Air Pollution Prediction using Attention Module with CNN -OptBiLSTM

Visu P[1], Rajesh Khanna M[2], Sundara Rajulu Navaneethakrishnan[3], Subramanian P[4]

[1] Department of Artificial Intelligence and Data Science, Velammal Engineering College, Chennai -600066, India. dr.visu.p@gmail.com

[2] Department of Information Tecnology, Vel Tech Multi Tech Dr Rangarajan Dr Sakunthala Engineering College, Chennai, India. rajeshkhanna@veltechmultitech.org

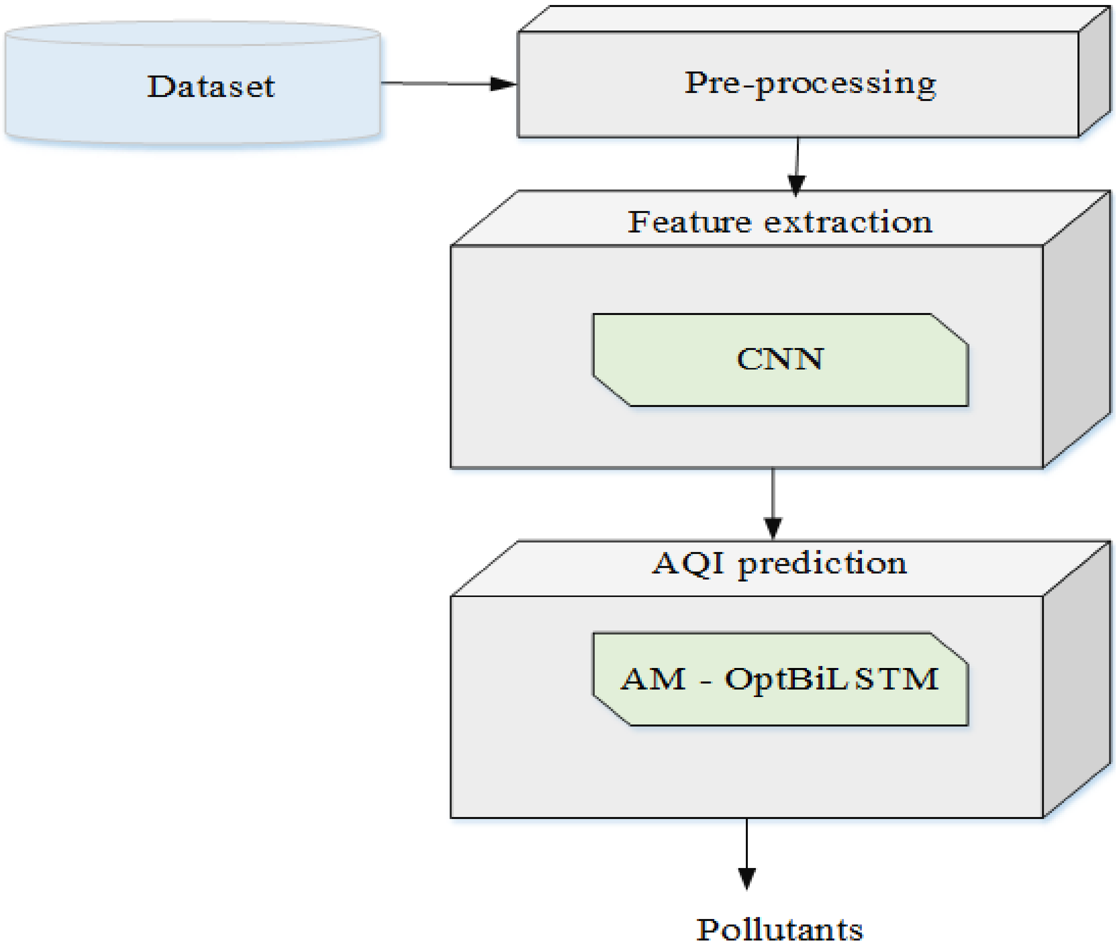
[3]Department of Computer Science and Engineering, School of Engineering and Technology, Dhanalakshmi Srinivasan University Trichy, Samayapuram Campus, Tiruchirappalli – 621112, India. drsundararajulu@gmail.com

[4] Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, 602105, India. subramanianp.sse@saveetha.com

*Corresponding author:

E-mail: dr.visu.p@gmail.com

GRAPHICAL ABSTACT



ABSTRACT

Predicting air pollution using environmental data assessment parameters becomes increasingly significant amid growing fears about climate change and the sustainability of urban areas. The use of sophisticated deep learning (DL) methods to model the intricate relation among these variables represents a promising area of research. However, current approaches have not effectively taken advantage of the temporal features derived from spatiotemporal correlations among air quality prediction systems, leading to poor long-term predictions. This work presents an AM (attention module) with CNN (convolutional neural network)-OptBiLSTM (optimal bidirectional long- and short-term memory) for AQIP (air quality index prediction). Here, the optimal process is carried out by the WSA (white shark algorithm). The analysis is demonstrated on the dataset and achieved better MSE and MAE values of 0.72 and 0.532

respectively. The developed model has the potential for application to other air pollutants. This proposed AM-CNN-OptBiLSTM has the capacity to substantially improve information services related to air quality prediction for the public. In addition, it provides support for regional pollution control and early warning systems.

Keywords: Air Pollution, Air Quality Index Prediction, Optimal Bidirectional Long-Short-Term Memory, White Shark Algorithm

1. Introduction

Air quality issues pose a serious threat to public health and constitute a widespread focus of research for researchers around the world. Due to the rapid growth of the world economy and the emergence of urbanization and industrialization, air pollution is a problem in various cities worldwide [1]. The challenges posed by air pollution are becoming increasingly evident, posing a serious threat to human productivity, life, and long-term social progress. As a prominent criterion of environmental pollution, air pollution has attracted global attention. Robust, reliable, and consistent AQIP (air quality index prediction) is essential for effective atmospheric environment management and public health management [2].

AQIP is an essential factor to judge the level of air pollution. Accurate prediction of air pollution levels is crucial for collaboration with governments and raising public awareness about the hazards of pollution [3]. Air pollution data is commonly described by identifying trends such as rising or falling patterns, seasonal variations, cycles, or erratic movements. AQIP play a crucial role in mitigating air pollution and addressing environmental degradation issues [4]. Consequently, a set of prediction factors is required to compile air quality statistics. However, there are numerous air pollution factors and they are complicated [5].

Existing research methods often do not provide effective predictions for air pollution. Notably, $O_3, PM_{2.5}, PM_{10}, CO, NO_2, SO_2$ emerges as the primary factors in air pollution, and the rising concentration of $PM_{2.5}$ directly impacts the health of human. The sub-index with the highest range is chosen as the AQI [6]. This quantitatively describes the AQI of the respective area in a specific manner. Currently, numerous air quality monitoring systems have been established in many places to monitor meteorological parameters and pollutants concentrations [7]. On the contrary, statistical approaches do not consider physical variation, transport, and chemical processes. They rely solely on data-driven exploration of the interior relation with the prior data. Consequently, the cost of computation for these methods is considerably less than that of numerical methods. Conventional statistical methods such as ARIMA (integrated adaptive moving average) and ARMA (integrated adaptive moving average) are very easy to evaluate [8]. But these approaches are well adaptable for small databases and modelling of single parameters. Additionally, they are based on linear consideration and impose high needs on data stationarity. Therefore, capturing non-linear relationships in the data becomes inherently challenging. These limitations significantly slow the efficiency and availability of conventional statistical methods in AQIP [9].

