Published by
NIGERIAN SOCIETY OF PHYSICAL SCIENCES
Available online @ https://journal.nsps.org.ng/index.php/jnsps

J. Nig. Soc. Phys. Sci. 7 (2025) 2473

Journal of the Nigerian Society of Physical Sciences

Air quality prediction enhanced by a CNN-LSTM-Attention model optimized with an advanced dung beetle algorithm

Xiaojie Zhou[a,b], Majid Khan Bin Majahar Ali[a,*], Farah Aini Abdullah[a], Lili Wu[a], Ying Tian[a], Tao Li[a], Kaihui Li[a]

[a]*School of Mathematical Sciences, Universiti Sains Malaysia, Pulau Pinang 11800, Malaysia*
[b]*Weifang University of Science and Technology, Shouguang 262713, China*

Abstract

Air pollution significantly impacts human health and socioeconomic development, making accurate air quality prediction crucial. This study proposes a hybrid CNN-LSTM-Attention model optimized with an improved Dung Beetle Optimization (IDBO) algorithm to enhance predictive performance. IDBO integrates multiple strategies to improve global search capabilities and overcome the limitations of conventional DBO. Experiments using PM2.5 data from Penang, Malaysia, demonstrate that the proposed model outperforms other models across multiple evaluation metrics R^2 = 0.904, RMSE = 2.677, MSE = 7.168, MAE = 1.982, MAPE = 44.1%). The findings validate the effectiveness of the proposed approach in improving air quality prediction, offering valuable insights for environmental monitoring and pollution control.

DOI:10.46481/jnsps.2025.2473

Keywords: Air quality prediction, DBO, CNN, LSTM, Attention

Article History :
Received: 29 October 2024
Received in revised form: 06 November 2024
Accepted for publication: 26 February 2025
Available online: 21 April 2025

Communicated by: B. J. Falaye

1. Introduction

With rapid industrialization, air pollution has become a critical global issue, posing severe threats to public health and socioeconomic development. PM2.5, a key pollutant, has been linked to increased respiratory and cardiovascular diseases [1, 2]. Given its complex and nonlinear nature, accurate air quality prediction is crucial for effective environmental monitoring and policymaking. However, air pollution data exhibits high variability and strong dependencies on meteorological conditions, making it challenging to develop precise predictive models [3].

Air quality prediction [4] has been approached through three primary modeling techniques: statistical models, machine learning models, and deep learning models. Statistical models, such as ARIMA [5] and MLR [6], rely on predefined assumptions about data distribution but struggle with nonlinear patterns. Machine learning methods, such as RF [7], SVM [8], and XGBoost [9], can improve predictive accuracy by leveraging large datasets, but they often require extensive feature engineering. Deep learning models, including MLP, CNNs [10], and LSTM [11], have demonstrated superior performance in handling spatiotemporal dependencies, yet suffer from high computational costs and overfitting risks.

*Corresponding author Tel. No.: +60-4534-782.
Email address: majidkhanmajaharali@usm.my (Majid Khan Bin Majahar Ali)

To overcome the limitations of individual models, hybrid models have been proposed, integrating different techniques to enhance predictive accuracy. Several recent studies have explored various hybrid architectures, such as ARIMA-CNN-LSTM and ACNN-ARIMA-QPSO-LSTM-XGBoost [12]. While these approaches improve prediction performance, they still face challenges such as suboptimal hyperparameter tuning, slow convergence, and sensitivity to input variations [13–15].

Table 1 summarizes key studies on air quality prediction, highlighting their methodologies, evaluation metrics, contributions, and limitations.

Despite significant advancements in air quality prediction, existing studies face several challenges:

1. Data Complexity: Air quality data exhibits high variability and strong nonlinear dependencies, making it difficult for conventional models like ARIMA to capture long-term patterns.
2. Hyperparameter Optimization: Many studies rely on standard optimization methods (e.g., PSO, ISSA), which are prone to local optima and fail to efficiently explore the search space.
3. Computational Cost: Complex hybrid models such as ACNN-ARIMA-QPSO-LSTM-XGBoost require extensive computational resources, limiting their practical applications.
4. Scalability and Generalization: Most hybrid models are tested on limited datasets, often failing to generalize effectively to diverse environmental conditions.

To address these gaps, this study proposes a CNN-LSTM-Attention hybrid model optimized using an enhanced Dung Beetle Optimization (IDBO) algorithm. Compared to previous works:

1. IDBO improves optimization efficiency, leveraging golden sine strategy, self-spiral strategy, Levy flight, and adaptive t-distribution to enhance global search and avoid local optima.
2. The CNN-LSTM-Attention architecture effectively captures spatial-temporal features, surpassing conventional hybrid models.
3. The proposed model achieves higher accuracy on real-world datasets, demonstrating superior predictive performance compared to existing approaches.

The key contributions of this study include:

1. Proposing a novel hybrid model: Integrating CNN, LSTM, and Attention mechanisms to enhance feature extraction and sequence learning for PM2.5 prediction.
2. Developing an improved Dung Beetle Optimization (IDBO) algorithm: Enhancing hyperparameter tuning through advanced optimization strategies, improving convergence speed and accuracy.
3. Comprehensive performance evaluation: Comparing the proposed model with multiple hybrid and individual models using real-world air pollution data from Penang, Malaysia, demonstrating superior predictive accuracy (R^2=0.904), RMSE = 2.677, MSE = 7.168, MAE = 1.982, MAPE = 44.1%).

2. Materials and methods

This paper proposes the IDBO-CNN-LSTM-Attention model. To tackle the issues of low accuracy and slow convergence speed, the DBO method is initially upgraded.

1. The golden sine strategy is incorporated into the rolling phase to increase the likelihood of discovering high-quality solutions.
2. The self-spiral strategy is used in the foraging stage to push the population to the optimal position.
3. During the stealing stage, the Levy flight technique is used to introduce random disturbances to the most promising position at that moment.
4. An adaptive t-distribution is applied to refine the global optimal solution, enhancing the algorithm's capability to avoid local optima. Subsequently, the IDBO is utilized to optimize the hyperparameters of the CNN-LSTM-ATT model. Figure 1 presents the flowchart of the composite model.

2.1. CNN

One of the most potent networks in deep learning is the CNN. The main advantage of CNN is to extract data features. It extracts local features of data through a series of convolution operations in the convolution layer, and then reduces the feature dimension through the pooling layer, reduces the computational complexity, and extracts the main features. In air quality prediction, CNN processes historical PM2.5 concentration values through convolution layers. At the same time, the filter detects the key features of the data, such as the concentration values of related pollutants, which are crucial for PM2.5 prediction. CNN performs feature extraction and dimensionality reduction through convolution layers and pooling layers, which speeds up the efficiency of extracting key features from air quality prediction data. Therefore, features are extracted from the normalized data using a CNN, followed by LSTM for analysis and prediction.

