CS365 Deep Learning
Project Report

MULTI-STEP AIR QUALITY FORECASTING USING ROBUST DEEP LEARNING MODEL

October 28, 2023

Anurag Deo (2101AI04) and Aritra Bhaduri (2101AI40)
Department of Computer Science Engineering
Indian Institute of Technology Patna
anurag_2101ai04@iitp.ac.in
aritra_2101ai40@iitp.ac.in

Abstract

Air quality forecasting is a critical task for environmental protection and public health. It can help people avoid exposure to harmful air pollutants and take necessary precautions. Traditional air quality forecasting methods, such as statistical models, are often limited in their ability to capture the complex relationships between air pollutants and meteorological factors.

Deep learning models have emerged as a promising new approach for air quality forecasting. They can learn complex patterns from data and make accurate predictions even in the presence of noisy or incomplete data.

This report presents a multi-step air quality forecasting model using a robust deep learning model called DAQFF [1]. The DAQFF model is a hybrid model that combines multiple convolutional neural networks and a bidirectional long short-term memory network to effectively capture the spatial-temporal dependencies of air quality data.

The DAQFF model was evaluated on a real-world air quality dataset. The results showed that the DAQFF model outperformed traditional air quality forecasting methods in terms of prediction accuracy.

The DAQFF model has the potential to be used as a powerful tool for multi-step air quality forecasting in various cities and regions around the world. It can help improve air quality monitoring and management efforts, and ultimately, protect public health.

Keywords: Air quality forecasting, deep learning, DAQFF model, spatial-temporal dependencies, prediction accuracy

1 Introduction

The goal of this project is to develop a robust deep learning model for multi-step air quality forecasting. The rationale of the project is that deep learning models have the potential to improve the accuracy and robustness of air quality forecasting, which can lead to better air quality monitoring and management efforts.

The main contributions of this project are:

- Development of a robust deep learning model for multi-step air quality forecasting, called DAQFF.
- Evaluation of the DAQFF model on a real-world air quality dataset.
- Demonstration of the DAQFF model's superior performance over traditional air quality forecasting methods.

The DAQFF model is a hybrid model that combines multiple convolutional neural networks and a bidirectional long short-term memory network to effectively capture the spatial-temporal dependencies of air quality data. The model was evaluated on a dataset of hourly air quality and meteorological data and was shown to outperform traditional air quality forecasting methods in terms of prediction accuracy [2].

The DAQFF model has the potential to be used as a powerful tool for multi-step air quality forecasting in various cities and regions around the world. It can help improve air quality monitoring and management efforts, and ultimately, protect public health.

2 Project Idea

The project idea is to develop a robust deep learning model for multi-step air quality forecasting, called DAQFF. The DAQFF model is a hybrid model that combines multiple convolutional neural networks (CNNs) and a bidirectional long short-term memory (BiLSTM) network to effectively capture the spatial-temporal dependencies of air quality data.

Advantages of the DAQFF Model

The DAQFF model has several advantages over traditional air quality forecasting methods, including:

- **Effectively captures spatial-temporal dependencies:** The combination of CNNs and BiLSTM allows the model to extract both local spatial patterns and long-term temporal dependencies from air quality data.
- **Improved prediction accuracy:** The DAQFF model has been shown to outperform traditional methods in terms of prediction accuracy, demonstrating its effectiveness in air quality forecasting.
- **Flexibility for various air quality parameters:** The model can be applied to forecast various air quality parameters, such as PM2.5, PM10, and ozone concentrations.

Potential Applications

The DAQFF model has the potential to be used in a variety of applications, including:

- Air quality monitoring: The model can be used to monitor air quality in real time and identify areas where air quality is poor.
- Air quality forecasting: The model can be used to forecast air quality for future time steps, which can help people avoid exposure to harmful air pollutants.
- Air quality management: The model can be used to inform air quality management decisions, such as when to issue air quality alerts or implement pollution control measures.

