A Novel Approach for Predicting Particulate Matter 2.5 and 10 Concentration using Modal Autoformer and Seq-2Seq Model

Rajendran Thavasimuthu [1], Vidhya PM[2], Arulkumar Varatharajan*[3], Sridhar Sekar[4], Ashokkumar Sighamani[5], Ajay Kumar Yadav[6]

[1]Department of Sustainable Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, Tamilnadu, India. rajendran.thavasimuthusamy@gmail.com

[2]Department of Information Technology, Rajagiri School of Engineering and Technology, Kochi, Kerala, India. vidhya_pm@proton.me

[3]Department of Database Systems, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India. arulkumaran.ckpc@gmail.com, +91 94860 78445

[4]Department of Research, Rajalakshmi Institute of Technology, Chennai, Tamilnadu, India. sridhar.sse@protonmail.com

[5]Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, Tamilnadu, India. ashokkumar_sse@proton.me

[6]Department of Computer Application, United Institute of Management, Allahabad, Uttar Pradesh, India. ajaykumar.uim@gmail.com

*Corresponding author: arulkumaran.ckpc@gmail.com

Graphical Abstract

Abstract: Precise and dependable forecasting of Particulate Matter 2.5 ($PM_{2.5}$) and PM_{10} levels hold significant importance for the public's ability to proactively mitigate exposure to air pollution and for informing governmental policy responses. Nonetheless, predicting PM2.5 and PM10 concentrations presents considerable challenges due to the complex dynamics of atmospheric flows. In existing mainstream research, most air pollution prediction models presently employ a single predictor, hence limiting the potential for enhancing stability and accuracy. This study proposes a pioneering methodology for forecasting $PM_{2.5\ and\ 10}$ concentration levels by integrating a Modal Auto former system with the Sequential-to-Sequential predictive model. The Seq-2Seq network model leverages sequential learning and square transformation of Long Short-Term Memory (LSTM) techniques for improved accuracy in $PM_{2.5\ and\ 10}$concentration prediction. Additionally, the incorporation of a Modal Auto former enhances the predictive capabilities by efficiently capturing nuanced variations in atmospheric conditions. The proposed Seq-2Seq LSTM network predictor is given a weight, and the Adaptive Beetle Feelers Optimization (ABFO) algorithm is utilized for weight optimization to attain the best prediction results. Through rigorous experimentation and validation, the proposed approach demonstrates superior performance compared to traditional methods using Air Quality Data in India from Kaggle, offering a promising avenue for precise $PM_{2.5\ and\ 10}$ concentration forecasting with practical implications for air quality management and public health initiatives. The Proposed seq-2seq LSTM model achieved 10.211 RMSE for PM2.5, 10.321 RMSE for PM10, 5.641 MAE for PM2.5, 5.764 MAE for PM10, 0.976 R^2 for PM2.5, and 0.945 R^2 for PM10.

Keywords: Air Quality, Adaptive Beetle Feelers Optimizer, Deep learning, Modal Auto Former, PM Concentration Forecasting.

1. INTRODUCTION

The exponential growth of the world's economy has increased concerns pertaining to the issue of air pollution [1]. Exhaust emissions, primarily resulting from the burning of fossil fuels, have significantly contributed to an increase in atmospheric pollutants. The term PM2.5 refers to particulate matter that has a diameter of 2.5 micrometres or less. These small, lightweight, and inhalable pollutants can last in the atmosphere for prolonged periods and provide a substantial risk to human health when found in substantial amounts. The World Health Organization (WHO) suggests that the average annual $PM_{2.5}$ concentrations should not surpass 5 µg/m³. The primary contributors of $PM_{2.5}$

emissions are the combustion of solid waste, road vehicles, and power plants [2]. As per the WHO, the global yearly mortality rate associated with air contamination is estimated to be more than seven million, and this rate is steadily increasing. WHO states that PM is a widely used proxy measure for air pollution. There exists substantial data supporting the adverse health effects linked to exposure to this contaminant. Sulfate, nitrates, ammonia, sodium chloride, black carbon, mineral dust, and water are the primary constituents of PM [3].

The top five countries with the highest levels of pollution in 2023 were:

- Bangladesh's $PM_{2.5}$ concentration (79.9 µg/m3) exceeds the WHO $PM_{2.5}$ yearly limit by more than 15 times. Bangladesh's elevated pollution levels can be attributed to the country's continuous traffic, construction operations, and industrial emissions, specifically from brick kilns that heavily depend on coal.
- Pakistan's $PM_{2.5}$ concentration (73.7 µg/m3) exceeds the WHO $PM_{2.5}$ yearly limit by more than 14 times.
- India's $PM_{2.5}$ concentration (54.4 µg/m3) exceeds the WHO $PM_{2.5}$ yearly limit by more than tenfold.
- Tajikistan's $PM_{2.5}$ concentration (49.0 µg/m3) exceeds the WHO $PM_{2.5}$ yearly limit by more than 9 times.
- The $PM_{2.5}$ concentration in Burkina Faso (46.6 µg/m3) exceeds the WHO $PM_{2.5}$ yearly limit by more than 9 times.

Out of a total of 134 nations and regions, 124 (92.5%) surpassed the annual PM2.5 recommendation value of 5 µg/m3 set by the WHO.

1.1. Air Pollution in India

India is one of the most rapidly expanding economies globally; however, its rapid process of urbanization and industrialization has had adverse effects on the country's environment and the well-being of its citizens. The country has seen significant water pollution, soil degradation, and poor air quality due to human activities, resulting in a substantial number of premature deaths annually. In the year 2023, New Delhi emerged as the capital city with the highest level of pollution globally, as evidenced by its average $PM_{2.5}$ concentration of 92.7 µg/m³. Subsequently, the capital city of Bangladesh, Dhaka, followed. Begusarai, located in northeastern India, has the highest $PM_{2.5}$ levels globally, with an average $PM_{2.5}$ concentration of around 119 µg/m³.

The monthly $PM_{2.5}$ concentrations in urban areas of India exhibited comparable trends from 2020 to 2023, with the winter months consistently exhibiting the greatest concentrations. During the specified time frame, the city of Delhi had the highest mean PM2.5 concentration, surpassing 255 µg/m³ in November 2023. According to data from 2023, Delhi had the third-highest mean PM2.5 concentration among cities in India, ranking below Begusarai and Guwahati. The prevalence of severe air pollution in India might have adverse health consequences for the nation's populace. Fine particulate contaminants can extensively infiltrate the pulmonary system, leading to respiratory complications and perhaps leading to premature mortality. By 2022, almost 96 percent of India's inhabitants were subjected to hazardous levels of atmospheric contamination as shown in Figure 1.

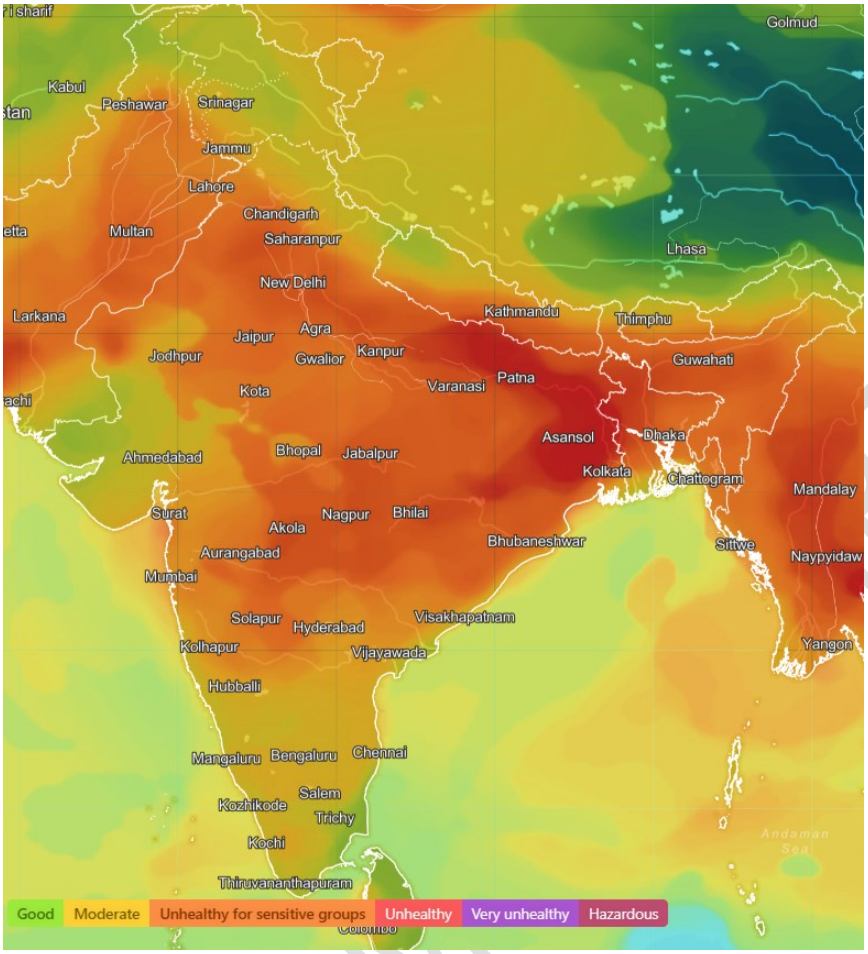


Figure 1. Map View of Air Quality in India (March 2024)

PM10 aerosols, which are PM with a size of 10 µm or less, are a kind of air pollutant that contributes to the decline in air quality. PM10 originates from a combination of natural and human activities, and its constituents are classified as primary (emitted directly) and secondary (produced in the atmosphere) in the natural environment. A considerable proportion of PM10 sources can be attributed to human activities. Several variables influence atmospheric PM10 concentrations, including local sources, dispersion, transportation, land-use patterns, geography, and meteorological conditions. The ambient air concentration of PM10 in Delhi, India's capital, was recorded at 181 micrograms per cubic meter in 2021. The pollutant levels remained consistently elevated for more than ten years. A higher quantity of particulate matter in the atmosphere has been linked to a wide range of physical, environmental, and health problems.