2.2. LSTM

LSTM are improved by recurrent neural networks (RNNs) and are mainly used to solve the gradient explosion and vanishing problems caused by the repeated use of weights. Unlike the original RNN's hidden layer, the LSTM introduces a cell state, and at the same time, the number of LSTM "gates" is larger and the structure is more complex. It comprises three gates: the output gate, the forget gate, and the input gate. as well as memory cells that are identical in shape to the hidden state, as shown in Figure 2. At the heart of LSTM is the state of the cell, as well as the various phylum structures within it. The following are each gate's structures: (a) The forgotten entrance mainly uses the

Table 1. Summary of previous studies on air quality prediction.

Study	Method	Evaluation Metrics	Key Contribution	Limitations
Shao et al. (2023) [16]	VMD + Informer + XGBoost	RMSE = 3.42, MAE = 2.19	Integrated time-series decomposition and XGBoost for PM2.5 prediction	Limited generalization ability on complex datasets
Duan et al. (2023) [12]	ARIMA-CNN-LSTM	R² = 0.86, RMSE = 4.12	Combined statistical models with deep learning to improve accuracy	Linear assumptions of ARIMA reduce flexibility in capturing nonlinear dependencies
Nguyen et al. (2024) [14]	ACNN-ARIMA-QPSO-LSTM-XGBoost	R² = 0.88, MSE = 9.35	Hybridized multiple methods for PM2.5 forecasting	High computational cost, long training time
Liu & Guo (2022) [15]	ISSA-LSTM	R² = 0.89, RMSE = 3.87	Used Sparrow Search Algorithm (ISSA) to optimize LSTM	Prone to overfitting, sensitive to parameter initialization
Wu et al. (2022) [17]	PSO-XGBoost	RMSE = 3.21, MAE = 2.14	Applied Particle Swarm Optimization (PSO) for hyperparameter tuning	Vulnerable to local optima, slow convergence
Zhu et al. (2023) [18]	CEEMDAN-VMD-ISCSO-LSTM	R² = 0.91, RMSE = 3.05	Combined decomposition techniques with Improved Sand Cat Swarm Optimization (ISCSO)	High computational cost, suboptimal performance on long time-series data

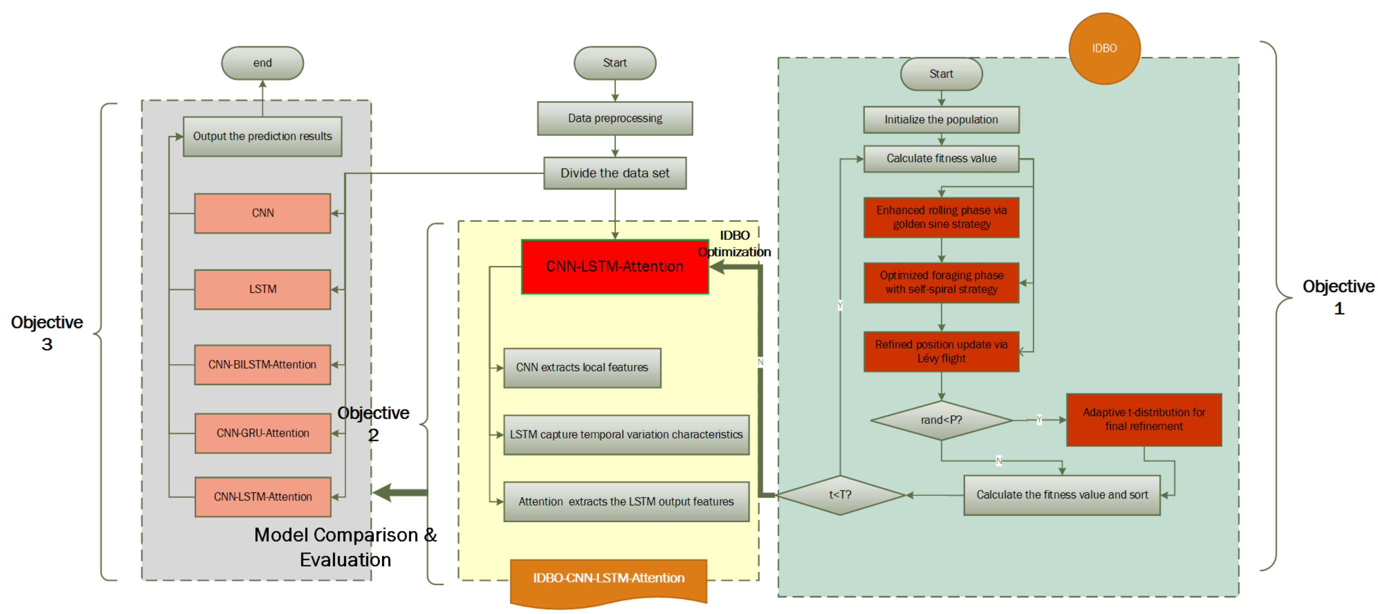

Figure 1. The workflow of the IDBO-CNN-LSTM-Attention hybrid model.

sigmoid structure to perform linear transformation to generate an output value between 0 and 1, so that it can selectively forget the input. This mechanism can retain or discard previous information as needed, thus more effectively capturing the long-term dependencies within the data. (b) The input gate consists of two components: the first uses a sigmoid function to select the value to be updated, while the second employs a tanh function to generate new candidate information. This update relationship can effectively process time series and Address the issues of gradient explosion and vanishing in RNNs. (c) The output gate uses the current input and state to generate new output. The math formulas for each door are as follows:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right). \tag{1}$$