DL (deep learning) models prove to be well suited for addressing air pollution prediction challenges, particularly those involving nonlinear, sequential, cyclical and seasonal dependencies within pollutants [10]. The DL models like LSTM (long-short-term memory) and BI-LSTM (bidirectional LSTM) models are developed for capturing long-range dependencies from time series dataset, outperforming the ML (machine learning) models. Challenges such as predicting pollution levels and the parameters influenced by sequential-based behavior align well with the ability of the DL model to retain internal memory [11-13]. That is, the prediction of the pollution levels for every gas is based on previous analyses, where similar behavioral patterns will appear in the future [14].

Motivation: The improvement of AQIP accuracy is of significant importance in the control of air pollution and the improvement of air quality. The conventional models like RNN (recurrent neural network), GRU and ARIMA approaches may find it complex to capture deep features.

To address this, the study introduces AM-CNN-OptBiLSTM and utilizes AQIP and for handling long time series. Consequently, the proposed AM-CNN-OptBiLSTM model, compared to its existing counterparts, exhibits improved prediction accuracy through comprehensive learning, analysis, and historical data processing across different models. Training the DL model involves a crucial step of finding its hyperparameters. The choice of hyper-parameters is pivotal, as inappropriate selections can result in overfitting or underfitting, impacting the overall performance. Meta-heuristic-based approaches are employed for hyperparameter optimization for enhancing prediction performance. These algorithms exhibit global search ability, generalization, and robustness, making them suitable for addressing various problems. The foremost contributions are as follows:

To present an enhanced DL model for AQIP (air quality index prediction) and long-range dependencies.

To enhance the performance of AQIP by AM (attention module) with CNN (convolutional neural network)-OptBiLSTM (optimal bidirectional long-term memory).

To enhance the prediction performance of the WSA (white shark algorithm).

The remainder of the work is unfolded as follows: Section 2 delves into an exploration of related work that includes various air quality prediction approaches. Section 3, elucidates the algorithmic process of the proposed air quality prediction approach. The implementation and results of our method are examined in Section 4. Finally, Section 5 encapsulates the conclusion, summarizing the work, and engaging in a discussion on result analysis.

2. Related Works

Zhang et al. [15] introduced SABT (sparse attention based Transformer) for predicting the $PM_{2.5}$ pollutant. It was an encoder with a decoder model for reducing the complexity and the complex relation from the $PM_{2.5}$ and the RMSE value achieved was 0.93. Ravindiran et al. [16] introduced different ML models to predict AQIP in the coastal city of Visakhapatnam, India. When comparing all ML models, Catboost achieved better MAE and RMSE of 0.6 and 0.76 respectively. Janarthanan et al. [17] presented SVR (support vector regression) SVR for classifying the AQI values. The texture features were then extracted by the GLCM and the RMSE value achieved was 7.8.

Gilik et al. [18] presented CNN with LSTM model to extract spatial and temporal features in AQIP. The goals of this existing work involve creating a supervised approach to predict air pollution utilizing actual sensor data and transferring the model across different cities. In addition, this work performed various pollutants in cities such as Barcelona, Istanbul, and Kocaeli.

Lakshmipathy et al. [19] developed ESCA (enhanced serial cascaded autoencoder) based LSTM -MVR (multivariate regression) model for AQIP. That is, ESCA was exploited for feature extraction and LSTM –MVR was exploited for AQIP. Then, the FIFDO (fitness-improved flow direction optimizer) was utilized for producing better prediction results. Drewil et al. [20] presented a DL model LSTM and GA (genetic algorithm) to predict air pollution. The objective of this existing work was to identify the optimal hyperparameters for LSTM and predict the level of pollution for the next day based on different pollutants. The MAE and RMSE values achieved were 19.1 and 9.5 respectively.

Zhang et al. [21] presented a DL model CNN with LSTM for AQIP. Initially, CNN was exploited to extract features, and LSTM was exploited for AQIP. When comparing other approaches, CNN with LSTM achieved better performance. Mao et al. [22] presented TS-LSTME (temporal sliding long-short-term memory extended approach) for AQIP. This existing work incorporated the optimized time lagging to enable sliding prediction using a multilayer bidirectional long-short-term memory (LSTM) network. This involved considering the hourly historical concentration of PM2.5, meteorological data, and temporal data.

Liao et al. [23] developed DM-ST-GNN (Dynamic Multi granularity-Spatiotemporal- Graph Neural Network) for AQIP. It was an encoder-decoder-based model; on the encoder side, the spatial features were identified, and on the decoder side the attention LSTM was for learning temporal features. Zhang et al. [14] introduced residual learning based CNN model for forecasting $PM_{2.5}$ and PM_{10}. A spatial temporal attention module was exploited for assigning weights and the residual learning-based CNN was utilized to extract features. Finally, the RMSE and MAE values achieved were 11.9 and 6.9 respectively.

3. Proposed methodology

This study aims to create a DL approach to forecasting air pollution using the transferability of the model between different cities. The proposed AQIP approach involves the integration of a DL approach AM-CNN-OptBiLSTM. This combination is designed to predict air pollutant concentrations across various places within a city by capturing spatiotemporal relationships in the data. Figure 1 defines the framework of the proposed AQIP which includes various stages like preprocessing, spatiotemporal feature extraction, and AQIP. Here, the AM-CNN-OptBiLSTM is utilized for extracting spatio-temporal features of AQI. Moreover, the AM is deployed to focus on the essential features that have a better relationship with the AQI outcomes. This section outlines the step-by-step procedure flow for generating the AQIP model.

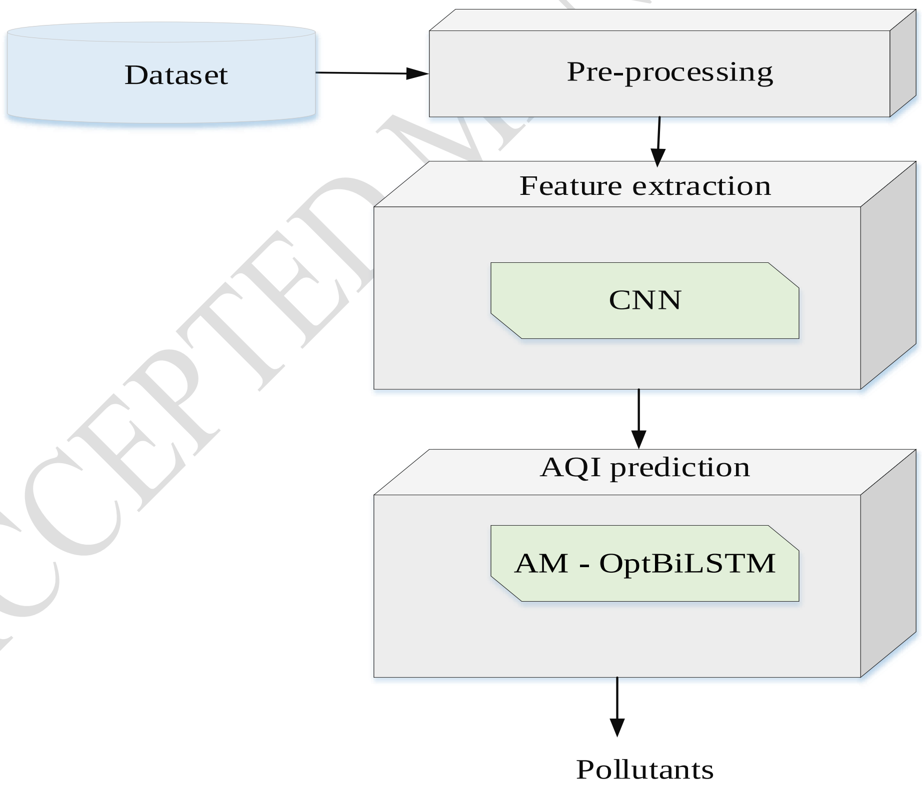


Figure 1. Framework of the proposed AQIP

3.1. Pre-processing

Initially, data is extracted from the data set and subjected to pre-processing to remove unnecessary information. At first, the missing values are removed from the dataset. Also, missing data imputation fills in any gaps or missing values in your dataset with estimated or calculated values. The preprocessed data is fed into the proposed classifier for performing the AQIP. Following the removal of missing values, a type conversion from object to floating data type was performed on the AQIP.

