Evidence in Earth Science

RESEARCH ARTICLE

DOI: 10.63221/eies.v1i01.1-16

Highlights:

- SSA was enhanced through nonlinear SCA integration and adaptive PS ratio adjustment.
- ISSA-BiLSTM accurately predicts seepage pressure dynamics with high precision.
- Input layer parameters were optimized via redundancy pruning to enhance predictive accuracy.
- ISSA-BiLSTM achieves significant error reduction versus SSA-BiLSTM/BiLSTM.

Keywords:

Sparrow search algorithm
Improve sparrow search algorithm
Bidirectional long-term and short-term neural network
Osmotic pressure prediction
Dam seepage

Correspondence to:

suo20202020@outlook.com

Citation: S. Chen, B. Liu, C. Xing *et al.*, 2025. Prediction of dike seepage pressure based on ISSA-BiLSTM. Evidence in Earth Science, 1(01), 1-16.

Manuscript Timeline:

Received	March 6, 2025
Revised	March 15, 2025
Accepted	March 18, 2025
Published	March 19, 2025

Academic Editor:

Huicong Jia

Prediction of dike seepage pressure based on ISSA-BiLSTM

Shoukai Chen [1,2], Beiying Liu [1,2], Chunpeng Xing [1,2,3], Mengdie Zhao[4,5,*], and Jiayang Zhou [6]

[1] School of Water Conservancy，North China University of Water Resources and Electric Power, Zhengzhou, 450046, China
[2] Henan Key Laboratory of Water Environment Simulation and Treatment, Zhengzhou, 450046, China
[3] Hebei Investigation Design & Research Institute of Water Conservancy & Hydropower, 300250, China
[4] School of Water Resources，North China University of Water Resources and Electric Power, Zhengzhou, 450046, China
[5] Henan Key Laboratory of Water Resources Conservation and Intensive Utilization in the Yellow River Basin, 450008, China
[6] Henan water conservancy investment group co., LTD, Zhengzhou, 450046, China

Abstract The existing traditional dam seepage pressure prediction models have problems such as falling into local optimum. The sparrow search algorithm (SSA) was improved as ISSA using both methods of nonlinear Sine Cosine optimization algorithm and adaptive producer and scrounger ratio. We combined the Bidirectional Long Short-Term Memory (BiLSTM) neural network model with ISSA to develop the ISSA-BiLSTM seepage pressure prediction model. The critical feature factors were extracted based on LightGBM to construct the input layer for seepage pressure prediction. The results show that the ISSA-BiLSTM model's fitting outcomes are generally consistent with the observed changes in seepage pressure observations, achieving an R^2 of 0.987. In comparison to SSA-BiLSTM and BiLSTM, the model exhibits a substantial reduction in errors, decreasing by approximately 20% and 30%, respectively. This model can provide technical support and insights for accurately predicting dam seepage, contributing to the advancement of this field.

1. Introduction

Dams play a crucial role in water resource management, irrigation, power generation, and flood control (Liu *et al.* 2024). The high-quality, efficient, and safe construction and long-term efficient and safe operation of dams are crucial to national welfare and people's livelihoods (Li, *et al.* 2022). More than 100,000 dam projects have been built around the world (Song *et al.*, 2023). Since the initiation of economic reforms and opening-up policies, along with the development of technological capabilities and improvements in the standard of living, the scale and quantity of dams in China have continuously expanded. According to statistics, the number of dams in operation has exceeded 23,000, accounting for 40.6% of the global total (For dams with a height exceeding 15 meters, or dams with a height greater than 5 meters and a reservoir capacity exceeding 3 million cubic meters)

(Huang *et al.*, 2021). Presently, many dams have been in service for numerous years, with some approaching or surpassing their design lifespan, placing them in a state of "ailing" or "high-risk" conditions. Compounded by frequent extreme weather events in recent years, potential hazards to the safe operation of dams have increased. Seepage issues represent a significant factor affecting the safety of dams (Chen *et al.* 2024; Li *et al.* 2024). According to statistics, dam failures, particularly those involving earth-fill dams, are often attributed to permeation-induced deformations. Between 2000 and 2018, 25.72% of dam failure cases in China were linked to seepage issues (Li *et al.* 2021). Given the increasing visibility of risks associated with a considerable number of dams in operation, research on dam monitoring and safety diagnostic technologies has garnered significant attention (Wang *et al.* 2018). Seepage, as a direct indicative factor of a dam's operational safety, has become a focal point and hotspot in this field (Beiranvand & Rajaee, 2022). Addressing dam seepage issues, uncovering their temporal patterns, understanding their seepage behaviors, and conducting predictive analyses are conducive to comprehensively grasping the engineering operation status and reducing the probability of failures. This research holds substantial practical significance in the current context where the risks associated with a considerable number of dams in operation are gradually becoming apparent (Li *et al.* 2023).

Numerous factors influence seepage pressure in dams, including upstream and downstream water levels, precipitation, temperature, etc. The study of dam seepage pressure prediction began in the previous century, and with the rapid development of artificial intelligence technology, various models have been employed for dam seepage pressure prediction (Wen *et al.* 2023), such as neural network models, support vector machine models, extreme learning machine models, among others. Among them, bidirectional long short-term neural networks have been widely applied due to their outstanding temporal processing capability and good fitting performance (Wang *et al.* 2023). Scholars continuously improve these models using two main approaches.

The first approach involves coupling other algorithms into the models. For instance, Qiu *et al.* (2016) utilized genetic algorithms to optimize backpropagation neural networks, Zhang *et al.* (2020) coupled wavelet analysis and neural network technology to establish a seepage pressure prediction model for earth-rock dams. Miao *et al.* (2019) introduced cloud model coupling with artificial fish swarm algorithm, addressing deficiencies in BP neural network weights and thresholds, constructing the CM-AFSA-BP prediction model. Yue *et al.* (2020) incorporated upstream and downstream water depth components, time components as input layers, combined with convolutional neural networks and long short-term memory neural networks to establish a dam seepage prediction model. Li *et al.* (2023) enhanced the local search capabilities of GBO by introducing cross operators and nonlinear parameters, proposing the IGBO-ELM model combining Improve Gradient-based optimizer and Extreme Learning Machine. Zhang *et al.* (2021) employed Extreme Gradient Boosting (XGBoost) to predict seepage in the dam body. Hybridizing grey wolf optimization (HGWO), which combines Differential Evolution (DE) with Grey Wolf Optimization (GWO), along with five-fold cross-validation, was utilized to optimize the hyper-parameters of XGBoost. Ishfaque *et al.* (2022) utilized Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) as time series algorithms for seepage flow prediction at the Tarbela Dam in Pakistan. Wang *et al.* (2022) established a seepage pressure prediction model for dams based on Grey Self-Memory Theory, utilizing the Grey Model GM (1,1) as the dynamic core. In the Grey Self-Memory model, information from multiple monitoring values is fully utilized.

