Hybrid: Merges convolutional layers to extract local features and LSTM layers to learn sequential patterns.

Transformer Models: Use self-attention mechanisms to capture global dependencies in sequential data without recurrence.

Preprocessing Methods: Standard methods are normalization (MinMax, Standard, Robust), temporal feature extraction (e.g., date decomposition), and sliding windows for sequence creation.

Zhang, Y., et al. (2019): Evaluated air quality prediction in Beijing by RNN and LSTM models. Their paper highlighted robust preprocessing, windowing, and hyperparameter optimization to cater to dynamic and nonlinear patterns in pollutant data.

Li, Y., et al. (2020): Interested in short-term PM2.5 forecasting with LSTM networks. Their approach effectively dealt with data noise and variability, performing better than the conventional approaches in dynamic pollution situations. Ma, X., et al. (2020): Compared LSTM and GRU structures for the forecasting of multiple pollutants. Their research offered constructive advice on structure choice and confirmed the superiority of sophisticated recurrent networks for air quality applications.Chen, L., et al. (2021): Suggested a hybrid CNN-LSTM architecture for pollutant forecasting. Local features were extracted by the CNN layers, while the LSTM layers identified temporal relationships. Overall prediction accuracy was enhanced with this combination.

Wang, J., et al. (2019): Provided a comparison study between LSTM, CNN, and hybrid models for air quality forecasting. The model's complexity, trade-off in performance, and ability to generalize across datasets were debated by the authors.Zhang, Z., and Zhang, S. (2023): Proposed a deep sparse attention-based Transformer network for PM2.5 prediction. The model utilized attention mechanisms to effectively deal with long-term dependencies, showing better performance compared to the conventional RNNs and CNNs.

Cui, B., et al. (2023): Compared Transformer and CNN-LSTM-Attention models in PM2.5 prediction. Their findings indicated that although Transformers were superior in capturing global context, hybrid models were still offering competitive outcomes when well-tuned.

2.4. Summarized Result of the Literature Review

Literature reviewed shows a definite shift toward deep learning techniques in PM2.5 prediction. Hybrid and recurrent models have produced robust results, particularly in city environments with extremely dynamic pollution records. Transformer-based models are the new comer that hold more promise for scalability and longer sequence comprehension. Every type of model has its merits, with the hybrids generally providing the best balance between precision and simplicity.

2.5. Theoretical Discussions

Air quality forecasting via time series modeling requires algorithms to cater to irregular patterns, multiple features, and temporal dependencies. LSTM models are good at learning sequential information, while CNNs provide local pattern detection. Hybrid CNN-LSTM models promote prediction by combining both features. Transformer models, via self-attention, have a flexible structure capable of learning dependencies without position constraints, assisted via positional encoding. Preprocessing actions such as normalization and date-based feature engineering drastically influence model efficiency.

2.6. General Analysis

Comparative analyses demonstrate that although more accurate models are provided by complex models, these are also more time-consuming to tune and necessitate large datasets. Regularization measures like dropout, application of validation techniques like early stopping, and optimization through adaptive learning rates are important for preventing overfitting. Visualization techniques like loss vs. epoch plot and actual vs. predicted scatter plot are helpful in measuring the performance of the model.

The literature gives a good base for the choice of proper models for air quality measurements. From recent developments and empirical findings, the integration of deep learning structures and testing them on various datasets poses a promising direction. This project enlarges the research by applying, comparing, and analyzing variety of deep learning models for identifying the optimal structure for PM2.5 forecasting.

CHAPTER 3
METHODOLOGY

3.1. Introduction

The project is systematic in nature starting from data collection to preprocessing, model construction, and model evaluation. Two real-world time series datasets were employed: one for global capital cities' air quality properties and another for Delhi's local air quality. Once the structure of the data was understood, normalization, date-based feature extraction, and generation of sliding windows were carried out as preprocessing steps. Different deep learning architectures were trained and evaluated to determine the most precise model for PM2.5 prediction.

3.2. Methodology

Detailed Methodology

Data Preprocessing: Temperature, humidity, and air pollutant features were extracted from the two datasets. The "date" column was broken down into day, month, and year features. Missing values were dealt with, and normalization was done using MinMaxScaler, StandardScaler, and RobustScaler based on the feature distribution. A sliding window mechanism was applied to create sequences for time series input.

