

## Article

# A Novel Long Short-Term Memory Seq2Seq Model with Chaos-Based Optimization and Attention Mechanism for Enhanced Dam Deformation Prediction

Lei Wang, Jiajun Wang \* , Dawei Tong and Xiaoling Wang

State Key Laboratory of Hydraulic Engineering Intelligent Construction and Operation, Tianjin University, Tianjin 300354, China

\* Correspondence: jiajun\_2014\_bs@tju.edu.cn

**Abstract:** The accurate prediction of dam deformation is essential for ensuring safe and efficient dam operation and risk management. However, the nonlinear relationships between deformation and time-varying environmental factors pose significant challenges, often limiting the accuracy of conventional and deep learning models. To address these issues, this study aimed to improve the predictive accuracy and interpretability in dam deformation modeling by proposing a novel LSTM seq2seq model that integrates a chaos-based arithmetic optimization algorithm (AOA) and an attention mechanism. The AOA optimizes the model's learnable parameters by utilizing the distribution patterns of four mathematical operators, further enhanced by logistic and cubic mappings, to avoid local optima. The attention mechanism, placed between the encoder and decoder networks, dynamically quantifies the impact of influencing factors on deformation, enabling the model to focus on the most relevant information. This approach was applied to an earth-rock dam, achieving superior predictive performance with RMSE, MAE, and MAPE values of 0.695 mm, 0.301 mm, and 0.156%, respectively, outperforming conventional machine learning and deep learning models. The attention weights provide insights into the contributions of each factor, enhancing interpretability. This model holds potential for real-time deformation monitoring and predictive maintenance, contributing to the safety and resilience of dam infrastructure.



**Citation:** Wang, L.; Wang, J.; Tong, D.; Wang, X. A Novel Long Short-Term Memory Seq2Seq Model with Chaos-Based Optimization and Attention Mechanism for Enhanced Dam Deformation Prediction.

*Buildings* **2024**, *14*, 3675. <https://doi.org/10.3390/buildings14113675>

Academic Editor: Hugo Rodrigues

Received: 16 October 2024

Revised: 15 November 2024

Accepted: 18 November 2024

Published: 19 November 2024

**Keywords:** dam deformation prediction; long short-term memory sequence-to-sequence model; attention mechanism; arithmetic optimization algorithm; chaotic optimization

## 1. Introduction

Health monitoring is crucial for the safe operation of a dam, with the structural performance as well as surrounding changes measured by monitors installed on the dam such as environmental factors, dam deformation, water seepage, etc. [1]. In engineering practice, dam deformation has been used as the most effective monitoring indicator of dam safety [2], and deformation prediction models have been proposed and extensively studied [3–6] in dam health monitoring, in order to accurately predict the deformation to assess the potential risk based on historical monitoring data.

Statistical models have been used traditionally to predict dam deformation [2], which take influencing factors (e.g., water pressure, temperature and time) as independent variables, and performance indicators (e.g., dam deformation and water seepage) as dependent variables. One of the commonly used statistical models is the hydrostatic-season-time (HST) model and its variants [7–10]. Despite the broad applications of statistical models, their simple form of linearly superimposing the components of water pressure, temperature, and time, without considering the nonlinear relationship between these components and the deformation [3], results in issues of accuracy, robustness, and generalization. In recent years, machine learning models have been applied in prediction tasks for dams such as the back



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

propagation (BP) neural network [11,12], support vector machine (SVM) [13–16], and extreme learning machine (ELM) [17–19]. Machine learning models have been demonstrated to achieve a higher prediction accuracy than statistical models as they have the capability of approximating the nonlinear relationship between influencing factors and monitoring data. Most traditional machine learning models primarily focus on capturing direct regression relationships between the input and output, which often leads to overlooking the inherent long-term temporal dependencies within the deformation data that refer to the influence that changes in certain variables at a given time can have on future time points. In complex dam deformation monitoring scenarios, neglecting these long-term temporal dependencies can limit the model's predictive accuracy, as such algorithms may fail to fully capture the underlying features in the deformation data [3].