Fine particulate matter (PM2.5) and Inhalable particulate matter (PM10), which have aerodynamical dimensions lower than 10 and 2.5 µm, are widely recognized as significant pollutants [4]. Increased levels of PM2.5 and PM10 in the environment provide substantial health hazards, which may result in respiratory infections and conditions related to cardiopulmonary dysfunction, hence providing serious consequences to human health [5].

The precise prediction of air pollution provides significant early indicators and assists in the decision-making process for both governmental and public entities in addressing instances of severe pollution [6,7]. Hence, there is a pressing requirement for reliable and accurate prediction of ambient PM2.5 and PM10 levels to improve air quality and protect public health. The accurate forecasting of PM levels, including PM10 and PM2.5, over an extended period is an essential element in endeavours to comprehend and address the widespread problem of air pollution. PM comprises microparticles floating in the atmosphere and originates from several origins including automotive emissions, industrial operations, construction activities, and natural occurrences such as dust storms and

wildfires. The presence of these particles presents notable health hazards due to their ability to infiltrate the respiratory system, resulting in respiratory and cardiovascular ailments, alongside several other harmful health consequences.

The significant threat posed by air pollution, particularly the concentrations of PM2.5 and PM10, to public health and environmental sustainability is acknowledged. Accurate forecasting of these concentrations is considered crucial for proactive mitigation strategies and informed policy responses. However, limitations in accuracy and stability are often encountered by existing forecasting models due to the complexities of atmospheric processes. It is observed that traditional forecasting approaches typically rely on single predictors, which may not adequately capture the intricate interactions that drive air pollution dynamics. Consequently, a pressing need is identified for novel methodologies that can improve the precision and reliability of PM2.5 and PM10 concentration forecasts. This research addresses this need by proposing a novel methodology that integrates a Modal Autoformer system with a Seq-2-Seq predictive model. This innovative approach is intended to overcome the shortcomings of traditional forecasting methods by leveraging advanced techniques in sequential learning and optimization.

The novelty of this research present in the integration of a Modal Auto-former system with a Seq-2Seq LSTM network, enhanced by ABFO for weight optimization. Unlike traditional models that rely on single predictors, this approach captures complex atmospheric dynamics and variations in PM levels, significantly improving the accuracy and stability of PM2.5 and PM10 forecasting, as validated on Indian air quality data. By combining the capabilities of the Seq-2-Seq model with the nuanced insights provided by the Modal Autoformer, the proposed methodology aims to increase the accuracy and robustness of PM2.5 and PM10 concentration anticipation. Furthermore, the utilization of the ABFO algorithm for weight optimization is intended to further refine the forecasting process, enabling more precise estimations of air pollutant levels. Through experimentation and validation using real-world Air Quality Data from India, the research seeks to indicate the superior performances of the research model to the conventional approaches. By providing a more reliable means of forecasting PM2.5 and PM10 concentrations, the developed methodology has the potential to have significant implications for air quality management and public health initiatives.

The main contributions of this work are:

- The study proposes a novel methodology by integrating a Modal Autoformer system with the Seq-2Seq predictive model. This integration aims to increase the accuracy of forecasting PM2.5 and PM10 contamination levels.
- Incorporating a Modal Autoformer to further enhance the predictive capabilities of the model by efficiently capturing nuanced variations in atmospheric conditions. This enables the model to capture subtle changes in atmospheric dynamics, leading to more accurate predictions.
- Employing the ABFO algorithm to optimize the weights of the proposed predictor. This adaptive optimization approach dynamically adjusts the weights based on the model's performance, leading to improved forecasting results.
- Through rigorous experimentation and validation using Air Quality Data from Kaggle, the proposed approach demonstrates superior performance compared to traditional methods. The improved accuracy of PM2.5 and PM10 concentration prediction offers a promising avenue for precise air quality management and public health initiatives.

Section 2 describes the research data and methodology. The research methodology is evaluated and explained in section 3. Finally, the work is concluded.

2. LITERATURE REVIEW

There are three main categories of existing approaches used for predicting air pollution concentrations: numerical models, statistical models, and artificial intelligence (AI) techniques. Numerical models are utilized to replicate the intricate differential equations that control the physical and chemical mechanisms of pollutants present in the atmosphere. Notable instances of such models encompass Weather Research and Forecasting coupled with Chemistry (WRF-Chem) and Community

Multi-scale Air Quality (CMAQ). Nevertheless, the efficacy of these models is strongly dependent on comprehensive and frequently conflicting pollutant emission data, necessitating significant computational resources due to their intricate nature. On the other hand, statistical models like autoregressive integrated moving average (ARIMA) and autoregressive moving average (ARMA) rely on data and need minimal processing resources. However, they may encounter difficulties when dealing with nonlinear relationships and stationary data assumptions. AI models, such as artificial neural networks (ANN), random forest (RF), extreme gradient boosting (XGB), and support vector regression (SVR) have the capability to effectively capture intricate nonlinear relationships. However, these methods frequently need manual feature engineering and may encounter difficulties when dealing with extensive datasets because of data redundancy problems. Deep learning (DL) algorithms have recently gained attention as a possible method for predicting air pollution. This is because they possess the ability to learn on their own and effectively handle intricate nonlinear mappings.

Zhang et al. [8] provided a reliable prediction method that enables precise multi-steps forward forecasting of PM10 and PM2.5 levels. Following this, the corrected inputs were modelled using the convolution neural networks (CNN) based on residuals and has the ability to extract features. Ultimately, the effectiveness of this system was thoroughly evaluated by using five accuracy measures and two extra statistical tests. The STA-ResCNN model demonstrated a significant reduction in root mean square error (RMSE), ranging from 5.595% to 15.247% and 6.827% to 16.906%, for the average of 1- to 4-hour forward forecasts of PM2.5 and PM10 in three prominent cities, respectively. Yu et al. [9] provided a DL architecture called SpatioTemporal (ST)-Transformer, which utilized multi-head attention. The purpose of this design was to enhance the accuracy of spatiotemporal forecasts for PM2.5 concentration in areas exposed to wildfires. This model utilized the sparse attention method that focused on useful data across variable, temporal, and spatial dimensions.

Faraji et al., [10] introduced a model that integrated 3D CNN and GRU to predict the concentration of PM2.5 on an hourly and daily basis. The model demonstrated superior performance in comparison to ANN, GRU, LSTM, ARIMA, and SVR. Ding et al. (2011) proposed a CNN-LSTM method to predict PM2.5 concentrations by leveraging spatiotemporal correlations. The methods were utilized for extracting the spatial characteristics and temporal relationships of the inputs. The method demonstrated superior performances than the multilayers perceptron (MLP) and individual LSTM models. The accuracy of prediction was improved by including spatiotemporal correlation.

The spatial and temporal information were gathered using either a CNN or DNN architecture in [12] and [13]. In brief, the models possessed notable capabilities in their capacity to integrate geographical and temporal data, hence enabling precise prediction of PM2.5 air quality and concentration. Nevertheless, the majority of the models exhibited limitations in terms of short-term projections, with many models being restricted to certain locations or contaminants. Moreover, certain models may incur significant computing costs because of their intricate designs. Additional investigation was required to cultivate more effective and precise models that can be implemented on a broader scope. The study [14] introduced a balanced methodology of sampling to mitigate the issue of unbalanced data to increase the accuracy of PM2.5 prediction. Most enhancements derived from the LSTM architecture mostly focussed on feature extraction.

The research in [15] employed the BiLSTM method, which deviated from conventional LSTM by incorporating two distinct hidden layers to process the sequences in both the backward and forward directions. This approach effectively addressed the temporal and spatial correlations presented in the information and facilitated the modelling of intricate nonlinear associations within meteorological factors and air quality parameters. The study [16] employed a hybrid approach using principal component analysis (PCA), an attention mechanism, and long short-term memory (LSTM). Peralta et al., [17] provided a technique that utilized the LSTM recurrent network model to forecast the concentration of PM2.5 at any given geographical location. This method can forecast PM2.5 concentration for the upcoming day in a newer place where data were unavailable by considering air pollutant historical values and meteorological parameters (wind speed, temperature, relative humidity, and direction) assessed at stations fixed for monitoring.