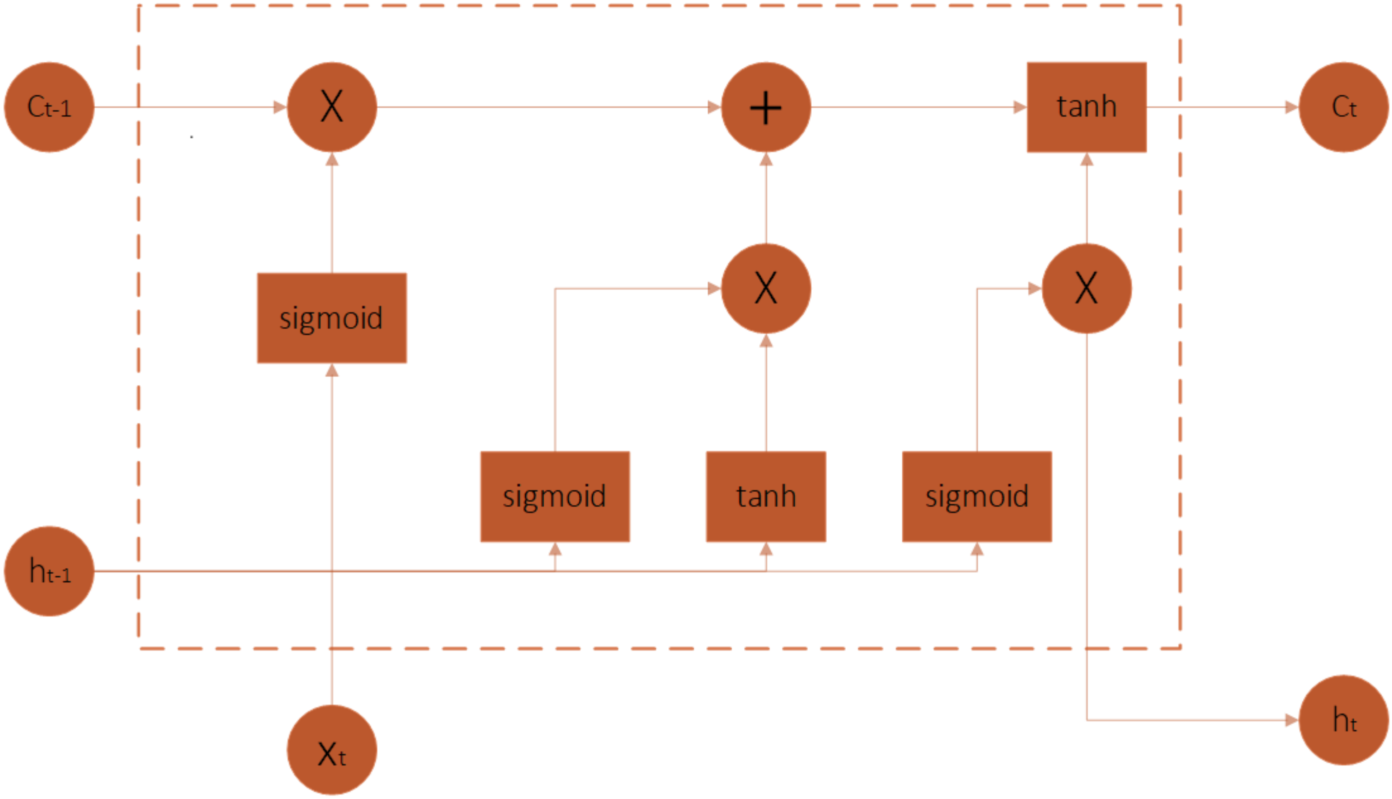

Figure 2. LSTM.

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right). \tag{2}$$

$$\tilde{C}_t = \tanh\left(W_c \cdot [h_{t-1}, x_t] + b_c\right). \tag{3}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t. \tag{4}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o). \tag{5}$$

$$h_t = o_t \cdot \tanh(C_c). \tag{6}$$

2.3. Attention

The primary functions of attention are to improve the model's predictive capability, focus on key segments of the sequence, and effectively handle long sequences. Following the LSTM, ATT is employed to weight the output data from the LSTM layer, highlight important features, and enhance the accuracy of PM2.5 predictions. ATT takes the output from the LSTM hidden layer as its input. This output is processed through a fully connected layer and normalized using the softmax function, producing a set of weights. These weights allow the model to assign greater importance to key elements during predictions, enabling it to identify and focus on significant portions of complex data. Attention performs calculations as follows:

$$S_i = \tanh(WH_i + b_i). \tag{7}$$

$$\alpha_i = soft\max(S_i). \tag{8}$$

$$C_i = \sum_{i=1}^{k} \alpha_i H_i. \tag{9}$$

The result of the LSTM hidden layer is H_i; S_i is the score output by each hidden layer; C_i is the weighted average summation value; b_i is the bias term.

2.4. DBO

DBO [19] primarily adjusts its position by mimicking four natural behaviors of dung beetles: rolling, egg-laying, foraging, and stealing.

2.4.1. Rollerball dung beetle

Rolling dung beetles are updated using equation (1).

$$\begin{cases} x_i(t+1) = x_i(t) + \alpha \times k \times x_i(t-1) + b \times \Delta x \\ \Delta x = |x_i(t) - X^{\omega}|. \end{cases} \tag{10}$$

When it comes across an obstacle,, it updates its position by dancing with the following formula.

$$x_i(t+1) = x_i(t) + \tan\theta |x_i(t) - x_i(t-1)|. \tag{11}$$

2.4.2. Spawning dung beetles

The positions of the hatching dung beetles are updated as follows, reflecting the dynamic characteristics of the spawning area:

$$B_i(t+1) = X^* + b_1 \times (B_i(t) - Lb^*) + b_2 \times (B_i(t) - Ub^*). \tag{12}$$

$B_i(t)$ denotes the positional information of the i th hatching ball dung beetle ; b_1 and b_2 represent two independent random vectors, each of size $1 \times D$.

2.4.3. Foraging dung beetles

The following shows the updated position of the little baby.

$$x_i(t+1) = x_i(t) + C_1 \times (x_i(t) - Lb^b) + C_2 \times (x_i(t) - Ub^b), \tag{13}$$

where C_1 and C_2 indicate a random scalar and a random vector, respectively.

2.4.4. Stealing dung beetles

The stealing behavior involves taking dung balls from others. Its position is adjusted in the following ways during the iterative process.

$$x_i(t+1) = X^b + S \times g \times (|x_i(t) - X^*| + |x_i(t) - X^b|), \tag{14}$$

where g represents a randomly generated vector; S denotes a constant.

2.5. IDBO algorithm

2.5.1. Golden sine optimization algorithm

The golden sine optimization algorithm [20] adopts the use of sinusoidal function for iterative optimization search, at the same time, The incorporation of the golden section coefficient in the updating process provides the method with robust global search capabilities during the pre-iteration phase and sufficient local search abilities in the final phases of the iteration. In the rolling phase, The following is the position update formula that utilizes the golden sine strategy.

$$x_i(t+1) = x_i(t) \cdot |\sin r_1| - r_2 \cdot \sin r_1 \cdot |c_1 x^b - c_2 x_i(t)|, \tag{15}$$

where r_1 is a random number for the distance travelled; r_2 is a random number for the update direction; g_1 is the golden section number, c_1 and $c_2$are a constant.

2.5.2. Self-spiral strategy

The delayed pursuit of dominance by dung beetles during foraging negatively impacts the algorithm's convergence in its later stages. Inspired by the spiral search strategy [21], it is improved and the following presents the position update formula:

$$x_i(t+1) = e^{zl} \cdot \cos(2\pi l) \cdot x_i(t) + c_1(x_i(t) - Lb^b) + c_2(x_i(t) - Ub^b), \tag{16}$$

where z is a constant for the spiral equation; the random number l falls within the interval [-1, 1].

2.5.3. Levy flying

Due to the lack of interactive behaviour between peers in the stealing behaviour, The algorithm can easily get trapped in a local optimum. To address this issue, we implement the Lévy flight strategy from the cuckoo algorithm [22] to improve the algorithm's capability for finding the global optimum and to enhance search space exploration. The Lévy flight strategy is defined as follows.

$$\text{Levy}(x) = 0.01 \times \frac{r^3 \times \sigma}{|r^4|^{(1/\varepsilon)}}. \tag{17}$$

$$\sigma = \left(\frac{\Gamma(1+\xi) \times \sin(\pi\xi/2)}{\Gamma((1+\varsigma)/2) \times \xi \times 2^{((\xi-1)/2)}} \right)^{(1/\xi)}, \tag{18}$$

where $\Gamma(x) = (x-1)!$.