The second approach involves filtering and processing seepage influencing factors at the input layer. Chen, *et al.* (2020) starting from the factors affecting dam seepage, applied multi-population Jaya algorithm to achieve adaptive optimization of the kernel extreme learning machine (KELM), obtaining a seepage prediction model. Qin *et al.* (2018) initially used a stepwise regression method to analyze the significance of 11 seepage influencing factors, using significant factors as input layers for seepage prediction in a wavelet neural network. Liu *et al.* (2020) emphasized the correlation between seepage flow and its influencing factors, selecting input layers based on this correlation, and introduced a biogeography-based optimization algorithm (BBO) to construct a seepage pressure prediction model. Wang *et al.* (2022) initially employing principal component analysis (PCA) for dimensionality reduction of seepage influencing factors, then enhanced the global exploration capability of long short-term neural networks using ant lion optimizer (ALO), establishing an ALO-LSTM-

based seepage pressure prediction model. Application instances indicate high fitting accuracy of the model, with an R^2 reaching 0.98. Liu *et al.* (2023) employed the HTRT (hydrostatic-thermal-rainfall-time) model to identify the time-series influencing factors of seepage flow. The HTRT model nested the Random Forest algorithm to establish a seepage flow prediction model for dams. The MIC algorithm was utilized for dual purposes of time delay analysis and sensitivity analysis. The aforementioned research has achieved notable improvements in enhancing model prediction accuracy and stability.

While the aforementioned studies have made progress in dam seepage pressure prediction, challenges such as susceptibility to local optima and low prediction efficiency persist (Shao *et al.* 2023). In light of these issues, this paper attempts to address them by introducing a multi-strategy improved Sparrow Search Algorithm (ISSA) to overcome deficiencies in global optimization and convergence speed. The bidirectional long short-term neural network (BiLSTM) is employed to construct the input layer for seepage pressure prediction based on the correlation between seepage influencing factors and seepage values, as well as its inherent temporal patterns. The LightGBM technique is introduced to extract critical feature factors. Utilizing the improved Sparrow Search Algorithm (ISSA), the parameters of the BiLSTM model are optimized. The objective is to establish an efficient and accurate dam seepage prediction model. Finally, the performance of the improved seepage prediction model is compared with a benchmark seepage prediction model, and an analysis is conducted to evaluate the accuracy of the enhanced seepage prediction model.

2. Materials and Methods

2.1. Multiple-Strategy Collaborative Improvement Sparrow Search Algorithm

2.1.1. Sparrow Search Algorithm

The Sparrow Search Algorithm (SSA) (Xue, *et al.*, 2020) is a novel multi-population algorithm inspired by the foraging and predator avoidance behaviors of sparrows. SSA exhibits advantages such as fast convergence speed, strong optimization capability, and stability (Yang *et al.*, 2024). When coupled with the Bidirectional Long Short-Term Memory Neural Network (BiLSTM), it enhances the overall optimization ability and computational efficiency of the model. During foraging, a portion of sparrows are responsible for providing foraging areas and directions, while the rest follow the former for foraging. Simultaneously, when an individual sparrow detects danger, it emits an alarm signal, prompting the population to immediately engage in anti-predatory behavior. Within the sparrow population, producers offer foraging directions to other individuals; scroungers accompany producers in foraging activities; and sentinels are tasked with monitoring and alerting (Zhou *et al.*, 2024). In SSA, sparrows are categorized into three roles: producers, scroungers, and sentinels.

The update formula for the position of producers is as follows:

$$x_{i,j}^{t+1} = \begin{cases} x_{i,j}^{t} \cdot \exp\left(\dfrac{-i}{\alpha \cdot T_{\max}}\right), & \text{if } R_2 < ST \\ x_{i,j}^{t} + Q \cdot L, & \text{if } R_2 \geq ST \end{cases} \quad (1)$$

where $x_{i,j}^{t}$ represents the position of the i^{th} sparrow in the j^{th} dimension, where t is the current iteration count, T is the maximum iteration count, α is a random number belonging to the interval (0,1], R_2 is the warning value ($0< R_2 \leqslant 1$), ST is the safety value ($0< \mathrm{ST} \leqslant 1$), Q is a random number following a normal distribution, L represents a $1 \times$ dim matrix, where dim denotes the dimension value.

The update formula for the followers' positions is as follows:

$$x_{i,j}^{t+1} = \begin{cases} Q \cdot \exp\left(\dfrac{x_{\text{worst}}^{t} - x_{i,j}^{t}}{\alpha \cdot T_{\max}}\right) & \text{if } i > \dfrac{n}{2} \\ x_{\text{best}}^{t+1} + \left| x_{i,j}^{t} - x_{\text{best}}^{t+1} \right| \cdot A^{+} \cdot L & \text{else} \end{cases} \tag{2}$$

where x_{worst}^{t} represents the worst position of the producers globally at the sub-iteration, x_{best}^{t} represents the optimal position of the producers, and A is a 1 × D dimensional matrix.

The updated positional coordinates of the sentinel are as follows:

$$x_{i,j}^{t+1} = \begin{cases} x_{\text{best}}^{t} + \beta \cdot \left| x_{i,j}^{t} - x_{\text{best}}^{t} \right| & \text{if } f_i > f_g \\ x_{i,j}^{t} + K \cdot \left[\dfrac{\left| x_{i,j}^{t} - x_{\text{worst}}^{t} \right|}{(f_i - \mathrm{f_w}) + \varepsilon} \right] & \text{if } f_i = f_g \end{cases} \tag{3}$$

where *fi, fw, fg* represent the current fitness of the sparrow, and takes values for the minimum and maximum fitness; β denotes the step size control parameter; K stands for a random number; ε is a sufficiently small value to prevent division by zero.

2.1.2. The Multi-Strategy Cooperative Improved Sparrow Search Algorithm

While the Sparrow Search Algorithm (SSA) demonstrates favorable performance, it tends to encounter challenges related to local optima in later stages, with certain deficiencies in optimization accuracy and stability (Xiong *et al.*, 2022). In order to enhance SSA's optimization effectiveness for BiLSTM and elevate model's prediction precision, a novel approach is introduced herein. This approach involves the formulation of a Multi-Strategy Cooperative Improved Sparrow Search Algorithm (ISSA) aiming at synergistically refining SSA through multiple strategies. The essence lies in compensating for the rigid defects of the original algorithm through a dynamic parameter adjustment mechanism, and its characteristics of low computational overhead and high environmental adaptability enable it to demonstrate significant advantages in optimization tasks targeting complex models such as BiLSTM. Compared to other strategies such as Cauchy mutation and Tent chaotic mapping, this combination achieves a delicate exploration-exploitation balance while ensuring algorithmic efficiency, providing a more reliable optimization foundation for time series prediction.

(1) The coefficient for the adaptive producer and scrounger ratio.

