

Forecasting Air Quality with Deep Learning

Gayathri M. ^{*1}, Kavitha V. ², Anand Jeyaraj ³

Submitted: 10/03/2024

Revised: 25/04/2024

Accepted: 02/05/2024

Abstract: Due to factors such as increased urbanization, growing populations, transportation, home activities, agricultural methods, and industrial processes, Air pollution has emerged as a major issue in the past several years. It is linked to several illnesses and has emerged as a substantial issue in several urban areas, particularly in developing nations such as India. As part of our research, we make use of the Air Quality Index (AQI) for assess the quality of the air in Mumbai, India. Our emphasis is on evaluating 13 different pollutants and 7 meteorological indicators for the period from July 2017 to September 2022 to be able to predict air pollution levels. We employed three deep learning models: LSTM, Bi-LSTM, and CNN-Bi-LSTM. Results show that the CNN-Bi-LSTM model had better accuracy compared to earlier models, as proven by a MAE of 0.45, a MSE of 0.58, a RMSE of 0.60, and RMSLE of 0.36. This study shows that deep learning model are effective in forecasting AQI and by using historic data and deep learning algorithms enable precise forecasts of urban air quality levels worldwide.

Keywords: Air Pollution, Air Quality Index, Deep Learning, Prediction.

1. Introduction

Polluted air is a significant problem worldwide and impacts most urban areas. Industry is one of the main causes of air pollution in India; it's responsible for 51% of all the pollution in the country [1]. It can be assessed using the air pollutants index (Nitric Oxide, Nitrogen Dioxide, Nitrogen Oxides, Ammonia, Sulphur Dioxide, Carbon Monoxide, Ozone, Benzene, Toluene, Ethylbenzene, m-Xylene and p-Xylene, PM2.5, PM10). Most of the airborne substances come from stable sources of main air pollutants, namely microscopic particles and dust of PM10 with a diameter lower than 10 μm , and PM2.5 particles, which are especially dangerous since their diameter is less than 2.5 microns. PM2.5 particles are generated from incomplete combustion of fuel and byproducts of industrial activities, as well as the emission of Sulphur dioxide (SO₂). Combusting fuel at elevated temperatures results in the generation of nitrogen oxides, carbon monoxide, and ozone via a reaction among oxygen and nitrogen [2]. Monitoring air pollution is gaining greater focus in recent years as it has a significant influence on health and well-being of people. The air pollution that is generated by industry is responsible for a variety of illnesses that affect the cardiovascular system and the respiratory system. It has been demonstrated that there is a connection between higher levels of PM2.5, nitrogen

dioxide, and Carbon monoxide and an increase in the frequency of respiratory diseases that are suffered by employees and their families [3]. Heart disease risk is increased by 1% to 3% after long-term exposure to fine matter, defined as particles with a diameter of less than 2.5 μm [4]. Air pollution significantly contributes to several health problems, including cardiovascular and respiratory conditions, premature mortality, and hospitalizations for cardiac and lung disorders [5][6]. Further, demographic transition, i.e., the percentage of individuals who moved to urban areas, can pose a number of challenges relating to health, transportation, and air quality. According to the Metropolitan Demographic web site, demographic transition was 56.15% in 2020. One way that governments keep the people informed about air pollution levels is through air quality indexes. The National Ambient Air Quality Guidelines were created by the Central Pollution Control Board in order to determine the threshold of air quality that must be maintained in order to protect the health of the general population within a calculated margin of safety. Table 1 illustrates the AQI classification for the air's quality into six distinct groups according to the level of pollutants present. Each category, ranging from "Good" (0-50 $\mu\text{g}/\text{m}^3$) to "Very Poor" (301-400 $\mu\text{g}/\text{m}^3$), represents a specific degree of possible health risk. Higher numbers indicate more severe air pollution. The AQI primarily quantifies the amounts of certain pollutants, such as PM10, PM2.5, atmospheric sulphur oxides, the gas nitrogen dioxide, ozone, and the gas carbon monoxide, sodium hydroxide, lead, nickel, arsenic, benzo(a)pyrene, and benzene. It is preferable to use an innovative approach which includes as many substances from the list of recognized contaminants as possible. PM10, PM2.5, SO₂, NO₂, CO, and O₃ are regularly monitored at several air quality stations and WS (Wind Speed), RH (Relative

¹ Department of Data Science and Business Systems, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Chennai-603203, Tamil Nadu, India.

ORCID ID: 0009-0004-7501-7824

² Department of Data Science and Business Systems, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Chennai-603203, Tamil Nadu, India.

ORCID ID: 0000-0002-2728-5646

³ Department of Information Systems, Wright State University, Dayton, United States.

ORCID ID: 0000-0003-3691-881X

* Corresponding Author Email: gm4462@srmist.edu.in

Humidity), WD (Wind Direction), BP (Barometric Pressure), AT (Ambient Temperature), RF (Rainfall), TOT-RF (Total Rainfall), Nicotine (Ni), arsenic (As), benzo(a)pyrene, and benzene are subject to every year limits, whereas carbon monoxide (CO) and ozone (O3) have short-term standards based on (01 and 08 hourly) averages.

Table 1. Air Quality Index (AQI) Categories and Concentration Ranges

AQI Category	Concentration Range (in $\mu\text{g}/\text{m}^3$)
Good	0 – 50
Satisfactory	51 – 100
Moderate	101 – 200
Poor	201 – 300
Very Poor	301 – 400
Severe	401 - 500

1.1 Research Objectives.

The purpose for this study is to use deep learning models to forecast the air quality index.