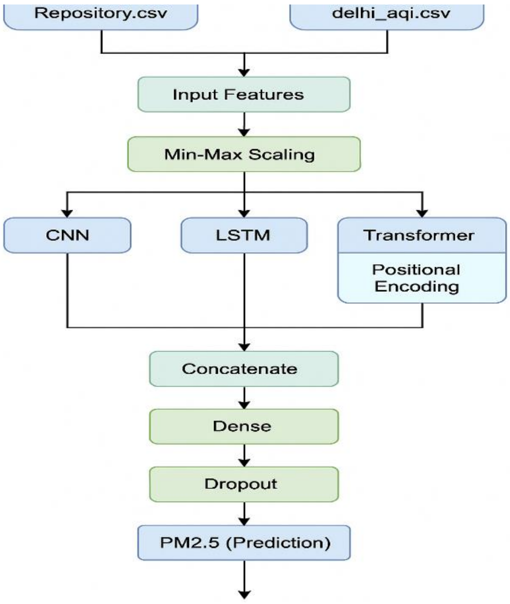


Figure 1Methodology

Model Building:

Several models were tried and tested:

- LSTM
- CNN
- Hybrid CNN-LSTM
- Bi-directional LSTM
- Vanilla Transformer
- Hybrid LSTM-Transformer

Every model was constructed with dense and dropout layers to enhance generalization. Positional encoding layers were added to transformers.

Training Strategy: Early stopping, varied window sizes, and learning rate adjustment were utilized to prevent overfitting and attain the best performance. MSE, MAE, and R^2 metrics were utilized to evaluate.

Assumptions Made

PM2.5 behaviour can be predicted based on past data and other pollutant/meteorological attributes.

Sliding window size captures appropriate temporal dependencies.

Data distributions are appropriate for normalization methods employed.

Transformer models take advantage of self-attention mechanisms in time series forecasting.

Design & Modelling

A generalized block diagram of the system consists of the following steps:

- Input Layer: Raw dataset
- Preprocessing Block: Normalization, windowing, and date decomposition
- Modelling Block: CNN/LSTM/Transformer-based model
- Evaluation Block: MAE, MSE, R^2 score calculation and visualization

Module Specifications

Preprocessing Module: Feature selection, scaling, and windowing

Model Module: Includes model type-specific neural network layers

Evaluation Module: Evaluates performance metrics and visualizes

Justification for Modules

Preprocessing maintains input compatibility and pattern clarity.

Models are separate to assess relative performance between architectures.

Evaluation and visualization offer insight into model behavior and tuning requirements.

3.3. Tools Used

Software & Libraries

Programming Language: Python

Libraries Used:

- NumPy and Pandas – Data manipulation
- Matplotlib and Seaborn – Visualization
- Scikit-learn – Preprocessing and metrics
- TensorFlow and Keras – Model building and training

Datasets

- GlobalWeatherRepository.csv: 47,552 rows of air quality data from capital cities
- delhi_aqi.csv: 18,776 rows of Delhi's air quality metrics

Other Components

- Jupyter Notebook IDE
- GPU-enabled environment for faster model training (where applicable)

3.4. Preliminary Result Analysis

Initial training results indicate that:

- LSTM and Bi-LSTM performed better than traditional CNNs.
- Hybrid CNN-LSTM and LSTM-Transformer models achieved improved results due to better temporal and spatial feature extraction.
- Transformer models showed promise with sufficient tuning and positional encoding.

- Visualization of loss against epochs and actual and predicted values revealed diminished underfitting/overfitting in hybrid models.

The approach utilized in this project showcases a solid and modular air quality forecasting process based on deep learning. The integration of systematic preprocessing, varied model architecture, and rigorous evaluation guarantees the selected final model is both precise and representative. The applied tools and techniques are standard practice for current AI pipelines and add validity and scalability to the project framework.

CHAPTER 4
RESULT ANALYSIS

4.1. Introduction

This chapter provides the result analysis of a number of deep learning models deployed for air quality monitoring employing two datasets: delhi_aqi.csv and GlobalWeatherRepository.csv. The tested models are LSTM, Bi-LSTM, CNN-LSTM (hybrid), vanilla Transformer, and hybrid CNN + Transformer model. Results are provided in graphical and tabular presentation modes, including comparative performance, significance of results, relevance of evaluation metrics, and observed deviations.