The proportion of producers to scroungers in the SSA remains constant throughout the entire optimization process. Consequently, during the early stages of optimization, the producers' proportion is relatively low, resulting in suboptimal global search performance. Conversely, in later stages, the number of producers becomes excessive when additional scroungers are more pertinent for enhancing local optimization capabilities. To address these shortcomings, an adaptive strategy is introduced involving coefficients for the discovery-to-scroungers ratio. This strategy aims to increase the proportion of producers during the early optimization stages and, in the later stages, dynamically adjust the number of scroungers, reducing the number of producers adaptively. This adaptive adjustment transitions the optimization focus from global exploration to precise local search, thereby improving the convergence accuracy of the SSA.

$$r = b\left(\tan\left(-\frac{\pi\, t}{4 \cdot iter_{\max}} + \frac{\pi}{4}\right) - k \cdot rand(0,1)\right) \tag{4}$$

$$pNum = r \cdot N \tag{5}$$

$$sNum = (1-r) \cdot N \tag{6}$$

where *pNum* represents the number of producers, *sNum* denotes the number of scroungers, *b* signifies the ratio between producers and scroungers, and *k* is the perturbation deviation factor.

(2) Nonlinear Sine Cosine Optimization Algorithm

In the early iterations of SSA, as the number of iterations increases, each dimension of the scroungers tends to enlarge, while the dimensions of the producers decrease. This results in a continuous reduction of the search space, making it susceptible to local optima. To address this issue, a Nonlinear Sine Cosine Algorithm is introduced to enhance the step-size search factor $r_1^{'}$ and the followers' position update formula (eq. 7). The transformation curve of $r_1^{'}$ is illustrated in Figure 1, and the formula for $r_1^{'}$ is provided in (eq. 8). The improvement ensures a slow decrease in the early stages when the weights are relatively small, allowing producers to maintain larger values and facilitating global search. In the later stages, when the weights increase, the decrease rate accelerates, leading to an increase in the number of scroungers. This enhancement contributes to improved local optimization capabilities, expediting the convergence to the optimal solution (Li, *et al.*, 2022).

$$r_1^{'} = a \times (1 - (\frac{t}{Iter})^{\eta})^{1/\eta} \tag{7}$$

$$X_{i,j}^{t+1} = \begin{cases} w \cdot X_{i,j}^{t} + r_1^{'} \cdot \sin r_2 \cdot \left| r_3 \cdot X_{i,j}^{t} \right|, \text{if i} > \dfrac{n}{2} \\ w \cdot X_{i,j}^{t} + r_1^{'} \cdot \cos r_2 \cdot \left| r_3 \cdot X_{i,j}^{t} \right|, \text{other} \end{cases} \tag{8}$$

2.2. The ISSA-BiLSTM Seepage Pressure Prediction Model

2.2.1. Bidirectional Long Short-Term Memory Neural Network Model

The Long Short-Term Memory (LSTM) neural network (Hochreite *et al.*, 1997) was proposed to address the limitations of traditional neural network techniques, specifically the "vanishing or exploding gradient" issue. LSTMs excel in addressing predictive or classification problems related to time-series data. There are two main variants: the Bidirectional Long and Short-Term Memory network model (BiLSTM) and the unidirectional Long and Short-Term Memory network model (LSTM). However, the unidirectional LSTM has certain limitations in terms of memory capacity. In this study, a more comprehensive approach is adopted by selecting the Bidirectional Long Short-Term Memory Neural Network Model, which comprises LSTM units in both forward and backward directions, as shown in Figure 1. The input sequence, denoted as X = (x_1, x_2, ..., x_t), is processed through the network, and the output layer is connected to both the forward and backward layers, as illustrated in Figure 1.

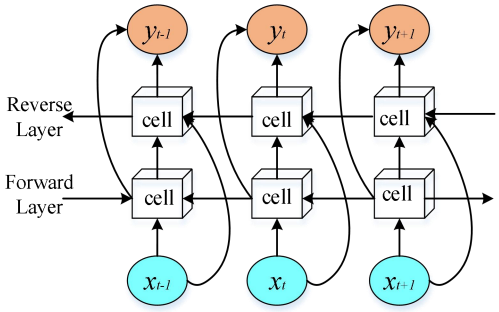


Fig. 1 The schematic diagram of the Bidirectional Long Short-Term Memory network model structure

2.2.2. Data Selection and Processing

In accordance with the overall sample, a random selection of 15% to 20% of the total samples is chosen as the testing dataset, while the remaining 80% to 85% constitutes the training dataset. Given the significant variations in the input parameter ranges for the infiltration prediction model and the disparate physical dimensions of different factors, it is essential to normalize the data of each indicator in the input layer of the infiltration prediction model. This paper employs the Min-max method, as indicated in (eq. 9).

$$y_i = \frac{x_i - x_{i\min}}{x_{i\max} - x_{i\min}} \tag{9}$$

where y_i represents the normalized value, x_i denotes the original data of the monitoring variable, x_{imax} is the maximum value in the long sequence monitoring dataset, and x_{imin} is the minimum value in the long sequence monitoring dataset.

2.2.3. Model Parameters and Input Variables

The seepage pressure head of a dam is mainly influenced by the time-varying characteristics of factors such as upstream and downstream water levels, temperature, rainfall, and the material properties of the dam body (Gu & Wu, 2006). In light of this, this paper establishes the initial input layer for seepage pressure prediction from both feature and time perspectives.

The expression for the seepage pressure effect quantity in the dam is as follows:

$$\delta = \delta_H + \delta_T + \delta_P + \delta_\theta \tag{10}$$

where δ represents the seepage pressure effect quantity in the dam, δ_H represents the water level component, δ_P represents the rainfall component, δ_θ represents the time-dependent component, and the rules for determining each component are as follows.

(1) δ_H : Considering the significant impact of reservoir water level on dam seepage and the existence of lag effects (Pang *et al.* 2016; Shi, *et al.* 2018), it is essential to not only consider the current upstream water level but also to take into account the variations in water level in the days preceding the monitoring date. In this study, the mean upstream water level for the 3 days preceding the monitoring date and the mean upstream water level for days 4-7 preceding the monitoring date are considered.

$$\delta_H = (h_u, \overline{h}_{1-3}, \overline{h}_{4-7}) \tag{11}$$

where h_u represents the upstream water level of the reservoir on the monitoring date, $\overline{h}_{1-3}$ represents the average upstream water level of the reservoir in the 3 days preceding the monitoring date, $\overline{h}_{4-7}$ represents the average upstream water level of the reservoir in the 7 days preceding the monitoring date.

(2) δ_T : The mechanism of temperature component on the seepage effect in dams lies in the thermal expansion and contraction of the dam body and bedrock caused by temperature changes, affecting the permeability coefficient of the soil. Generally, the permeability coefficient of soil decreases with an increase in temperature, exhibiting periodic variations. In this study, the representation of the temperature component is adopted from (Wang *et al.* 2022).

$$\delta_T = (\sin\frac{2\pi it}{365}, \cos\frac{2\pi it}{365}) \tag{12}$$

(3) δ_P : The impact of the rainfall component on the seepage characteristics of the dam involves two aspects (Yu *et al.* 2010). One part is the increase in water level caused by surface runoff generated by rainfall flowing into

the reservoir. The other part is the direct infiltration of rainfall into the dam body, influencing the unsaturated seepage field and causing changes in the seepage pressure values.

$$\delta_P=(P,\overline{P}_{1-3},\overline{P}_{4-7}) \tag{13}$$

where P represents the rainfall amount on the monitoring date, $\overline{P}_{1-3}$ represents the average rainfall amount in the 3 days preceding the monitoring date, and $\overline{P}_{4-7}$ represents the average rainfall amount in days 4-7 preceding the monitoring date.