- To provide an assessment methodology for evaluating air quality in Mumbai, India, taking into account aspects such as urbanization, population expansion, and industrial activity.
- To utilize LSTM, Bi-LSTM, and CNN-Bi-LSTM deep learning models to enhance the accuracy of air pollution level forecasting approaches, exceeding earlier models in precision. This improvement will be measured using performance measures such as MAE, MSE, RMSE, and RMSLE.

2. Literature Survey

Table 2 shows recent illustrates and the different approaches to air pollution analysis and prediction. The review of prior literature on air pollution prediction methods reveals several key insights and gaps. While various techniques, such as LSTM networks, artificial neural networks (ANNs), and hybrid models, have shown promise in forecasting air quality, limitations persist in terms of handling nonlinear relationships, ensuring computational efficiency, and preventing overfitting. Traditional models like ARIMA and SVMs, while effective for certain applications, are constrained by their inability to capture complex interactions and their computational inefficiency, respectively. Moreover, deep learning models, including CNNs, face scalability issues and can suffer from overfitting, as evidenced by previous studies. Additionally, the need for robust feature extraction methods to enhance model accuracy and generalization across different regions

is highlighted.

Our study addresses these gaps by integrating advanced feature extraction techniques and employing a hybrid model that combines CNN and Bi-LSTM. This approach leverages CNN's strength in extracting local patterns and Bi-LSTM's ability to capture long-term dependencies, thereby improving predictive accuracy and robustness. We focus on computational efficiency and prevent overfitting through regularization techniques, ensuring the model's performance remains robust across different datasets. By evaluating 13 pollutants and 7 meteorological indicators over five years in Mumbai, India, our study provides a comprehensive, scalable, and generalizable solution for real-time air quality prediction. This contributes significantly to the field by offering precise forecasts of urban air quality levels, enhancing our understanding and management of air pollution worldwide.

3. Methodologies

3.1 Data Description

The data was collected from the India's Central Pollution Control Board (CPCB) and also followed the Central Pollution Control Board's (CPCB) guidelines to calculate the AQI. This involved monitoring various air pollutants like fine particles (PM2.5 and PM10), nitrogen oxides, along with common air contaminants like carbon monoxide, ozone, and aromatic hydrocarbons such as benzene, toluene, and xylene. The AQI is calculated by measuring a specific kind of particulate matter along with three different types of gaseous contaminants. The highest value among the three pollutants is then taken into consideration when determining the AQI. The AQI is not only affected by climate conditions but also the presence of air pollutants inside a certain region. The meteorological parameters that were observed were temperature, relative humidity (RH), wind speed (WS), wind direction (WD), Barometric pressure (BP), ambient temperature (AT), rainfall (RF), and total rainfall (TOT-RF). The data set had a total of 74,400 observations spanning January 1, 2021 to August 30, 2023. The data set included 20 components including 13 air pollutants and 7 meteorological factors.

3.2 Data Pre processing

Traditional data processing methods consist of two main procedures. Firstly, the datasets are enhanced by imputing missing data and changing or eliminating outlier data points to provide greater accuracy and completeness. Secondly, the data is made evenly distributed via the processes of normalization and standardization. More than 50% of the information from the Central Pollution Control Board's data was found to be missing. By using a second-order polynomial approximation and utilizing the data points in closest proximity, we effectively replaced the missing data. The outcomes were better to those achieved by the use of

linear interpolation or series mean. The methodology for calculating AQI is consistent with both the CPCB. AQI is determined by using individual formulae for each parameter. To compute the AQI based on four parameters, you need to apply the formula four times. The AQI is determined by the sub-index that represents the poorest air quality among the four parameters. A sub-index is a linear function that is derived from the concentration of pollutants, representing two distinct but interconnected concepts in the Equation (1).

The equation may be expressed as follows:

$$P_i = \frac{A_{Hi} - A_{Lo}}{BT_{Hi} - BT_{Lo}} \cdot (P_c - BT_{Lo}) + A_{Lo} \quad (1)$$

where

P_i = pollutant of index p

P_c = represents the concentration level of a pollutant (p) after truncation.

BT_{Hi} = concentration threshold (breakpoint) i.e. if the pollutant concentration (P_c) is greater than or equal to BT_{Hi} , a specific AQI value (A_{Hi}) applies.

BT_{Lo} = concentration threshold. If P_c is less than or equal to BT_{Lo} , a different AQI value (A_{Lo}) applies.

BT_{Lo} = concentration breakpoint i.e. less than or equal to P_c

A_{Hi} = AQI value that corresponding to BT_{Hi}

A_{Lo} = AQI value that corresponding to BT_{Lo}

In the field of data analysis, the problem of missing values in datasets can be addressed by the use of a technique known as data imputation. The purpose of this process is to increase the completeness of the data and to make it possible to conduct an analysis that is more accurate. It entails replacing missing data points with estimated values. Imputation helps to use more data points for more accurate results. Multiple techniques exist for data imputation including the type of data (numerical, categorical), K- Nearest Neighbor imputation is a method that is used to estimating missing values in a dataset by utilizing the characteristics of similar data points. It involves several steps, including data preparation, distance calculation, k-nearest neighbors' selection, missing value imputation, and output.

In the data preparation phase, missing values are identified, and any necessary data preprocessing is performed to handle outliers, inconsistencies, and data type conversions. Then Euclidean distance is chosen to measure the similarity between data points. Pairwise distances are then computed between each data point and all other data points.

$$z_{(p,q)} = \sqrt{\sum_{a=1}^t (p_{xj} - q_{yj})^2} \quad (2)$$

Where, $z(p, q)$: Euclidian distance between two data points,

a : indicates a specific data attribute, ranging from 1 (first attribute) to t (total number of attributes).

t : signifies the total number of dimensions (attributes) considered in the data.

p_{xj} : refers to the value in the j^{th} dimension (attribute a) of data point p , but with potentially missing data.

q_{yj} : refers to the value in the j^{th} dimension (attribute a) of data point q , which is assumed to have complete data.