3.2. Feature extraction

In this research, a DL approach AM-CNN-OptBiLSTM is introduced to predict air quality in India. The method combines the advantages of CNN and BiLSTM, with AM for assessing each feature significance in the input data. CNN is utilized for the extraction of spatial features and AM with OptBiLSTM is utilized for AQIP. Figure 2 shows the structure of the proposed DL approach AM-CNN-OptBiLSTM. The network has two blocks, such as the feature extraction block and the AQI prediction block. Within the feature extraction block, the CNN is employed for extracting spatial features, and the AM-OptBiLSTM is utilized for capturing long-range dependency features and identifying the AQIP. The outcomes from the BiLSTM are subsequently input into the AM, and it assigns diverse weights to the model's feature input, emphasizing the impact of essential features, thereby aiding the model in making more accurate predictions. Subsequently, the AQI prediction block incorporates the FC (fully connected) layer and an output layer to identify the outcomes. Every layer has training variables like size of filter, loss function, kernels, and total neurons that minimize the error.

CNN: CNN is constructed by layering three fundamental components: convolutional, pooling, and FC layer. Within every convolutional layer, there exists a set of adaptable filters designed to extract local features from the input matrix in an automated manner. These filters execute convolution operations, according to the concepts of weights and local connectivity, to alleviate the computational complexity and enhance the efficiency of the network. The result of the convolutional operation is given as:

$$Z_{k,l}^{o} = \sigma_{con}\left(\sum_{n=0}^{m}\sum_{p=0}^{m} W_{n,p} \times Z_{k+n,l+p}^{in} + b\right) \tag{1}$$

where $Z_{k+n,l+p}^{in}$ and $Z_{k,l}^{o}$ are the input and output of the feature map, $W_{n,p}$ is the convolutional kernel, b is the bias and σ_{con} is the activation function.

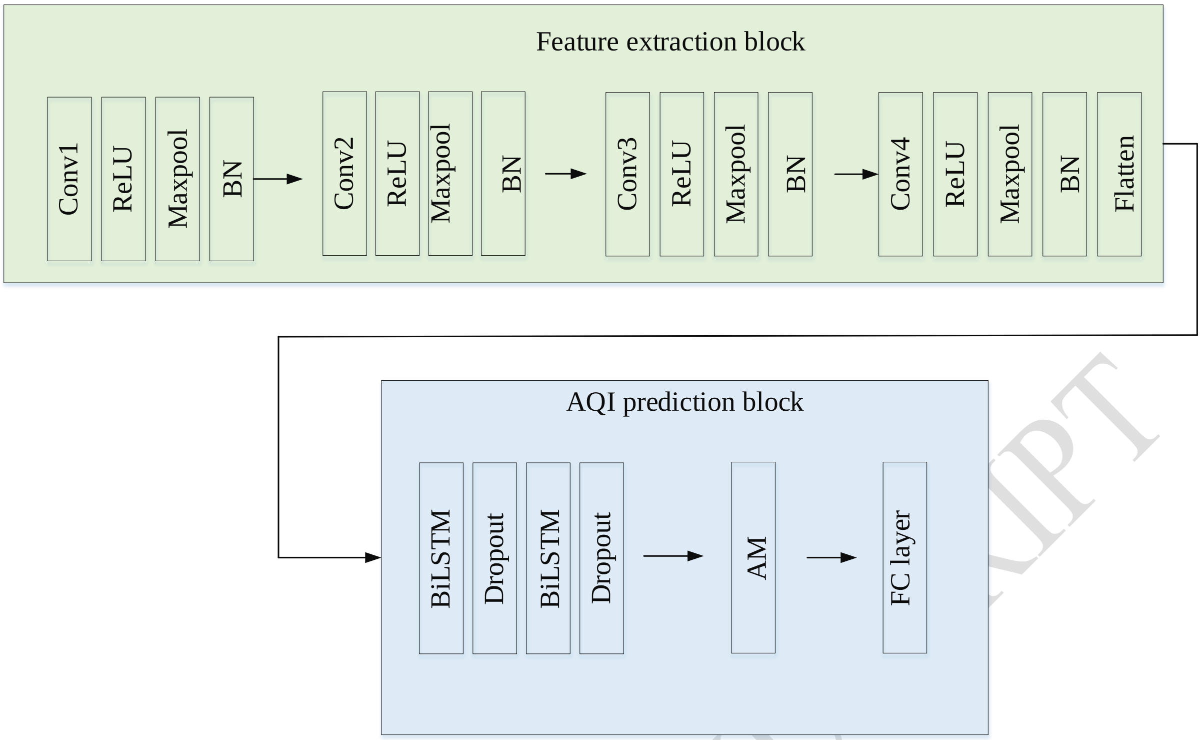


Figure 2. Structure of the proposed AM-CNN-OptBiLSTM

Following the convolution, the pooling layer executes the down-sampling operation. The advantage of the pooling layer is its ability to reduce the dimensionality of the feature map, thus preventing overfitting. In this feature extraction process, the ReLU is utilized as the activation term, and the BN (batch normalization) is utilized as the regularization term. Typically, the FC layer is incorporated finally and its role is to comprehend the non-linear combined features obtained using the convolution layer for generating the final outcomes.

BI-LSTM: The LSTM can only use information in one single direction (forward). On the other hand, the BI-LSTM structure consists of two LSTM layers, one of which functions forward and the other backward. The standard LSTM has the layers like input i_j, forget f_j and output o_j gates. Let the input data y_t at the present stage and h_{j-1} output from the hidden stage of the prior layer. The f_j is utilized for deciding what feature must be retained or eliminated, and it is given as:

$$f_j = \sigma\left(W_f \left[h_{j-1}, x_j\right] + b_f\right) \tag{2}$$

The i_j is utilized for deciding which features are updated, and it is given as:

$$i_j = \sigma\left(W_j \left[h_{j-1}, x_j\right] + b_i\right) \tag{3}$$

At last, the o_j is represented as:

$$o_j = \sigma\left(W_f \left[h_{j-1}, x_j\right] + b_o\right) \tag{4}$$

The hidden phase h_j is given as:

$$h_j = o_j \circ \tanh(c_j) \tag{5}$$

where c_j is memory cell.

On the contrary, the BILSTM network comprises two LSTM layers, positioned in both forward and backward directions. The forward LSTM is capable of assimilating information from the past information of the input sequence $\vec{h}_j$, while the reverse LSTM captures details regarding the future information of the input sequence $\overleftarrow{h}_j$. Subsequently, the results of both h_j are integrated and it is represented as

$$h_j = \vec{h}_j \oplus \overleftarrow{h}_j \tag{6}$$

where $\oplus$ is the summation.