(4) δ_θ : The effects of variables such as water level and rainfall on the seepage pressure values in dams exhibit a lag. Here, these effects are transformed into time-dependent components, typically expressed by the following equation:

$$\delta_\theta=(\theta,\ln\theta) \tag{14}$$

where θ represents the number of days from the monitoring date to the reference date divided by 100.

Thus, from an overall perspective, the initial input layer variables for the seepage pressure prediction model are represented as in (eq.15):

$$\delta=(h_u,\overline{h}_{1-3},\overline{h}_{4-7},P,\overline{P}_{1-3},\overline{P}_{4-7},\cos\frac{2\pi it}{365},\sin\frac{2\pi it}{365},\theta,\ln\theta) \tag{15}$$

2.2.4. Dimensionality reduction of input variables based on LightGBM technology

The variation in dam seepage pressure exhibits strong regularity, influenced by various factors such as rainfall, temperature, water level, and time-dependent factors at both feature and time levels (Zhang *et al.*, 2021). Section 2.3 has preliminarily determined the input layer of the seepage pressure prediction model. However, if all the above factors are input as feature parameters into the seepage pressure prediction model, it may lead to excessively high data dimensionality, redundant feature information, and a decrease in prediction accuracy. To address this, LightGBM (Light Gradient Boosting Machine) technology is introduced (Yuan *et al.* 2022) to perform dimensionality reduction on the input layer of the seepage pressure prediction, accurately extracting essential features and eliminating redundant factors.

2.2.5. The steps and workflow for implementing the prediction model are as follows:

This article utilizes MATLAB (2022a) to implement the construction of the ISSA-BiLSTM seepage pressure prediction model and the training and prediction of the seepage pressure model. The implementation steps are shown in Figure 2, and the specific process is as follows:

Step 1: Normalizing the monitoring data of dam seepage and influencing factors.

Step 2: Introducing the LightGBM technology to perform dimensionality reduction on the model's input layer, extracting important features, and partitioning the dataset into training and testing sets.

Step 3: Setting parameters such as the sparrow population size, maximum iteration count, etc. Initializing the sparrow population.

Step 4: Constructing the BiLSTM seepage prediction model, determining the optimization range of its parameters.

Step 5: Utilizing (eq. 5) and (eq. 6) to adaptively adjust the number of producers and scroungers. Employing a nonlinear Sine Cosine optimization to enhance the update mechanism of the improved producer population. Finally, using the nonlinear positive cosine algorithm to improve the scrounger positions.

Step 6: Checking if the maximum iteration count has been reached. If so, outputting the optimal parameters. If not, returning to Step 5 to continue the optimization process.

Step 7: Using the optimized parameters from ISSA to construct the seepage pressure prediction model with BiLSTM and performing predictions.

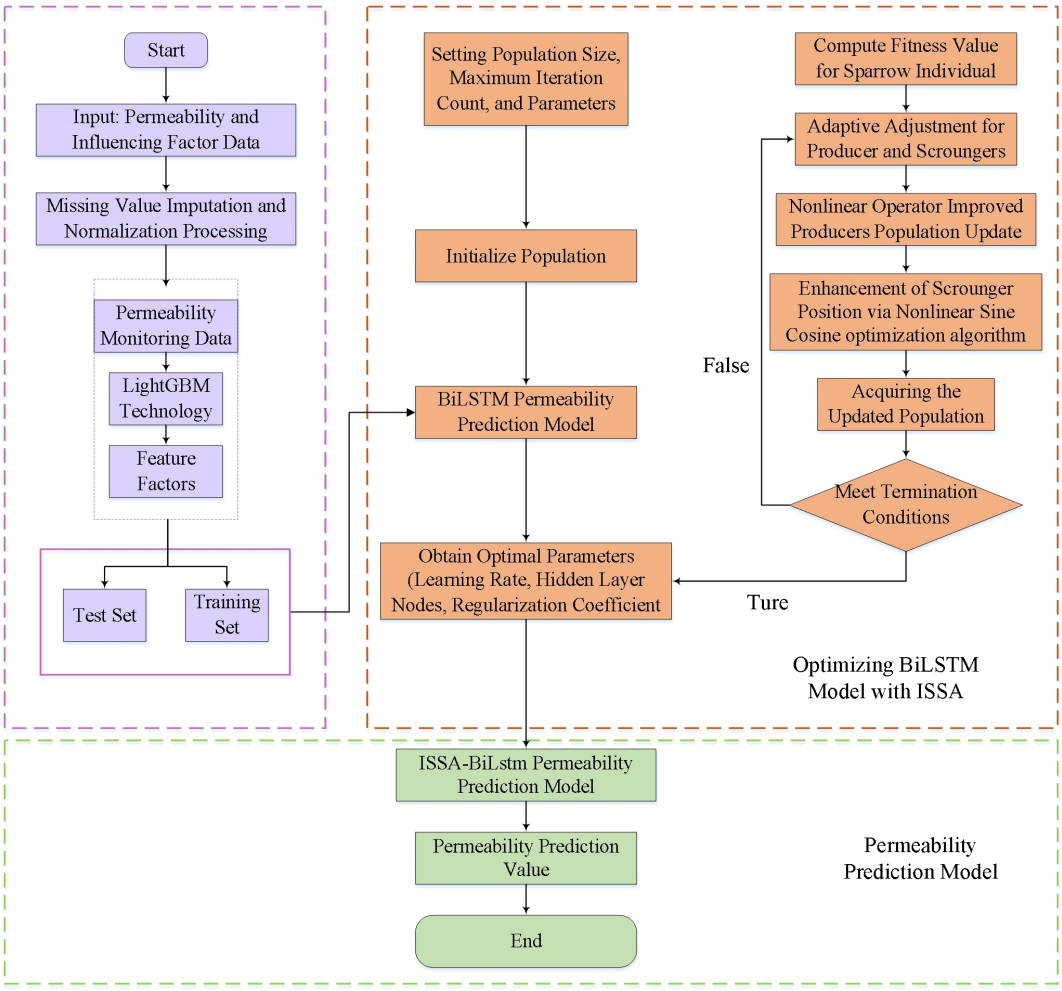


Fig. 2 Application steps of ISSA-BiLSTM seepage pressure prediction model

3. Results and Discussion

The panel-stacked rockfill dam is classified as a Grade I large-scale (1) project. As shown in Figure 3,the dam crest elevation is 413.8 m, the maximum dam height is 99.8 m, the crest width is 10 m, the crest length is 540.46 m, and an L-shaped reinforced concrete wave wall with a height of 4.0 m is installed on the upstream side of the dam crest. The reservoir's normal storage water level is 410 m³, with a storage capacity of 11.076 million m³. A comprehensive prototype dam monitoring system has been implemented at this reservoir, consisting of deformation monitoring, seepage monitoring, and environmental quantity (rainfall, reservoir water level) monitoring. For seepage monitoring, piezometers are embedded and installed, with a total of 4 piezometers arranged at the dam foundation, and data collection intervals range from 1 to 3 days.