In the k-nearest neighbors selection phase, the number of nearest neighbors (k) is specified. For each data point with missing values, the k nearest neighbors is identified based on their computed distances in the Equation (2). The missing value imputation phase involves selecting an appropriate imputation strategy is weighted imputation. The chosen strategy is then applied to estimate the missing values using the k nearest neighbors' values. Finally, the imputed data is generated and evaluated to assess the quality of the imputation process. This evaluation may involve comparing the distribution of imputed values to the original data. KNN is the best imputation method for handling the missing data.

Data normalization is a method for bringing all dataset values up to a consistent scale. Normalizing the data is essential since many ML algorithms are sensitive to the input characteristics' size and perform better with it. The min-max normalization method normalizes values. This approach re-scales the values of a feature to a range bounded by 0 and 1. To do this, calculate the difference between each value and the feature's lowest value, and then divide the result by the feature's range in the Equation (3).

$$zn' = \frac{zo - \min(zo)}{\max(zo) - \min(zo)} \quad (3)$$

In this context, zn' represents the normalized value, zo represents the original value, $\min(zo)$ represents the lowest value in the dataset, and $\max(zo)$ represents the maximum value in the dataset. In Equation 3 normalizes the value z to a range bounded by 0 and 1, with 0 being the least value and 1 denoting the highest value. This might be beneficial when using machine learning algorithms that are sensitive to the amount of data available. These normalization values of the measured parameters will be utilized in the computations to enhance forecast accuracy. In the context of air quality, the correlation matrix represents the relationships between different air quality parameters and metrological characteristics in Figure 1.

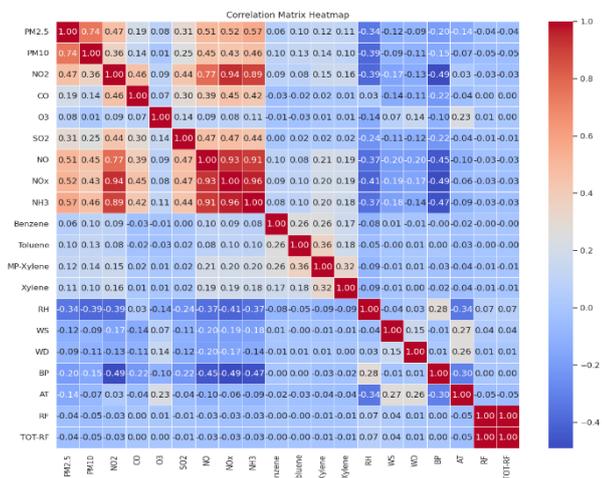


Fig. 1. Correlation Matrix

3.3 Long-Short Term Memory (LSTM) models

Recurrent Neural Networks (RNNs) have challenges when it comes to acquiring knowledge about long-term relationships. LSTM-based models are an extension of RNNs that effectively solve the issue of vanishing gradients. LSTM models significantly improve the stored-value capacity of RNNs, enabling them to efficiently store and understand complex relationships between inputs over lengthy periods of time. This memory extension has the capacity to retain information for an extended duration, hence facilitating the processes of reading, writing, and removing data from one's memory. A "gated" cell is a frequent name for the LSTM memory. The concept of "gate" originates from the fact that the LSTM memory has the ability to decide whether or not to keep the information stored in the memory. The LSTM model efficiently captures prominent features from inputs and retains this information for a prolonged period. The decision to eliminate or retain the information is contingent upon the weight values given to the data throughout the training phase. Therefore, an LSTM model acquires the ability to determine which information is valuable and should be preserved as well as which information can be eliminated. The forget gate, the input gate, and the output gate are the three gates that are often included in an LSTM implementation. The forget gate decides whether to keep or delete current data, Input gate regulates memory to take in new data, and the output gate controls the cell's current value contribution to the

output. Figure 2 illustrates the structure of the LSTM. A "gated" cell is a frequent name for the LSTM memory. The concept of "gate" originates from the fact that the LSTM memory has the ability to decide whether or not to keep the information stored in the memory. The LSTM model efficiently captures prominent features from inputs and retains this information for a prolonged period. The decision to eliminate or retain the information is contingent upon the weight values given to the data throughout the training phase [17].

