Proceedings of CONF-CDS 2025 Symposium: Data Visualization Methods for Evaluation
DOI: 10.54254/2755-2721/2025.PO24524

Air Quality Prediction Based on CSSA-LSTM

Yuqi Xu

Colleges of medical technology and information Engineering, Zhejiang Chinese Medicine University, Hangzhou, China
18305869869@163.com

Abstract. Urban air pollution can result in respiratory and cardiovascular diseases, immune system damage, and various other health issues, jeopardizing citizens'health and the ecological environment, escalating social and economic burdens, and significantly impacting public health.There is still room for enhancement in the precision of current air quality prediction technologies.Consequently, utilizing the daily hourly air pollutant data from the National Environmental Monitoring Station in Binjiang District, Hangzhou City, Zhejiang Province, China, for the year 2024, this paper adopts the Long Short-Term Memory (LSTM) neural network model as the primary prediction framework. The chaotic sparrow search algorithm (CSSA) is introduced as the optimization algorithm for the hyperparameters of the LSTM model to predict the air quality of Hangzhou, and the RMSE, MAE, MAPE and R^2 indicators are used to evaluate the model effect, providing a high-precision prediction tool for dynamic environmental monitoring.The findings indicate that the CSSA-LSTM model produces more accurate predictions compared to both the standalone LSTM model and the SSA-LSTM model, which is beneficial for public health management and environmental enhancement.

Keywords: Long short-term memory neural network, chaotic sparrow search algorithm, air quality, hyperparameter optimization

1. Introduction

Urban air pollution significantly impacts public health, necessitating proactive measures like air quality forecasting. Such predictions enhance environmental management across key areas: pollution control, traffic regulation, urban planning, emergency response, public health strategies, policy formulation, research, and international collaboration.

Given the complexity of influencing factors, traditional statistical methods (e.g., regression, clustering, BP models) fall short. Current approaches include:Numerical models (e.g., NAQPMS by Chinese Academy of Sciences); Machine learning (e.g., SVM, ELM)

Developing high-performance prediction models remains critical to address these challenges effectively.

Existing air quality prediction models have limitations when dealing with highly complex, nonlinear, and non-stationary air quality data.In order to increase the prediction accuracy of this

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.PO24524

paper, the optimization and improvement of the air quality prediction model are the main topics of this work.

Deep learning has been gradually created in the last few years, which has the benefit of being able to self-adjust, learn from, and work toward more effective solutions to both complicated and nonlinear problems.At present, major achievements have been produced in image, speech and natural language processing, include convolutional neural network(CNN), recurrent neural net(RNN) and long short-term memory (LSTM) network models. To enhance the accuracy of air quality prediction, this paper utilizes the deep learning model prediction method introduced by Zhu Yanmin et al. [1], and after comparing the advantages and disadvantages and characteristics of different prediction models, the LSTM neural network model is selected as the basic prediction model.Improved Intelligence Optimisation Algorithm, Chaos Sparrow search algorithm(CSSA) is used to achieve more efficient and accurate parameter optimisation.

To assess the model's effectiveness, this paper performs experiments using the daily hourly air pollutant data from the China Environmental Monitoring Station in the Binjiang District of Hangzhou City, Zhejiang Province, in 2024. The results are compared against those obtained from a standard LSTM model and an unoptimized SSA-LSTM model, employing various evaluation metrics to support the analysis. The findings indicate that the proposed model demonstrates greater prediction accuracy.

The CSSA-LSTM model proposed in this paper improves the accuracy of air quality prediction results. which not only provides more reliable decision-making basis for the government environmental protection department to formulate haze emergency response plans and industrial emission control measures. but also helps the public health management department to optimize the allocation of medical resources and the early warning mechanism of diseases. For ordinary citizens, accurate air quality prediction can guide them to rationally arrange outdoor activity time and choose effective protective measures to better protect respiratory health.

2. Methodology

2.1. Problem analysis and solution

The nonlinear and non-stationary characteristics of air quality data are the core challenges in time series analysis, which directly affect the construction and prediction effect of the model.