The research in [18] introduced the CE-AGA model, which integrated the attention-based GRU with the convolutional encoders with adaptive gated activations. This model was specifically designed for predicting air quality. Several studies have employed transfer learning techniques to exploit pre-trained methods for associated operations, hence enhancing the efficacy of prediction models. The LSTM model was employed by Gul et al. [19] and Waseem et al. [20], who employed both partial fine-tuning of the parameters or structure. The utilization of the LSTM model facilitated the representation of temporal relationships, while the extensive array of trials conducted on diverse real-time monitoring of air quality data sets from many stations enhanced the applicability of the findings. Nevertheless, many research works failed to offer a comprehensive examination of the LSTM model's interpretability or the characteristics it acquired from air pollutant data were crucial for predicting based on DL models.

The methodology presented by Wang et al. [21] utilized LSTM, RF, and PSO. Particle Swarm Optimization (PSO) facilitated rapid convergence of the model to the optimal solutions, especially in search space with a high number of dimensions and intricate, nonlinear interactions among variables. The integration of PSO and ESN has the potential to achieve superior performance in the prediction of time series by capitalizing on ESN's proficiency in handling sequential information and PSO's test in conducting global search. The research in [22] proposed a DL hybrid approach that integrated the Particle Swarm Optimization and slime mould algorithm (SMA) into the adaptive neuro-fuzzy inferences system (ANFIS) for predicting PM2.5 levels.

Table 1. Analysis of Reviewed Studies

Study	Approach	Application	Advantages	Disadvantages
[8]	Residual-based CNN with feature extraction	PM10 and PM2.5 prediction	- Improved accuracy - Comprehensive assessment of performance	- Computational complexity - Reliance on detailed pollutant emission data
[9]	Multi-head attention-based deep learning	Spatiotemporal predictions of PM2.5 concentrations in wildfire-prone areas	- Sparse attention mechanism for useful contextual information	- Limited to specific scenarios - May require large datasets to train effectively
[10]	3DCNN-GRU	PM2.5 concentration forecasting	- Best results compared to other models	- Computational complexity
[11]	CNN-LSTM	PM2.5 prediction based on spatiotemporal correlations	- Enhanced prediction accuracy through spatial and temporal feature extraction	- Complexity in architecture
[12]; [13]	CNN or DNN architecture	PM2.5 concentration predictions	- Incorporates temporal and spatial data for precise predictions	- Limited to short-term predictions - Computational expense due to complex architectures
[14]	LSTM with a balanced sampling approach	PM2.5 concentration prediction with imbalanced data	- Addresses imbalanced data for improved prediction	- Limited analysis of interpretability - Complexity in feature extraction
[15]	BiLSTM model	Handling temporal and spatial correlations in data for air pollution prediction	- Handles spatial and temporal correlations effectively	- Computational complexity
[16]	PCA, attention mechanism, LSTM	PM2.5 concentration prediction with feature extraction and attention mechanism	- Incorporates principal component analysis for feature reduction - Attention mechanism for improved focus on relevant features	- Complexity in architecture

[17]	LSTM recurrent network model	Space-time prediction of PM2.5 concentration considering historical air pollutant and meteorological data	- Predicts PM2.5 concentrations at new locations using historical data	- Relies on the availability of historical data - Limited to fixed monitoring station locations
[18]	Attention-based GRU and convolutional encoder	Air quality prediction using adaptive gated activation and transfer learning	- Utilizes attention mechanism and transfer learning for improved prediction	- May require large datasets for transfer learning - Computational complexity
[19]; [20]	LSTM model	Air quality prediction using LSTM neural networks	- Models temporal dependencies effectively - Generalizable results due to experiments on multiple datasets	- Limited interpretability of the LSTM model - Complexity in training and tuning parameters
[21]	LSTM, RF, PSO	PM2.5 prediction using PSO, RF, LSTM.	- PSO aids in fast converging to optimal solutions - Combined strengths of PSO and ESN for time series forecast.	- Computational complexity
[22]	DL hybrid method	PM2.5 prediction using SMA, PSO, and ANFIS	- Novel hybrid model for PM2.5 predictions	- Complexity in combining different algorithms

From the above review, Traditional methods struggle for capturing the complex, nonlinear relations between input variables and PM concentrations. LSTM can automatically extract relevant features from raw data, learning complex representations that improve prediction accuracy. However, LSTMs are more computationally expensive compared to simpler recurrent architectures like the Elman RNN due to their additional gating mechanisms and cell state management. These increased complexities could result in longer training period and higher resource demand, making them less suitable for deployment on resource-constrained devices or in real-time applications. To solve this, the seq-2seq model is introduced here to improve the PM concentration prediction accuracy.

2.1. Research Gap

The research into PM2.5 and PM10 forecasting has predominantly concentrated on predictive modelling and algorithmic optimization, although it is lacking in an integrated combination of approaches that consider pollutant emission characteristics, spatial distribution dynamics, and real-world fuel impact analyses. Analyses of the morphology and nanostructure of emissions, as examined in studies of air pollution production, alongside the reduction properties of sustainable aviation fuels, may enhance comprehension of pollutant sources and variability [Chen et al. 2024; Gong et al. 2024; Meng et al. 2023; Xu et al. 2024]. Furthermore, employing methods for effective small-target identification in noisy environments and comprehending the factors influencing spatial pollutant distribution could improve the precision of forecasting models by addressing the complexities of dynamic atmospheric conditions and varied pollutant sources. This highlights the necessity of incorporating advanced emission analysis, spatial dynamics, and robust detection approaches into PM forecasting systems to effectively address these gaps.

3. RESEARCH METHODOLOGY

The proposed model employs a Seq-2seq LSTM network architecture to forecast the PM2.5 and PM10 concentrations, as seen in Figure 2. Initially, the Kaggle dataset is subjected to pre-processing procedures, which involve filling in missing values and normalizing the data. Subsequently, the dataset is divided into training and test sets to facilitate model training and assessment. During the training process, the Seq-2seq LSTM model is trained using each batch of

training data for a predetermined number of iterations. The loss value during forward propagation inside the network is minimized by employing the Adaptive Beetle Feelers Optimization (ABFO) technique via backpropagation. After training, the test data is loaded into the seq-2seq LSTM model to get predicted values for PM2.5 and PM10. To acquire the actual expected values, the predictions are de-normalized. Ultimately, the efficiency of the seq-2seq LSTM model is assessed using diverse measures, whereby the actual projected values are compared with the true values.

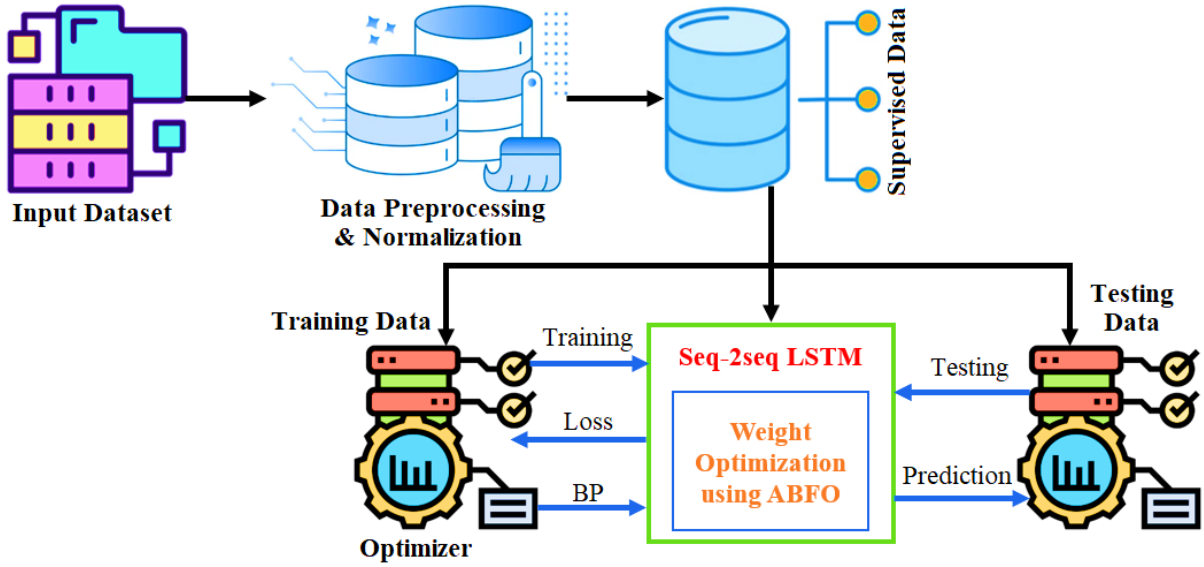


Figure 2. Seq-2seq LSTM-based PM concentration prediction

3.1. Dataset Collection

The dataset was obtained from the website https://www.kaggle.com/datasets/fedesoriano/air-quality-data-in-india Fedesoriano (2022). Providing data on significant air pollutants like particle matter (PM2.5 and PM10), carbon monoxide (CO), sulfur dioxide (SO2), nitrogen dioxide (NO2), and ozone (O3), across several cities in India. The dataset often contains timestamps that correlate to the day and time of measurement, as well as pollutant concentrations measured in quantities such as micrograms per cubic meter (µg/m³). This dataset is an essential source for analysts and researchers to analyze air quality trends, investigate the impact of pollution on public health, develop predictive models for predicting air quality, and evaluate the effectiveness of air quality management strategies and policies. The metadata includes information on monitoring locations and quality control. The map view of monitoring stations in India is displayed in Figure 2.