With the availability of large datasets and GPU-accelerated computational performance, deep learning, a branch of machine learning, has been widely applied and studied [20], achieving breakthrough results in fields such as healthcare [21,22], education [23], economics [24], renewable energy [25–28], computer technology [29], and industry [30,31]. In the field of dam deformation prediction, models like recurrent neural networks (RNNs) [4], convolutional neural networks (CNNs) [32], graph neural networks (GNNs) [33], and autoencoders [34] have been extensively studied. Among these methods, the long short-term memory network (LSTM), a variant of RNNs [35], has been proposed to solve the long-term dependence issue of time-series data [36], and deformation prediction models based on LSTM have been developed ever since. Li et al. [37] built an LSTM prediction model for the decomposed deformation data, and Liu et al. [38] further used principal component analysis and the moving average method to reduce the dimensionality of the input variables of the LSTM model. To consider the global and local connections between the deformation and the long sequences of influencing factors, Yang et al. [39] added an attention mechanism ahead of the output layer of LSTM to improve the prediction performance. For the sequence data, the sequence-to-sequence (Seq2Seq) structure, a variant of RNN [40], is commonly used when the input and output lengths differ. This approach is especially beneficial for dam deformation prediction, where input factors such as water level, temperature, and time are sequential, but the output (deformation prediction) may vary in length. The Seq2Seq structure effectively captures the complex relationships between these factors and deformation, addressing the challenge of differing input–output lengths and enhancing prediction accuracy. Additionally, incorporating an attention mechanism in the Seq2Seq structure is crucial for both prediction accuracy and model interpretability in dam deformation prediction. It enhances accuracy by allowing the model to dynamically focus on the most relevant parts of the input sequence, adjusting attention to specific time steps or factors that are critical for each prediction. This ability to emphasize key features improves the model's capacity to capture essential patterns and dependencies in the deformation data. Moreover, the attention mechanism enhances interpretability by assigning weights to input features, making it possible to visualize and understand which factors most significantly impact the prediction [3,6].

Although LSTM and its related models have been successfully applied in dam deformation prediction, some challenging issues are still to be tackled that may otherwise affect the prediction accuracy. The learnable parameters (i.e., weights and biases of the model) are trained with gradient-based algorithms by default in most available prediction models, which would inevitably fall into the local optima when the data exhibit high nonlinearity. To mitigate this issue, researchers have attempted to apply the global optimization approach, especially meta-heuristic algorithms (e.g., genetic algorithm, particle swarm optimization algorithm, and whale optimization algorithm) to obtain globally optimized learnable parameters [41,42], which has been demonstrated to significantly improve the prediction accuracy. Compared with other meta-heuristic algorithms, the arithmetic optimization algorithm (AOA) is more computationally efficient, has better convergence performance [43], and has been applied in learning models with real-world data. Li et al. [44] optimized the hyperparameters of the support vector regression model using AOA to accurately predict

tunnel crown displacement caused by blasting excavation, and Xu et al. [45] used AOA to optimize the initial weights and thresholds of the BP neural networks, improving the accuracy of cluster failure prediction.

Apart from the optimization of learnable parameters, the importance of the influencing factors in the dam deformation is not straightforward to quantify. Attempts have been made to use the attention mechanism to dynamically quantify the relative importance of each influencing factor in the prediction [6]. Nonetheless, the influence of the spatial correlation of monitors on the dam has not been studied, while the monitoring data show that the long-term deformation at different locations on the dam are highly correlated. Hence, we also aimed to analyze the influence of the spatial correlation, in addition to other influencing factors, in the prediction model with an attention mechanism.

In this paper, we propose an LSTM-seq2seq model equipped with an attention mechanism and AOA for dam deformation prediction with high accuracy. The novelties of this study are summarized as follows:

(1) Training deep learning models using gradient-based algorithms can lead to the learnable parameters likely trapped in local optima, thereby causing inadequate accuracy for dam deformation prediction. To mitigate this challenge, we propose a chaos-based AOA to optimize the parameters further, thus greatly improving the prediction accuracy of dam deformation.

(2) In order to explain the effect of different factors on the prediction of deformation, an attention mechanism was integrated in the model, which was located between the encoder and decoder networks to quantify the dynamic impact of time, water level, and temperature on the deformation prediction.

(3) The model proposed in this paper was applied to predict deformation in an earth-rock dam, achieving a high prediction accuracy that outperformed conventional machine learning and deep learning models. The results illustrate the innovation and superiority of the proposed model.

For the notational convenience, the proposed model is referred to as LSTM-seq2seq-AA in the following. An LSTM-seq2seq model was first established using LSTM cells in the seq2seq structure, and then the attention mechanism was introduced in the seq2seq structure to form the LSTM-seq2seq-A model. The final LSTM-seq2seq-AA model was built using the AOA to train learnable parameters. The rest of the paper is organized as follows. Section 2 presents the research framework of this study. A brief description of the HST statistical model and the details of the LSTM-seq2seq-AA model are introduced in Section 3. The case study and results are given in Section 4. Finally, our conclusions are shown in Section 5.