AM: The AM selectively concentrates on essential features, ignores unnecessary details, and amplifies relevant information. The essence of the attention function lies in its definition as a map from a Q (Query) to values of the pairs of K (Key) and V (Value). As depicted in Figure 3, the calculation of AM has three stages. Initially, in the first stage, the correlation between the Q and every K is computed as:

$$A_t = \tanh\left(V_h h_t + a_h\right) \tag{7}$$

where A_t, V_h, h_t and a_h are the attention value, weight, bias, and input value. The value of the first phase is standardized in the second stage, and the softmax is exploited for converting the A_t.

$$b_t = \frac{\exp(A_t)}{\sum_t \exp(A_t)} \tag{8}$$

The third stage is obtained by the weighted summation of b_t and h_t and it is given as:

$$A = \sum_t b_t h_t \quad (9)$$

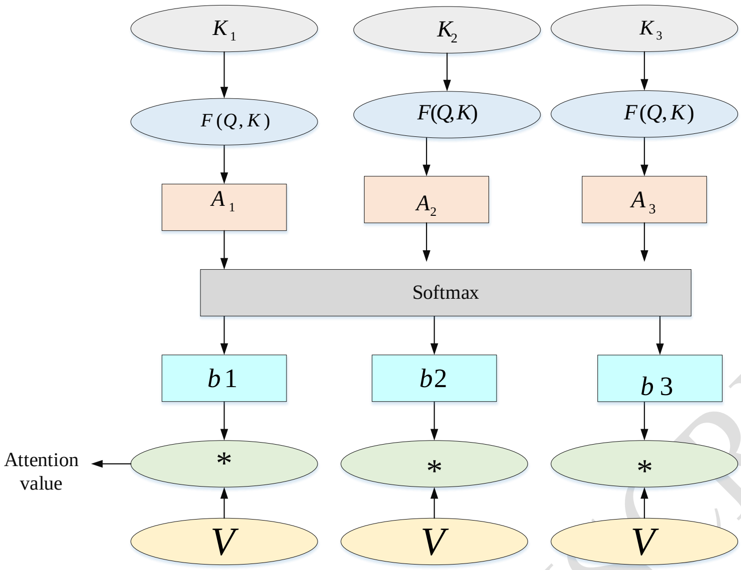


Figure 3. AM model

WSA: Determining the hyperparameters of AM-CNN-OptBiLSTM is a crucial step in the DL model. The training and overfitting issues are all greatly impacted by the selection of these hyperparameters, which in turn affects the final model's accuracy. In many instances, the selection of hyperparameters takes more time for attaining the best hyperparameters. To address the challenge of hyperparameter selection, particularly concerning the size of window and number of AM-CNN-OptBiLSTM units, the metaheuristic approach WOA is utilized. The AM-CNN-OptBiLSTM is trained with WSA to find the best window size and the number of AM-CNN-OptBiLSTM units and predict the level of air pollution.

This optimizer mimics the remarkable characteristics observed in the WS (white shark), particularly its exceptional olfaction and hearing used in foraging and navigation. These distinctive traits can be effectively modelled numerically and scrutinized mathematically, establishing a balanced approach for studying and deploying this scheme. This methodology helps search agents to explore and exploit various zones within the search area systematically, facilitating a better process. Simultaneously, the search agents in the WSA have the capability to adjust the positions randomly. The position of WS is given as

$$v = \begin{bmatrix} v_1^1 & v_2^1 \cdots v_d^1 \\ v_1^2 & v_2^1 \cdots v_d^1 \\ \vdots & \vdots \\ v_1^m & v_2^m \end{bmatrix} \quad (10)$$

where v is the position of shark and the it is computed by the upper ul_j and lower ll_j limits at the j^{th} dimension is given as:

$$v_j^t = ul_j + rand \times (ul_j - ll_j) \tag{11}$$

where $rand$ is the random number; When the high WS identifies its prey's location through wave frequency detection, it can approach the target using oscillating movements, guided by the expressed velocity.

$$u_{k+1}^t = \alpha\left(u_k^t + p_1\left[v_{gbk} - v_k^t\right]\times c_1 + p_2\left[v_b^u - v_k^t\right]\times c_2\right) \tag{12}$$

$$u = \left[m \times rand(1, m)\right] + 1 \tag{13}$$

where m is the random number, v_k^t is the location of WS, v_{gbk} is the best parameter, α is the WS term, u_{k+1}^t and u_k^t are the previous and present velocities; p_1 and p_2 are the parameters; c_1 and c_2 are the random parameters. The parameters p_1 and p_2 are given as:

$$p_1 = p_{\max} + (p_{\max} - p_{\min}) \times \exp\left(-\frac{4k}{K}\right) \tag{14}$$

$$p_1 = p_{\min} + (p_{\max} - p_{\min}) \times \exp\left(-\frac{4k}{K}\right) \tag{15}$$

The progression toward locating optimal prey involves WS detecting the scent of the target, observing the motion of the prey, or potentially identifying the waves generated by the prey's actions. Continuously advancing towards the prey, the WS persistently tracks its motions. Even if the prey relocates or escapes its initial position, the lingering scent remains in that area. As a result, the WS updates its position accordingly.