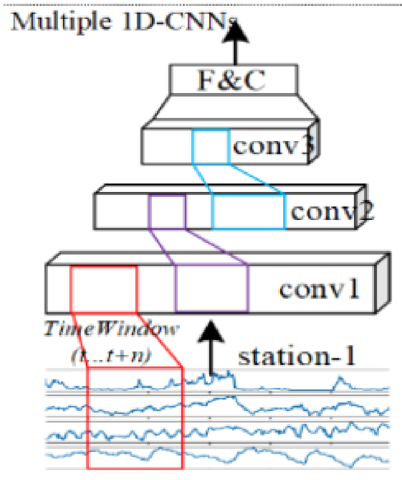


Figure 1: Multiple 1-D CNN layers

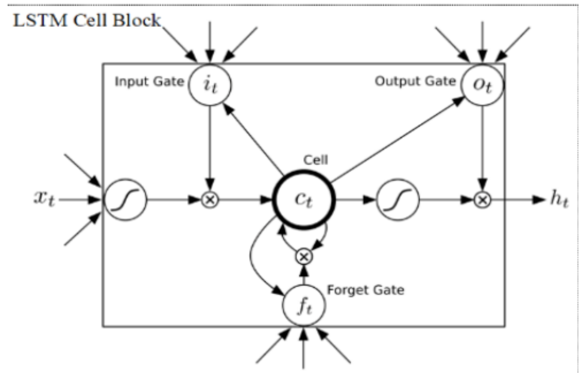


Figure 2: LSTM Cell Block

3 Model Architecture

The DAQFF model can be divided into three main sections:

1. Feature Extraction with CNNs

Multiple CNNs are employed to extract local trend features and possible spatial correlation features from air quality data collected from multiple stations. CNNs are well-suited for this task as they can learn to identify spatial patterns in the data. A typical CNN has three layers: convolutional layer, activation layer, and pooling layer.
Unlike the classical CNN model (also traditional two-dimensional CNN used for images), we propose to use multiple one-dimensional filters convolved (1D-CNNs) over all time steps of air quality time series data. Moreover, one-dimensional CNN's local perception and weighted sharing features can reduce the number of parameters for processing multivariate time series data, thereby improving learning efficiency. Thus, our method can learn more deep representation features of air quality related data.

2. Temporal Dependency Learning with BiLSTM

Although traditional statistical methods like ARIMA and shallow learning models similar to deep neural networks can process time series, the efficiency is not so good, because it does not take into account the long-term temporal dependence of time series data. In order