## 2. Research Framework

Figure 1 shows the framework of the LSTM-seq2seq-AA model. The LSTM-seq2seq-AA model can be divided into a temporal feature prediction module and an influencing factor prediction module, and the final prediction is obtained by merging the results of the two modules. For the temporal feature prediction module, the input of influencing factors is in a sequence of time steps, and the influence of key time steps can be highlighted with the attention mechanism, in order to deal with the long-term dependency and improve the prediction accuracy. For the influencing factor prediction module, the input of influencing factors is sorted by category, and the contribution of each influencing factor to the predicted deformation is also quantified with the attention mechanism, offering certain interpretability of the model. Finally, the AOA enhanced with the chaos theory is used for optimizing the learnable parameters (weights and biases) of the model, attempting to obtain a better optimized model.

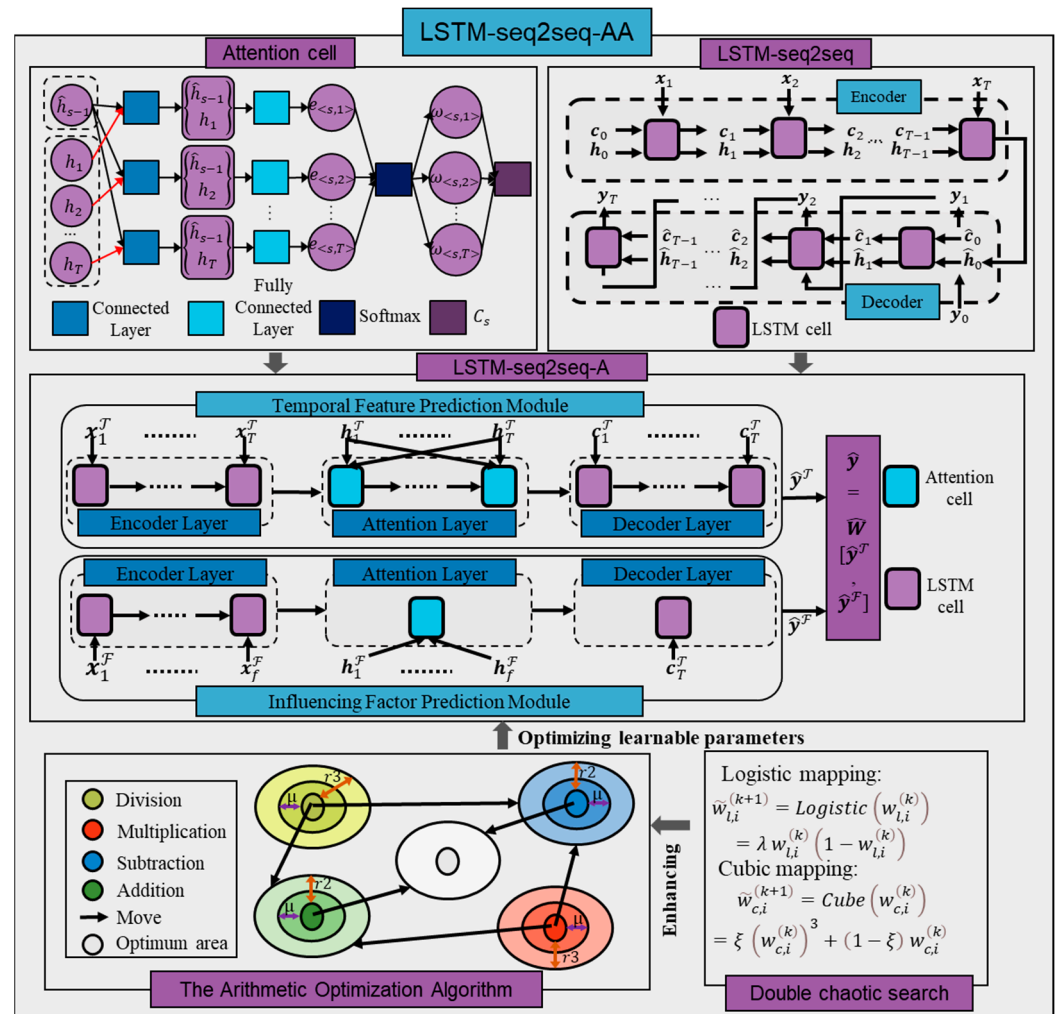


Figure 1. LSTM-seq2seq-AA framework.