<i>S.no</i>	<i>Author</i>	<i>Methods</i>	<i>Inferences</i>
1.	S. Al-Janabi et. al. [7].	Utilization of deep learning techniques using Recurrent Neural Networks (RNNs), particularly LSTM networks	LSTM networks are able to capture the relationships between events that occur over time in sequential data, allowing for accurate modelling of the dynamics of air pollutants.
2.	S.M.S. Cabaneros et. al. [8]	Artificial Neural Networks (ANNs) are utilized to forecast ambient air quality.	The prediction of air quality depends on several factors, such as the size, accuracy, and computing expenses of the data.
3.	J. Ma, Y. Ding et. al. [9]	Lag layer-LSTM-Fully Connected network model based on Bayesian Optimization (BO).	Multivariate air quality prediction with a focus on PM2.5.
4.	D Saravanan et.al. [10].	Transferred Bi-directional Long Short-Term Memory (TL-BLSTM) model for air pollution prediction	The TL-BLSTM model effectively captures complex temporal correlations in PM2.5 data.
5.	Mao. et. al. [11]	Graph Transformer Long Short-Term Memory	The GT-LSTM model effectively captures intricate spatiotemporal correlations of air contaminants.
6.	Maleki, et. al. [12].	Artificial Neural Network (ANN)	The ANN technique proves to be useful in evaluating hourly air pollution thresholds and forecasting air quality indices.
7.	Castelli M . et. al [13]	Utilizing the ARIMA model to examine and predict air pollution levels by utilizing historical time series data.	Predicting air pollution and evaluating air quality using historical time series data is made easier using the ARIMA model.
8.	L Ge, K Wu et.al. [14].	Long Short-Term Memory (LSTM) to effectively capture temporal relationships in air quality data. Support Vector Regression (SVR) to perform regression analysis for the purpose of air quality prediction. GLCM to extract characteristics in order to improve the accuracy of predictions.	Air quality prediction was effective using the DL-based model (LSTM and SVR). GLCM feature extraction enhanced model accuracy by giving statistical characteristics.
9.	A Dun, Y Yang et.al.[15].	Convolutional Neural Network	The CNN-based model had an overfitting issue, resulting in a negative impact on its performance.
10,	Shahzad et.al [16].	Support Vector Regression (SVR)	Nonlinear time series prediction model for stock market analysis and trend prediction.

Table 2. Literature Review on Methods for Predicting Air Pollution

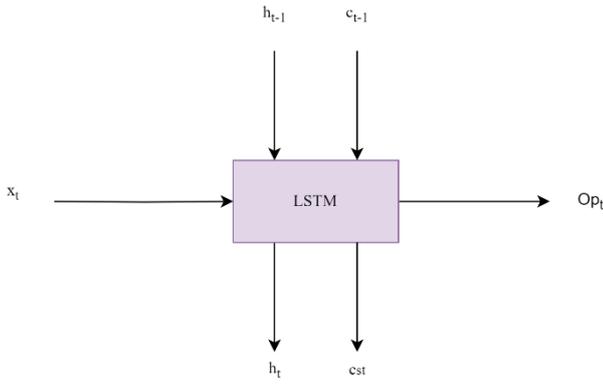


Fig 2. Structure of LSTM

The Mathematical equation may be expressed as follows:

$$f_{gt} = \sigma_g(W_{fg} \times x_t + U_{fg} \times h_{t-1} + b_{fg}) \quad (4)$$

$$ip_t = \sigma_g(W_{ip} \times x_t + U_{ip} \times h_{t-1} + b_{ip}) \quad (5)$$

$$op_t = \sigma_g(W_{op} \times x_t + U_{op} \times h_{t-1} + b_{op}) \quad (6)$$

$$c's_t = \sigma_{cs}(W_{cs} \times x_t + U_{cs} \times h_{t-1} + b_{cs}) \quad (7)$$

$$c_{st} = f_t \cdot c_{st-1} + i_t \cdot c's_t \quad (8)$$

$$hs_t = op_t \cdot \sigma_c(c_{st}) \quad (9)$$

Where,

f_{gt} – indicates the forget gate

ip_t – indicates the input gate

op_t – represents the output gate

c_{st} - cell state

hs_t – Hidden state

σ_{cs} – represents the tanh activation function

σ_g – represents the sigmoid activation function

3.4 Bidirectional LSTMs (Bi-LSTM)

The Bidirectional Long Short-Term Memory (Bi-LSTM) is an enhanced version based on the LSTM model. It utilizes both forward and backward layers of LSTM to capture information from both current and succeeding contexts. The outputs from these two layers are then combined to provide the final result. The structure of the Bi-LSTM is shown in Figure 3. The primary LSTM layer is used to compute the sequential data of the present moment, while the next layer is utilized to subsequently analyze the same series and include reverse sequence information. Furthermore, the settings of each LSTM layer is different. The hidden outputs generated by LSTM layers are sent to adjacent units and to the input of the subsequent LSTM layer. As there is no interaction among neurons in both states, the network weights may be modified by propagating the neurons forward and backward [18].

The forward LSTM sequentially analyzes input in a left-to-right manner, and its hidden state may be seen in Equation (10).

$$\vec{hd}_t = LSTM(x_{ft}, \vec{hd}_{t-1}) \quad (10)$$

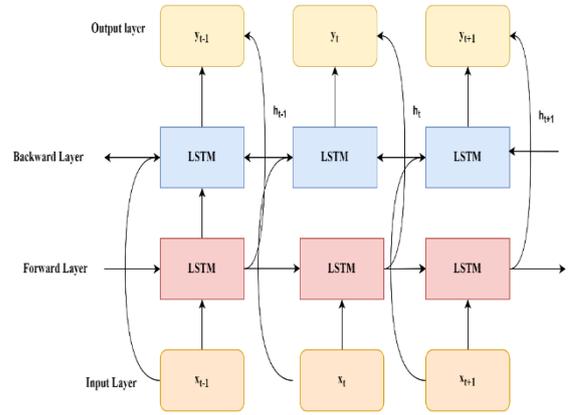


Fig 3. Structure of Bi-LSTM

The reverse LSTM, on the other hand, analyzes data moving right to left, and the hidden state may be described as in Equation (11).

$$\begin{aligned} \leftarrow \\ hd_t \\ = LSTM(x_{ft}, \overleftarrow{hd}_{t+1}) \end{aligned} \quad (11)$$

Equation (12) shows the output of the Bi-LSTM may be summed by connecting the forward and backward layer.

$$hd_t = \left[\begin{array}{c} \vec{hd}_t \\ \overleftarrow{hd}_t \end{array} \right] \quad (12)$$