2.5.4. Adaptive t-distribution

For high dimensionality and high complexity objective function, in the late iteration, it is very easy to ignore the global optimal position, adaptive t-distribution can be perturbed to the current position, to Strengthen the algorithm's resistance to getting stuck in local optima, and to strengthen the algorithm's convergence speed and solving efficiency, and the individual updates are as follows:

$$x_i^{t+1} = x_i^t + x_i^t \cdot t(iter). \tag{19}$$

Meanwhile, The dynamic selection probability p is employed to automatically adjust the equation as follows, aiming to reduce computation time:

$$p = w_1 - w_2 \cdot \frac{T-t}{T}, w_1 = 0.5, w_2 = 0.1. \tag{20}$$

To address the limitations of the DBO algorithm, this study introduces four key improvements: the Golden Sine Strategy, the Self-Spiral Strategy, Levy Flight, and the Adaptive tDistribution. These enhancements aim to improve global search capability, convergence speed, and robustness against local optima.

Table 2 summarizes the major challenges in the original DBO algorithm and the corresponding improvements introduced in IDBO. Each enhancement is designed to overcome a specific limitation and collectively contributes to a more efficient and effective optimization process.

From Table 2, we can see that the Golden Sine Strategy enhances early-stage exploration, preventing premature convergence. The Self-Spiral Strategy focuses on improving local exploitation, ensuring faster and more efficient fine-tuning in later iterations. The Levy Flight mechanism introduces stochastic jumps, which significantly reduce the chances of getting trapped in local optima. Finally, the Adaptive t-Distribution provides dynamic step-size adjustments, making the algorithm more robust when handling high-dimensional and complex optimization problems.

By integrating these enhancements, IDBO achieves a better balance between exploration and exploitation, leading to improved performance in hyperparameter tuning for the CNNLSTM-Attention model.

Table 2. Enhancements in IDBO algorithm and corresponding improvements.

DBO Limitation	IDBO Enhancement	Improvement Achieved
Slow convergence in early iterations	Golden Sine Strategy	Improves global search by leveraging sinusoidal perturbations
Poor local exploitation in later iterations	Self-Spiral Strategy	Enhances fine-tuning in local regions to improve solution quality
Easily trapped in local optima	Levy Flight	Introduces stochastic jumps to escape local optima
Insufficient precision in final search	Adaptive t-Distribution	Adjusts search radius dynamically to refine the optimal solution

From Table 2, we can see that the Golden Sine Strategy enhances early-stage exploration, preventing premature convergence. The Self-Spiral Strategy focuses on improving local exploitation, ensuring faster and more efficient fine-tuning in later iterations. The Levy Flight mechanism introduces stochastic jumps, which significantly reduce the chances of getting trapped in local optima. Finally, the Adaptive t-Distribution provides dynamic step-size adjustments, making the algorithm more robust when handling high-dimensional and complex optimization problems.

By integrating these enhancements, IDBO achieves a better balance between exploration and exploitation, leading to improved performance in hyperparameter tuning for the CNN-LSTM-Attention model.

3. Experimental results and analysis

3.1. Data sources and pre-processing

This study selected Penang from the Malaysian Ministry of Environment as the experimental object. The dataset under analysis consists of 8,760 hourly data points, covering the period from January 1, 2022, to December 31, 2022. The dataset includes air pollutants such as PM2.5, PM10, SO_2, NO_2, NO, O_3, CO, and NOx, In addition, the dataset includes meteorological factors such as temperature(ET), humidity(RH), wind direction(W_d), wind speed(W_s), and solar radiation(SA). The dataset is split into a test set (30%) and a training set (70%).

In this study, Pearson correlation(PC) was utilized to select the most significant features for predicting PM2.5 levels. Pearson's r quantifies the strength of a linear relationship between two variables, with a range from -1 to 1; values nearer to 1 or -1 signify a stronger correlation.

Based on the correlation matrix (Figure 3), the six features most strongly correlated with PM2.5 were selected. These include PM10, NO_2, CO, SO_2, NO and AT. These features were chosen because of their relatively higher correlation with PM2.5, suggesting that they are influential in predicting its concentration.

By selecting these top 6 variables, we aim to improve model performance by focusing on the most impactful features while reducing dimensionality and potential noise from less relevant variables.

$$PC = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}, \tag{21}$$

where $\bar{x}$ is the mean of the x data, $\bar{y}$ is the mean of the y data, and r ranges between -1 and 1.

Because the air quality data collection device will be affected by noise and human errors, the collected data will have abnormal values and missing values, so this paper first needs to preprocess the data.

- Missing values: Adoption of the mean value treatment
- outliers: The interquartile range method (IQR method) was used to identify outliers, calculate the mean, and then replace the outliers with the mean.

Interquartile Range (IQR) is an outlier detection method based on statistical distribution, mainly used to identify and process outliers in data. The IQR method determines which data points may be abnormal by analyzing the quartiles of the data. The following is a detailed introduction to the interquartile range method: Definition of quartiles

First quartile (Q1): When the data is arranged in ascending order, Q1 corresponds to the value at the 25th percentile, showing that 25% of the data points are at or below the first quartile (Q1).

Third quartile (Q3): When the data is sorted in ascending order, Q3 is the value at the 75th percentile, meaning that 75% of the data points are at or below this value.

Interquartile Range (IQR): The IQR represents the spread of the middle 50% of the data and is calculated by subtracting Q1 from Q3.

$$IQR = Q3 - Q1. \tag{22}$$

Outlier Detection:
Values that significantly differ from the majority of data points in the dataset are typically regarded as outliers.Through the interquartile range method, The upper and lower bounds for identifying outliers can be defined as follows:
Upper Bound:

$$UB = Q3 + 1.5 \times IQR. \tag{23}$$

Lower Bound:

$$LB = Q1 - 1.5 \times IQR. \tag{24}$$

Identify outliers: A data point is classified as an outlier if it exceeds the upper bound or falls below the lower bound. In Figure 4, we use a box plot to visually demonstrate the definition of IQR.