### 3. Methodologies

#### 3.1. Brief Description of HST Statistical Model of Earth-Rock Dam Deformation

The HST statistical model gives the statistical relationship between deformation and environmental quantities, which we rely on to determine the input of the deformation prediction model. In the HST statistical model, dam deformation is computed with components induced by hydrostatic pressure, temperature, and time [46], which can be expressed as

$$y_t = y_p + y_\tau + y_\eta + y_0, \tag{1}$$

where  $y_t$  represents the deformation at a monitoring point at time  $t$ ,  $y_p$  denotes the hydrostatic pressure component,  $y_\tau$  denotes the deformation induced by temperature change,  $y_\eta$  indicates the time-dependent aging effect, and  $y_0$  is a constant term.

The hydrostatic pressure component represents the deformation under the reservoir hydrostatic pressure, which is given by the polynomial of the upstream water level as [15]

$$y_p = \sum_{i=1}^3 \alpha_i p^i, \tag{2}$$

where  $p$  represents the upstream water level and  $\alpha_i$  ( $i = 1, 2, 3$ ) denotes the corresponding coefficients.

Temperature change has a relatively small impact on earth-rock dams except in alpine regions [47]. The resulting dam deformation stems from periodic changes in air and water temperature, which is given by [48],

$$y_\tau = \sum_{i=1}^2 \beta_i \cos\left(\frac{2\pi id}{365}\right), \quad (3)$$

where  $d$  indicates the cumulative number of days from the starting date, and  $\beta_i$  ( $i = 1, 2$ ) are the corresponding coefficients.

The deformation of earth-rock dams is mainly caused by soil consolidation. It changes rapidly during the initial period within one to two years and then stabilizes [31], which is indicated by the time component expressed as follows [6],

$$y_\eta = \gamma_1 \eta + \gamma_2 \ln \eta, \quad (4)$$

where  $\eta = t/100$ , and  $\gamma_1$  and  $\gamma_2$  are the corresponding coefficients.

Finally, the HST statistical model (1) of the earth-rock dam deformation can be written as

$$y_t = \sum_{i=1}^3 \alpha_i p^i + \sum_{i=1}^2 \beta_i \cos\left(\frac{2\pi id}{365}\right) + \gamma_1 \eta + \gamma_2 \ln \eta + y_0, \quad (5)$$

Therefore, for the deformation prediction model, the parameters of the statistical model in Equation (5) (i.e.,  $\{p, p^2, p^3, \cos\left(\frac{2\pi id}{365}\right), \cos\left(\frac{4\pi id}{365}\right), \eta, \ln \eta\}$ ) are taken as the input factors, and the complex nonlinear relationship of these factors are computed with the prediction model.

### 3.2. The Proposed Methods

The proposed LSTM-seq2seq-AA model was built upon the LSTM sequence-to-sequence model with an attention mechanism (LSTM-seq2seq-A). The LSTM-seq2seq-A model was divided into a temporal feature prediction module and an influencing factor prediction module, and the final prediction was obtained by merging the results of the two modules. For the temporal feature prediction module, the input of influencing factors was in a sequence of time steps, and the influence of time steps can be highlighted with the attention mechanism. For the influencing factor prediction module, the input of influencing factors was sorted by category, and the contribution of each influencing factor to the predicted deformation was also quantified with the attention mechanism. The AOA enhanced with the chaotic optimization was used to train the learnable parameters (i.e., weights and biases of the model) to improve the prediction accuracy.

#### 3.2.1. LSTM Sequence-to-Sequence Model (LSTM-seq2seq)

Figure 2 shows the seq2seq structure consisting of an encoder and a decoder. In the encoder neural network, the input sequence  $\{x_1, x_2, \dots, x_T\}$  with the number of time steps  $T$  is read one time step at a time, and the last hidden state  $h_T$  produces a high-dimensional vector  $D$  that is encoded to represent the information of the input sequence. The decoder neural network structure takes the vector  $D$  as the input to obtain the output sequence  $\{y_1, y_2, \dots, y_T\}$  through a directed loop. The computation involved in the seq2seq structure is as follows,

$$h_t = \psi(x_t, h_{t-1}), \quad (6)$$

$$D = \phi(h_1, \dots, h_t), \quad (7)$$

$$\hat{h}_t = \theta(y_{t-1}, \hat{h}_{t-1}, D), \quad (8)$$

where  $h_t$  and  $\hat{h}_t$  denote the hidden state in the encoder and decoder at the time step  $t$ , respectively;  $\psi$ ,  $\phi$ , and  $\theta$  represent nonlinear activation functions.