3.2. Data Pre-processing

The proposed approach considers the prediction of PM2.5 as a regression issue of time series, requiring continuous time series data as input. Still, it is frequently observed in practical situations that there are disruptions in the chronological order of data, resulting in the presence of data gaps. The issue pertaining to the absence of multi-modal and multi-site air quality information can be classified into two distinct classes: the complete absence of time series recordings, and the absence of feature values within individual records. Missing data-filling software mostly uses functional design to address the two categories of missing data. To address missing values in PM concentration data, this system must calculate the coefficients of cubic polynomials that interpolate between neighbouring known data points using cubic spline interpolation. The polynomials are subsequently employed to approximate the absent values.

The proposed approach produces cubic spline functions for every interval p_i, p_{i+1} using the known data points specified as p_i, q_i, where p_i indicates the time points and q_i indicates the PM concentration values.

$$CS_i(p) = w_i(p - p_i)^3 + (p - p_i)^2 + y_i(p - p_i) + z_i \quad (1)$$

The coefficients w_i, x_i, y, z_i need to be calculated. The original dataset was filled with missing values and subsequently normalized. Min-max scaling is a method that adjusts the data to fit inside the predetermined level, usually ranging from 0 to 1. The Min-Max scaling formula is as follows:

$$P_n = \frac{P - P_{min}}{P_{max} - P_{min}} \quad (2)$$

Here, P represent the initial PM concentration value, P_n represent the normalized value of P, and P_{max} and P_{min} indicate the highest and lowest values of P in the dataset. Autoformer is a transformer-based DL model. The system comprises an internal sequences decomposition unit, an enhanced decomposition structure based on the encoder-decoder, and a self-correlation mechanism. The decomposition unit utilizes the sliding average concept to extract and deconstruct the seasonal elements of time-series information. The primary aim of this process was to examine the intricate temporal patterns exhibited by a lengthy time series. To mitigate periodic oscillations and emphasize enduring trends, moving average lines are strategically incorporated. The incorporation of the sliding average effects will be accomplished by the sliding average window size manipulation.

3.3. Seq-2seq LSTM Model for PM Prediction

To acquire the time sequence attention of the multi-variable input data, the module of time sequence attention was employed to examine the various time data steps. Next, the input should be updated, and feature coding should be performed based on the attention received for each input data. The final projected value is obtained by fusing the matrix of encoded features with the historic data of PM2.5 concentrations and inputting it into the decoding features for decoding. A solitary LSTM unit consists of a memory cell and three gates, namely the input gate, the output gate, and the forget gate. At these gates, activation functions are utilized. A higher activation rate at the input gate indicates the need to store the input information in the memory. Conversely, a higher value at the output gate prompts the stored data release to the subsequent neurons. Lastly, a higher value at the forget gate eliminates data from the memory units. Although originally designed for neural machine translation, the Seq-2seq architecture has demonstrated its efficacy in a range of machine learning applications, such as time series prediction. The Seq-2seq model has modules, including the encoder, intermediate vector mechanism, and decoder. The input sequences are processed, and features are extracted in the encoder using LSTM cells. The context vector, which encompasses information from the complete input data, is derived from the final hidden state. The decoder consists of many LSTM units, with each unit performing calculations on its hidden state and producing output data. This paper presents a concept for the utilization of the Seq-2seq model in the prediction of PM concentration.

3.3.1. Encoder

In a Seq-2seq model, the encoder usually includes one or many LSTM modules. Every unit has responsibility for processing input sequences, gathering pertinent information, and transmitting it to the next unit. The mathematical expression that characterizes the functioning of the encoder is as outlined below:

$$hid_t = fn\left(wt^{hidhid} * hid_{t-1} + wt^{hip} * p_t + b_{hid}\right) \quad (3)$$

In this context, the variable hid_t denotes the hidden state at time step t, p_t represents the input at time step t, wt^{hihi} represents the weight matrix for the recurrent connections, wt^{hip} represents the weight matrix for the input connections, and fn represents the activation function.

3.3.2. Intermediate Vector

In the Seq-2seq model, the decoder component utilizes the final hidden state generated by the encoder as its beginning hidden state. The intermediate vector, which represents the hidden state, is calculated using equation (8). The primary objective of the intermediate vector is to integrate the knowledge acquired from the complete source sequence, serving as the initial hidden state of the decoder.

3.3.3. Decoder

The system comprises one or many LSTM units. The trailing hidden state is passed to each LSTM cell, which then creates both the output and the current hidden state. The equation provided was utilized to calculate the hidden state at the current step *t*, denoted as hid_t.

$$cs_t = fn\left(wt^{hidhid} * cs_{t-1}\right). \quad (4)$$

The cell state at time step *t* and the prior time step *t*-1 are denoted as cs_t and cs_{t-1} respectively. The equation [23] provides the output at each time step.

$$O_t = fn(wt^s * cs_t). \quad (5)$$

The Softmax function is utilized in the sequence-to-sequence (Seq-2seq) paradigm to produce the output sequence. Furthermore, it is possible to employ an attention mechanism, such as the Bahdanau attention mechanism, to capture the correlation between the input and output sequences. This approach ensures that the input and output sequences are aligned and that important information in the input sequence is given proper attention throughout the decoding process by giving alignment values. Through the utilization of attention, the model may choose to concentrate on pertinent segments of the input sequence, hence enhancing its capacity to produce precise and contextually appropriate output sequences [24].

$$cv_t = \sum_{i=1}^{T}(\zeta_{ti} * hid_i). \quad (6)$$

The context vector at time step *t* is denoted as cv_t, where *T* represents the length of the input sequence. The alignment scores among the current decoder hidden state hid_i and all the encoder hidden states hid_i are represented as ζ_{ti}.

$$\zeta_{ti} = \frac{\exp(es_{ti})}{\sum_{k=1}^{T}\exp(es_{tk})}. \quad (7)$$

Here, es_{ti} denotes the alignment energy score among the current hidden state of the decoder and the i-th hidden state of the encoder.

$$es_{ti} = cv^T * \tanh(wt[cs_{t-1}, hid_i]). \quad (8)$$

$$cs_t = \tanh(wt[cs_{t-1}, O_{t-1}, cv_t]). \quad (9)$$

The objective of ABFO is to enhance the LSTM's learning rate. To normalize, the moving average of the gradient square is employed. This allows for the augmentation of the step size under the vanishing gradient condition, as well as the reduction of the step size for larger gradients.

3.4. Adaptive BFO Algorithm for Weight Optimization

Initially, a weight selection objective function is established. The training dataset is denoted as $P_{tr} \in \mathbb{R}^{r\times m}$, where *r* is the sample size and m represents the number of features. The objective values that correspond to this are represented as $Q_{tr} \in \mathbb{R}^r$. The proportion among the fitting and validation sets is determined by the parameter *x*, which ranges from 0.3 to 0.95. The fitting model is assigned to the first $r_1 = x$. *r* samples of P_{tr}, while the remaining $r_2 = r - r_1$ samples are utilized for validation. The fitting set is denoted as $P_f \in \mathbb{R}^{r_1\times m}$, while the validation set is denoted as $P_{va} \in \mathbb{R}^{r_2\times m}$. The objective values for these sets are $Q_f \in \mathbb{R}^{r_1}$ and $Q_v \in \mathbb{R}^{r_2}$, respectively. It is crucial to acknowledge that validation plays a significant role in guaranteeing the model's ability to generalize beyond the training set. This is achieved by separating the fitting set P_f from the weights of the neural network *W*, which are directly derived using $M(1{:}r_1)$ and Q_f. The neural network predictions Q_v for the validation set P_v are derived using $M(r_1+1{:}r)$ and Wt. The formula used to compute the mean absolute error (MAE) between the objective Q_v and the predicted values is as follows:

$$MAE = \frac{1}{r_2}\sum_{k=1}^{r_2}\left|Q_k - \hat{Q}_k\right|. \quad (10)$$

The Mean Absolute Error (MAE) is frequently utilized in machine learning as a loss function, especially in regression tasks, due to its ability to quantify the average amount of mistakes between paired observations that reflect the same situation. Let us consider the vector is $[x, v, N]^T$, where N denotes a vector that encompasses the power values of the neurons in the hidden layer, and c represents a vector that encompasses the indices of the best activation function chosen from Table 1 for each neuron in the hidden layer. Algorithm 1 presents the approach in the form of an objective function.