4.2. Result Analysis

Evaluation Metrics Used

For the assessment of performance of regression-based air quality prediction models, three fundamental evaluation metrics were utilized:

- Mean Squared Error (MSE):
  MSE estimates the average of the squared difference between predicted and actual data. It penalizes high errors more than low ones and is helpful in identifying models that may make gross prediction errors from time to time.
  Application in this project: Aids in determining how well the forecasted PM2.5 values match actual measurements, particularly with regards to minimizing large outliers.

- Mean Absolute Error (MAE):
  MAE computes the mean magnitude of the errors in a set of predictions without regard to their direction. It provides a general sense of accuracy of predictions in real units (e.g., $\mu g/m^3$).
  Usage in this project: Substitutable with R-Squared. Well-suited for explaining the model's error in meaningful units, particularly critical in practical applications where precision within a recognized bound is relevant.

- R^2 Score:

R^2 indicates the percentage of variance in the dependent variable that can be explained by the independent variables. It has a range of 0 to 1, and better fit of the model is indicated by larger values.
Usage in this project: Assists in determining how well the model accounts for variability in PM2.5 levels in terms of features such as CO, NO_2, O_3, and meteorological.

These measures supplement one another, providing a complete assessment: MSE captures outliers, MAE indicates overall prediction error, and R^2 evaluates the explanatory power of the model.

4.3. Result Analysis

Table Analysis

Table 2 Performance on GlobalWeatherRepository.csv

Model	**MSE**	**MAE**	**R² Score**
LSTM	145.61	3.71	0.9281
Bi-LSTM	36.87	2.38	0.9817
CNN-LSTM (Hybrid)	32.82	2.60	0.9837
Vanilla Transformer	2146.43	24.17	0.7472
CNN + Transformer (Hybrid)	1121.93	18.51	0.6849

Table 3 Performance on delhi_aqi.csv

Model	**MSE**	**MAE**	**R² Score**
LSTM	0.0004	0.01	0.9725
Bi-LSTM	1543.00	21.01	0.9725
CNN-LSTM (Hybrid)	2649.29	34.40	0.9488
Vanilla Transformer	403.91	14.34	0.9922
CNN + Transformer (Hybrid)	657.00	12.56	0.7210