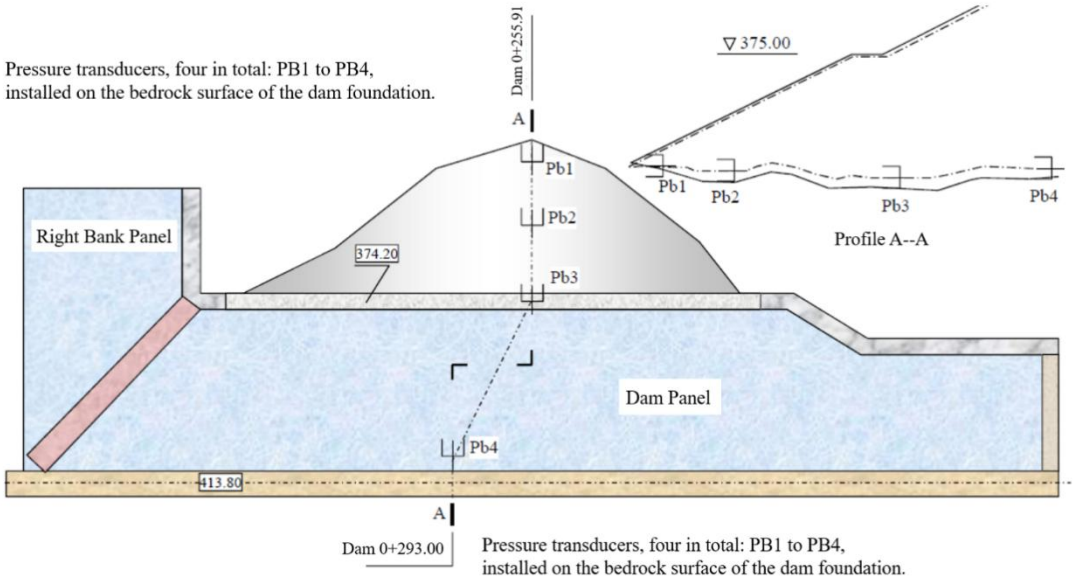


Fig. 3 Layout Diagram of Dam Base Seepage Pressure Monitoring

3.1. The simulation results and analysis

Using the actual measurement data of the panel-stacked rockfill dam from the past 10 years as a case study, an analysis was conducted using the ISSA-BiLSTM prediction model. The raw feature data for the input and output layers are presented in Table 1.

Table 1. Raw Feature Data for Input and Output Layers

	Parameter Names (Units)	*Minimum Value*	*Maximum Value*	*Mean Value*	*Median*
Input Layer	h_u (m)	385.51	409.70	405.38	406.08
	$\overline{h}_{1-3}$ (m)	385.52	409.50	405.35	405.95
	$\overline{h}_{4-7}$ (m)	385.52	409.49	405.33	405.83
	P (mm)	0	95.50	1.85	0
	$\overline{P}_{1-3}$ (mm)	0	74.10	1.83	0
	$\overline{P}_{4-7}$ (mm)	0	82.38	1.76	0
	$\delta_T 1$	-1.00	1.00	0.02	0.03
	$\delta_T 2$	-1.00	1.00	-0.03	-0.06
	$\ln\theta$	-0.84	3.47	2.56	2.81
	θ	0.43	32.36	16.54	16.55
Output Layer	Permeability Value (mm)	318.72	319.79	319.14	319.12

By employing LightGBM technology to perform dimensionality reduction on Table 1, individual importance scores were obtained for each feature factor. The importance scores for each feature factor are presented in Table 2. From the table, it is evident that the most influential feature factor on the seepage pressure value is the upstream water level on the same day, followed by the rainfall on the same day. To extract important features, we effectively eliminated redundant

information and parameters, and based on the importance scores of feature factors, we sequentially added them to the ISSA-BiLSTM seepage pressure prediction model, with the root mean square error (RMSE) as the evaluation metric.

Table 2. The importance ranking of 11 permeability influencing factors based on LightGBM technology.

Importance ranking.	*Permeability influencing factors.*	*Feature factor importance scores.*
1	h_u	55.64
2	P	47.38
3	$\overline{h}_{1-3}$	32.56
4	$\overline{P}_{1-3}$	30.23
5	$\ln\theta$	26.17
6	$\delta_T 1$	15.13
7	$\delta_T 2$	13.25
8	$\overline{P}_{4-7}$	5.15
9	$\overline{h}_{4-7}$	3.28
10	θ	2.69

*The higher the importance score of significant factors, the greater the significance of the corresponding factors, indicating a more pronounced impact on the percolation behavior.

Figure 4 depicts the variation of root mean square error (RMSE) with different numbers of significant factors. From the graph, it is observed that as the number of significant factors inputted into the model accumulates from 1 to 7, the RMSE exhibits an approximately linear decreasing trend. When significant factors ranked 8th to 10th in importance are introduced into the model, the RMSE shows little change. This suggests that the first seven significant factors exert a significant influence on prediction accuracy, whereas the 8th to 10th significant factors have minimal impact on predicting permeability values. In light of this observation, the model identifies upstream water level (h_u), daily precipitation (P), average upstream water level for the previous 1-3 days ($\overline{h}_{1-3}$), average precipitation for the previous 1-3 days ($\overline{P}_{1-3}$), time lag ($\ln\theta$), temperature component ($\delta_T 1$), and temperature component ($\delta_T 2$) as the input layer for permeability prediction.

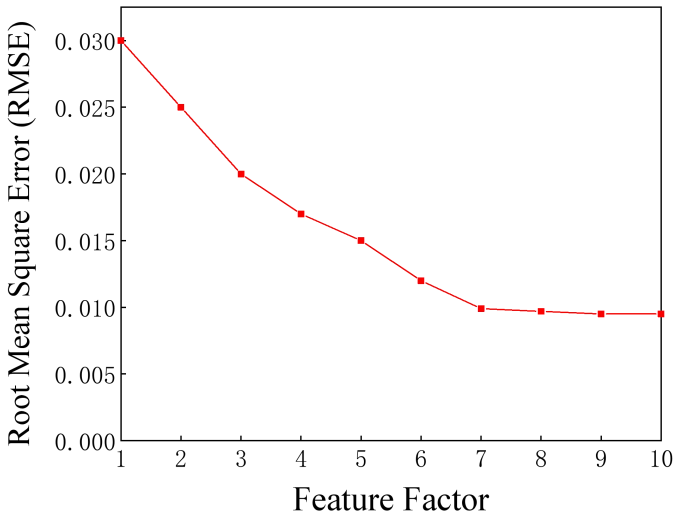


Fig. 4 The graph depicts the curve of the root mean square error as a function of the number of characteristic factors

The training samples of prolonged-duration permeability data are fitted using the ISSA-BiLSTM permeability prediction model. The simulation analysis environment involves an I7-7700U processor with a clock speed of 3.60 GHz, 16GB of RAM, and a Windows 10 operating system. Concurrently, the sparrow population is set at 50, the number of sample groups is set at 300, and the maximum iteration count is set at 1000. Figure 5 illustrates the fitting results of the ISSA-BiLSTM model. As depicted in Figure 5, the fitting curve of the current prediction aligns closely with the observed variations in permeability values. The coefficient of determination (R^2) is 0.987, indicating that the model effectively captures the temporal and feature information of long sequences of permeability values, resulting in a high fitting accuracy. The root mean square error (RMSE) is 0.0993, the mean absolute error (MAE) is 0.0725, and the mean absolute percentage error (MAPE) is 0.0227. These metrics collectively underscore the model's capability to accurately predict prolonged-duration permeability values.