Algorithm 1 Objective function.
Requirement: The vector p, the input data P and the target Q.
1: process objective fn(P, Q, p)
2: Divided p into x, v and N, and set r the rows number of P.
3: Ensure that only the nonnegative elements are retained in N, and in v, keep only their corresponding activation function numbering.
4: Compute the matrix K under the N and v.
5: Set $r_1 = pr, r_2 = r - r_1, P_{fi} = P(1:r_1, :), Q_{fi} = Y(1:r_1), P_{va} = P(r_1 + 1{:}r, :)\ and\ Q_{va} = Q(r_1 + 1{:}r)$.
6: calculate Wt utilizing $K(1:r_1)$ and Q_{fi} by using LSTM.
7calculate $\hat{Q}_{va}$ utilizing $M(r_1 + 1{:}r)$ and Wt.
8: assign the MAE using eqn (10).
9: **end**

The optimization technique outlined in Algorithm 1 involves the minimization of the objective function by the utilization of beetle behaviour. The optimization process entails the minimization of the objective function in relation to a vector $\theta = [x, v^T, N^T]^T$, where x is a parameter and v is a vector variable consisting of integer values 1, 2, 3, and 4, which correspond to the activation functions outlined in method 1. Furthermore, it should be noted that the vector N possesses an equivalent magnitude to that of v, with its elements spanning from $0\ to\ n_{max} - 1$, where n_{max} represents the upper limit of hidden layer neurons as determined by the user. The power of the activation functions for each neuron in the hidden layer is represented by the $n_{max} + 1$ values.

In the suggested methodology, the position of the beetle, namely its weight value, is represented by the vector p. The objective function *f(p)* in algorithm 1 is used to represent the concentration of odour at position p. The lowest value of *f(p)* indicates the origin of the odour. Furthermore, the notation p^t is employed, where t ranges from 1 to t_{max}, and t denotes the number of iterations. Thus, the lower bound $L = [0.3, 1^T, 0^T]$, where $1,0 \in \mathbb{R}^{(n_{max}+1)}$ represents the vectors all-ones and all-zeros, respectively. The upper boundary, denoted as (U)=$[\,0.3, 1^T.[4,1]^T.n_{max}]^T$. To make sure that $L \leq p \leq U$ is satisfied, the element-wise function provided for the element $i = 1, \ldots 2n_{max} + 1$ will be employed.

$$g(p_i) = \begin{cases} \mathrm{U}_i, & p_i > \mathrm{U}_i \\ p_i, & \mathrm{L} \leq p_i \leq \mathrm{U} \\ \mathrm{L}_i, & p_i < \mathrm{L}_i \end{cases} \tag{11}$$

Therefore, the beetle's chaotic search route defines a model of searching behaviour as follows:

$$Cs = \frac{\gamma}{\epsilon + \|\gamma\|'} \tag{12}$$

The expression $\gamma \in Ra^{2n_{max}+1}$ denotes a random vector consisting of $2n_{max} + 1$ elements, while $\epsilon = 2^{-52}$. The left (p_L) and right (p_R) feeler are created using the following formulae to imitate the seeking behaviors of the beetle's feeler:

$$p_R = g\big(Ra(p^t + \eta^t Cs)\big), \quad p_L = g\big(Ra(p^t - \eta^t Cs)\big). \tag{13}$$

In this context, the sensing breadth of the feeler, denoted as η^t, represents the capability of the exploit at the t-th instant. Additionally, consider the probable optimal solution (i.e. weight) (p_x):

$$p_x = g\left(Ra\left(p^t + \xi^t \eta^t \operatorname{sign}(f(p_L) - f(p_{Ra}))\right)\right), \tag{14}$$

The notation ξ^t denotes a step size, which signifies the rate of convergence after an increment in t during the search procedure. Following this, the behaviour of detection may be characterized as follows:

$$p^{t+1} = \begin{cases} p_x, & f(p_x) \le f(p^t) \\ p^t, & f(p_x) > f(p^t) \end{cases} \tag{15}$$

The subsequent section delineates the updating regulations pertaining to η and ξ.

$$\eta^{t+1} = 0.991\eta^i + 0.001, \quad \xi^{t+1} = 0.991\xi^i \tag{16}$$

The fundamental requirements must be known for the aforementioned methodology are as follows:

$$p^0 = [1 - y, 2 - y, \ldots., 2n_{\max} + 1 - y]^T, \tag{17}$$

Where $y = Ra\left((2n_{\max} + 1)/2\right)$.

Subsequently, the Seq-2seq LSTM method utilizes the whole training data set to identify and produce the optimal ratio x^* among the fitting and validation sets, the ideal weight (Wt), the optimal power value (N^*), and the optimal activation function for each neuron in the hidden layer (v^*). The proposed algorithm's pseudocode is illustrated in algorithm 2, while the flowchart is presented in Figure 3.

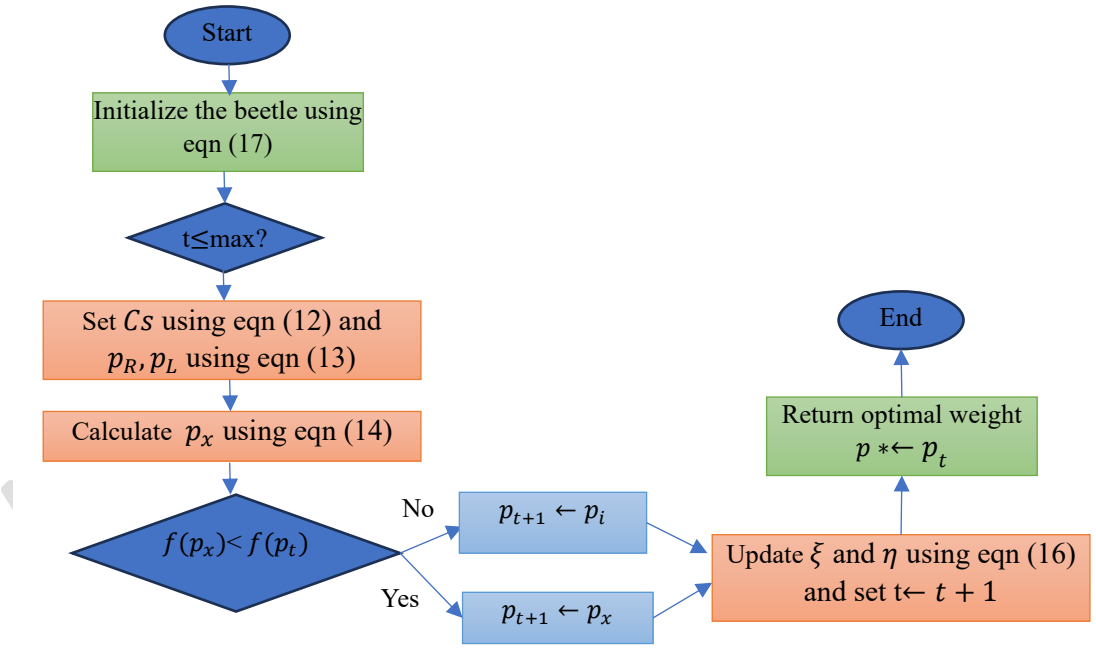


Figure 3. Flow chart of ABFO for weight selection

Algorithm 2: weight selection for seq-2seq LSTM
Input: objective fn(P, Q, p) , utilize the notation p^t, where t=1,2, 3,…,t_{max}
Output: optimal solution
Initialize the beetle using eqn (17)
While(t≤max?)
Set Cs using eqn (12) and p_R, p_L using eqn (13)
Calculate p_x using eqn (14) Update the value

If $f(p_x) < f(p_t)$ then
Update $p_{t+1} \leftarrow p_i, p_{t+1} \leftarrow p_x$
Update ξ and η using eqn (16) and set t$\leftarrow t+1$ potential solution
Return optimal weight $p* \leftarrow p_t$
The best solution is considered for Optimal weight in seq-2seq LSTM.
End

4. RESEARCH RESULTS ANALYSIS AND DISCUSSION

4.1. Experimental Setup

To assess the superiority and generalization of the method proposed seq-2seq LSTM in this paper for long-term prediction, comparative experiments were conducted using several control groups, including the general GRU module, the informer model (known for its superior short-term forecasting capabilities), the Autoformer method, and a model combining empirical mode decompositions with the GRU module (modal GRU). The completed dataset was divided into a training dataset comprising 85% of the data and a test dataset comprising the remaining 15%. For the models using non-modal decomposition and those employing modal decomposition, automatic parameter adjustment functions were incorporated. The key difference lies in the timing of parameter adjustment: for the non-modal decomposition model, adjustment happens after every prediction process completes, whereas, for the modal decomposition model, modification takes place after the predictions of all the components. Initial settings of parameters were based on the features of various models for hyperparameter tuning. The proposed seq-2seq LSTM performance is evaluated and the performance is compared with existing STA-ResCNN [8], CNN-LSTM [11] and LSTM [16] schemes. In the experiment, the model hyperparameters are configured as follows: The training process comprises 1,000 iterations with a batch size of 128 and a rejection rate set at 0.1. The model architecture includes GRU (Gated Recurrent Unit) layers with 64 hidden units, and the input data incorporates sequences of eight long-term historical data points. Training is facilitated using the Adam optimizer, with the MSE serving as the chosen loss function. The total objective function aims to minimize the MSE across the training iterations, thus optimizing the model's predictive performance.

$$O\left(X^{t\sim t+k}, \hat{X}^{t\sim t+k}\right) = \frac{1}{Ns}\sum_{i=1}^{Ns}(X^{t\sim t+k}, \hat{X}^{t\sim t+k}) \quad (18)$$

Where Ns was the total training samples. Three parameters were utilized to compute the differences among the actual value $\hat{X}_t$ of PM2.5 and the projected value X_t . The PM2.5's mean value was defined as $\bar{X}$, which includes:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_t - \bar{X}_t)^2} \quad (19)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|X_t - \bar{X}_t| \quad (20)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(X_t - \bar{X}_t)^2}{\sum_{i=1}^{m}(X_t - \bar{X}_t)^2} \quad (21)$$

Among these metrics, MAE and RMSE were employed to quantify the disparity among the original and predicted values. RMSE indicates the model's sensitivity to huge errors, while MAE indicates its reliability. Smaller values of both RMSE and MAE signify better predictive performance. Additionally, R^2 the predictive accuracy of the model relative to the actual data. A higher R^2 value indicates a more effective forecasting outcome.