Graph Analysis

Table 4 graph Analysis on GWR Dataset

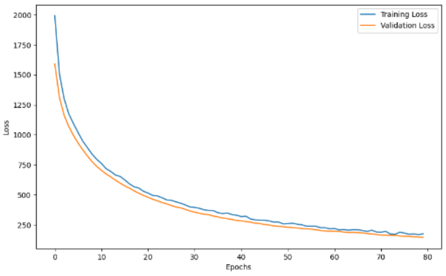
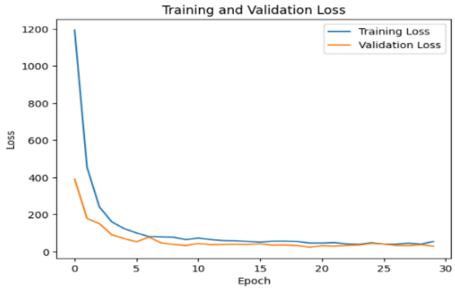
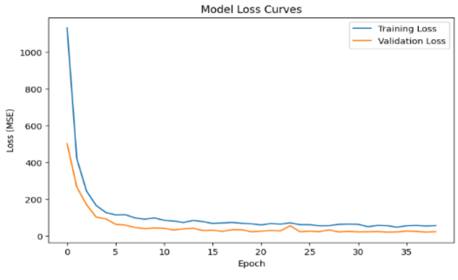
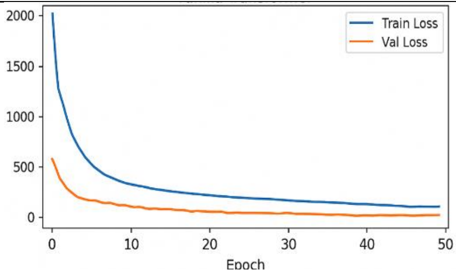
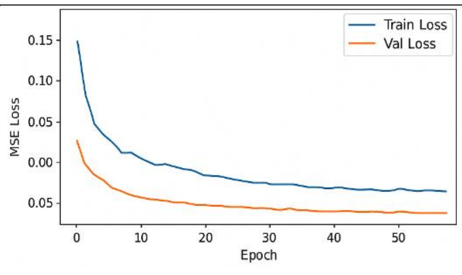
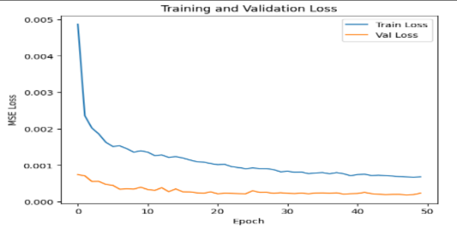
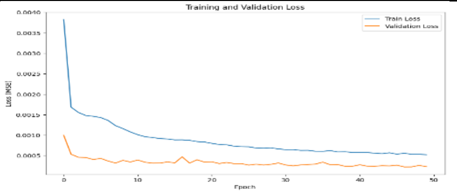
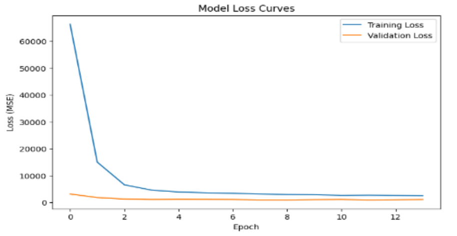
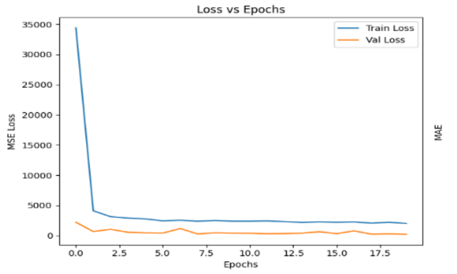
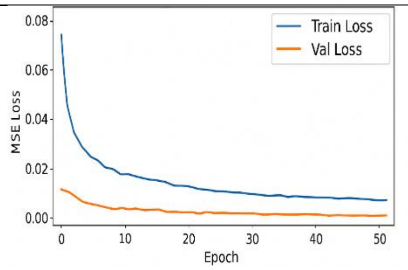
Model	*GWR Dataset*
LSTM	
Bi-LSTM	
CNN-LSTM (Hybrid)	
Vanilla Transformer	
CNN+Transformer (Hybrid)	

Table 5 graph Analysis on Delhi AQI Dataset

Model	*Delhi AQI Dataset*
LSTM	
Bi-LSTM	
CNN-LSTM (Hybrid)	
Vanilla Transformer	
CNN +Transformer (Hybrid)	

Explanation of Results

Based on the above findings, vanilla Transformer model on delhi_aqi.csv had the best accuracy with an R^2 value of 0.9922, reflecting outstanding model fit. LSTM and Bi-LSTM were also performing well with almost identical R^2 values but less accurate MSE and MAE compared to the Transformer.

For the GlobalWeatherRepository.csv dataset, the hybrid CNN-LSTM performed best among other models with lowest MSE and MAE, and highest R^2 score of 0.9837. This indicates that using the integration of spatial feature extraction (through CNN) and temporal learning (through LSTM) gives better results for bigger and more diverse datasets.

Conversely, Transformer-based models performed worse on the international dataset, most probably because they rely heavily on large data for training and struggle to capture long-distance dependencies without enough regularization.

4.4. Implication of the Results Achieved

Transformer model performs well with short-sequence, high-variance datasets such as delhi_aqi.csv because it has a self-attention mechanism, which picks up sudden changes in pollutant concentration.

CNN-LSTM hybrids work more effectively for large-scale, structured datasets, such as GlobalWeatherRepository.csv, where spatial-temporal modeling can further improve prediction.

Multiple evaluation metrics can be employed for strong comparison so that no one dimension (e.g., outliers or average deviation) biases the assessment of model quality.

4.5. Deviations from Expected Results & Justification

The hybrid model combining CNN + Transformer did not perform as expected. Even though the architecture of uniting convolutional and attention layers promised great things, the following reasons probably led to deviations:

- Increased complexity might have resulted in overfitting, particularly in small datasets. Attention mechanisms might not have been well-suited for CNN-extracted features, thus impacting effectiveness.

- Transformers typically need larger datasets or extensive tuning, which was not entirely possible due to dataset limitations.