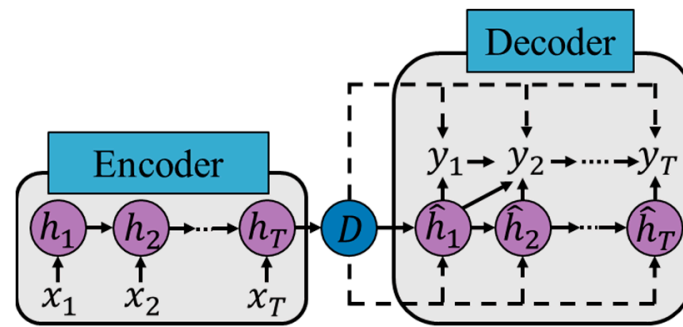


Figure 2. Seq2seq structure consisting of an encoder and decoder.

The LSTM-seq2seq model is illustrated in Figure 3. Assuming that  $x^{m \times n}$  are the influencing factor sequence data and  $y^{m \times 1}$  are the monitoring deformation data, with  $m$  the number of samples and  $n$  the number of influencing factors (e.g., water level, temperature, time, etc.), the two data sequences are first rearranged into  $x^{(m/T) \times n \times T}$  and  $y^{(m/T) \times 1 \times T}$ , respectively, by time steps  $T$ . After normalization, they are taken as the input of the LSTM-seq2seq structure. In the encoder, the input sequence data  $x$  and  $y$  are encoded with hidden layer outputs of all of the LSTM networks corresponding to time steps, which contains all of the sequence features of the input. The output of the last hidden layer  $h_T$  is then fed into the decoder. The output of each time step in the decoder is given by

$$y_t = \sigma(\hat{h}_t), \quad (9)$$

and they are combined in a sequence to form the final fitting result, and the deformation  $y$  is finally obtained after denormalization.

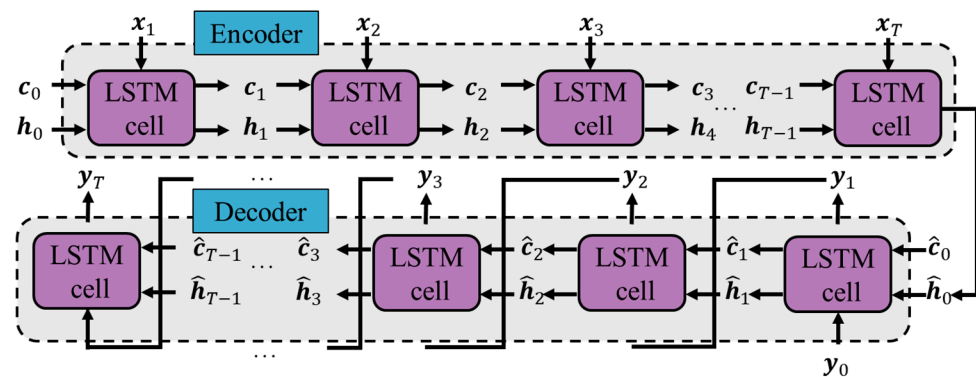


Figure 3. LSTM-seq2seq model with sequence data.

### 3.2.2. Optimization of Learnable Parameters with Chaos-Based AOA

Deep learning models typically use the gradient-based algorithm to train the learnable parameters, and the existence of local minima is one of the key issues that influence the prediction accuracy. Here, we adopted the AOA [33] to further optimize the learnable parameters of the trained LSTM-seq2seq model. The AOA is a meta-heuristic algorithm that takes advantage of the distribution behaviors of the four main mathematical operators (i.e., addition, subtraction, multiplication, and division) to perform optimization in the search space. We also applied logistic mapping and cubic mapping in chaotic optimization to better avoid the local optima in the optimization process. In this study, the fitness function of the optimization was the mean absolute error (MAE) of the prediction model, and the initial candidate solutions were 10 sets of learnable parameters of LSTM-seq2seq obtained with the gradient-based training.