Where

hd_t – Hidden state

x_{ft} – Input of forward layer

x_{dt} – Input of backward layer

3.5 Bidirectional LSTM – Convolutional Neural Network

The Convolutional Neural Network (CNN) is robust artificial neural network structure which utilizes deep supervised learning. Its architecture has been specifically designed for handling sequential data and image data. Convolutional Neural Networks (CNN) have been effectively used for preparing two-dimensional images, it is feasible to extend the same concept to analyze data in one dimension. CNN utilizes only a limited set of features for extract the attributes of the incoming data and then combine them to produce intricate features for the input. By incorporating these intricate features of the air quality data into a fully connected layer within the model, we enable it to perform more accurate forecasting tasks, whether through

regression analysis for continuous values like AQI or classification for categorizing pollution levels. A CNN follows a specific architecture for processing data like images. It starts with an input layer receiving the data, then applies convolution layers with filters to extract features, followed by pooling layers to reduce size, then uses fully connected layers to combine the features, and finally outputs the result. In the main function of the convolution layer is to perform convolution processes on samples utilizing the kernel of convolution in order to provide the input data for the next layer. The pooling layer is a crucial component of CNNs, since it efficiently reduces the model's parameters and optimizes operations, while preserving the relevant data from the feature map. CNN uses layer-by-layer convolution and pooling procedures to extract data features. The filter has the ability to adjust the window size and its movement (step size) based on the amount of data it's given (input data size) [19]. This optimization helps the filter effectively extract the necessary features from the data. It is possible for the filter to find the ideal size of the window and window sliding size for each step by taking into account the size of the input data as well as the specifications for feature extraction it shows in Equation (13).

$$z_b^y = f \left(\sum_{a=1}^l z_a^{y-1} \otimes ck_{ab}^y + ne_j^y \right) \quad (13)$$

where the outcome feature map that represents the bth neuron in the currently active layer (layer y) by the representation z_b^y ; y represents the total amount of features that are input in the y-th convolutional layer; In expression z_a^{y-1} represents the input parameter for the current layer y; The process of convolution is represented by the symbol \otimes . The kernel of convolution connecting an ath neuron within the y-1 layers into the bth neuron on the y layer is represented by ck_{ab}^y ; ne_j^y is the standard deviation of bth neuron in y layer. The formula that is used to determine the activation function, which is indicated by the f , is as follows:

$$f(x) = \max(0, x) \quad (14)$$

The pooling layer, functioning as a subsampling layer, ensures the consistency in the mapping.

The mathematical representation of max-pooling is as below:

$$x_a^y = \max \text{pooling} (z_a^{y-1}, sp_{scale}, sl_{stride}) \quad (15)$$

The value of a-th neuron in the current layer y which is denoted as x_a^y . Max-pooling is a function used for down-sampling, where the maximum value in a specific range is chosen. The term sp_{scale} refers to the size of the pooling operation, whereas sl_{stride} refers to the distance between each pooling operation. The combination of CNN with Bi-LSTM is known as CNN-Bi-LSTM. The CNN – Bi-LSTM is a hybrid model that combines a convolutional neural network (CNN) for extracting features with a bidirectional LSTM model for predicting sequences. For the purpose of improving the Bi-LSTM, the novel approach is $1 - \tanh()$ function has been implemented to the output gate. This modification ensures that the output gate has a value range of about (0.24, 1). During the process of training the model, the enhanced Bi-LSTM not only possesses a highly effective learning capability, but it also possesses a fitting effect that is superior to that of the typical Bi-LSTM. As an outcome of this, enhanced Bi-LSTM is an appropriate method for evaluating the link among time series data. Bi-LSTM can identify the long-term temporal relationship between meteorological factors and contaminants. In order to simulate the bidirectional temporal structure, the Bi-LSTM is fed data from the output of an CNN location and trained using formulae (16) – (21).

$$ip_t = \sigma(Wt_i \cdot [hd_{t-1}, x_t] + b_{ip}) \quad (16)$$

$$\overline{CD}_t = \tanh(Wt_c \cdot [hd_{t-1}, x_t] + b_{cd}) \quad (17)$$

$$fg_t = \sigma(Wt_f \cdot [hd_{t-1}, x_t] + b_{fg}) \quad (18)$$

$$CD_t = f_t * CD_{t-1} + i_t * \overline{CD}_t \quad (19)$$

$$op_t = 1 - \tanh(\sigma(Wt_o \cdot [hd_{t-1}, x_t] + b_{op})) \quad (20)$$

$$hd_t = op_t * \tanh(CD_t) \quad (21)$$

Where,

fg_t - forgetting gate,

σ - sigmoid function

This model has both temporal and spatial depth, making it highly flexible to solve a wide range of prediction problems. The convolution process generates the correlation between two functions. This particular type of LSTM, known as Conv1D-Bi-LSTM, can be identified by its encoding forecasting architecture. The Conv1D-Bi-LSTM layer was constructed by including a Conv1D layer, two Bi-LSTM layers, and two dense layers in sequence. Figure 4. Illustrates the block diagram of Bi-LSTM -CNN model. This model applies padding before to the convolution operation to make sure that the dimensions of the state match those of the input data in terms of both rows and columns. The conv1D layer utilizes several settings, such as 32 filters, the Rectified Linear Unit (relu) activation function, a kernel size of 5, and strides of 1. Figure 5 shows the architecture of the Bi-LSTM -CNN model.