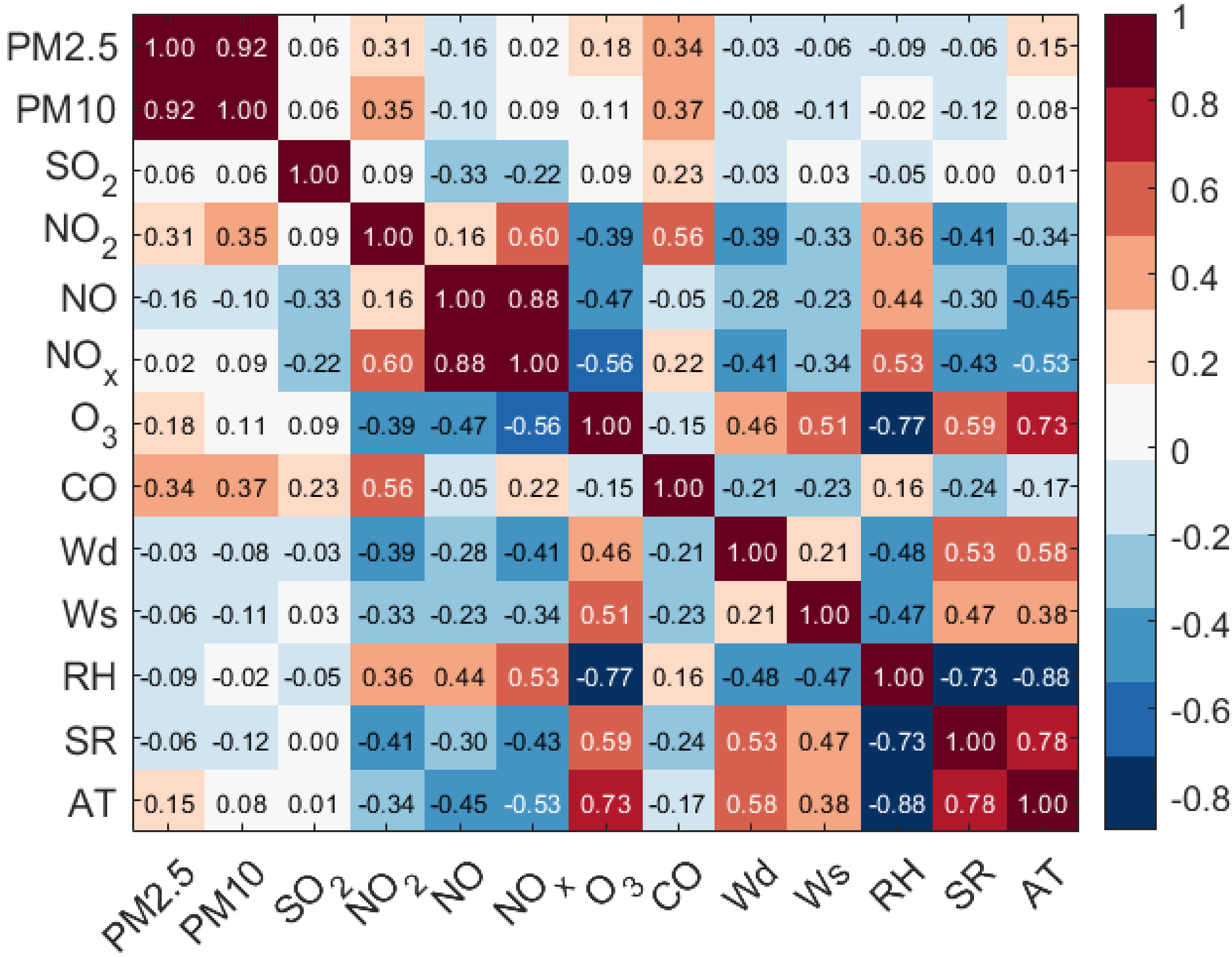

Figure 3. Pearson correlation coefficient plot.

- Standardization: Data normalization is applied for weighted processing. The following presents the normalization formula:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}. \tag{25}$$

3.2. Tuning the hyperparameters of the model.

To ensure the experiment's impartiality and fairness, the same dataset and test environment were used, and the relevant algorithms and models adopted the original default parameters.

3.3. Evaluation criteria for experimental results

The indicators for the air quality prediction model used in this work include R^2, RMSE, MSE, MAE and MAPE.The following are the calculation formulas:

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{N} (\bar{y}_i - y_i)^2}. \tag{26}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2}. \tag{27}$$

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2. \tag{28}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i||. \tag{29}$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%. \tag{30}$$

When the expected value is $\hat{y}_i$, the mean value is $\bar{y}_i$, and the sample value is y_i .

3.4. Analysis of results

For the prediction of air quality PM2.5, we selected two single models and four currently popular Hybrid prediction models. Specifically, the models include CNN(model-1),

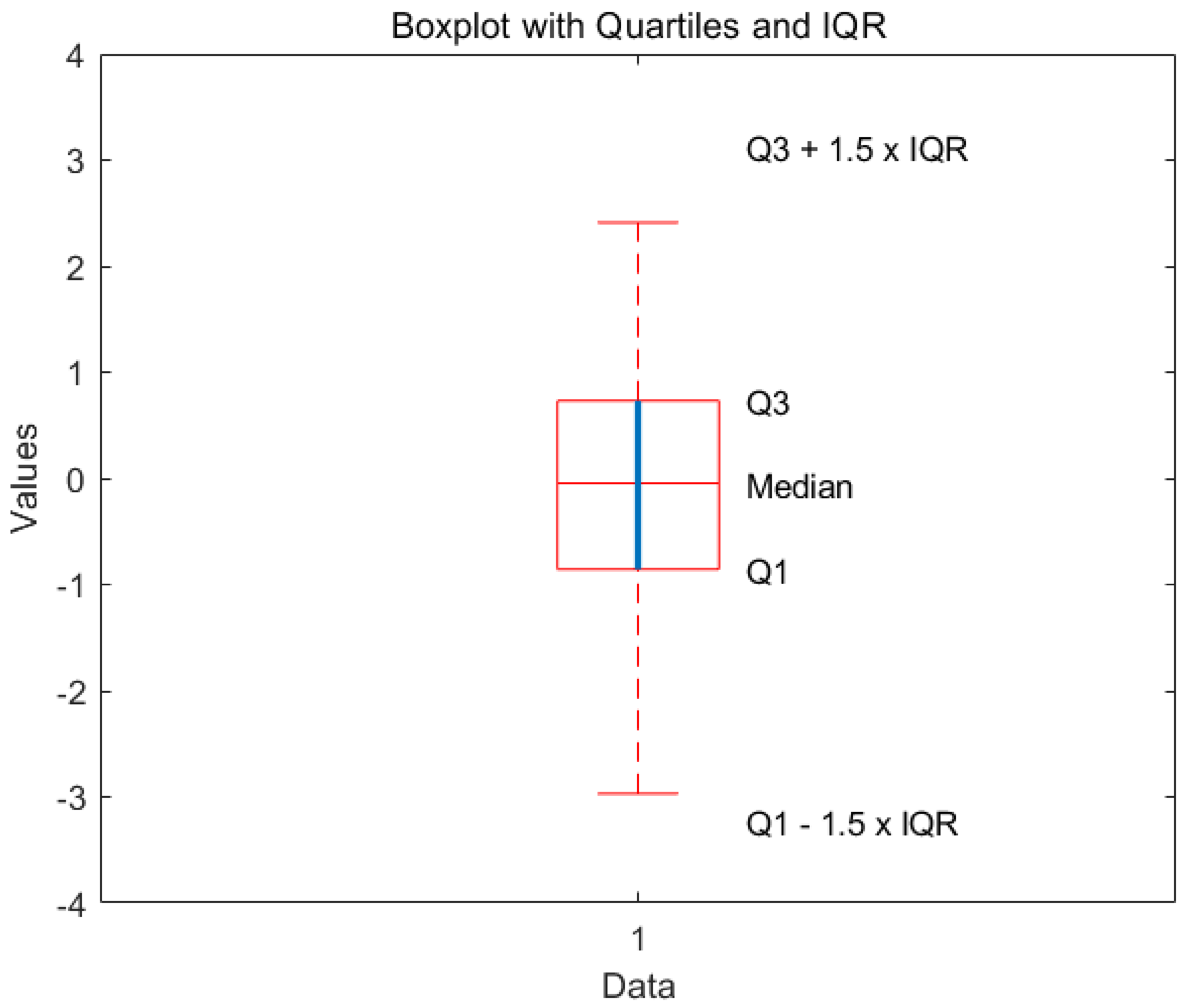

Figure 4. Box plot with IQR.