The solution update in AOA involves two stages (i.e., the exploration stage and the exploitation stage). A mathematical optimization accelerator is defined as follows to determine the stage at the  $k$ -th iteration, in other words,

$$A^{(k)} = A_{\min} + k \times \frac{A_{\max} - A_{\min}}{M}, \quad (10)$$

where  $A_{\min}$  and  $A_{\max}$  are the prescribed minimum and maximum values of the accelerator, respectively, and  $M$  is the maximum number of iterations.

Consider a random number  $r_1 \in (0, 1)$ , where the exploration stage is activated if  $r_1 \geq A^{(k)}$ . The division operator and the multiplication operator are used as the search strategy in this stage, in other words,

$$w_{i,j}^{(k)} = \begin{cases} w_j^* \div (P^{(k)} + \epsilon) \times [(w_j^U - w_j^L) \times \mu + w_j^U], & r_2 < 0.5 \\ w_j^* \times P^{(k)} \times [(w_j^U - w_j^L) \times \mu + w_j^L], & r_2 \geq 0.5 \end{cases}, \quad (11)$$

where  $r_2$  is another random number between 0 and 1, which is used to switch the exploration between multiplication and division operators;  $w_{i,j}^{(k)}$  represents the  $j$ -th position of the  $i$ -th solution in the  $k$ -th iteration;  $w_j^*$  is the  $j$ -th position of the optimal solution currently obtained;  $\epsilon$  is the floating-point relative accuracy;  $w_j^U$  and  $w_j^L$  represent the upper and lower limits of the  $j$ -th position, respectively;  $\mu$  is a control parameter to adjust the search process, which is fixed to 0.5 in our model;  $P^{(k)}$  is the math optimizer probability defined as

$$P^{(k)} = 1 - \frac{k^\rho}{M^\rho}, \quad (12)$$

where  $\rho$  is a sensitive parameter that controls the exploitation accuracy, which was fixed to 5 in our model.

The exploitation stage is activated if  $r_1 < A^{(k)}$ , and the search strategy uses the addition operator and the subtraction operator, so that the search target can be approached with a low dispersion, in other words,

$$w_{i,j}^{(k)} = \begin{cases} w_j^* - P^{(k)} \times [(w_j^U - w_j^L) \times \mu + w_j^L], & r_3 < 0.5 \\ w_j^* + P^{(k)} \times [(w_j^U - w_j^L) \times \mu + w_j^L], & r_3 \geq 0.5 \end{cases}, \quad (13)$$

where  $r_3$  represents a third random number between 0 and 1, which is used to switch between the addition and the subtraction search.

The AOA optimization of learnable parameters terminates when the maximum number of iterations  $M$  is reached, and the optimized parameters are returned. Though the AOA aims to avoid the local optima, the fitness function may still converge to a local optimum in complex optimization problems. To further improve the optimization, when the values of the fitness function in consecutive iterations are less than a threshold (e.g.,  $10^{-5}$ ), double chaotic mappings is introduced to try to remove a local optimum. In the following, the local search based on chaotic optimization is discussed.

To train a deformation prediction model that involves quite a few local optima, the AOA may terminate without searching further in the solution space. A chaotic optimization mechanism is integrated into the local search strategies of the AOA to improve the optimization result, in which two completely different chaotic mappings (i.e., logistic mapping and cubic mapping) are used for independent searches. Studies have demonstrated that with the same initial value and the same number of iterations, the chaotic variable values of

the two mappings are different in most cases [49,50], which helps the optimization process converge to a better solution.

Suppose the population size is  $p$ , with each individual an  $n$ -dimensional vector. Consider an optimal individual  $w^{(k)}$  at the iteration  $k$ , which is transformed to two chaotic values  $w_{l,i}^{(k)}$  and  $w_{c,i}^{(k)}$ , respectively, as the input of the logistic mapping and the cubic mapping, in other words,

$$w_{l,i}^{(k)} = \frac{w_i^{(k)} - l_i}{u_i - l_i}, \quad (14)$$

$$w_{c,i}^{(k)} = \frac{2(w_i^{(k)} - l_i)}{u_i - l_i} - 1, \quad (15)$$

with  $u$  and  $l$  the upper and lower bounds of the search domain, respectively, and the index  $i = 1, 2, \dots, n$ . The logistic and cubic mappings are then defined as follows,

$$\tilde{w}_{l,i}^{(k+1)} = \text{Logistic}(w_{l,i}^{(k)}) = \lambda w_{l,i}^{(k)} (1 - w_{l,i}^{(k)}), \quad (16)$$

$$\tilde{w}_{c,i}^{(k+1)} = \text{Cubic}(w_{c,i}^{(k)}) = \zeta (w_{c,i}^{(k)})^3 + (1 - \zeta) w_{c,i}^{(k)}, \quad (17)$$