Air quality data's nonlinear and non-stationary nature poses key challenges for time series modeling:

Nonlinear Characteristics:

1)Threshold effects (e.g., wind speed-PM2.5 relationship shows critical transition points)

2)Multi-factor synergies (e.g., ozone formation requiring NOx/VOCs + sunlight/temperature)

3)Abrupt changes from pollution events (sandstorms, accidents) that defy linear fitting

Non-stationary Features:

1)Long-term trends (policy-driven emission reductions)

2)Seasonal periodicity (winter heating/summer photochemistry effects)

3)Varied fluctuation ranges (winter stability vs. summer diffusion)

4)External shocks (typhoons, holidays disrupting patterns)

When employing a traditional time series model, like ARIMA, for forecasting, it's essential that the data is stationary. This necessitates preprocessing non-stationary data through methods such as differencing, detrending, or seasonal decomposition(e.g., STL decomposition), which can be quite cumbersome.

DOI: 10.54254/2755-2721/2025.PO24524

According to the above characteristics of air quality data, the LSTM nonlinear model is selected for experiments in this paper. In the LSTM model establishment method used by Xu Haichao, the accuracy of air quality assessment reaches 89% [2]. Xu Shenghua has built an LSTM model to predict the air quality index AQI, and the assessment indicators obtained are RMSE 10.47, MAE 10.16, MAPE 15.85%,and R2 75.19% [3]. It can be shown that the LSTM model has a certain effect on air quality prediction.

Nonetheless, the accuracy of predictions made by the LSTM model is influenced by the initial settings of its hyperparameters, including the number of iterations, learning rate, and the number of neurons in the hidden layers [4]. The process of optimizing the parameters of LSTM is one that is essential to raising the model's performance. However, artificial selection of hyperparameters may lead to poor prediction results, resulting in non-convergence, such as unstable faults.

To get around this disadvantage, the SSA-LSTM model is built by combining the SSA algorithm with the LSTM network,and then the Sparrow Search Algorithm(SSA)is used to optimize the parameters [5].An SSA-LSTMPM2.5 concentration prediction model was also suggested by Cao Jun [6]. In comparison to the individual LSTM, PSO-LSTM, and WOA-LSTM models, the goodness-of-fit R-squared results were enhanced by 45.93%, 31.55%, and 19.12%, respectively. (MAE)dition, the performance of Root Mean Square Error (RMSE)and Mean Absolute Error (MAE) is better, indicating that the model has high accuracy and efficiency in PM2.5 concentration prediction, and can also be used to the prediction of air pollutants.

The SSA algorithm and subsequent research revealed that the original random initialization population of the SSA has certain drawbacks. In high-dimensional or multimodal scenarios, it tends to get trapped in local optima, and the distribution of the initial population can result in low search efficiency. In contrast, chaotic sequences possess ergodicity, randomness, and sensitivity to initial values, making them capable of effectively generating diverse solutions and aiding in escaping local optima.

Therefore, in this paper, we choose to use the Chaotic Sparrow Search Algorithm (CSSA), which is improved from SSA for parameter optimization [7]. CSSA innovatively introduces a tent chaotic mapping mechanism, which systematically improves the population initialization process and position update strategy of traditional SSA through the ergodic and random characteristics of chaotic system. In the initialization step, CSSA uses a better tent chaotic map to generate the diversity and uniformity of the initial population distribution, which effectively avoids the population aggregation problem that may be caused by the traditional random initialization, and establishes a more ideal starting point for the subsequent optimization search. In the iterative optimization process, CSSA incorporates a chaotic disturbance mechanism into the position update formula, significantly improving its global exploration capabilities and enhancing its ability to escape local optima, all while preserving effective local search performance. These advantages make CSSA have a wide application prospect in the field of engineering optimization, and provide a new idea for the development of intelligent optimization algorithms.

2.2. Data description

2.2.1. Source of data

We chose use the scholars such as Wang Xiaolei since China's environmental monitoring station(www.cnemc.cn)released by the national urban air quality platform of real-time data released by the national air quality(https://quotsoft.net/air/#archive), From 0:00 on January 1, 2024 to 23:00 on December 31, 2024, the hourly concentration data of each air pollutant in Hangzhou were

DOI: 10.54254/2755-2721/2025.PO24524

selected to ensure the continuity and granularity of the data, so as to make the experimental prediction results more real-time. A total of 8683 data records were collected.The first 80%of the data is used as the training set and the last 20%of the data is used as the test set.