LSTM(model-2), CNN-BiLSTM-Attention(model-3), CNN-GRU-Attention(model-4), CNN-LSTM-Attention(model-5), and IDBO-CNN-LSTM-Attention(This paper). Figure 5 shows the prediction results for the test set, while Table 3 presents the evaluation indicators.

The air quality data was input into the four trained models, and Figure 5 compares the forecasted hourly PM2.5 concentrations with the actual observed values. The predicted values of the new model are close to the actual values, indicating low prediction error and high accuracy for PM2.5. To further validate the advantages of the new model in predicting PM2.5, it is compared with individual models: model-1 and model-2, as well as hybrid models model-3, model-4, model-5, and new model. The prediction results are shown in Figure 5. The chart illustrates the extent to which a single model may underperform compared to the overall model's level of fitting. Among the model-5 and new models, the latter demonstrates the best degree of fitting. However, when compared to the peak stage, the optimized new model shows a stronger ability to explain the data and achieves better prediction results. The predicted values of the model-3 are lower than the actual values at the peak stage, while the model-4 exhibits significant errors for some data points.

Table 3 presents the prediction evaluation indicators for the following models concerning the test set: model-3, model-4, model-5 and new model. MAE represents the mean absolute difference between the predicted and actual values, while RMSE is the square root of MSE. The model's error decreases as the values of RMSE, MSE, and MAE decrease. R^2 assesses how well the model explains the data, with values approaching 1 indicating a stronger fit, the greater the model's capacity to explain the data. Table 4 indicates that the Hybrid model demonstrates superior predictive ability compared to the individual models among the six examined. The new model is the most accurate for predicting PM2.5 concentrations, with R^2 of 0.904, RMSE of 2.677, MSE of 7.168, MAE of 1.982, and MAPE of 0.441. The model-3 has the worst prediction effect, with R^2(0.798), RMSE(3.888), MSE(15.117), MAE(3.076), and MAPE(0.467). The correlation coefficients of the four models are between 0.798 and 0.904. This may be

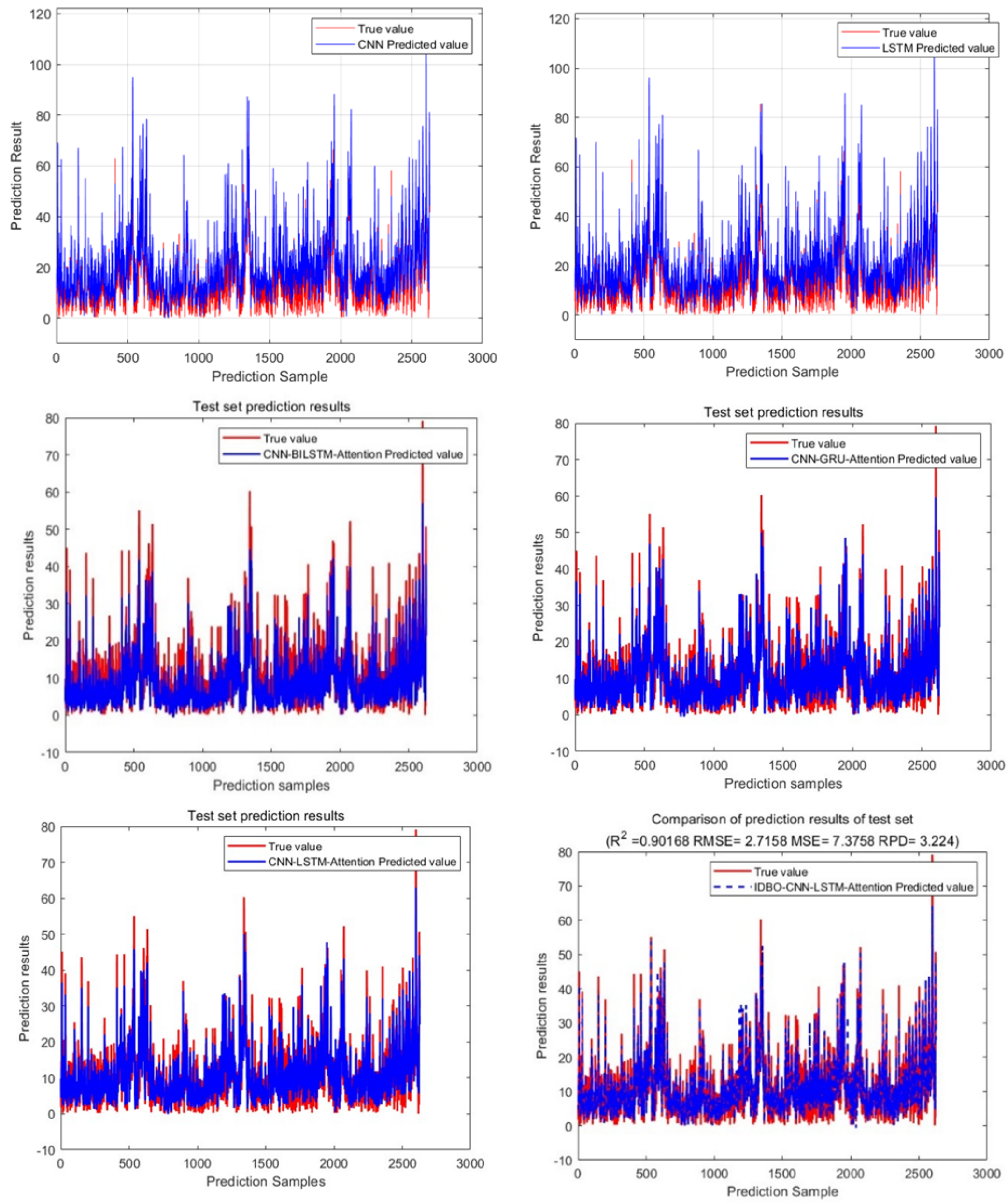

Figure 5. PM2.5 hourly concentration forecast.

due to the lack of data such as other industrial pollution sources and longitude and latitude geographic environment data, which leads to the lack of further improvement in model accuracy.

This paper integrates the strengths of multiple models and develops the new model. Firstly, since CNN excels at extracting spatial features, LSTM has an advantage in time series, and the Attention focuses on important features, the CNN-LSTM-Attention model is developed to enhance the interpretability of the model. Next, the DBO is improved to generate IDBO, which leverages its powerful global optimization ability to search for the hyperparameters of the CNN-LSTM-Attention model. Finally, the constructed new model is applied to air quality prediction. The results show that the hybrid model has greater advantages over single models in air quality prediction, and the impact of the hybrid model optimized by the enhanced dung beetle optimization method is more significant.