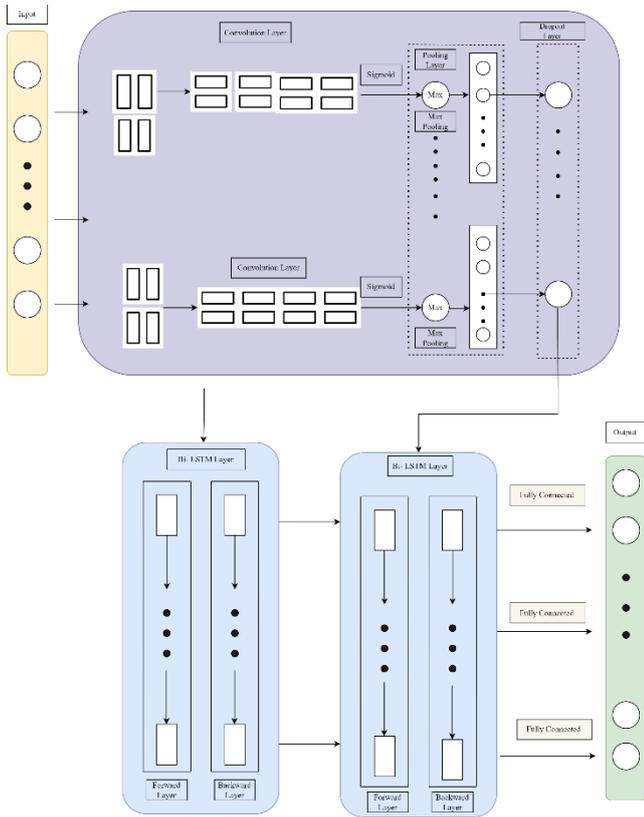


Fig 4: A block diagram of CNN – Bi-LSTM

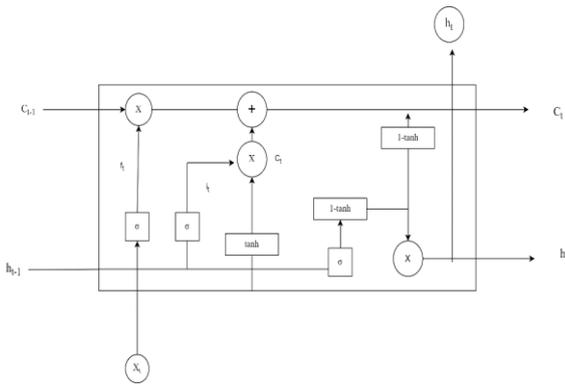


Fig 5. The architecture of the Bi-LSTM-CNN model

4. Results And Discussiong

4.1 CNN-Bi-LSTM parameter Settings

Table 3. Comparison of error metrics for different AQI forecasting models

Error Metrics	LSTM	Bi-LSTM	CNN-Bi-LSTM
MSE	0.68	0.62	0.58
RMSE	0.70	0.65	0.60
MAE	0.55	0.51	0.45
RMSLE	0.46	0.40	0.36

- Conv1D Filter: 32

This parameter signifies the quantity of filters, sometimes referred to as kernels, in the Convolutional layer. Every filter identifies distinct patterns or characteristics within the provided data. Increased number of filters enables the model to acquire complex

hierarchical characteristics.

- Conv1D kernel size: 1

The kernel size indicates the dimensions of the sliding window that passes the input data. When the kernel size is 1, the convolution operation is done to each element in the sequence separately. It is frequently utilized to capture trends within a certain area.

Conv1D activation function: Sigmoid

The model integrates non-linearity through the utilization of the

activation function. Sigmoid function is frequently utilized in binary classification tasks or when the objective is to constrain the output within the range between 0 and 1. It is appropriate for the last layer in forecasting a single output.

Conv1D padding: Same

Padding is utilized to maintain the spatial dimensions of the input volume. Padding with "same" ensures that the output sequence matches the length of the original input sequence.

Max Pooling 1D pool size:1

Max Pooling is a technique that decreases the spatial dimensions of the input volume by down sampling. A pool size of 1 indicates that no pooling action is performed, and each element is sent without modification.

MaxPooling1D padding: Same

Padding is used to preserve the spatial dimensions of the input after pooling, similar to convolutional layers.

Bi-LSTM units: 64

Bi-LSTM is a type of recurrent neural network (RNN) it performs analysis on sequence of input in both directions (forward as well as backward). The "units" parameter determines the quantity of memory cells or neurons within the LSTM layer. Increasing the number of units enables the model to capture complex temporal relationships.

Bi-LSTM activation function: tanh

To activate the LSTM units, the activation function is utilized. The hyperbolic tangent (tanh) activation function is frequently utilized in LSTM layers. It compresses the output values within the range of -1 to 1, which assists the network in addressing issues related to vanishing gradient.

Dense units: 1

Additionally, the Dense layer is a layer that is fully connected, and it is responsible for producing the final result of the computational model. This unit signifies that the model is intended for regression, which involves predicting a continuous numerical value.

Learning rate: 0.0001

The learning rate, as a hyperparameter, defines the degree of the optimization step. Although a slower convergence is possible with a lesser learning rate, it may aid in locating a more precise minimum. An extremely tiny step size is indicated by a learning rate of 0.0001.

The dependability of various models is assessed by means of many statistical measures, including MSE, RMSLE, MAE, and RMSE.

Mathematically, they are defined in the following manner:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_{ai} - \hat{y}_{pi})^2$$

Where: (22)

n - quantity of observations in our dataset.

y_{ai} - actual value.

\hat{y}_{pi} is the predicted value.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{ai} - \hat{y}_{pi})^2}$$

(23)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_{ai} - \hat{y}_{pi}|$$

(24)

$$RMSLE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log(1 + y_{ai}) - \log(1 + \hat{y}_{pi}))^2}$$

(25)

Where:

\log - natural logarithm.

y_{ai} - the actual value.