4.2. Comparison of Different Time Window Sizes for PM Concentrations 2.5 and 10

PM2.5 and 10 concentration data are influenced by various relevant time series, although variations in all the time series values do not impact the PM2.5 concentration values. This indicates a lag effect, where the variable value at the previous moment affects the PM2.5 concentration values at

the following moment with a lag. While the lag effect might be pronounced in the short period, it diminishes in the long term.

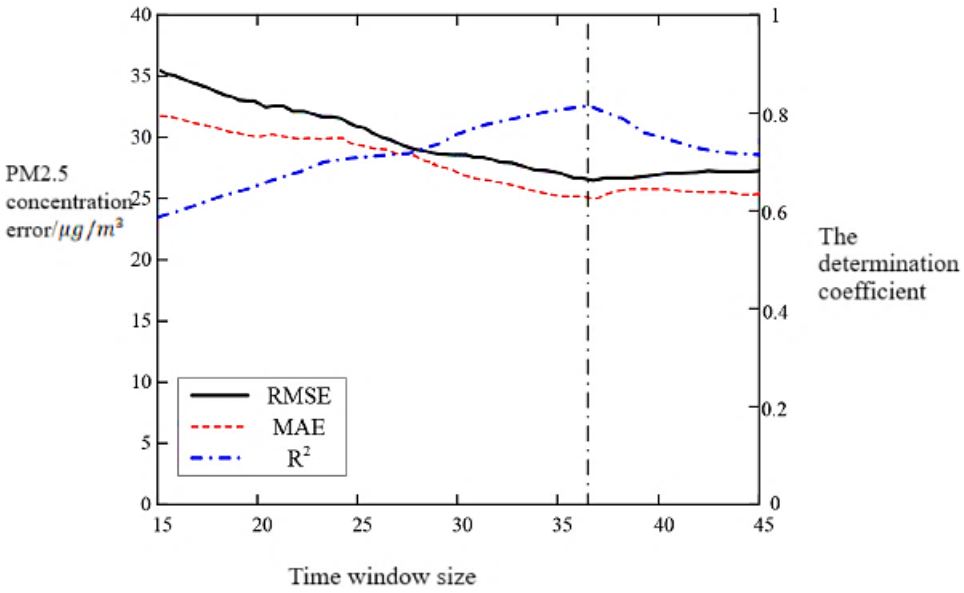


Figure 4. Different Time Window Sizes Performance for PM Concentrations 2.5

A smaller window size may not provide sufficient long-term memory input for the Seq-2seq LSTM model, while a larger window size may introduce irrelevant information, increasing unnecessary computational complexity. Hence, determining the optimal window size is crucial. A sliding window strategy was employed to create relative time series samples for all the records. To establish a suitable historical time window size, various values from the candidate sets [12, 16, 20, 24, 28, 32, 36, 40, 44] are selected. The changes in MAE, RMSE, and R^2 of the research model were depicted in Figures 4 and 6 for PM 2.5 and 10 to guide the selection process. As illustrated in figure 4 and 5, if the size of the window was less than 36, both the MAE and RMSE evaluations decreased while the R^2 evaluation value improves with increasing window size. This trend is attributed to the limited historical feature information inputted to the model when the window size is too small, resulting in lower prediction performance. Conversely, as the size of the window increases gradually, the research model receives more historical data as input, enabling it to capture additional nonlinearities and dependencies within the sequence, thereby enhancing predictive ability. However, when the size of the window surpasses 36, the values of MAE and RMSE start to increase, while the evaluation value of R^2 decreases before stabilizing. This phenomenon occurs due to the excessive input of unnecessary information with larger window sizes, leading to increased noise and interference with the model's performance. Consequently, in the experiment, the historical time window size was set to 36 to achieve the optimal balance between capturing relevant historical features and mitigating noise interference.

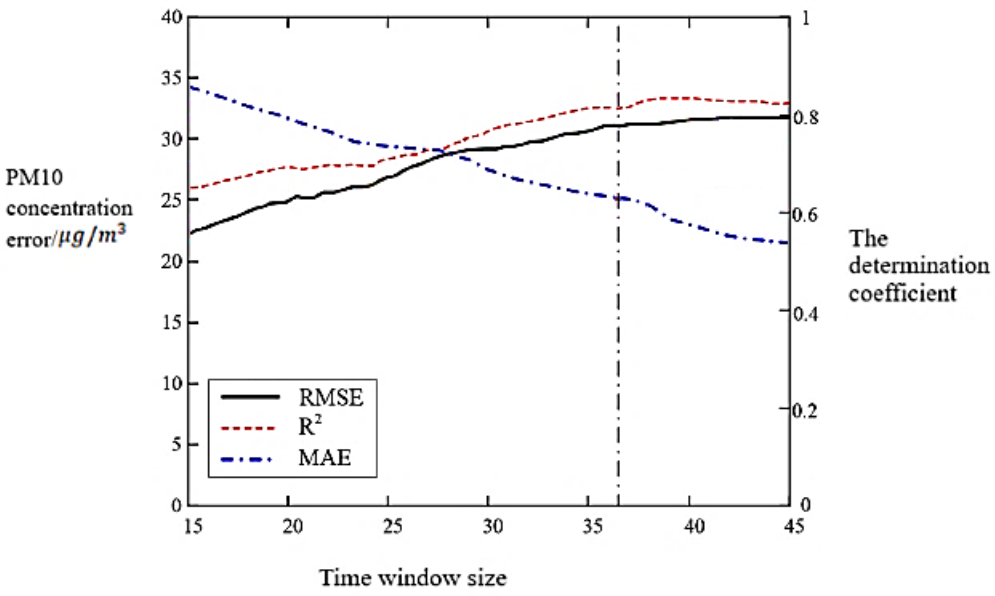


Figure 5. Different Time Window Sizes performance comparison for PM concentrations 10

The overall performance of the proposed scheme in terms of RMSE, MAE and R^2 are depicted in table 2. It shows the performance numeric evaluation and is compared with the current scheme's numeric values. It shows the proposed seq-2seq LSTM attained better performance results compared to current schemes. The proposed seq-2seq LSTM models are relatively easier to train and tune compared to complex convolutional architectures like STA-ResCNN. This simplicity in model design and training process may lead to faster convergence and better generalization performance and it's known for their ability to handle noisy data and missing values effectively. LSTM's robustness to such noise can result in more reliable predictions compared to STA-ResCNN.

Table 2. Overall Performance Comparison Among PM Concentration Prediction Schemes

Methods	RMSE ($\mu g/m^3$)		MAE($\mu g/m^3$)		R^2	
	PM 2.5	PM 10	PM 2.5	PM 10	PM 2.5	PM 10
Proposed seq-2seq LSTM	10.211	10.321	5.641	5.764	0.976	0.945
STA-ResCNN	11.971	12.232	6.938	7.24	0.828	0.834
CNN-LSTM	12.659	12.896	7.296	7.542	0.762	0.745
LSTM	13.536	13.876	7.781	8.122	0.668	0.675

4.3. RMSE performance comparison

Figure 6 shows the RMSE performance comparison among the proposed seq-2seq LSTM model and compared with existing PM concentration prediction schemes like STA-ResCNN, CNN-LSTM and LSTM. It shows the RMSE of proposed and existing schemes for PM 2.5 and PM 10, and the results show that the proposed scheme attained less RMSE compared to others. The proposed seq-2seq LSTM is designed to handle sequential data with varying lengths and time lags. PM concentration prediction involves forecasting future values based on historical observations, which aligns well with the sequential nature of LSTM models. The model's ability to retain relevant information over time enables it to make accurate predictions, resulting in lower RMSE. As well as it has a high capacity to learn complex temporal patterns present in PM concentration data. They can capture both short-term fluctuations and long-term trends, allowing them to adapt to the dynamic nature of air quality data. This capacity to learn intricate patterns contributes to the model's ability to achieve lower RMSE.