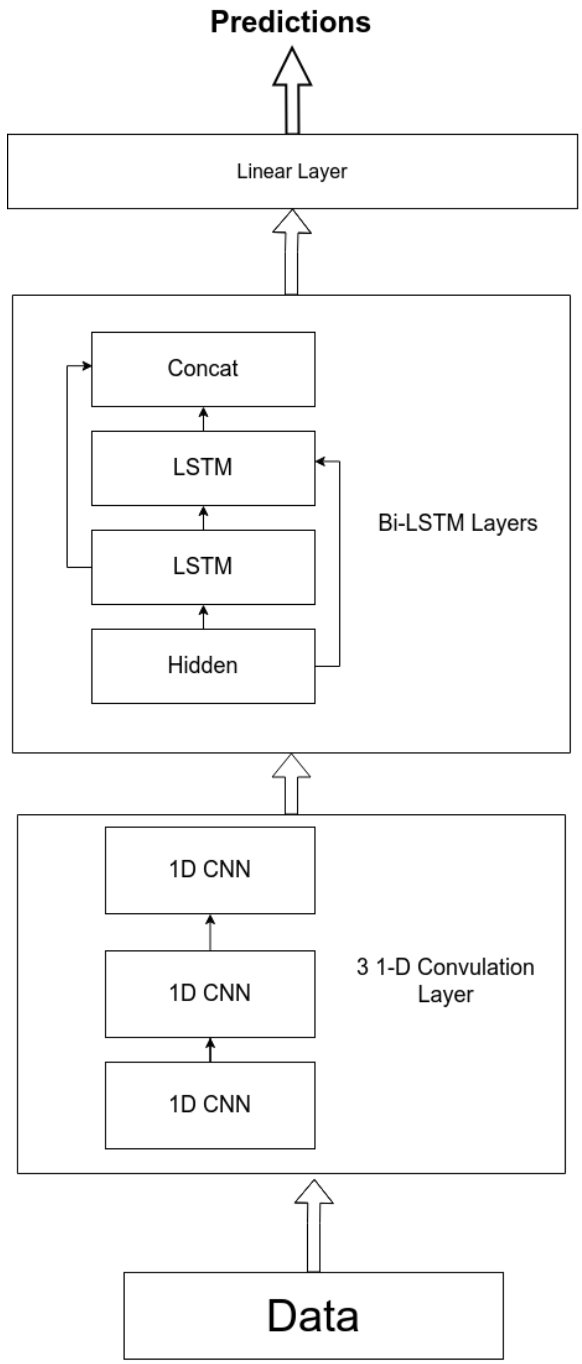


Figure 3: Model Architecture

to overcome this shortcoming, Long Short-term Memory network (LSTM) is a good option, which is a popular dynamic model for handling sequence tasks. As shown in the upper right corner of Fig. 2, the LSTM Cell Block represents a typical LSTM diagram. The memory cell of each LSTM block contains four main components. The collaboration of these components enables cells to learn and memory long dependency features.
One disadvantage of traditional LSTMs is that they can only utilize the previous context of sequence data, and Bi- directional LSTM can process the time series data in two directions simultaneously through two independent hid den layers [30], and these data are concatenated and fed forward to the output layer. In other words, Bi-directional LSTM processes the time series data in two directions iter atively (forward layer from t = 1 to T, backward layer from t = T to 1).

3. Prediction Layer

The extracted features and learned temporal dependencies are combined to make air quality predictions for future time steps. The prediction layer is a simple feed-forward neural network that learns to map the extracted features to the predicted air quality values.