\hat{y}_{pi} - the predicted value.

During the pre-processing step of this research, AQI was calculated. After that, the calculated AQI was used the primary parameter, and the pollutant data (included 13 air pollutants, 7 meteorological factors, and 1 AQI) were used as feature variables for training the models. The measurement of the AQI entails a two-step procedure, as stated in Equation 1. In order produce these forecasts, a range of machine learning and deep learning models are used, including air pollution data as input variables to predict AQI values. A comparative analysis is conducted on the performance of three distinct models: LSTM, Bi-LSTM and CNN-Bi-LSTM. The evaluation is based on several performance measures. The performance metrics measure the difference between the predicted value and the actual value. MAE and RMSE values are nearly zero, illustrating

the maximum level of prediction accuracy. CNN-Bi-LSTM outperforms various deep learning models in terms of performance measures. Table 3 presents a comparison of the error rates among the three deep learning models.

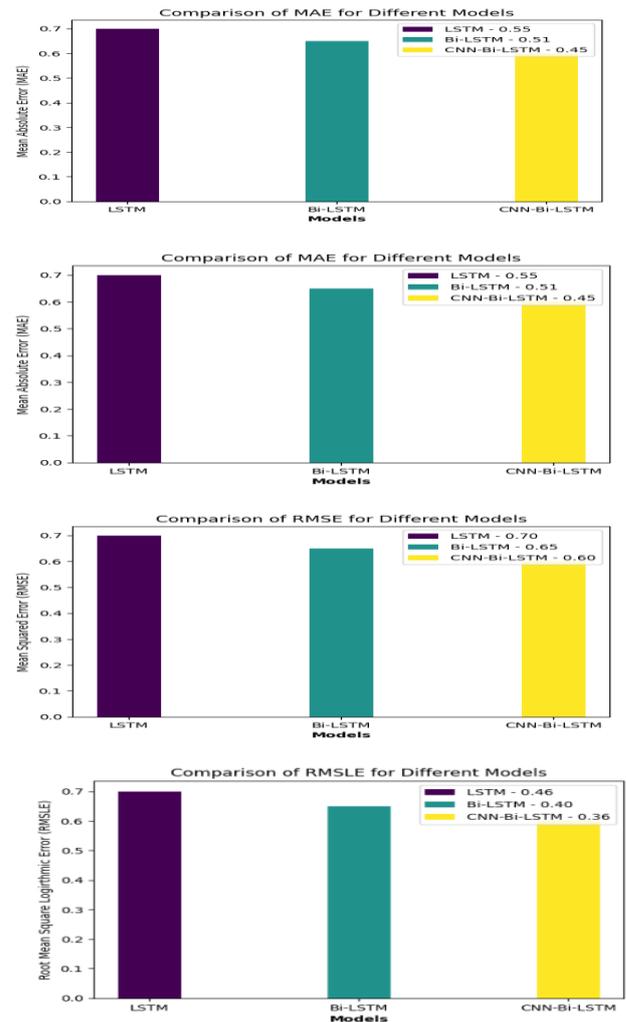


Fig 6. Comparison of Evaluation metrics

Figure 6 illustrates the all four-error metrics (MSE, RMSE, MAE, and RMSLE), the CNN-Bi-LSTM model consistently achieves the lowest errors, indicating its superior performance compared to LSTM and Bi-LSTM. The hybrid CNN-Bi-LSTM design, which integrates convolutional layers with bidirectional LSTM, has notable efficacy in identifying intricate patterns and relationships to AQI data, resulting in enhanced prediction accuracy.

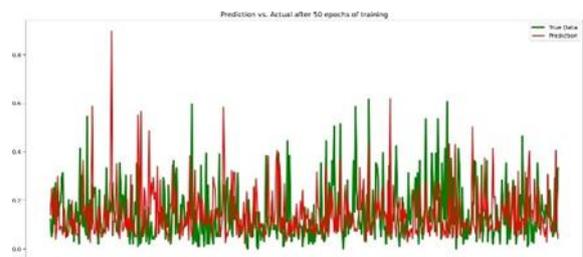


Fig 7: Prediction Using LSTM Model

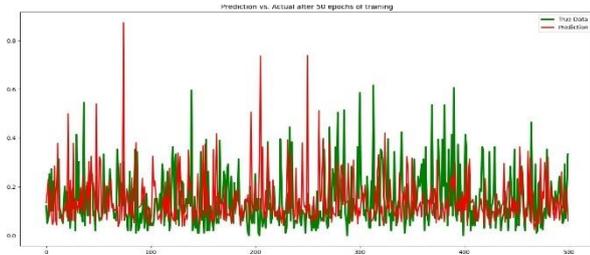


Fig 8: Prediction Using Bi- LSTM Model

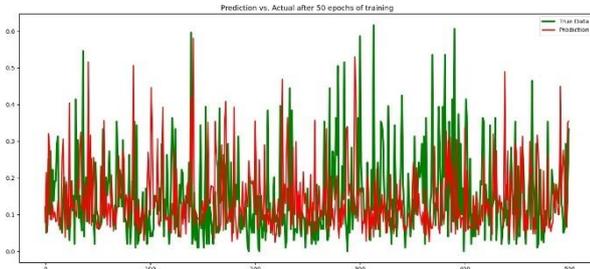


Fig 9: Prediction Using Bi-LSTM-CNN Model

Figure. 7,8 and 9 provide visual representations of the prediction accuracy of the LSTM, Bi-LSTM, CNN-Bi-LSTM models after 50 training epochs. It enables the identification of specific areas where the model struggles to accurately predict the true data. True data refers to the actual measurements of AQI that have been gathered and recorded over a period of time. The purpose of this data is to train and verify the forecasting model. The data usually consists of past AQI values for different sites, together with measurements of the corresponding pollutants. To generate a prediction, actual data (i.e., previous AQI measurements) are input into the model, and the model's underlying logic is then used to forecast future AQI values. The model's accuracy is assessed by comparing the expected results with the real AQI values, also known as true data. Visualizing predictions of LSTM, Bi-LSTM, and CNN-Bi-LSTM after 50 epochs reveals their strengths and weaknesses in early data pattern capture. Closest red-green alignment signifies best performance, while deviations highlight errors or missed trends. The CNN-Bi-LSTM's hybrid architecture performs well when comparing its prediction accuracy with LSTM and Bi-LSTM.