3.5. Discussion

In this study, we applied the new model to predict PM2.5 concentrations and conducted a comparative analysis with several commonly used models, including model-1, model-2, model-3, model-4, and model-5. The results indicate that the new model demonstrates superior prediction accuracy, particularly in handling complex time-series data and capturing nonlinear relationships between different variables.

Table 3. Model evaluation indicators.

Model	R^2	RMSE	MSE	MAE	MAPE
model-1	0.776	5.810	33.765	4.541	0.608
model-2	0.785	6.077	36.929	4.842	0.588
model-3	0.798	3.888	15.117	3.076	0.467
model-4	0.896	2.788	7.773	2.060	0.566
model-5	0.900	2.703	7.305	1.992	0.528
This paper	0.904	2.667	7.168	1.982	0.441

Table 4. Statistical significance test of model performance.

Comparison	Mean RMSE Difference	t-value	p-value
This paper vs. model-1	3.143	9.72	$p < 0.001$
This paper vs. model-2	3.410	10.53	$p < 0.001$
This paper vs. model-3	1.221	5.31	$p = 0.004$
This paper vs. model-4	0.121	1.41	$p = 0.173$ (not significant)
This paper vs. model-5	0.036	0.57	$p = 0.489$ (not significant)

3.5.1. Statistical analysis of model performance

To comprehensively evaluate model performance, we analyzed the RMSE differences and relative improvement rates. Our proposed model demonstrated a substantial reduction in RMSE, outperforming model-1 and model-2 by 54.1% and 56.1%, respectively. This significant improvement highlights the effectiveness of our approach in minimizing prediction errors. Furthermore, while the RMSE reduction compared to model-4 and model-5 was 4.3% and 1.3%, respectively, even these marginal gains indicate that our method provides additional refinements beyond state-of-the-art hybrid deep learning models. These results collectively validate the robustness and superiority of our model, establishing it as a more reliable and efficient solution for the given task.

3.5.2. Comparative analysis of models

Currently, models used for predicting PM2.5 levels in Penang are primarily based on machine learning techniques, such as RF and SVM, as well as independent models such as CNN and LSTM. While these models perform well in certain scenarios, their accuracy declines when dealing with highly complex, large datasets exhibiting nonlinearity. They often fail to capture the intricate spatiotemporal characteristics present in the data.

For instance, LSTM suffers from gradient vanishing, which reduces prediction accuracy over long time sequences. In contrast, IDBO effectively mitigates local optima issues and enhances the model's search capability, improving predictive performance.

- model-1 processes spatial features well but lacks strong temporal modeling capabilities.

- model-2 captures time-dependencies but still struggles with long-term dependencies.

- model-3 improves bidirectional dependencies, while The model-4 enhances computational efficiency.

- The model-5 integrates spatial and temporal features, making it a strong baseline.

- IDBO-CNN-LSTM-Attention further optimizes the loss function through enhanced search capabilities, achieving the lowest RMSE and MSE.

In this study, the RMSE optimized via IDBO effectively reduces errors, demonstrating superior predictive accuracy. The R^2 of 0.904 confirms the model's strong explanatory power. The RMSE of 2.677 is significantly lower than other models, particularly CNN and LSTM, showing clear improvements. The MSE reduction of 80% compared to single models validates the effectiveness of the optimization algorithm. Across MAE and MAPE metrics, IDBO-CNN-LSTM-ATT consistently outperforms other models.

3.5.3. Potential limitations

Despite its strong performance, the proposed model has some limitations:

- Dependence on dataset characteristics: The model was trained and tested on Penang air quality data, which may limit its generalizability to other regions. Future research should include cross-location validation.
- Computational cost: The hybrid deep learning model, combined with IDBO optimization, requires significantly more computational power compared to traditional machine learning models. Real-time deployment may require model compression techniques.
- Hyperparameter sensitivity: While IDBO enhances optimization, it still relies on predefined search spaces. A more adaptive search approach, such as reinforcement learning-based hyperparameter tuning, could further improve efficiency.

3.5.4. Practical implications for air quality management

The study's findings have significant real-world applications in air quality management and policy:

- Early warning system: Accurate PM2.5 predictions enable proactive measures, such as issuing air quality alerts and adjusting transportation policies.
- Industrial pollution monitoring: Improved forecasting accuracy allows government agencies to track and regulate emissions more effectively.
- Public health and urban planning: More precise long-term predictions help design green infrastructure, reduce pollution exposure, and protect vulnerable populations.

Overall, this study presents a novel hybrid deep learning approach for PM2.5 prediction, demonstrating that IDBO-optimized models provide state-of-the-art forecasting capabilities. Future research should focus on further optimizing computational efficiency and validating the model across different locations.

4. Conclusion

This work develops an air quality prediction model utilizing IDBO, CNN, LSTM, and Attention.

First, in order to solve the problem that the existing air quality prediction model relies on empirical knowledge for parameter selection, the golden sine strategy, self-spiral strategy, Levy Flight, and DBO method are enhanced through the incorporation of adaptive t-distributions.

Then, the critical parameters of the CNN-LSTM-Attention model (model-5) are tuned using the IDBO method.

It has been verified that the proposed model significantly outperforms single models (model-1, model-2) and hybrid models (model-3, model-4, model-5) in terms of R^2, RMSE, MSE, MAE, MAPE, and demonstrates good robustness. This confirms that the proposed model in this study is effective for predicting air quality due to its strong generalization performance.

4.1. Application to other regions and pollutants

Although the model is trained on PM2.5 data from Penang, its architecture and optimization method are applicable to other regions and pollutants. Since air pollution characteristics vary across different locations, future studies should apply the proposed model to diverse geographical areas, including urban and industrial regions, to assess its adaptability and generalizability.

Moreover, this approach can be extended beyond PM2.5 to predict other major air pollutants, such as:

- NO_2 (Nitrogen Dioxide): A key pollutant from vehicular emissions.
- SO_2 (Sulfur Dioxide): Primarily emitted from industrial activities.
- O_3 (Ozone): Crucial for smog formation in urban areas.

By training the model on multi-pollutant datasets, it is possible to develop a comprehensive air quality forecasting system that enhances environmental decision-making.

4.2. Future research directions

Although the proposed model achieves state-of-the-art predictive performance, several areas for future improvement remain:

- Integration of additional environmental factors: Incorporating meteorological parameters (temperature, humidity, wind speed) and traffic emissions to further enhance prediction accuracy.
- Real-time air quality prediction: Implementing the model with streaming data processing techniques, allowing for continuous real-time forecasting and immediate pollution alerts.
- Cross-region validation and transfer learning: Applying domain adaptation and transfer learning techniques to enhance the model's adaptability to different geographic regions and climates.
- Computational efficiency improvement: Exploring lightweight versions of the model (e.g., knowledge distillation or pruning) to enable real-time deployment on edge computing devices and mobile platforms.

4.3. Final remarks

In conclusion, this study presents a novel hybrid deep learning approach, optimized by an enhanced DBO algorithm, that significantly improves air quality prediction accuracy. The findings suggest that optimization-based deep learning models have strong potential for practical air quality forecasting applications.