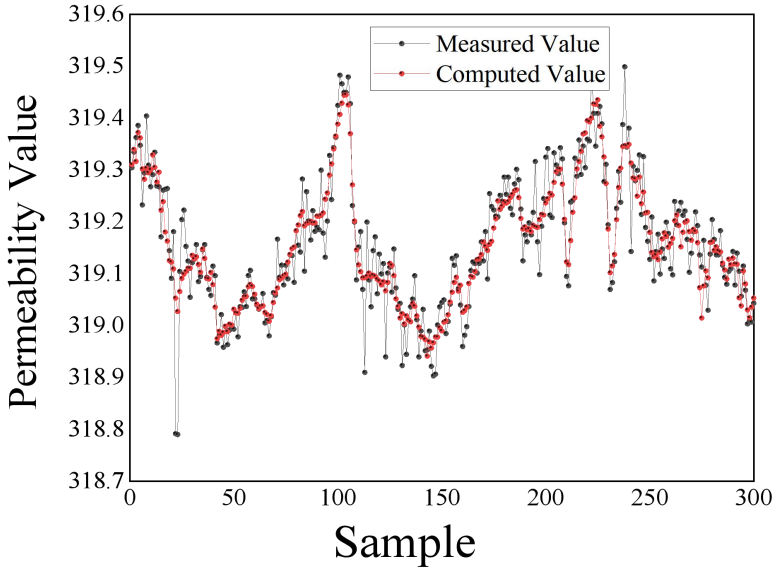
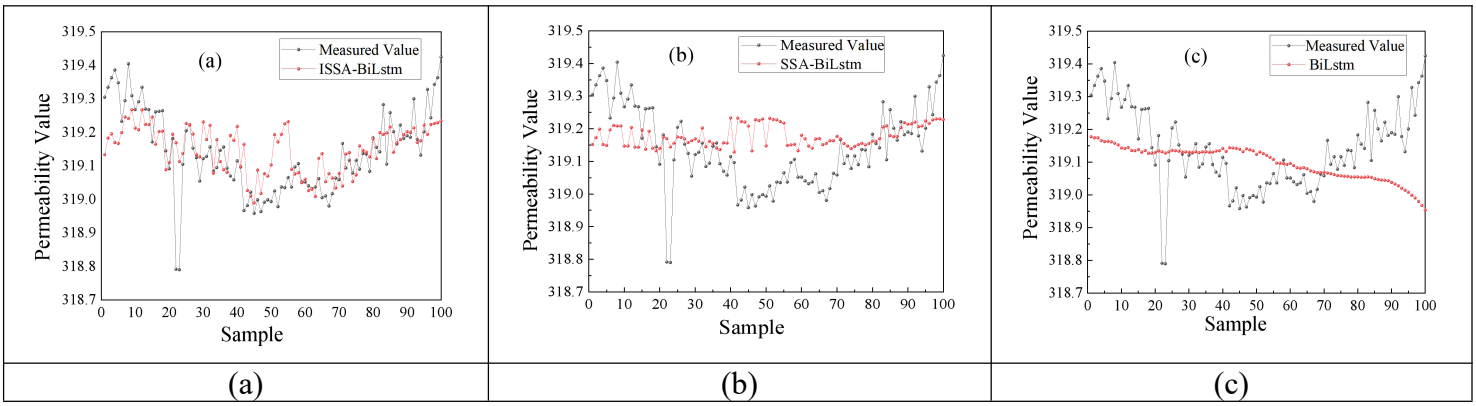


Fig. 5 The comparison graph of the actual and fitted values for the training set of the ISSA-BiLSTM model

3.2. Comparison analysis

To validate the superiority and permeability prediction accuracy of the ISSA-BiLSTM model, SSA-BiLSTM, and BiLSTM models were selected as controls. Figure 6 illustrates the comparative results of the predicted values and actual values for the ISSA-BiLSTM (Figure 6(a)), SSA-BiLSTM (Figure 6(b)), and BiLSTM(Figure 6(c)) models with 100 sample sets. As observed from the graph, the prediction curve of ISSA-BiLSTM aligns more closely with the observed curve, indicating a better match between predicted and actual values. Following in precision is SSA-BiLSTM, with BiLSTM exhibiting the lowest predictive accuracy.



(a) (b) (c)

Fig. 6 The comparison graph of predicted values and actual values for different permeability prediction models

Figure 7 displays the statistical results of RMSE, MAE, and MAPE for the ISSA-BiLSTM, SSA-BiLSTM, and BiLSTM models. As evident from the graph, the ISSA-BiLSTM model exhibits higher accuracy, with RMSE, MAE, and MAPE values of 0.10%, 0.07m, and 0.02m, respectively. Compared to the BiLSTM model, these values are reduced by 31.9%, 35.7%, and 35.7%, respectively. The SSA-BiLSTM model shows a smaller improvement, with reductions of only 10.8%, 8.4%, and 8.4%. This indicates that SSA contributes to some optimization of BiLSTM but is not sufficient. ISSA achieves the best optimization effect on BiLSTM, with a substantial decrease in errors, with all three types of errors reduced by over 30%.

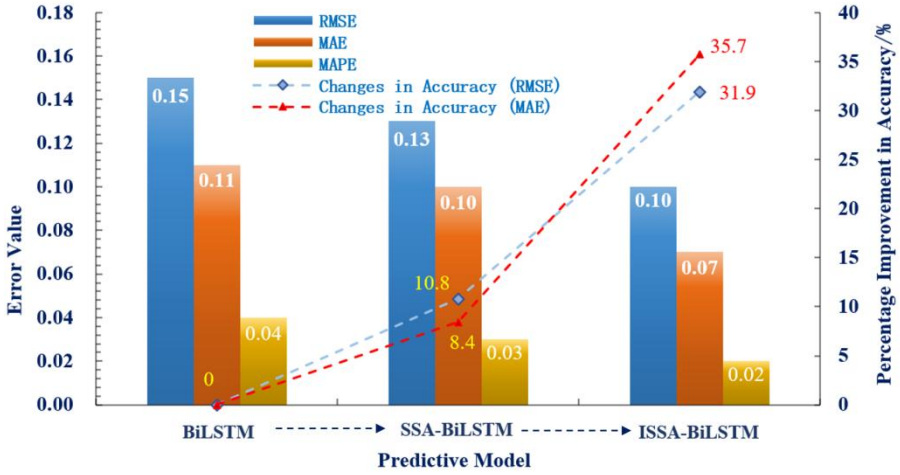


Fig. 7 The comparative graph of the statistical results for three types of errorsamong the three prediction models.