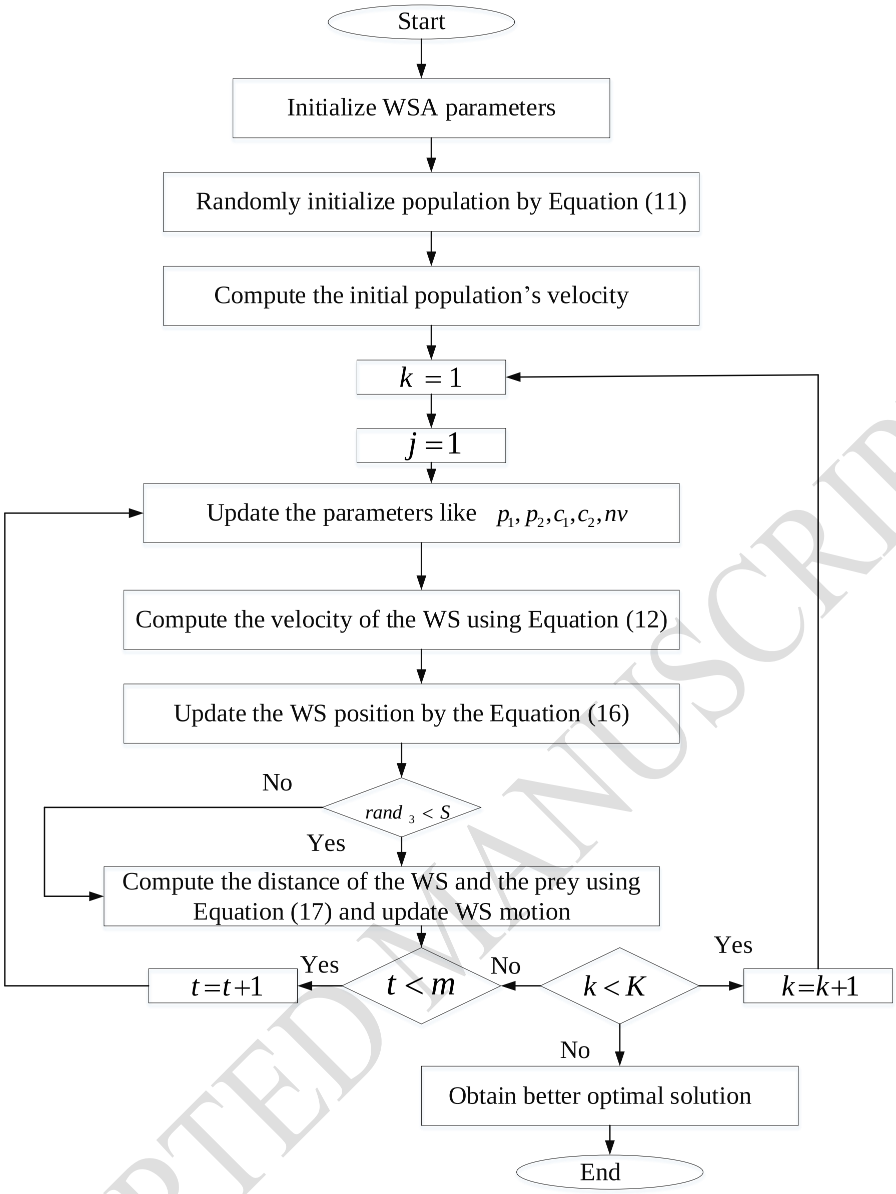


Figure 4. Flowchart of WSA

$$v_{k+1}^{t} = \begin{cases} v_k^t \neg \oplus v_0 + ul \times a + ll \times b & when\ rand < nv \\ v_k^t + \dfrac{u_k^t}{f} & when\ rand \geq nv \end{cases} \tag{16}$$

where nv is the motion force and the motion to the best WS is given as:

$$v_{k+1}^{t} = v_{gbk} + rand_1 \times \vec{D}_v \times \mathrm{sgn}\left(rand_2 - 0.5\right) \quad when\ rand_3 < S \tag{17}$$

where $rand_1$, $rand_2$ and $rand_3$ are the random numbers; S is the strength of the WS, $\vec{D}_v$ is the distance among the prey and the WS. Figure 4 shows the WSA flow chart and Algorithm 1 defines the pseudocode of the overall AQIP.

Algorithm 1: Pseudocode of the overall AQIP
Input: AQI dataset
Output: Air pollutants and AQI
Pre-processing phase
For
Missing values
Type conversion
End for
Feature extraction and AQIP
For (overall samples)
Split data based on the 10-fold cross validation
End for
For every training set
Evaluate the parameters of the AM-CNN-OptBiLSTM
End for
For every testing set
Test the AM-CNN-OptBiLSTM
Evaluate the performance measures
End for
End for

4. Analysis of results

The implementation of smart contract-based malicious detection and mitigation is employed using PYTHON programming language and is assessed based on various measures. Table 1 presents the parameters used for the experimental analysis.

Table 1. Parameters

Parameters	Values
Learning rate	0.0001
Size of batch	64
Epochs	100
BiLSTM nodes	16
Dropout	0.3
Optimizer	Adam
Loss function	Cross entropy

4.1. Performance measures

The evaluation measure for regression methods is employed to measure the efficiency of the model in forecasting output values according to the inputs. Measures like MSE (mean square error), RMSE (root MSE), MAE (mean absolute error), MAPE (mean absolute percentage error) and R-squared (R2).

MSE: It is the variation of predicted y_k and actual values $\hat{y}_k$ and RMSE is the square value of MSE. These two expressions are given as

$$MSE = \frac{1}{m}\sum_{k}^{m}\left(y_k - \hat{y}_k \right)^2 \tag{18}$$

$$RMSE = \sqrt{\frac{1}{m}\sum_{k}^{m}\left(y_k - \hat{y}_k \right)^2} \tag{19}$$

MAE: This measure serves as an alternative metric for quantifying the disparity between y_k and $\hat{y}_k$. Its computation involves determining the absolute value of the variation among y_k and $\hat{y}_k$, followed by averaging these absolute variations.

$$MAE = \frac{1}{m}\sum_{k}^{m}\left| y_k - \hat{y}_k \right| \tag{20}$$

MAPE: It is a metric that defines the performance of the model, represented as a percentage. Its computation involves taking the absolute value of the variation among y_k and $\hat{y}_k$, dividing it by the y_k, and then averaging these resulting percentages.

$$MAPE = \frac{100}{m}\sum_{k}^{m}\frac{y_k - \hat{y}_k}{y_k} \tag{21}$$

R2: It is an indicator of how effectively a model aligns with the data set. Its calculation involves summing the squares of the differences between y_k and $\hat{y}_k$. This sum is then divided by the sum of the squares of the variation among the average of the y_k and $\vec{y}_k$.

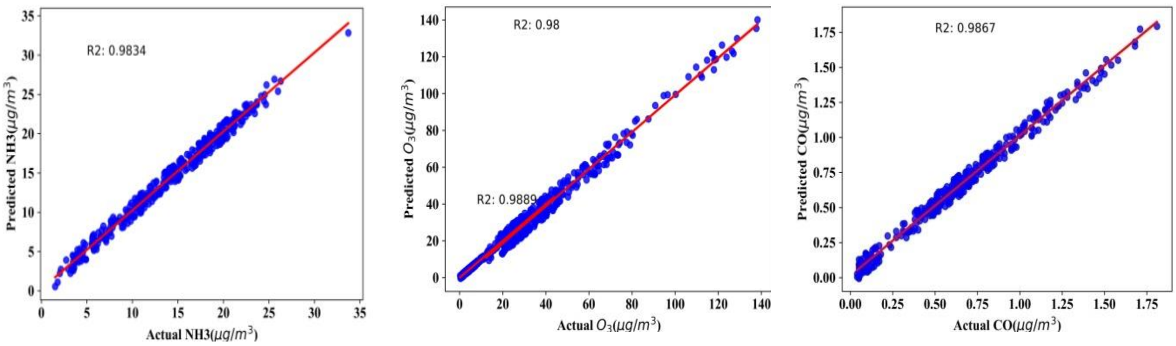
$$R^2 = 1 - \frac{\sum \left(y_k - \hat{y}_k \right)^2}{\sum \left(y_k - \vec{y}_k \right)^2} \tag{22}$$

4.2. Dataset Description

The data set considered for this work is collected between 2015 and 2023. <u>from</u> various metrological cites. This dataset includes the pollutants like $PM_{2.5}, PM_{10}, CO, NO_2, NO, NH_3, O_3, NO_x, SO_2$.

4.3. Performance Analysis

This section examines and compares the evaluation metrics of the proposed AM-CNN-OptBiLSTM model with some DL approaches. Evaluation is a critical phase of any model evaluation, providing insights to determine the optimal process with respect to the performance outcomes. The study conducts a comprehensive evaluation analysis that compares the performance of various approaches using five metrics.