According to References [49,51], the logistic mapping is considered as chaotic when  $\lambda = 4$  and  $w_{l,i}^{(k)} \in (0, 1)$ , and the cubic mapping is chaotic when  $\zeta \in [3.3, 4]$  and  $w_{c,i}^{(k)} \in (-1, 1)$ . The obtained chaotic variables are then transformed back into the search space, in other words,

$$w_{l,i}^{(k+1)} = \tilde{w}_{l,i}^{(k+1)} (u_i - l_i) + l_i, \quad (18)$$

$$w_{c,i}^{(k+1)} = \frac{(\tilde{w}_{c,i}^{(k+1)} + 1)(u_i - l_i)}{2} + l_i, \quad (19)$$

and their corresponding fitness function values  $g(w_{l,i}^{(k+1)})$  and  $g(w_{c,i}^{(k+1)})$  are compared with the fitness function value  $g(w^{(k)})$  in the previous iteration. The optimal solution at iteration  $k + 1$  takes the best among the three candidates, which has the smallest fitness function value.

The chaotic mapping iterations continue until the difference between the two chaotic values  $w_{l,i}^{(n)}$  and  $w_{c,i}^{(n)}$  is less than a prescribed value, that is,

$$\|w_{l,i}^{(n)} - w_{c,i}^{(n)}\| < \nu \|u - l\|, \quad (20)$$

$$l_i^v = \min(w_{l,i}^{(n)}, w_{c,i}^{(n)}) - \zeta \nu \|w_{l,i}^{(n)} - w_{c,i}^{(n)}\|, \quad (21)$$

$$l_i^v = \max(w_{l,i}^{(n)}, w_{c,i}^{(n)}) + \zeta \nu \|w_{l,i}^{(n)} - w_{c,i}^{(n)}\|, \quad (22)$$

with  $\zeta \in [1, 2]$ . To avoid the new search step outside the bounds, the following treatment is also considered,

$$\begin{cases} l_i^v = l_i, & l_i^v < l_i \\ u_i^v = u_i, & u_i^v > u_i \end{cases} \quad (23)$$

to update the bounds of the search domain (i.e.,  $l = l^v$ ,  $u = u^v$ ). The optimization process of Equations (14)–(19) is repeated until Equation (20) is satisfied, and the optimal solution is obtained. The entire process of the optimization of learnable parameters with the chaos-based AOA is illustrated in Figure 4.



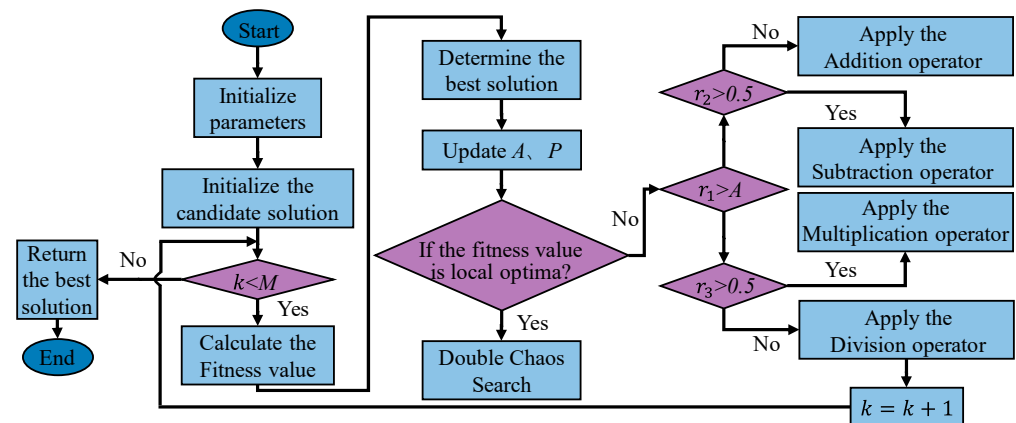


Figure 4. Optimization of learnable parameters with chaos-based AOA.