5. Conclusion

The AQI, derived from 13 air pollutant and 7 metrological factors functions as a dependable measure for evaluating the impact of pollutants in the air on the health of the people. Prolonged exposure to contaminants, such as those found in the environment, might result in respiratory disorders. This emphasizes the importance of accurate forecasting models. Our study shows results that across all metrics, the hybrid CNN-Bi-LSTM model is superior to the LSTM and Bi-LSTM models. The addition of convolutional layers improves the model's capacity to gather relevant features from the data, leading to more accurate predictions. These findings highlight the capacity of hybrid deep learning

methods to provide accurate estimates of AQI. This study enhances air quality management efforts by providing a strong foundation to protect human health from the detrimental impacts of outdoor air pollution.

Author contributions

M. Gayathri initiated the research topic and provided guidance throughout the project. **V. Kavitha** actively participated in the design and implementation of the modelling system. **Anand Jeyaraj** provided helpful insights and suggestions on various aspects of writing the paper. All authors read and approved the final version of the article.

Conflicts of interest

The authors declare no conflicts of interest.

References

- [1] Mohite, J., Sawant, S., Pandit, A., Pappula, S, "Impact of lockdown and crop stubble burning on air quality of India: a case study from wheat-growing region" *Environmental Monitoring and Assessment*, 194(2), 77. (2022).
- [2] Ghosh, Buddhadev, Barman, H. C., Padhy, P. K. "Analysis of spatiotemporal distribution of air quality index (AQI) in the state of West Bengal, India from 2016 to 2021". *Discover Atmosphere*, 1(1), 1. (2023).
- [3] Leong, W., Kelani, R., Ahmad, Z, "Prediction of air pollution index (API) using support vector machine (SVM). " *Journal of Environmental Chemical Engineering*, 8(3), 103208. (2020).
- [4] Ejaz, A., Malik, A. O., Lopez-Candales, A, "Ambient air pollution and cardiovascular disease: learnt from the COVID-19 pandemic." *Postgraduate Medicine*, 133(6), 587–588, (2021).
- [5] He, G., Pan, Y., Tanaka, T. "The short-term impacts of COVID-19 lockdown on urban air pollution in China". *Nature sustainability*, 3(12), 1005–1011 (2020).
- [6] Manisalidis, I., Stavropoulou, E., Stavropoulos, A., Bezirtzoglou, E. "Environmental and health impacts of air pollution: a review." *Frontiers in public health*, 8-14 (2020).
- [7] Al-Janabi, Samaher, Mustafa Mohammad, and Ali Al-Sultan. "A new method for prediction of air pollution based on intelligent computation." *Soft Computing* 24.1 (2020): 661-680.
- [8] Cabaneros, Sheen Mclean, John Kaiser Calautit, and Ben Richard Hughes. "A review of artificial neural network models for ambient air pollution prediction." *Environmental Modelling & Software* 119 (2019): 285-304.

- [9] Saravanan, D., and K. Santhosh Kumar. "Improving air pollution detection accuracy and quality monitoring based on bidirectional RNN and the Internet of Things." *Materials Today: Proceedings* 81 (2023): 791-796.
- [10] Mao, Wenjing, et al. "A hybrid integrated deep learning model for predicting various air pollutants." *GIScience & Remote Sensing* 58.8 (2021): 1395-1412.
- [11] Maleki, Heidar, et al. "Air pollution prediction by using an artificial neural network model." *Clean technologies and environmental policy* 21 (2019): 1341-1352.
- [12] Castelli, Mauro, et al. "A machine learning approach to predict air quality in California." *Complexity* 2020 (2020).
- [13] Ge, Liang, et al. "Multi-scale spatiotemporal graph convolution network for air quality prediction." *Applied Intelligence* 51 (2021): 3491-3505.
- [14] Dun, Ao, Yuning Yang, and Fei Lei. "A novel hybrid model based on spatiotemporal correlation for air quality prediction." *Mobile Information Systems* 2022 (2022).
- [15] Shahzad, Farrukh. "Does weather influence investor behavior, stock returns, and volatility? Evidence from the Greater China region." *Physica A: Statistical Mechanics and its Applications* 523 (2019): 525-543.
- [16] Wang, Jingyang, et al. "Air quality prediction using CT-LSTM." *Neural Computing and Applications* 33 (2021): 4779-4792.
- [17] Muzaffar, Shahzad, and Afshin Afshari. "Short-term load forecasts using LSTM networks." *Energy Procedia* 158 (2019): 2922-2927.
- [18] Zhang, Bo, et al. "Constructing a PM2.5 concentration prediction model by combining auto-encoder with Bi-LSTM neural networks." *Environmental Modelling & Software* 124 (2020): 104600.
- [19] Kavianpour, Parisa, et al. "A CNN-BiLSTM model with attention mechanism for earthquake prediction." *The Journal of Supercomputing* 79.17 (2023): 19194-19226.