(a) (b) (c)

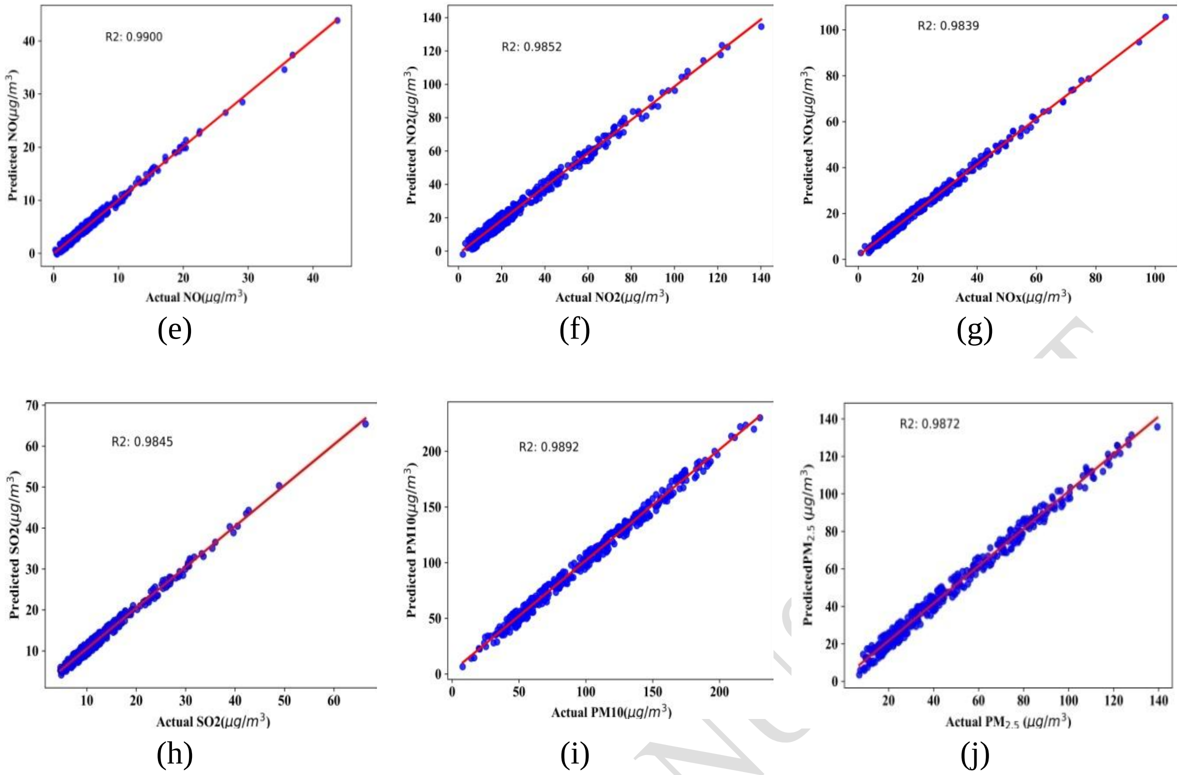


Figure 5. Correlation of various pollutants

Figure 5 presents the correlation among the actual and predicted values. Pollutants like $PM_{2.5}, PM_{10}, CO, NO_2, NO, NH_3, O_3, NO_x, SO_2$ are compared with respect to R2 values. It is observed from the Figures that the pollutant $PM_{2.5}$ achieved better R2 value of 0.9872.

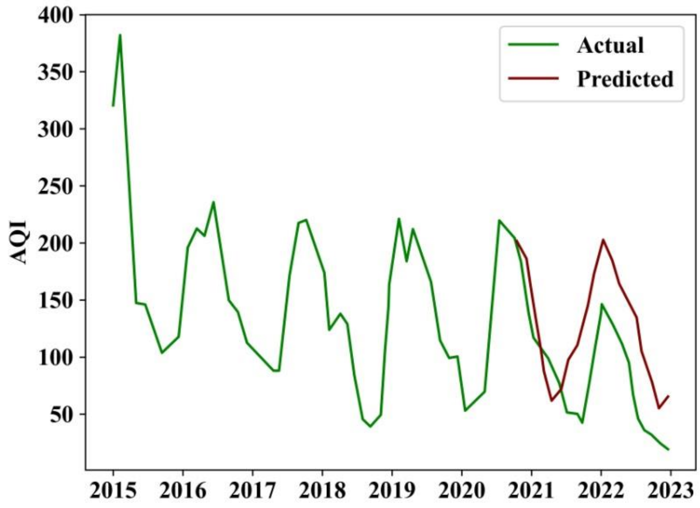


Figure 6. AQI trends between 2015 and 2023

Figure 6 presents the AQI trends between 2015 and 2023 in India. It is observed that the actual and the predicted values are the same.

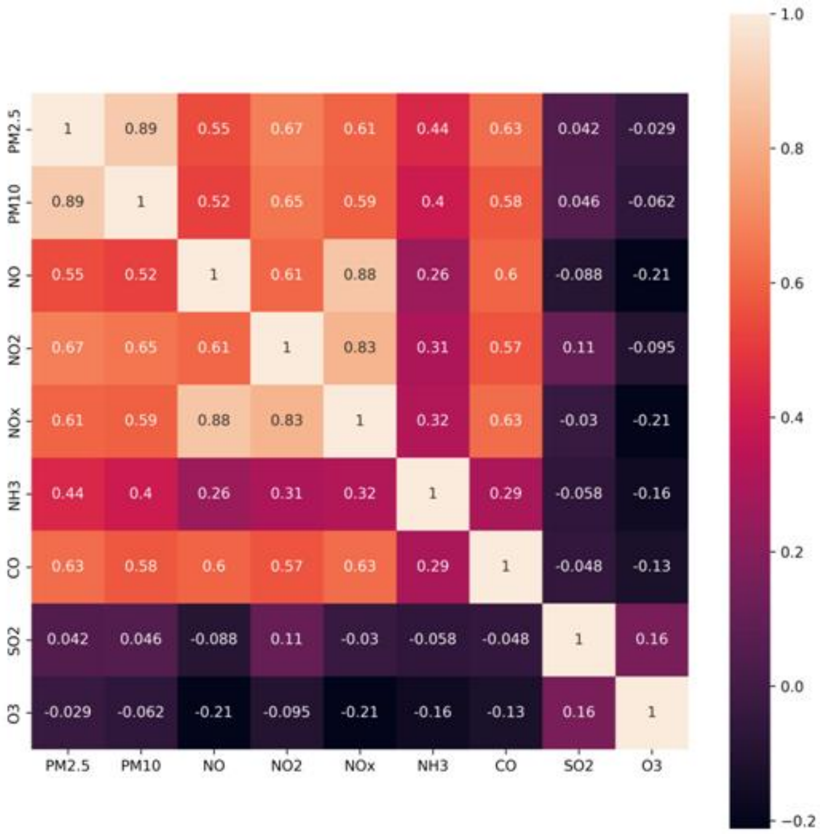
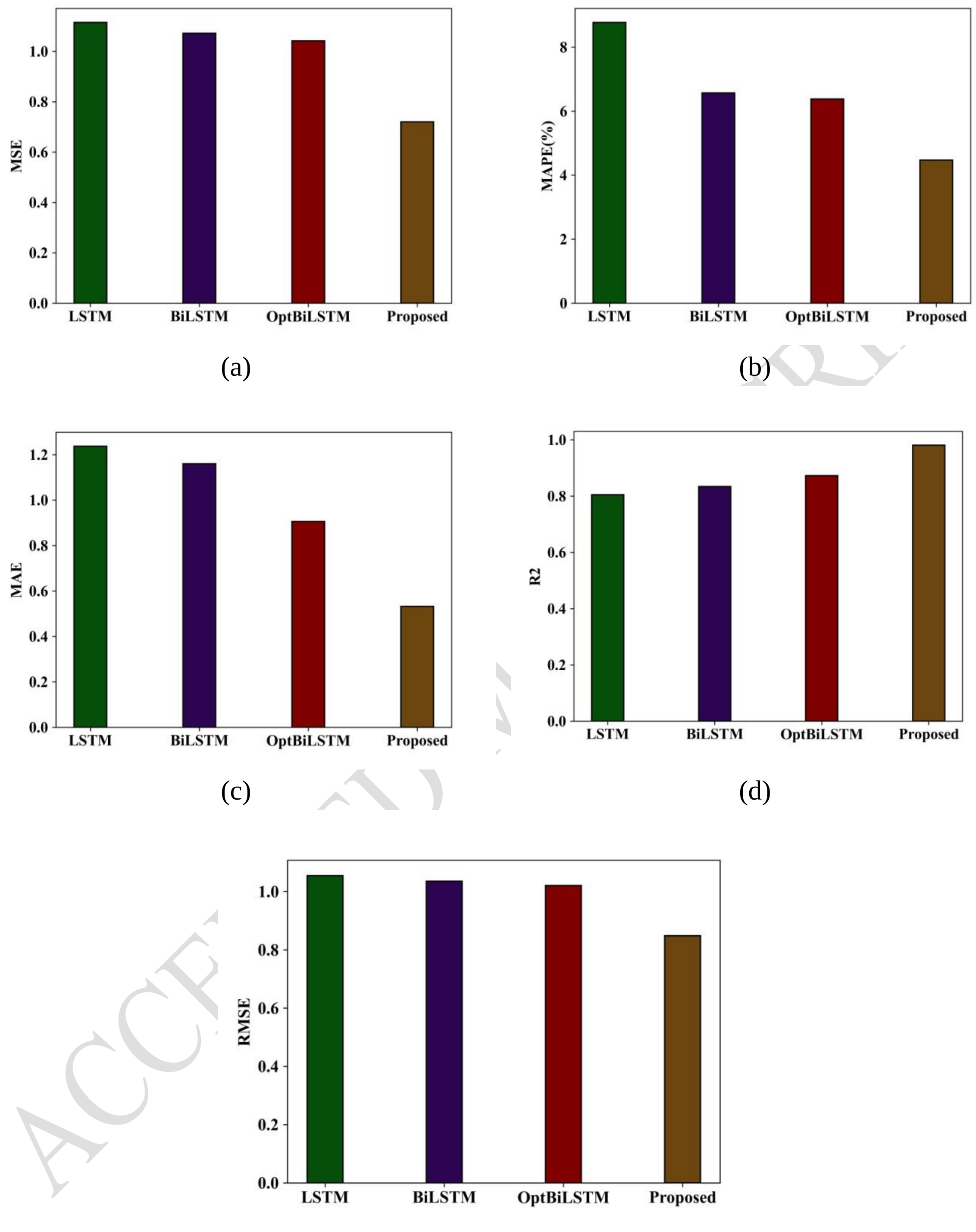