Six main air pollutants, including PM2.5, PM10, SO_2, NO_2, O_3 , CO, and the concentration values for the air quality dataset are included in the dataset.

2.2.2. Normalization processing

A key preprocessing step, data normalization reduces feature scale differences to enhance model performance and stability. This study employs min-max scaling, a linear method that transforms data into a fixed range (e.g., [0,1] or [-1,1]) while preserving distribution shape.The basic formula is as follows:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

In the original data, X represents the original data, Xmin for the characteristics of the minimum, Xmax for the maximum of characteristics, Xnorm for the results of the normalized.

2.3. LSTM network

LSTM, a specialized RNN variant, solves traditional RNNs' gradient vanishing problem in long sequences. Widely used in machine translation, speech recognition, and time-series forecasting, its core features—memory cell state and gate mechanisms (forget, input, output gates)—enable dynamic information flow control. This allows selective memory, long-term dependency handling, and gradient stability across hundreds of time steps, enhanced by Sigmoid and tanh functions' expressive power.

2.3.1. Structural decomposition of LSTM cells

The current input vector X_t is processed by each LSTM cell, which also includes the previous time step's hiddenstate and cell state as well as the current time step’s hidden state and the cell state.

Three layers of gated structures are used to finish the information flow, and two states cooperate inside the LSTM cell:

In the LSTM, which determines the amount of historical information todiscard from the cell state C_{t-1} according to the current input and the past state, the forget gate is a key part of the system. The input needs to concatenate the hidden state h_{t-1} at the last instant and the current input x_t into a long vector.

The weight matrix W_f is a two-dimensional matrix whose dimension is $(d_h + d_x) \times d_h$, where d_h is the hidden layer dimension and d_x is the input dimension. Each element of the output f_t can be between 0 and 1, which represents the probability that the cell state information at the corresponding position is retained.

Then there is the candidate memory and the input gate. What information to store in the cell state is determined by the input gate.

The input gate is calculated by controlling how much new information is written, similar to a"learning switch".

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

DOI: 10.54254/2755-2721/2025.PO24524

The formula for candidate memory generation is

$$\widetilde{C_t} = \tanh\left(W_c \bullet [h_{t-1}, x_t] + b_c\right) \tag{3}$$

While compressing the candidate memory such that the interval [-1, 1] is a nonlinear transformation, a new information $\widetilde{C_t}$ is provided, which is used to add the original unfiltered memory to the cell state at the current instant.

During the information flow, there is a cell state update, which combines the results of the forget gate and the input gate to update the long-term memory.

$$C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C_t} \tag{4}$$

It is utilized for element-wise multiplication, where $f_t \odot C_{t-1}$ signifies that the forget gate regulates the retention of past memories, and $i_t \odot \widetilde{C_t}$ represents that the input gate manages the incorporation of new memories.

The essence of gradient stabilization resides in the addition operation; during the back propagation process, the gradient can pass directly through this addition pathway, thereby circumventing the gradient vanishing issue typically caused by chain multiplication in traditional RNNs.

The output gate finally chooses which data to the hidden state in the cell state output.

The way the output gate is calculated is to control how exposed the cell state is

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

Its hidden state is generated by

$$h_t = o_t \odot \tanh(C_t) \tag{6}$$

The tanh compression function compresses the cell state C_t to the range[-1, 1]to prevent numerical explosion, while the gated filter ot determines the information output by the hidden state h_t at the current time step, as well as the information that is carried over to the next time step or provided as the current output.

2.3.2. Number of parameters and calculation details of LSTM

Each LSTM cell contains 4 weight matrices: W_f , W_i , W_C , W_o ,the dimension of each matrix is ($d_h + d_x$)× d_h , and the total number of parameters is 4×($d_h + d_x$)× d_h .