4. Conclusion

This study addresses the challenges of traditional seepage prediction models for dams, including susceptibility to local optima and inefficient prediction efficiency. Various strategies were introduced to enhance the baseline sparrow search algorithm, and the resulting ISSA was coupled with a Bidirectional Long Short-Term Memory (BiLSTM) neural network to form the ISSA-BiLSTM seepage prediction model. The model was applied to predict seepage in a panel dam, and simulation results indicate improved performance of the enhanced seepage prediction model. The key conclusions are as follows:

1. Improvement of Sparrow Search Algorithm (ISSA): To overcome limitations of Sparrow Search Algorithm (SSA), such as a tendency to be engaged in local optima and slow convergence, adaptive producer and scrounger ratio coefficients and a nonlinear Sine Cosine optimization were introduced to cooperatively improve SSA, resulting in ISSA. This enhanced algorithm demonstrated improved overall performance, providing effective support for accurate dam seepage prediction.

2. Feature Selection and Model Input Layer Establishment: The study initially identified seepage influencing factors and established the input layer from both temporal and feature perspectives. The LightGBM technique was then introduced to reduce the dimensionality of feature factors, selecting seven significant factors (upstream water level for the current day, daily precipitation, average upstream water level for the previous 1-3 days, average precipitation for the previous 1-3 days, time lag, temperature component 1, and temperature component 2) as the input layer for the seepage prediction model.

3. ISSA-BiLSTM Seepage Prediction Model: The ISSA algorithm was employed for parameter optimization in the BiLSTM model, resulting in the construction of the ISSA-BiLSTM seepage prediction model. When applied to predict seepage values in the selected engineering case, the model demonstrated a good fitting effect with an R^2 of 0.987. Moreover, the RMSE, MAE, and MAPE were 0.10%, 0.07m, and 0.02m, respectively. Compared to BiLSTM and SSA-BiLSTM, the errors were significantly reduced by over 30% and 20%, respectively.

In summary, this study demonstrates that the ISSA-BiLSTM seepage prediction model effectively overcomes the limitations of traditional models, offering valuable insights for accurate seepage prediction in dam engineering.

How to further improve the prediction accuracy of ISSA and its practical application effectiveness remains to be further investigated.

Author Contribution

C.SK provided the research idea, carried out the experimental design, and completed some of the writing and revisions, L.BY and X.CP organised the data and wrote some of the manuscript text and figures, Z.MD analysed the data and obtained the research funding and contacted journals for the submission of the manuscript, and Z.JY reviewed and revised the paper. All authors reviewed the manuscript.

Acknowledgements

This work was supported by the North China University of Water Resources and Electric Power. The authors also would like to acknowledge the support of The National Natural Science Foundation of China (51979169); Henan Province Innovation Talent Support Plan (24HASTIT017).

Conflict of Interests

The authors declare no conflicts of interest.

Data Availability

The data supporting the findings of this study are available upon request from the corresponding author.

References

Beiranvand B., Rajaee T.,2022,Application of artificial intelligence-based single and hybrid models in predicting seepage and pore water pressure of dams: A state-of-the-art review: *Advances in Engineering Software*, v. 173:,103268, https://doi.org/10.1016/j.advengsoft.2022.103268.

Chen, S., Gu, C., Lin, C., Wang, Y. and Hariri-Ardebili .M., 2020, Prediction monitoring and interpretation of dam leakage flow via adaptive kernel extreme learning machine: *Measurement*, v. 166, 108161, https://doi.org/10.1016/j.measurement.2020.108161.

Chen, X., Xu, Y., Guo, H., Hu, S., Gu, C., Hu, J., Qin, X. and Guo, J., 2024, Comprehensive evaluation of dam seepage safety combining deep learning with Dempster-Shafer evidence theory: *Measurement*, v. 26, 114172, https://doi.org/10.1016/j.measurement.2024.114172.

Gu, C. and Wu, Z., 2006, Theories and methods of dam and dam foundation safety monitoring and their applications: *Hohai University Press*.

Huang, Q., Liu, D. and Wei, X.,2021, Reasons for China owning largest number of water dams in the world: *Journal of Hydroelectric Engineering*, v. 40, p. 35-45, https://doi.org/10.11660/slfdxb.20210904.

Hochreite S. and Schmidhuber J., 1997, Long short-term memory: *Neural Computation*, v. 9, no. 8, p. 1735-1780, https://doi.org/10.1162/neco.1997.9.8.1735.

Ishfaque M., Dai, Q., Haq N., Jadoon K., Shahzad S. and Janjuhah H.,2022, Use of recurrent neural network with long short-term memory for seepage prediction at Tarbela Dam, KP, Pakistan: *Energies*, v. 15, no. 9, 3123, https://doi.org/10.3390/en15093123.

Li, A., Quan, L., Cui, G. and Xie, S.,2022, Sparrow Search Algorithm Combining Sine-Cosine and Cauchy Mutation: *Computer Engineering and Applications*, v.58, no.03, p.91-99, https://doi.org/10.3778/j.issn.1002-8331.2106-0148.

Li, H., Ma, G., Wang, F., Rong, W. and He, Y.,2021, Analysis of dam failure trend of China from 2000 to 2018 and improvement suggestions: *Hydro-Science and Engineering*, no. 5, p. 101-111, https://doi.org/10.12170/20201119001.

Li M., Xue B. , Gao J., Li B, Du M., Zhang S., Fang H., Wang F.,2023, Single-factor sensitivity analysis of earth dams with polymer cutoff wall under stress and seepage fields coupling: *Structures*, v. 57, 105145, https://doi.org/10.1016/j.istruc.2023.105145.

Li, Q., Ma, R., Hu Y., Huangfu, Z., Shen, Y., Zhou, S., Ma, J., An, Z. and Guo, G.,2022, A review of intelligent dam construction techniques: *Journal of Tsinghua University(Science and Technology*, v. 62, no. 08, p.1252-1269, https://doi.org/10.16511/j.cnki.qhdxxb.2022.25.018.

Li, L., Zhou, Y., Huang, H. and Luo, Q.,2023, Extreme Learning Machine Using Improved Gradient-Based Optimizer for Dam Seepage Prediction: *Arabian Journal for Science and Engineering* v. 48, p. 9693–9712, https://doi.org/10.1007/s13369-022-07300-8.

Li Y., Xu L., Ma Z., Ma B., Zhang J.,2024, Numerical simulation of the critical hydraulic gradient of granular soils at seepage failure by discrete element method and computational fluid dynamics: *Journal of Hydro-environment Research*, v. 53, p. 1-14, https://doi.org/10.1016/j.jher.2024.02.001.

Liu B, Cen W., Zheng C., Li D., Wang L.,2024, A combined optimization prediction model for earth-rock dam seepage pressure using multi-machine learning fusion with decomposition data-driven: *Expert Systems with Applications*, v. 242, 122798, https://doi.org/10.1016/j.eswa.2023.122798.