4 Testing and Experiments

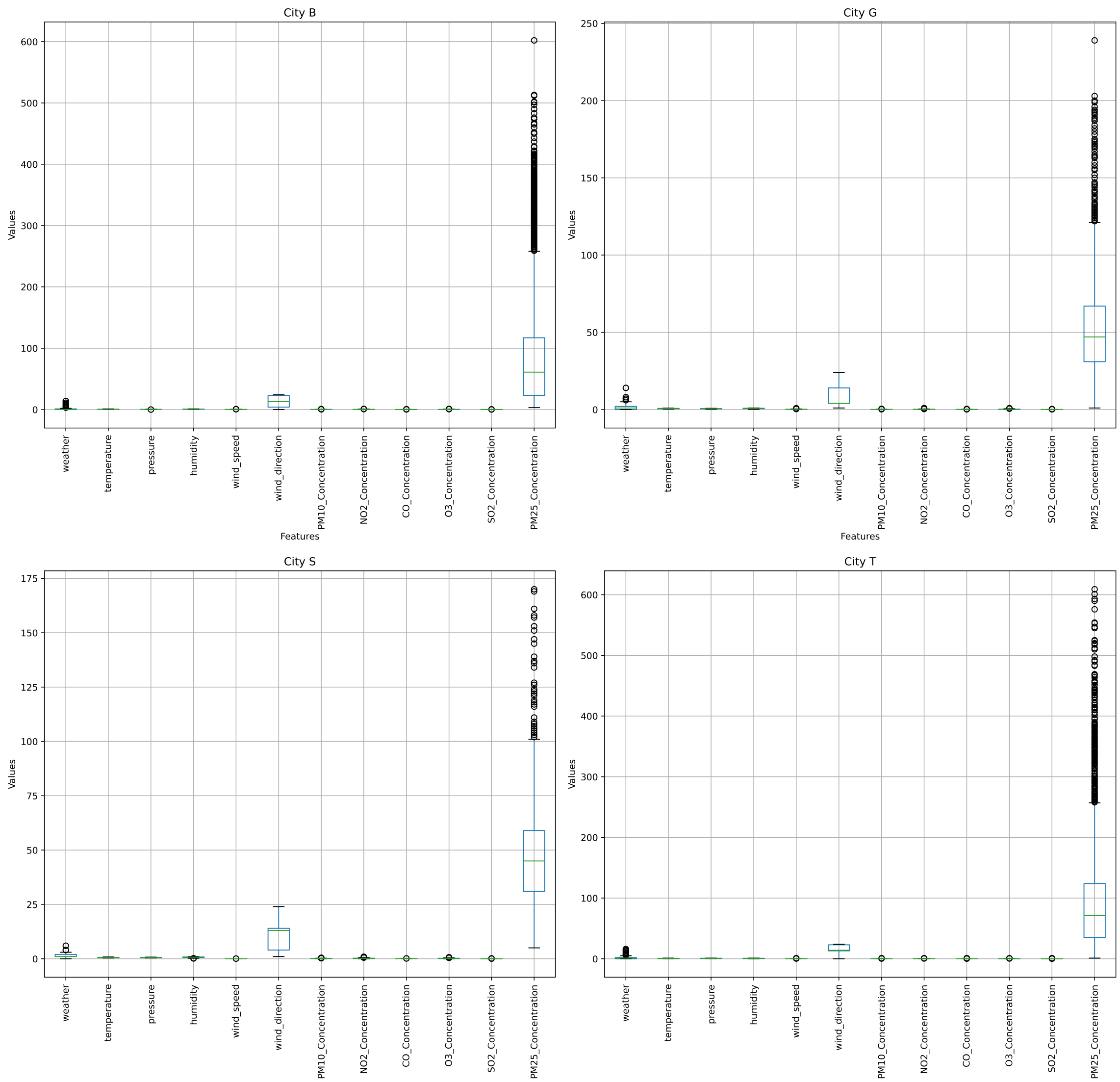
Multiple experiments are conducted to find out the right set of of the hyper-parameters. The table below summarizes the set of hyper-paramenter used in the training of the model.

Optimizer	Adam
Learning Rate	0.001
Loss	Mean Squared Error
Batch Size	32
Epochs	120
LSTM Hidden Units	128