In the calculation process, when the time step is t, the calculation steps are as follows:

Step 1, concatenate the input: [h_{t-1}, x_t]→dimensions ($d_h + d_x$)×1

Step 2, calculate the forget gate:

$$f_t = \sigma(W_f \bullet [h_{t-1}, x_t] + b_f) \tag{7}$$

Step 3, calculate the input gates and candidate memories:

$$i_t = \sigma(W_i \bullet [h_{t-1}, x_t] + b_i) \tag{8}$$

DOI: 10.54254/2755-2721/2025.PO24524

$$\widetilde{C_t} = \tanh\left(W_c \bullet [h_{t-1}, x_t] + b_c\right) \tag{9}$$

Step 4, update the cell state:

$$C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C_t} \tag{10}$$

Step5, compute the input gate and hidden state:

$$o_t = \sigma(W_o \bullet [h_{t-1}, x_t] + b_o) \tag{11}$$

$$h_t = o_t \odot \tanh(C_t) \tag{12}$$

2.4. Implementation of CSSA algorithm

2.4.1. SSA algorithm

SSA (Sparrow Search Algorithm), proposed by Xue Jiankai, is a swarm intelligence-based optimizer inspired by sparrow collective behaviors (discoverers, followers, and vigilance) [8]. It mimics sparrow foraging and anti-predator strategies to balance global exploration and local exploitation, effectively solving optimization problems with multiple local optima. By idealizing sparrow behaviors, SSA establishes key operational rules to optimize LSTM configurations and address real-world challenges.

(1)Searchers (Finders) locate food-rich areas and guide recruiters. Their energy reserves depend on fitness.

(2)Upon detecting predators, sparrows emit alarms. If alarm levels exceed safety thresholds, foragers move to safer areas.

(3)Any sparrow can become a finder if it discovers better food, but the finder-recruiter ratio stays constant (one finder means another becomes a participant).

(4)Participants with low energy occupy worse foraging positions; some may leave to seek more energy.

(5)Foraging participants follow top finders for food or scout nearby. Some may compete for resources to increase predation rates.

(6)When danger arises, outer sparrows quickly relocate to safety for better positioning, while central ones move randomly.

In our experiment, we employ virtual sparrows to seek out food, and a group of n sparrows can be depicted in the following manner.

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,d} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,d} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,d} \end{bmatrix} \tag{13}$$

Following the variable dimension d to be optimized in the problem and the number n of sparrows,the fitness values of all sparrows can be expressed as follows.

DOI: 10.54254/2755-2721/2025.PO24524

$$F_x = \begin{bmatrix} f([x_{1,1} & x_{1,2} & \cdots & x_{1,d}]) \\ f([x_{2,1} & x_{2,2} & \cdots & x_{2,d}]) \\ & \vdots & & \\ f([x_{n,1} & x_{n,2} & \cdots & x_{n,d}]) \end{bmatrix} \tag{14}$$

Where f denotes the fitness value.

In SSA, it is the finder with the superior fitness value that takes precedence in searching for food.Furthermore, the finders are tasked with locating food for the entire group and offering foraging guidance to all members;as a result, they have access to a broader foraging search area compared to the participants. (1)in accordance with rules(1)and(2), during each iteration, the finders whose positions are updated are described as follows:

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^{t} \bullet exp\left(-\frac{i}{\alpha \bullet item_{max}}\right) & if\ R_2 < ST \\ X_{i,j}^{t} + Q \bullet L & if\ R_2 \geq ST \end{cases} \tag{15}$$

The current iteration count t is denoted by Where, j =1, 2, 3,..., d. $item_{max}$ is a constant that specifies the maximum iteration count. X_{ij} indicates the position information of the ith sparrow in the jth dimension. Let $\alpha \in (0,\ 1]$ denote a random number. $R_2(R_2 \in [0,1])$ and $ST(ST \in [0.5,1])$ stand for alert and safe values, which denote the thresholds for warning and safe, respectively.A normal distribution is followed by a random number Q.A $1 \times d$ matrix is denoted by L, and each element in the matrix has a value of 1.If $R_2 < ST$, it suggests that the foraging area is free of predators, allowing the discovering sparrows to perform broader exploration. Conversely, if $R_2 \geq ST$, it signifies that certain sparrows within the group have sensed a predator's presence and warned others. Consequently, all individuals must relocate to alternative safe locations to continue foraging.