Future research should focus on integrating real-time monitoring, expanding model applications to different regions and pollutants, and optimizing computational efficiency for real-world deployment.

Data availability

Due to confidentiality/proprietary restrictions, the datasets used in this study are not publicly available but can be obtained from the corresponding author upon reasonable request.

Acknowledgment

The authors thank the Department of Environment (DOE) of Malaysia for providing ambient air quality data.

References

[1] A. Binbusayyis, M. A. Khan, A. M. M. Ahmed & W. R. S. Emmanuel"A deep learning approach for prediction of air quality index in smart city", Discover Sustainability **5** (2024) 272. https://doi.org/10.1007/s43621-024-00272-9.

[2] E. X. Neo, K. Hasikin, K. W. Lai, M. I. Mokhtar, M. M. Azizan, H. F. Hizaddin, S. A. Razak & Yanto, "Artificial intelligence-assisted air quality monitoring for smart city management", PeerJ Computer Science **9** (2023) e1306. https://doi.org/10.7717/peerj-cs.1306.

[3] A. Tella & A.-L. Balogun, "GIS-based air quality modelling: Spatial prediction of PM10 for Selangor State, Malaysia using machine learning algorithms", Environmental Science and Pollution Research **29** (2022) 86109. https://doi.org/10.1007/s11356-021-16150-0.
[4] N. H. Van, P. Van Thanh, D. N. Tran & D.-T. Tran, "A new model of air quality prediction using lightweight machine learning", International Journal of Environmental Science and Technology **20** (2023) 2983. https://doi.org/10.1007/s13762-022-04185-w.
[5] H. Xu, Y. Chong, Y. Liu, W. Cao, D. Xu, J. Zhang & J. Zhao, "Prediction study of ARIMA-BPNN model in tuberculosis incidence rate", Journal of Medical Pest Control **10** (n.d.) 970. https://www.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2024&filename=YXDZ202410008&uniplatform=OVERSEA&v=xU3b51NBa_m19ZNudj4ZIa0sjORnAfzn5_4FvZvv0msOeQ-6iZdfOFOfNiOhWhg1.
[6] H. Zhou, T. Wang, H. Zhao & Z. Wang, "Updated prediction of air quality based on Kalman-attention-LSTM network", Sustainability **15** (2023) 1. https://doi.org/10.3390/su15010356.
[7] R. Zhou, X. Ma, H. Dong, J. Zeng, L. Yao & Z. Zhao, "Comparison of the effectiveness of ARIMA model and GM (1,1) in predicting the epidemic trend of other infectious diarrheal diseases", Journal of Medical Pest Control 10 (n.d.) 955–960. https://www.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2024&filename=YXDZ202410005&uniplatform=OVERSEA&v=xU3b51NBa_n3s_tDp6hLQnahstDiuyjjtGb7Qi0902PSiJw6uBk2LkcbQ0UZbvgF.
[8] K. Wang & H. Li, "Air pollution index prediction model based on multiple linear regression", Information Recording Materials **25** (2024) 1. https://doi.org/10.16009/j.cnki.cn13-1295/tq.2024.02.008.
[9] C. Ren & G. Xie, "Prediction of PM2.5 concentration level based on random forest and meteorological parameters", Computer Engineering and Applications **2** (2024) 213. https://www.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2019&filename=JSGG201902034&uniplatform=OVERSEA&v=M9qWRE8bApgTBqHdcQIIN8H-kZWrVv6hKI4_Oz4b_9FiylcX34qSwe2fWgfFidmU.
[10] Y. Song, Y. Xu & Y. Zhang, "Seasonal PM2.5 prediction based on SARIMA-SVM model", Computer Engineering **01** (2024) 51. https://doi.org/10.19678/j.issn.1000-3428.0068372.
[11] S. Sheng & W. Yuan, "Short-term wind speed prediction based on quadratic decomposition and BiLSTM-XGBoost-ARIMA combination model", Journal of North China Electric Power University **2024** (2024) 1. https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CAPJ&dbname=CAPJLAST&filename=HBDL20240725003.
[12] J. Duan, Y. Gong, J. Luo & Z. Zhao, "Air-quality prediction based on the ARIMA-CNN-LSTM combination model optimized by dung beetle optimizer", Scientific Reports **13** (2023) 12127. https://doi.org/10.1038/s41598-023-36620-4.
[13] Q. Wu & H. Lin, "Daily urban air quality index forecasting based on variational mode decomposition, sample entropy and LSTM neural network", Sustainable Cities and Society **50** (2019) 101657. https://doi.org/10.1016/j.scs.2019.101657.
[14] A. T. Nguyen, D. H. Pham, B. L. Oo, Y. Ahn & B. T. H. Lim, "Predicting air quality index using attention hybrid deep learning and quantum-inspired particle swarm optimization", Journal of Big Data **11** (2024) 926. https://doi.org/10.1186/s40537-024-00926-5.
[15] X. Liu & H. Guo, "Air quality indicators and AQI prediction coupling long-short term memory (LSTM) and sparrow search algorithm (SSA): A case study of Shanghai", Atmospheric Pollution Research **13** (2022) 101551. https://doi.org/10.1016/j.apr.2022.101551.
[16] Q. Shao, J. Chen & T. Jiang, "A novel coupled optimization prediction model for air quality", IEEE Access **11** (2023) 69667. https://doi.org/10.1109/ACCESS.2023.3293249.
[17] K. Wu, Y. Chai, X. Zhang & X. Zhao, "Research on power price forecasting based on PSO-XGBoost", Electronics **11** (2022) 22. https://doi.org/10.3390/electronics11223763.
[18] J. Zhu, S. Zhang, T. Zhang, J. Sun & Z. Zhang, "Air quality prediction based on two-layer decomposition and improved sand cat swarm optimization", Foreign Electronic Measurement Technology **5** (2024) 190. https://doi.org/10.19652/j.cnki.femt.2305800.
[19] X. Wang, Y. Wei, Z. Guo, J. Wang, H. Yu & B. Hu, "A Sinh-Cosh-enhanced DBO algorithm applied to global optimization problems", Biomimetics **9** (2024) 271. https://doi.org/10.3390/biomimetics9050271.
[20] Tanyildizi, E., & Demir, G. (2017). Golden sine algorithm: a novel math-inspired algorithm. Advances in Electrical & Computer Engineering, **17** (2017) 71. http://dx.doi.org/10.4316/AECE.2017.02010.
[21] J. Liu, L. Cong, Y. Xia, G. Pan, H. Zhao & Z. Han, "Short-term power load prediction based on DBO-VMD and an IWOA-BILSTM neural network combination model", Dianli Xitong Baohu Yu Kongzhi/Power System Protection and Control **52** (2024) 123. https://doi.org/10.19783/j.cnki.pspc.231402.
[22] R. Rajabioun, "Cuckoo optimization algorithm", Applied Soft Computing **11** (2011) 5508. https://doi.org/10.1016/j.asoc.2011.05.008.