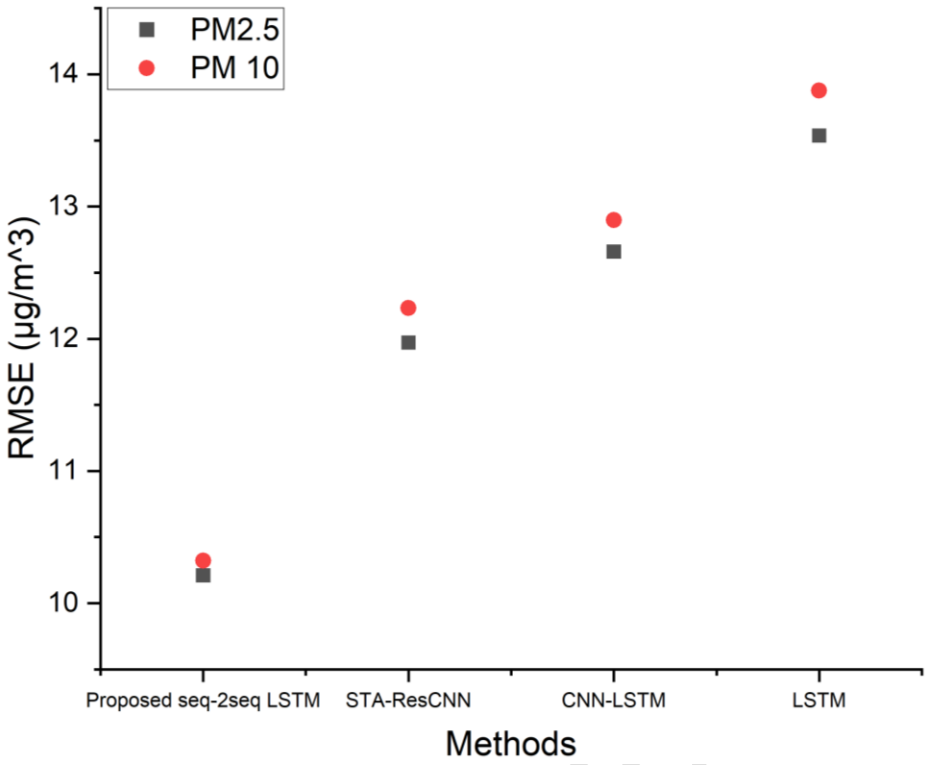
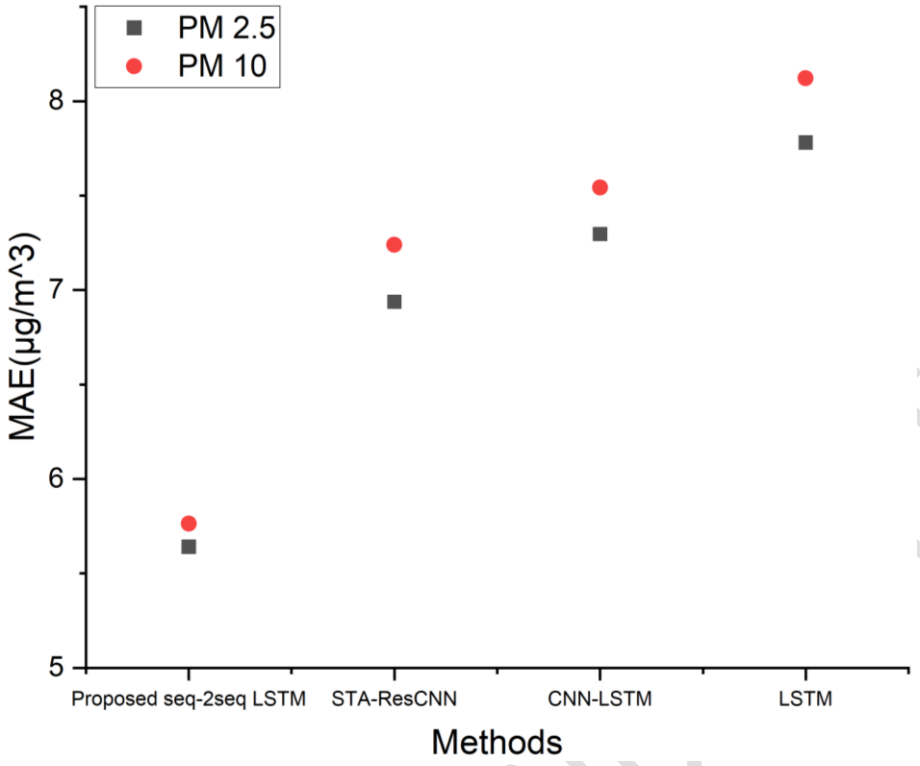


Figure 6. RMSE performance comparison among PM concentration schemes

4.4. MAE performance comparison

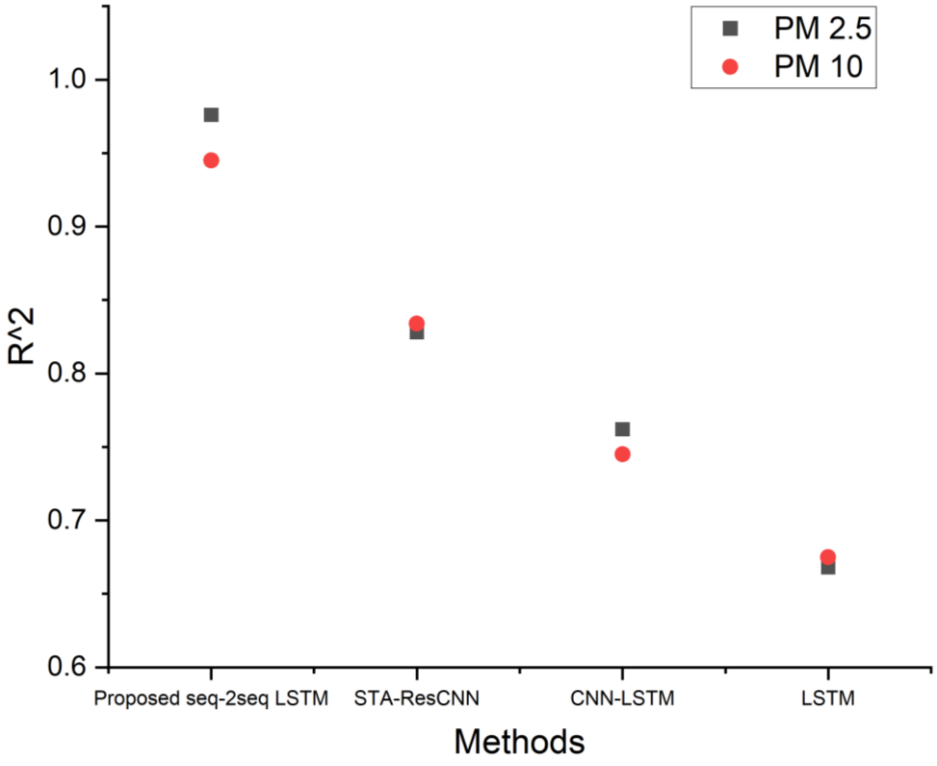
Figure 7 shows the MAE performance comparison among the proposed seq-2seq LSTM model and compared with existing PM concentration prediction schemes like STA-ResCNN, CNN-LSTM and LSTM. It shows the MAE of proposed and existing schemes for PM 2.5 and PM 10, and the results show that the proposed scheme attained less MAE compared to others. The proposed model has a high capacity to learn complex relationships between input features and target variables. PM concentration prediction often involves capturing intricate relationships between various environmental factors, such as weather conditions, geographic features, and pollutant emissions. The LSTM's ability to learn these relationships can lead to more accurate predictions and hence lower MAE. As well as it offers interpretability by allowing analysts to understand the importance of different features in predicting PM concentrations. PM concentration data may have irregular time intervals between observations due to factors such as sensor sampling frequency or data collection schedules. The proposed seq-2seq LSTM model can handle irregular time intervals effectively, allowing them to maintain predictive accuracy without requiring interpolation or resampling of the data.



519

520 **Figure 7. MAE performance comparison among PM concentration schemes**

521 **4.5. R^2 performance comparison**



522

523 **Figure 8. R^2 Performance comparison among PM concentration schemes**

524 Figure 8 shows the R^2 performance comparison among proposed seq-2seq LSTM model and
525 compared with existing PM concentration prediction schemes like STA-ResCNN, CNN-LSTM and
526 LSTM. It shows the MAE of proposed and existing schemes for PM 2.5 and PM 10, and the results

show that the proposed scheme attained high R^2 compared to others. PM concentration data may have irregular time intervals between observations due to factors such as sensor sampling frequency or data collection schedules. The proposed model can handle irregular time intervals effectively, allowing them to maintain predictive accuracy without requiring interpolation or resampling of the data. This flexibility in handling irregular time intervals contributes to higher R^2 values. As well as it is known for its capability for generalizing well to unknown data. By capturing initial data patterns, it can make predictions accurately even on data points not seen during training, leading to R^2 values on test or validation datasets. For the above reasons, the proposed scheme attained better performances compared to others.