Rules (4) and (5) need to be enforced through the observer. During foraging, some of the adders will observe the explorer at all times. They believe that the explorer has found a better food, they will immediately leave their current place to compete for food. If they win, they can get the finder of the food immediately, else Rule(5) needs to be continued. The location update for attendees is described below:

$$X_{i,j}^{t+1} = \begin{cases} Q \bullet exp\left(\frac{X_{worst} - X_{i,j}^{t}}{i^2}\right) & if\ i > n/2 \\ X_{p}^{t+1} + \left|X_{i,j}^{t} - X_{p}^{t+1}\right| \bullet A^{+} \bullet L & otherwise \end{cases} \tag{16}$$

X_p is the treasure hunter's current best position, and X_{worst} represents the current global worst position.A resents a $1 \times d$ matrix where each elementis assigned a random value of 1 or-1, while $A^{+} = A^{T}\left(AA^{T}\right)^{-1}$. When i>n/2, it signifies that the contestant with the lower fitness value has not yet received any food, is extremely hungry, and needs to seek energy elsewhere.

The experiment is based on the assumption that 10 to 20 percent of the population recognizes the danger. Initial positions of these sparrows were randomly created among the population. Following rule(6), the expression is as follows.

DOI: 10.54254/2755-2721/2025.PO24524

$$X_{i,j}^{t+1} = \begin{cases} X_{best}^{t} + \beta \bullet \left| X_{i,j}^{t} - X_{best}^{t} \right| & if\ f_i > f_g \\ X_{i,j}^{t} + K \bullet \left(\frac{\left| X_{i,j}^{t} - X_{worst}^{t} \right|}{(f_i - f_w) + \varepsilon} \right) & if\ f_i = f_g \end{cases} \tag{17}$$

In this context, X_{best} represents the current global best position.The step size control parameter,βis a random number drawn from a normal distribution with a mean of 0 and a variance of 1.K is a random value within the range of[-1,1], and f_i denotes the fitness value of the current sparrow individual. f_g and f_w are the current global best and worst fitness values, respectively.Additionally, ε is a small constant introduced to prevent the denominator from becoming zero.

The sparrow at this point is very susceptible to predators and is on the periphery of the population when $f_i > f_g$, which is indicated by the fact that it is currently on the edge of the population. According to X_{best} , the sparrow in this location is very safe and in the best of its population. The sparrows in the middle of the population are aware of the risk and have to travel closer to other sparrows in order to reduce the chance of predation, as shown by the fact that $f_i = f_g$. K is the step size control parameter and indicates the direction in which the sparrows are moving.

2.4.2. CSSA algorithm

The Sparrow Search Algorithm (SSA) achieves near-optimal global solutions but suffers from population diversity loss and local optima trapping. To address this, the Chaotic Sparrow Search Algorithm (CSSA) was developed with two key improvements: (1) Enhanced Tent chaotic mapping for population initialization boosts solution quality and global search capability; (2) Dynamic chaotic perturbation disrupts stagnant solutions, enabling escape from local optima.

The expression of Tent map is

$$Z_{i+1} \begin{cases} 2Z_i, \ 0 \le Z \le \frac{1}{2} \\ 2\left(1 - Z_i\right), \ \frac{1}{2} < Z \le 1 \end{cases} \tag{18}$$

Tent chaotic sequences exhibit points of small and unstable periods.(0,1)in order to prevent entering the small period or unstable period point and not destroy the three characteristics of the chaotic variable, random variable rand(0,1)/NT is introduced into the expression of the original Tent map.

The improved expression is

$$Z_{i+1} \begin{cases} 2Z_i + rand(0,1) \times \frac{1}{NT}, \ 0 \le Z \le \frac{1}{2} \\ 2\left(1 - Z_i\right) + rand(0,1) \times \frac{1}{NT}, \ \frac{1}{2} < Z \le 1 \end{cases} \tag{19}$$

rand(0,1) is a random value between[0,1], where NT is the number of particles in the chaotic sequence.

2.5. Model construction

1. Data Preprocessing:Split dataset into training/test sets. Normalize data to [0,1] range.

DOI: 10.54254/2755-2721/2025.PO24524

2. Population Initialization:Generate initial population using Tent chaotic mapping. Set parameters: population size, max iterations, etc.

3. Fitness Evaluation:Calculate and rank individual fitness values. Identify best/worst individuals.

4. Role Assignment:High-fitness individuals become discoverers, others as followers. Update position information.

5. Guard Mechanism:Randomly select guards and update their positions.

6. Adaptive Mutation:Trigger transient disturbance when individual fitness ≥ population average. Keep improved mutation results.