Figure 7. Confusion matrix of the proposed AM-CNN-OptBiLSTM model

Figure 7 presents the confusion matrix of the proposed AM-CNN-OptBiLSTM model that predicts the pollutants like $PM_{2.5}, PM_{10}, CO, NO_2, NO, NH_3, O_3, NO_x, SO_2$.



(a) (b) (c) (d) (e)

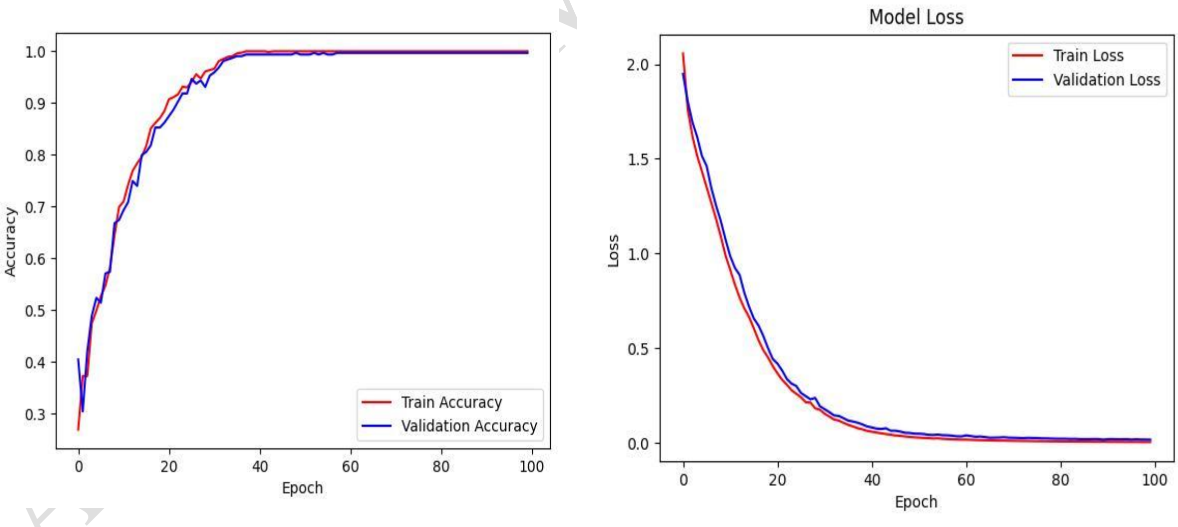
Figure 8. Comparison of (a) MAE, (b) MAPE, (c) MSE, (d) R^2 and (e) RMSE

Figure 8 and Table 2 presents Figure 7 illustrates the performance comparison among various techniques, including LSTM, BiLSTM, OptBiLSTM and the proposed AM-CNN-OptBiLSTM. Evaluation metrics such as MSE, RMSE, MAE, MAPE and R2 are computed.

Table 2. Comparative analysis

Methods	MSE	RMSE	MAE	MAPE	R^2
LSTM	1.11	1.05	1.23	8.77	0.80
BiLSTM	1.07	1.03	1.16	6.57	0.83
OptBiLSTM	1.04	1.02	0.90	6.38	0.87
Proposed	0.72	0.84	0.53	4.47	0.98

In Figure 8 (a), the MSE performance of different DL approaches is presented. It is evident from the graph that the MSE value (0.72) of the proposed AM-CNN-OptBiLSTM is significantly lower than other approaches. Similarly, for an effective weather prediction model, a lower MAPE value is preferable, and in Figure 8 (b), the proposed AM-CNN-OptBiLSTM exhibits a lower MAPE value of 4.47. Then, in Figure 8 (c), the MAE values achieved by the LSTM, BiLSTM, OptBiLSTM and the proposed AM-CNN-OptBiLSTM are 1.23, 1.16, 0.9 and 0.53 respectively. Additionally, for an improved AQIP model, a higher R2 value is desired, and in Figure 8 (d), the proposed AM-CNN-OptBiLSTM achieves a high R2 value (0.98). Finally, in Figure 8 (e) also, the proposed AM-CNN-OptBiLSTM achieved a better RMSE value of 0.84. Across all comparisons, the proposed model outperforms others, attributed to better weight selection by CNN with BiLSTM and WSA.



(a) (b)

Figure 9. Accuracy-loss curves of the proposed AM-CNN-OptBiLSTM

Figure 9 illustrates the performance of the proposed AM-CNN-OptBiLSTM instrument with respect to accuracy and loss curves. The evaluation covers variations in values over 100 epochs, and the graphs depict the relationship between the training and validation samples. In particular, the model does not exhibit under- or over-fitting, indicating its superior generalization capability. This substantiates the proposition that the proposed AM-CNN-OptBiLSTM can be effectively utilized in the AQIP process. Table 3 analyzes the comparative analysis and all comparative measures, the proposed AQIP model outperformed recent research works with respect to measures like MSE, RMSE, MAE and R2.