The proposed model offers significant advantages in air pollutant concentration prediction, exhibiting improved accuracy and strong predictive capability compared to existing models. By integrating advanced techniques such as Modal Autoformer and Seq-2Seq LSTM, the model achieves comprehensive forecasting by capturing detailed variations in atmospheric conditions. However, computational complexity and potential challenges in model interpretability may limit its scalability and utility in decision-making processes. The limitation of the research is its dependence on the Indian air quality dataset, which could limit the model's generalizability to regions with different atmospheric conditions and pollutant sources. As compared to the current models results, the developed research model has gained lower RMSE and MAE, and better R^2 results. However, there is potential to improve the results by capturing variations in concentrations of PM. The obtained results indicate that the research model's predictive capability could slightly reduce in highly or extreme dynamic weather conditions. Additionally, the computational complexity of integrating Modal Auto-former and Seq-2Seq LSTM networks, along with ABFO optimization, could pose challenges for real-time applications in resource-constrained environments. Additionally, the model's performance could be perceptive to the quality and integrity of input data, requiring robust preprocessing and handling of missing values for accurate predictions. Addressing these limitations through further research and refinement could enhance the model's applicability in real-world air quality forecasting scenarios.

5. CONCLUSION

This research presented a novel system for predicting PM2.5 and PM10 concentration levels by combining a Modal Autoformer with the Seq-2Seq predictive model. The Seq-2Seq model employed sequential learning and square transformation methods to improve accuracy in concentration prediction, while the integration of a Modal Autoformer capture subtle variations in atmospheric conditions. The developed model optimized by the ABFO algorithm exhibited enhanced performance relative to conventional techniques. The optimized research model was validated through extensive testing with Air Quality Data from Kaggle repository. The results demonstrated the model's potential for accurate forecasting of PM2.5 and PM10 concentrations with significant implications for air quality management and public health activities. The Proposed seq-2seq LSTM model achieved 10.211 RMSE for PM2.5, 10.321 RMSE for PM10, 5.641 MAE for PM2.5, 5.764 MAE for PM10, 0.976 R2 for PM2.5, and 0.945 R2 for PM10. The achieved results demonstrated the proposed model's superiority in PM2.5 and PM10 concentration prediction compared to existing methods. This was evidenced by lower RMSE and MAE values, alongside higher R^2 scores, signifying enhanced accuracy and predictive power.

In future, the research will investigate additional enhancements to the predictive model, such as incorporating attention mechanisms or exploring alternative deep learning architectures to capture complex patterns in the data more effectively. The research will focus on improving the generalizability by evaluating various datasets with different atmospheric conditions. Further the research will address the data quality issues with advanced preprocessing methods and the model can be optimized with advanced optimization technique for enhancing overall prediction accuracy.

Declarations:

Acknowledgements

Not applicable

Contributions

All authors have made substantial contributions to conception and ideas. All authors participated actively in the interpretation of data. This manuscript was initially prepared by the corresponding author and critically reviewed by the other co-authors. All the authors have revised and approved the final version of the manuscript.

Ethics approval statement

Not applicable

Competing Interests

The authors declare no competing interests.

Funding

There was no external/internal funding for this study.

References

Ahmad, W., Arulkumar, V., Parthiban, K., Bhuvaneswari, E., Arif, M., & Guruprakash, K. S. (2024). Convergence of Modern Technologies for Data Architectures. In Wireless Communication Technologies, CRC Press, **26**, 82-100.

B. Deep, I. Mathur, N. Joshi, An approach to forecast pollutant concentrations with varied dispersions, Int. J. Environ. Sci. Technol. **19** (2022) 5131e5138.

B. Peraltta, T. Sepúlvida, O. Nicolis, L. Caro, Space-Time Predictions of PM2.5 Concentration in Santiago de Chile Using LSTM Network, Appl. Sci. **12,** 113-127.

C. Wang, J. Zhang, J. Du, G. Wang, J. J. Klemes, B. Wang, Q. Liao, Y. Lang, Weather conditions-based hybrid model for multiple air pollutant forecasting and minimisations, J. Clean. Prod. (2022), **352**, 131-143.

Chen, L., Cui, B., Zhang, C., Hu, X., Wang, Y., Li, G., Chang, L., & Liu, L. (2024). Impacts of Fuel Stage Ratio on the Morphological and Nanostructural Characteristics of Soot Emissions from a Twin Annular Premixing Swirler Combustor. Environmental Science & Technology, **58**, 10558–10566.

Ding, C et al., 2021. A hybrid CNN-LSTM model for predicting PM2.5 in Beijing based on spatiotemporal correlations. Environ. Ecol. Stat, **28**, 503-522.

Dun, Y. Yang, F. Lei, A novel hybrid model based on spatiotemporal correlations for air quality predictions, Mobile Inf. Syst, **23,** 156-172.

Faraji, M., Nadi, S., Ghafarpasand, O., Homoyoni, S., Downey, K., 2022. An integrated 3D CNN-GRU deep learning methods for short-term predictions of PM2.5 concentrations in urban environments. Sci. Total Environ, **834**, 155-324.

Gong, H., Hu, J., Rui, X., Wang, Y., & Zhu, N. (2024). Drivers of change behind the spatial distribution and fate of typical trace organic pollutants in fresh waste leachate across China. Water Research, **263**, 122170.

Gul, S., Khan, G. M., & Yousaf, S. (2022). Multi-steps short-term PM 2.5 forecasting for enactments of proactive environmental regulations strategies. Environmental Monitoring and Assessments, **194**, 386-398.

K. H. Wasim, H. Mushtaq, F. Abid, A. M. A. Mahfouz, A. Shaikh, M. Turan, J. Rashid, Forecasting of air quality using an optimized recurrent neural networks, Processes, **10,** (2022), 239-253.

K. Zhang et al., Multi step forecasts of PM2.5 and PM10 concentration using convolutional neural networks integrated with spatialetemporal attentions and residual learning, Environ. Int. **171**, 345-367.

L. Shi et al., A balanced social LSTM for PM2.5 concentrations predictions based on local spatiotemporal correlations, Chemosphere, **291,** 133-144,

Li, C., Hamer, M. S., Zheng, B., & Cohen, R. C. (2022). Accelerated reductions of air pollutant in China, 2017-2020. Science of the Total Environments, **803**, 187-211.

Lin, Y. C. et al., (2020). Chemical characterizations of PM2.5 emission and atmospheric metallics elements concentration in PM2.5 emitted from mobile sources gasoline-fueled vehicle. Science of the Total Environments, **739**, 237-252.

Liu, Y., Zhou, Y., & Lu, J. (2020). Exploring the relationships between air pollutions and meteorological condition in China under environmental governances. Scientific report, **10**, 235-276.

M. A. A. Al-qaness, H. Fan, A. A. Ewees, D. Yousri, M. A. Elaziz, (2021) Improved ANFIS models for forecasting Wuhan City Air Quality and analysis COVID-19 lockdown impact on air quality, Environ. Res, **194**, 543-562,

M. Yu, A. Masarur, C. B. Boxe, (2023) Predicting hourly PM2.5 concentration in wildfires-prone area using a Spatio Temporal Transformers model, Sci. Total Environ. **860**, 345-364.

Mapok, A., Hyde, K. D., Hasan, K., Kemkuigno, B. M., Čamoková, A., Surup, F., ... & Stadler, M. (2022). Ten decadal advances in fungal biology leading toward human well-being. Fungal Diversity, **116**, 547-614.

Meng, S., Zhang, C., Shi, Q., Chen, Z., Hu, W., & Lu, F. (2023). A robust infrared small target detection method jointing multiple information and noise prediction: Algorithm and benchmark. IEEE Transactions on Geoscience and Remote Sensing, **61**, (1-17).

643 Pruthi, D., & Liu, Y. (2022). Low-cost natures-inspired deep learning systems for PM2.5 forecasts
644 over Delhi, India. Environment International, **166**, 107373.

645 Sutskever, O. Vinyals, and Q. V. Le (2014), "Sequences to sequence learning with neural network," In
646 Advance in neural information processing systems, NIPS, Cambridge, MA, United States: MIT
647 Press,**12**, (3104–3112).

648 W. Ding, Y. Zhu, Predictions of PM2.5 concentrations in ningxia hui autonomous regions based on
649 PCA-attentions-LSTM, Atmosphere, **13**, 143-168.

650 X. Wu, C. Zang, J. Zhu, X. Zang (2022), Research on PM2.5 concentrations predictions based on the
651 CE-AGA-LSTM models, Appl. Sci, **12**, 7009.

652 Xu, Z., Wang, M., Chang, L., Pan, K., Shen, X., Zhong, S., Xu, J., Liu, L., Li, G., & Chen, L. (2024).
653 Assessing the particulate matter emission reduction characteristics of small turbofan engine
654 fueled with 100% HEFA sustainable aviation fuel. Science of The Total Environment, **945**,
655 174128.

656 Yang, B. Y., Fan, S., Thiering, E., Seisler, J., Nowak, D., Dong, G. H., & Heinrich, J. (2020), Ambient
657 air pollutions and diabetes: a systematic review and meta-analysis. Environmental Research, **180**,
658 (138-153).

659 Yu, C. H. et al., (2016). A novel mobile monitoring approach to characterize spatial and temporal
660 variations in traffics-related air pollutant in an urban community. Atmospherics Environment,
661 **141**, (161-173).