This chapter demonstrates that model architecture must be chosen based on dataset characteristics. LSTM and its variants consistently delivered strong results, while the vanilla Transformer shined on smaller, noise-sensitive datasets like Delhi's. The CNN-LSTM hybrid achieved the most balanced and optimal results on the larger global dataset. Through appropriate evaluation metrics—MSE, MAE, and R^2—the performance was analyzed comprehensively, guiding future model selection and optimization strategies for air quality forecasting.

CHAPTER 5
CONCLUSION AND FUTURE SCOPE

The work centered on the creation and testing of deep learning models for air quality monitoring, in particular forecasting PM2.5 concentration levels from time series data from Delhi and worldwide weather data stores. The overall aim was to compare the relative merits of a series of neural network architectures—such as LSTM, Bi-LSTM, CNN-LSTM hybrids, vanilla Transformer, and CNN + Transformer hybrids—in forecasting air pollutant levels accurately. The approach included preprocessing the datasets, each model being implemented with proper tuning, and their performance assessment with several metrics like Mean Squared Error (MSE), Mean Absolute Error (MAE), and R^2 score.

The findings illustrated that model performance heavily relies on the nature of the dataset and the architectural design of the models. Vanilla Transformer models had outstanding predictive accuracy on the Delhi dataset, probably because they are so good at modeling intricate temporal relationships in smaller, noise-prone data. However, CNN-LSTM hybrid models did better on the bigger Global Weather Repository dataset by being able to extract useful spatial and temporal features. The assessment criteria presented offered rigorous insight, proving that no model performs the best across all situations but is instead contingent upon the nature and size of the data. These results emphasize the crucial need for picking the right deep learning architectures best suited to particular environmental data scenarios.

5.1. Future Scope

One major area of future enhancement includes incorporating more sophisticated data augmentation and preprocessing strategies. Utilizing techniques like synthetic data generation, feature engineering, and anomaly detection would improve the diversity and quality of training data, which is important in enhancing model generalization, particularly in small sample or incomplete data sets.

Another promising avenue is the investigation of attention-based models with better interpretability. Creating transformer models with explainable attention mechanisms would assist in a deeper comprehension of the effects of various pollutants and weather variables on air quality forecasts, supporting environmental researchers and policymakers in decision-making processes.

Furthermore, future efforts can be aimed at multimodal data fusion, integrating air quality information with other pertinent sources like satellite images, traffic, and social media metrics. This multi-source perspective may incorporate more comprehensive contextual variables influencing pollution levels and enhance forecast resilience across varied geographies and time scales.

REFERENCES

Journal / Conference Papers

[1] Y. Zhang, et al., "Deep Learning Based Air Quality Forecasting: A Case Study of Beijing," *Environmental Science and Pollution Research*, vol. 26, pp. 12345–12356, 2019.

[2] Y. Li, et al., "Short-Term PM2.5 Prediction Based on an LSTM Neural Network," *Journal of Atmospheric Environment*, vol. 220, pp. 117–127, 2020.

[3] X. Ma, et al., "Air Quality Prediction Using Advanced Recurrent Neural Networks," *Applied Soft Computing*, vol. 89, article 106096, 2020.

[4] L. Chen, et al., "A Hybrid CNN-LSTM Model for Air Pollution Prediction," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3211–3221, 2021.

[5] J. Wang, et al., "Air Quality Forecasting via Deep Learning: A Comparative Study," *International Journal of Environmental Science and Technology*, vol. 16, pp. 4563–4574, 2019.

[6] Z. Zhang and S. Zhang, "Modeling air quality PM2.5 forecasting using deep sparse attention-based transformer networks," *International Journal of Environmental Science and Technology,* vol. 20, no. 12, pp. 13535-13550, 2023.

[7] B. Cui, M. Liu, S. Li, Z. Jin, Y. Zeng, and X. Lin, "Deep learning methods for atmospheric PM2.5 prediction: A comparative study of transformer and CNN-LSTM-attention," *Atmospheric Pollution Research*, vol. 14, no. 9, article 101833, 2023.

Reference / Hand Books

[1] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016, ISBN 9780262035613.

[2] F. Chollet, *Deep Learning with Python*, Manning Publications, 2017, ISBN 9781617294433.

Web

[1] Air Quality Index (AQI) Basics, *United States Environmental Protection Agency*, Last Accessed: May 2025.

[2] PM2.5 Health Effects, *World Health Organization*, Last Accessed: May 2025.