Liu, Y., Zheng, D., Wu, X., Chen, X., Georgakis C. and Qiu, J.,2023, Research on Prediction of Dam Seepage and Dual Analysis of Lag-Sensitivity of Influencing Factors Based on MIC Optimizing Random Forest Algorithm: *KSCE Journal of Civil Engineering* v. 27, no. 2 , p. 508-520, https://doi.org/10.1007/s12205-022-0611-6.

Liu, Z., Zhang, G., Li, W. and Hu, S.,2020, Prediction model of dam seepage based on MIC − BBO − SVM: *Journal of Safety Science and Technology*, v. 16, no 11, p. 12-18, https://doi.org/10.11731/j.issn.1673-193x.2020.11.002.

Miao, C., Shi, B., Zheng, X., and Zhang, C.,2019, Seepage Pressure Prediction of Earth -rock Dam Based on CM-AFSA-BP Neural Network: *Water Resources and Power*, v. 37, no. 02, p. 82-85.

Pang, Q., Wang, S., Gu, Y., Wang, Y. and Wu, Y.,2016, Application of Earth rock Dam Seepage Water Level Model Based on Hysteresis Effect Function: *Journal of Soil and Water Conservation*, v. 30, no. 2, p. 225-229, https://doi.org/10.13870/j.cnki.stbcxb.2016.02.039.

Qin, J., Wu, Y., Gu, Y.,2018, A seepage pressure prediction model based on the stepwise regression and wavelet neural network for the embankment dam: *Journal of Safety and Environment*, v. 18, no. 05, p. 1670-1674, https://doi.org/10.13637/j.issn.1009-6094.2018.05.004.

Qiu, J., Zheng, D. and Zhu, K.,2016, Seepage monitoring models study of earth-rock dams influenced by rainstorms: *Mathematical Problems in Engineering*, p. 1-11, http://dx.doi.org/10.1155/2016/1656738.

Shao X., Yu J., Li Z., Yang X., Sundén B.,2023, Energy-saving optimization of the parallel chillers system based on a multi-strategy improved sparrow search algorithm: *Heliyon*, v. 9, no.10, e21012, https://doi.org/10.1016/j.heliyon.2023.e21012.

Shi Y., Zhao C., Peng Z., Yang H., He J.,2018, Analysis of the lag effect of embankment dam seepage based on delayed mutual information: *Engineering Geology*, v. 234, p. 132-137, https://doi.org/10.1016/j.enggeo.2018.01.009.

Song J., Yuan S., Xu Z., Li X.,2023, Fast inversion method for seepage parameters of core earth-rock dam based on LHS-SSA-MKELM fusion surrogate model: *Structures*, v, 55, p. 160-168, https://doi.org/10.1016/j.istruc.2023.06.049.

Wang, A., Yang, X. and Guo, D.,2022, The Application of Seepage Flow Prediction in Nuer Dam Based on the Grey Self-Memory Model: *Geofluids,* v. 2022,6211685, https://doi.org/10.1155/2022/6211685.

Wang S., Xu Y., Gu C., Bao T.,2018, Monitoring models for base flow effect and daily variation of dam seepage elements considering time lag effect: *Water Science and Engineering*, v. 11 no. 04, p. 344-354, https://doi.org/10.1016/j.wse.2018.12.004.

Wang, X., Li, K., Zang, Z., Yu, H., Kong, L. and Chen, W.,2022, Coupled ALO-LSTM and feature attention mechanism prediction model for seepage pressure of earth-rock dam: *Journal of Hydraulic Engineering*, v. 53, no. 04, p. 403-412, https://doi.org/10.13243/j.cnki.slxb.20210936.

Wang, X., Zhu, K., Yu, H., Cai, Z. and Wang, C.,2023, Combinatorial deep learning prediction model for dam seepage pressure considering spatiotemporal correlation: *Journal of Hydroelectric Engineering*, v. 42, no. 11, p. 78-91, https://doi.org/10.11660/slfdxb.20231108.

Wen L., Li Y., Zhao W., Cao W., Zhang H.,2023, Predicting the deformation behaviour of concrete face rockfill dams by combining support vector machine and AdaBoost ensemble algorithm: *Computers and Geotechnics*, v. 161, 105611, https://doi.org/10.1016/j.compgeo.2023.105611.

Xiong, L., Miao, Y., Fan, X., and Yao, Y.,2022, Energy Saving Control of Central Air Conditioning System Based on an Improved SSA: *Journal of Shanghai Jiao Tong University*, p. 1-10, https://doi.org/10.16183/j.cnki.jsjtu.2022.018.

Xue, J., and Shen, B.,2020, A novel swarm intelligence optimization approach: sparrow search algorithm: *Systems Science & Control Engineering*, v. 8, no. 1, p. 22-34, https://doi.org/10.1080/21642583.2019.1708830.

Yang J., Gao S., Zhao X., Li G., Gao Z., 2024, Enhanced sparrow search algorithm based on improved game predatory mechanism and its application: *Digital Signal Processing*, v. 145, 104310, https://doi.org/10.1016/j.dsp.2023.104310.

Zhou Y., Li C., Pang R., Li Y., Xu Y., Chen J., 2024, A new approach for seepage parameter inversion of earth–rockfill dams based on an improved sparrow search algorithm: *Computers and Geotechnics*, v. 167, 106036, https://doi.org/10.1016/j.compgeo.2023.106036.

Yu, H., Bao, T. and Xue, L.,2010, Numerical simulation of the hysteretic effects of rainfall: *Journal of Hydroelectric Engineering*, v. 29, no. 04, p. 200-206.

Yuan, J., Lin, D., Mei, A. and Wei, Z.,2022, Study on key technology of identification of mine water inrush source by PSO-LightGBM: *Water Supply*, v. 22, no. 10, p. 7416-7429 , https://doi.org/10.2166/ws.2022.323.

Yue, M., Chen, X. and Li, J.,2020, Seepage Prediction of Concrete Dams Based on CNN-LSTM: *Water Resources and Power*, v. 38, no, 09, p. 75-78.

Zhang, H., Song, Z., Peng, P., Sun, Y., Ding, Z. and Zhang, X.,2021, Research on seepage field of concrete dam foundation based on artificial neural network: *Alexandria Engineering Journal*, v. 60, no. 1, p.1-14 , https://doi.org/10.1016/j.aej.2020.03.041.

Zhang, K, Gu C., Zhu, Y., Chen, S., Dai, B., Li Y., and Shu, X.,2021, A novel seepage behavior prediction and lag process identification method for concrete dams using HGWO-XGBoost model: *IEEE Access*, v. 9, p. 23311-23325, https://doi.org/10.1109/ACCESS.2021.3056588.

Zhang, X., Chen, X. and Li, J.,2020, Improving dam seepage prediction using back-propagation neural network and gen etic algorithm: *Mathematical Problems in Engineering*, v. 2020, no. 11, p. 1-8, https://doi.org/10.1155/2020/1404295.