7. Optimal Solution Update:Record current best position and fitness value.

8. Termination Check:Output optimal parameters if conditions met. Otherwise return to Step 3 for further iterations.

9. Prediction Validation:Feed test data to optimized LSTM model. Denormalize and evaluate prediction results.

3. Results Analysis

3.1. Evaluation metrics

This paper utilizes the root mean square error(RMSE), mean absolute error(MAE), mean percentage error(MAPE), and the goodness-of-fit R-squared as the performance evaluation metrics for the model, which are calculated as follows.

$$\mathrm{RMSE}=\sqrt{\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}\left(\widetilde{\mathrm{Y}_{\mathrm{i}}}-\mathrm{Y}_{\mathrm{i}}\right)^{2}} \tag{20}$$

$$\mathrm{MAE}=\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}\left|\widetilde{\mathrm{Y}_{\mathrm{i}}}-\mathrm{Y}_{\mathrm{i}}\right| \tag{21}$$

$$\mathrm{MAPE}=\frac{100\%}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}\left|\frac{\widetilde{\mathrm{Y}_{\mathrm{i}}}-\mathrm{Y}_{\mathrm{i}}}{\mathrm{Y}_{\mathrm{i}}}\right| \tag{22}$$

$$\mathrm{R}^{2}=1-\frac{\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(\widetilde{\mathrm{Y}_{\mathrm{i}}}-\mathrm{Y}_{\mathrm{i}}\right)^{2}}{\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(\widetilde{\mathrm{Y}_{\mathrm{i}}}-\overline{\mathrm{Y}_{\mathrm{i}}}\right)^{2}} \tag{23}$$

$\mathrm{Y_i}$ is the actual gas concentration value that was seen at the ith time point; $\mathrm{Y_i}$ is the model-estimated gas concentration value;and N is the total length of the observed sequence.Lower values of RMSE,MAE, and MAPE typically signify smaller prediction errors;as the values of R^2 approach 1, the model's predictive power and fit improve.

3.2. Performance comparison and analysis

In the air quality prediction task, we compare the performance of the standard LSTM model with its two optimized variants—SSA-LSTM and CSSA-LSTM.

3.2.1. Prediction Accuracy Improvement

RMSE:The CSSA-LSTM achieves an RMSE of 11.85, which is 23.4% lower than the standard LSTM (RMSE=15.48) and 8.2% lower than SSA-LSTM (RMSE=12.91). This indicates that CSSA-

DOI: 10.54254/2755-2721/2025.PO24524

LSTM's predictions deviate less from the true values, demonstrating superior overall stability.The improvement is attributed to the chaotic optimization mechanism in CSSA, which enhances the search capability of SSA, allowing for better hyperparameter tuning and avoiding premature convergence.

MAE:The MAE of CSSA-LSTM (9.53) is 6.2% lower than that of LSTM (10.16) and 2.3% lower than SSA-LSTM (9.75), suggesting that CSSA-LSTM further reduces absolute prediction errors. This improvement stems from the more refined weight optimization in CSSA-LSTM, which adjusts the model's learning process to minimize large deviations.

MAPE:The MAPE of CSSA-LSTM (16.98%) is 25.6% lower than LSTM (22.82%) and 7.3% lower than SSA-LSTM (18.32%), confirming that CSSA-LSTM has the smallest relative prediction error. The chaotic perturbation strategy in CSSA helps escape local optima, leading to more accurate and robust predictions compared to standard SSA-LSTM.

3.2.2. Model Fitting Capability

R^2:The $R^2$of CSSA-LSTM (93.62%) is significantly higher than that of SSA-LSTM (91.28%) and LSTM (87.45%), indicating that CSSA-LSTM has the strongest ability to explain data variations. This improvement is due to the enhanced global search ability of CSSA, which optimizes LSTM's parameters more effectively, leading to better alignment with real-world air quality trends.

3.2.3. Reasons for CSSA-LSTM's Superior Performance

Chaotic Optimization in SSA:The introduction of chaotic mapping prevents the standard SSA from getting trapped in local optima, allowing for a more thorough exploration of the parameter space.