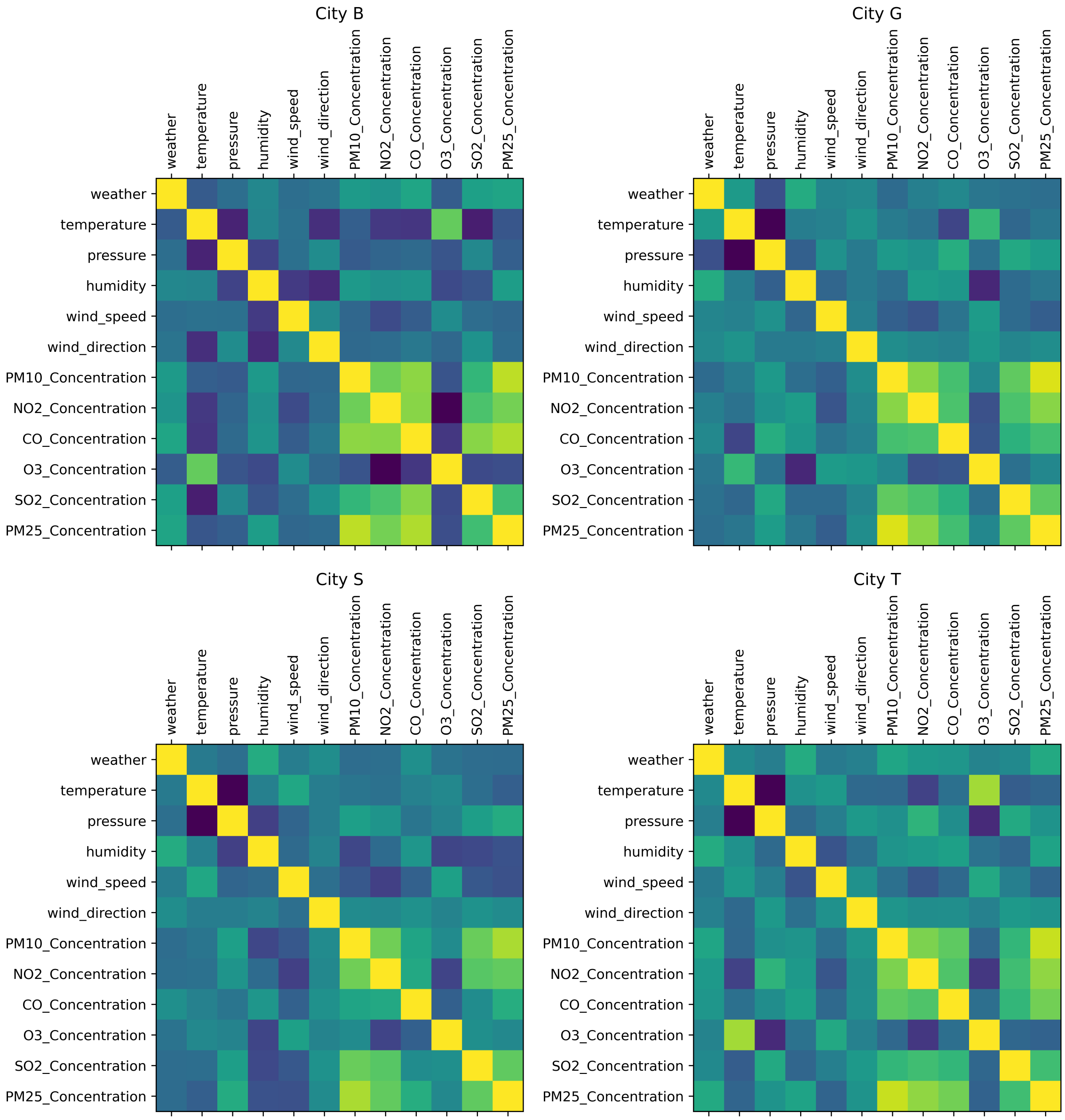
	Filter Size	Kernel Size	Activation Function
Layer 1	64x64	5x5	Relu
Layer 2	32x32	3x3	Relu
Layer 3	16x16	1x1	Relu

5 Data Analysis

The dataset used in the training and testing is divided into 4 cities named B, G, S and T. Each of the city have a air quality dataset containing 12 columns.The columns include weather, temperature, pressure, humidity, wind speed, wind direction, PM10 Concentration, NO2 Concentration, CO Concentration, O3 Concentration, SO2 Concentration, PM25 Concentration. The box plots of the cities are as follow:



The co-relation between the features in the dataset is as follows:

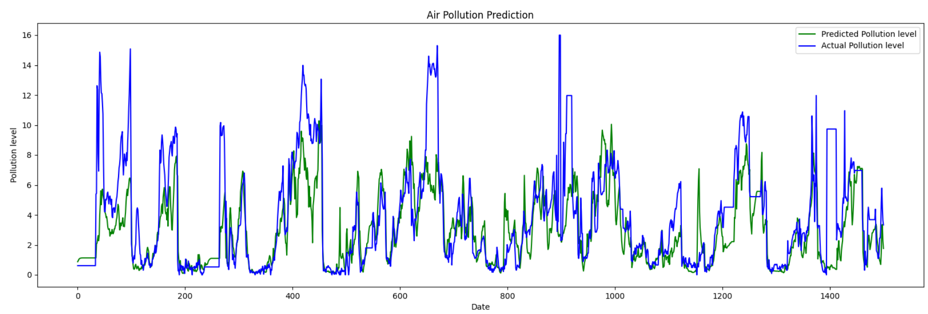


6 Results

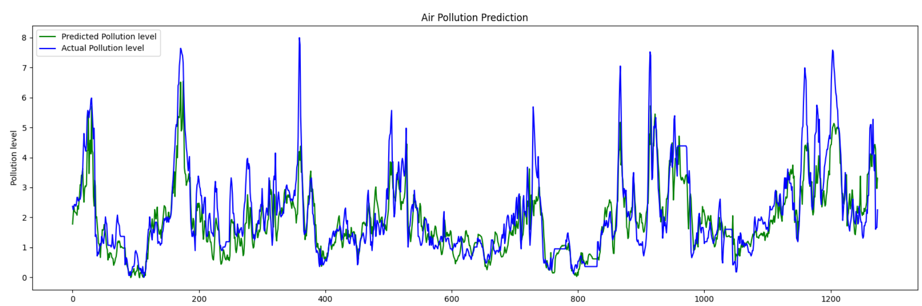
The results of the model is in term of the mean absolute error and the mean squared error. The errors that we get after training on the test set is shown in the table below:

	Mean Squared Error	Mean Absolute Error
City B	0.0233	0.0966
City G	0.0088	0.0675
City S	0.0139	0.0901
City T	0.0347	0.1430

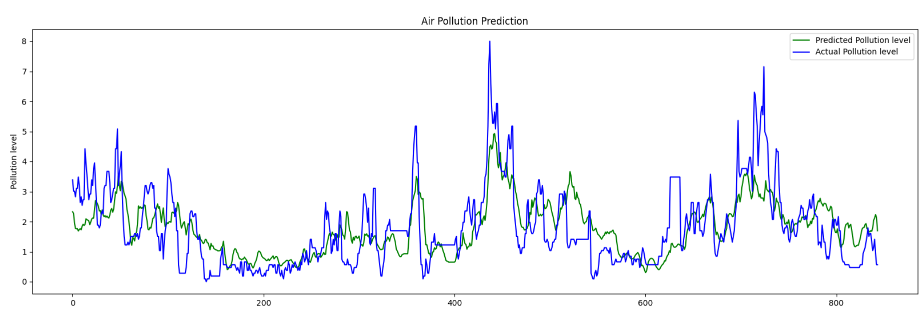
Table 1: Results of the model



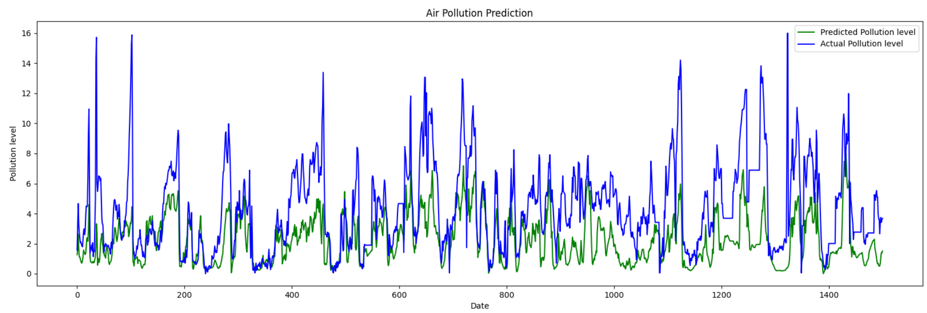
Graph of actual vs predicted PM-25 Concentration of City B



Graph of actual vs predicted PM-25 Concentration of City G



Graph of actual vs predicted PM-25 Concentration of City S



Graph of actual vs predicted PM-25 Concentration of City T

7 Conclusion

The DAQFF model is a promising new approach for multi-step air quality forecasting. It has the potential to improve air quality monitoring, forecasting, and management efforts, and ultimately, protect public health.

References

[1] S. Du, T. Li, Y. Yang, and S. Horng, "Deep air quality forecasting using hybrid deep learning framework," *IEEE Transactions on Knowledge amp; Data Engineering*, vol. 33, no. 06, pp. 2412–2424, jun 2021.

[2] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "The performance of lstm and bilstm in forecasting time series," pp. 3285–3292, 2019.