Table 3. Comparative analysis

References	MSE	RMSE	MAE	R^2
Maltare et al. [8]	-	4.9		0.82
Wu et al. [9]	-	2.3	2.1	0.94
Zhang et al. [14]	11.3	11.9	6.9	0.93
Zhang et al. [15]	-	0.93	-	-
Ravindiran et al. [16]	0.5	0.76	0.6	-
Janarthanan et al. [17]	-	7.8	-	0.63
Drewil et al. [20]	-	9.5	19.1	-
Mao et al. [22]	-	17.9	12.3	0.87
Proposed	**0.72**	**0.84**	**0.53**	**0.98**

5. Conclusions

The focus on monitoring air pollution is on the rise, with an increasing emphasis on understanding its impacts on human health. Presently, most air quality investigations transition from quantitative approaches to DL models. The AQIP experiences significant fluctuations influenced by pollutant concentrations. This study concentrates on establishing the AM-CNN-OptBiLSTM approach, reevaluating the AQI through the application of time series and DL methodology. The proposed AM-CNN-OptBiLSTM successfully extracted the spatio-temporal features, predicted all pollutants, and attained better performance. In the future, addressing sudden fluctuations in time series data related to air pollution poses an intriguing and challenging task for AQIP. Successful prediction of sudden changes in air pollution in advance holds significant advantages for the protection of the environment, government decisions, and the daily health of individuals.

References

[1] Terroso-Saenz, Fernando, Juan Morales-García, and Andres Muñoz. "Nationwide Air Pollution Forecasting with Heterogeneous Graph Neural Networks." ACM Transactions on Intelligent Systems and Technology (2023).

[2] Surendran R., Tamilvizhi T., and Lakshmi, S. (2021), Integrating the Meteorological Data into a Smart City Service Using Cloud of Things (CoT). In Emerging Technologies in Computing: 4th EAI/IAER International Conference, iCETiC 2021, Virtual Event, August 18–19, 2021, *Springer International Publishing,* **4,** 94-111..

[3] Heydari, Azim, Meysam Majidi Nezhad, Davide Astiaso Garcia, Farshid Keynia, and Livio De Santoli. "Air pollution forecasting application based on deep learning model and optimization algorithm." Clean Technologies and Environmental Policy (2022): 1-15.

[4] Gugnani, Veena, and Rajeev Kumar Singh. "Analysis of deep learning approaches for air pollution prediction." Multimedia Tools and Applications 81, no. 4 (2022): 6031-6049.

[5] Sachdeva, Shelly, Hitendra Singh, Shailee Bhatia, and Puneet Goswami. "An integrated framework for predicting air quality index using pollutant concentration and meteorological data." Multimedia Tools and Applications (2023): 1-30.

[6] Iskandaryan, Ditsuhi, Francisco Ramos, and Sergio Trilles. "Application of deep learning and machine learning in air quality modeling." In Current Trends and Advances in Computer-Aided Intelligent Environmental Data Engineering, pp. 11-23. Academic Press, 2022.

[7] Santhanaraj R. K., Rajendran S., Romero C. A. T., and Murugaraj, S. S. (2023), Internet of Things Enabled Energy Aware Metaheuristic Clustering for Real Time Disaster Management. *Comput. Syst. Sci. Eng.,* **45**, 1561-1576.

[8] Maltare, Nilesh N., and Safvan Vahora. "Air Quality Index prediction using machine learning for Ahmedabad city." Digital Chemical Engineering 7 (2023): 100093.

[9] Wu, Huiyong, Tongtong Yang, Hongkun Li, and Ziwei Zhou. "Air quality prediction model based on mRMR–RF feature selection and ISSA–LSTM." Scientific Reports 13, no. 1 (2023): 12825.

[10] Surendran R., Alotaibi Y. and Subahi A.F. (2023), Wind Speed Prediction Using Chicken Swarm Optimization with Deep Learning Model. *Computer Systems Science & Engineering*, **46,** 3.

[11] Zhang, Zhendong, Yongkang Zeng, and Ke Yan. "A hybrid deep learning technology for PM 2.5 air quality forecasting." Environmental Science and Pollution Research 28 (2021): 39409-39422.

[12] Surendran R., Alotaibi Y. and Subahi, A.F. (2023), Lens-Oppositional Wild Geese Optimization Based Clustering Scheme for Wireless Sensor Networks Assists Real Time Disaster Management. *Comput. Syst. Sci. Eng.*, **46,** 835-851.

[13] Ma, Jun, Jack CP Cheng, Changqing Lin, Yi Tan, and Jingcheng Zhang. "Improving air quality prediction accuracy at larger temporal resolutions using deep learning and transfer learning techniques." Atmospheric Environment 214 (2019): 116885.

[14] Zhang, Kefei, Xiaolin Yang, Hua Cao, Jesse Thé, Zhongchao Tan, and Hesheng Yu. "Multi-step forecast of PM2. 5 and PM10 concentrations using convolutional neural network integrated with spatial–temporal attention and residual learning." Environment International 171 (2023): 107691.

[15] Zhang, Z., and S. Zhang. "Modeling air quality PM2. 5 forecasting using deep sparse attention-based transformer networks." International Journal of Environmental Science and Technology (2023): 1-16.

[16] Ravindiran, Gokulan, Gasim Hayder, Karthick Kanagarathinam, Avinash Alagumalai, and Christian Sonne. "Air quality prediction by machine learning models:

A predictive study on the Indian coastal city of Visakhapatnam." Chemosphere 338 (2023): 139518.

[17] Janarthanan, R., P. Partheeban, K. Somasundaram, and P. Navin Elamparithi. "A deep learning approach for prediction of air quality index in a metropolitan city." Sustainable Cities and Society 67 (2021): 102720.

[18] Gilik, Aysenur, Arif Selcuk Ogrenci, and Atilla Ozmen. "Air quality prediction using CNN+ LSTM-based hybrid deep learning architecture." Environmental science and pollution research (2022): 1-19.

[19] Lakshmipathy, Munirathnam, Shanthi Prasad Mysore Jeevandharakumar, and Goddamachinnehalli Narayanappa Kodandaramaiah. "An adaptive serial cascaded autoencoder and LSTM with multivariate regression for ambient air quality prediction with improved flow direction algorithm." Energy Sources, Part A: Recovery, Utilization, and Environmental Effects 45, no. 4 (2023): 10304-10329.

[20] Drewil, Ghufran Isam, and Riyadh Jabbar Al-Bahadili. "Air pollution prediction using LSTM deep learning and metaheuristics algorithms." Measurement: Sensors 24 (2022): 100546.

[21] Zhang, Jiaxuan, and Shunyong Li. "Air quality index forecast in Beijing based on CNN-LSTM multi-model." Chemosphere 308 (2022): 136180.

[22] Tamilvizhi T., Surendran R., Romero C.A.T. and Sadish M. (2022), Privacy preserving reliable data transmission in cluster based vehicular adhoc networks, *Intelligent Automation & Soft Computing*, **34**, 1265–1279.

[23] Liao, Haibin, Li Yuan, Mou Wu, and Hongsheng Chen. "Air quality prediction by integrating mechanism model and machine learning model." Science of the Total Environment 899 (2023): 165646.

[24] https://www.kaggle.com/datasets/fedesoriano/air-quality-data-in-india