Improved Hyperparameter Tuning:The adaptive adjustment of learning rates and network weights in CSSA-LSTM leads to faster convergence and higher prediction accuracy.

Better Handling of Long-Term Dependencies:The optimized LSTM structure retains critical historical information more effectively, improving its ability to model complex air quality patterns.

3.2.4. Practical Implications

The superior performance of CSSA-LSTM makes it a more reliable tool for air quality forecasting, providing higher accuracy and stability compared to traditional LSTM and SSA-LSTM. This improvement is crucial for environmental monitoring, pollution control, and public health decision-making, where precise predictions are essential.

The experimental results, presented in Table 1, demonstrate that CSSA-LSTM outperforms both LSTM and SSA-LSTM across all evaluation metrics, confirming the effectiveness of its improvement strategy.

Table 1. A comparison of the evaluation metrics for LSTM and its two optimized models

Model	RMSE	MAE	MAPE/%	R^2/%
LSTM	15.47	10.16	22.81	87.45
SSA-LSTM	12.91	9.75	18.31	91.28
CSSA-LSTM	11.85	9.53	16.98	93.62

DOI: 10.54254/2755-2721/2025.PO24524

4. Conclusion

The air quality prediction model of the long-term short-term memory neural network, which was improved by the sparrow search algorithm, is put forth in this paper. By constructing the CSSA-LSTM model, the initialization problem of the hyperparameters of the LSTM model is effectively solved, and the concentration of various air pollutants is predicted. Based on the data in 2024, a comparative experiment was carried out on the concentration of air pollutants in Binjiang District of Hangzhou. The experimental results demonstrate that the CSSA-LSTM model outperforms both the single LSTM model and the unoptimized SSA-LSTM model in terms of prediction accuracy.

Although the CSSA-LSTM model proposed in this study has made some progress in the field of air quality prediction, from the experimental comparison results, the performance advantage of the model is still not obvious compared with the improved models (such as SSA-LSTM). The key indicator RMSE is only 8.2% lower than that of SSA-LSTM, and the improvement rate of MAE is only 2.3%. While these improvements are statistically significant, their practical value may be limited, particularly in scenarios involving sudden pollution events.Therefore, the model's response speed and prediction stability require further enhancement.

Future research can start from the direction of optimizing the network structure and enhancing the transfer learning ability to further improve the prediction performance and generalization ability of the model. At the same time, more kinds of environmental data will be combined, more complex external influence factors will be considered, and the model will be extended to the prediction of other pollutants and other meteorological data, so as to better serve the actual needs of environmental governance and public health protection.

References

[1] Yanmin Zhu, Ailan Xu, Qiang Sun.New progress of air quality prediction methods based on deep learning [J]. China Environmental Monitoring, 2020, 31 (2) : 68-62. 36 (03) : 10-18.10.19316 / j.i SSN. 1002-6002.2020.03.02.
[2] Xu Haichao, Chen Liang.Air Quality Prediction Based on LSTM [J].Electrotechnical Technique, 2020(6): 127-128.
[3] Xu Shenhua.Prediction of air quality Based on LSTM [J].Cleaning World, 2025(1): 1671-8909(2025)1-0064-003
[4] Lou Z H.Air quality forecasting modeling method based on CNN-LSTM [J].Yangtze River Information and Communication, 2015, 25 (1) : 71-8909. 10.20153 / j.i SSN. 2096-9759.2024.12.013.
[5] Wang Yukun, Liu Hong, Prediction of PM2.5 concentration in Wuhan City based on SSA combination model [D].Zhongnan University of China. Economics and Law, 2020.05.10.
[6] Cao Huanjun, Li Changyun.Research on Air quality prediction based on SSA-LSTM model [J]. 2024(2): 10.19850/ J. carroll nki.2096-4706.2024.04.030
[7] Liu Xingqing, Zhao Guoshuai, Han Sumin.IGBT Module Degradation Trend Prediction Based on CSSA-LSTM [J].Electronic Science and Technology of China, 2011, 26 (2) : 71-79. 2024, 37 (8): 60-67.
[8] XUE Jiankai.Research and Application of a New Swarm Intelligence Optimization Technology [D].Donghua University, 2020.