### 3.2.3. Quantifying Dynamic Contributions of Influencing Factors by Embedding Attention Mechanism

With the LSTM-seq2seq model only, the contributions of different influencing factors in the input sequence to the output cannot be quantified, and the effect of time step lengths of the input sequence on the output is also not clear. Hence, the attention mechanism was adopted for the interpretability of the LSTM-seq2seq model, in which a temporal feature prediction module and an influencing factor prediction module have been established. The output of the model contains information in both the time and feature dimensions. By dynamically measuring the contribution of each factor to the output, key features are figured out automatically during the training process.

In the LSTM-seq2seq structure introduced in Section 3.2.1, the contribution of each LSTM cell in the encoder network to the output of each LSTM cell in the decoder network is assumed to be the same. As a matter of fact, the influencing factors at different time steps have a different influence on the prediction, and different influencing factors also contribute differently to the prediction. The different contributions can be related to the adaptive weights of influencing factors in the attention mechanism.

The schematic diagram of the attention mechanism is shown in Figure 5. The weights  $\omega_{\langle s,t \rangle}$  ( $t = 1, 2, \dots, T$ ) measure the influence of the hidden layer output at time step  $t$  in the encoder on the output of the decoder at time step  $s$ , which are computed as

$$e_{\langle s,t \rangle} = \text{relu}(W^e [\hat{h}_{s-1}, h_t] + b^e), \quad (24)$$

$$\omega_{\langle s,t \rangle} = \frac{\exp(e_{\langle s,t \rangle})}{\sum_{t=1}^T \exp(e_{\langle s,t \rangle})}, \quad (25)$$

where  $W^e$  and  $b^e$  are the weights and biases of each attention cell, respectively,  $e_{\langle s,t \rangle}$  is the attention score of the  $t$ -th hidden layer state in the encoder, and  $\omega_{\langle s,t \rangle}$  is the weight of the  $t$ -th hidden layer state in the encoder. Afterward, the weighted summation of the output of each cell in the encoder is computed as the output of the attention mechanism layer, which is a context vector

$$C_s = \sum_{t=1}^T \omega_{\langle s,t \rangle} h_t, \quad (26)$$

and it is also the input of the corresponding cell of the decoder.

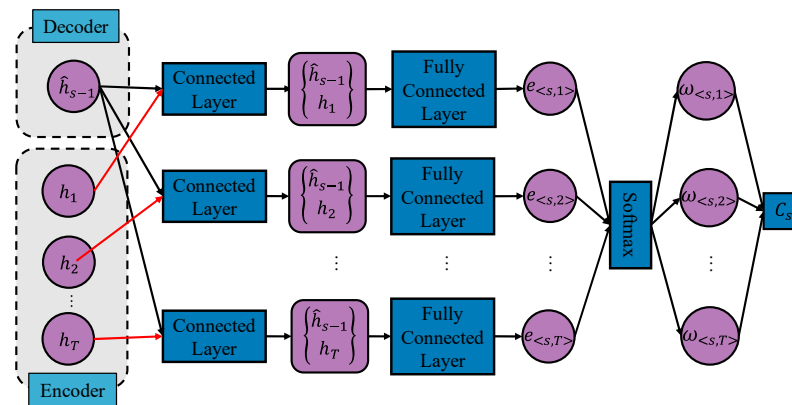


Figure 5. Schematic diagram of the attention mechanism.

Figure 6 depicts the framework of the LSTM-seq2seq-A model. The temporal feature prediction module captures deep temporal patterns within the deformation data, while the influencing factor prediction module dynamically evaluates the impact of each factor (e.g., water level, temperature) on the output by using attention weights. By analyzing data from both perspectives—time-based dependencies and factor-specific influences—the model achieves a more comprehensive feature extraction, leading to enhanced prediction accuracy and reliability in dam deformation forecasting. The original dataset  $\{x, y\}$  is first classified into a temporal feature dataset  $\{x^T, y^T\}$  and an influencing factor dataset  $\{x^F, y^F\}$ , and they are the input of a temporal feature prediction layer and an influencing factor prediction layer, respectively. In the first cell of the encoder, the cell state  $c_0$  and the hidden layer state  $h_0$  are initialized with zero matrices, and subsequently, the hidden layer states  $h_t^T$  ( $t = 1, 2, \dots, T$ ) of the temporal feature prediction layer and the hidden layer states  $h_f^F$  ( $f = 1, 2, \dots, m$ ) of the influencing factor prediction layer are obtained, and they are further combined with the hidden layer states at the previous time step of the decoder as the input of the individual attention cell. The attention weights of different states can then be obtained, and the output of the attention cell is the input of each cell in the decoder.

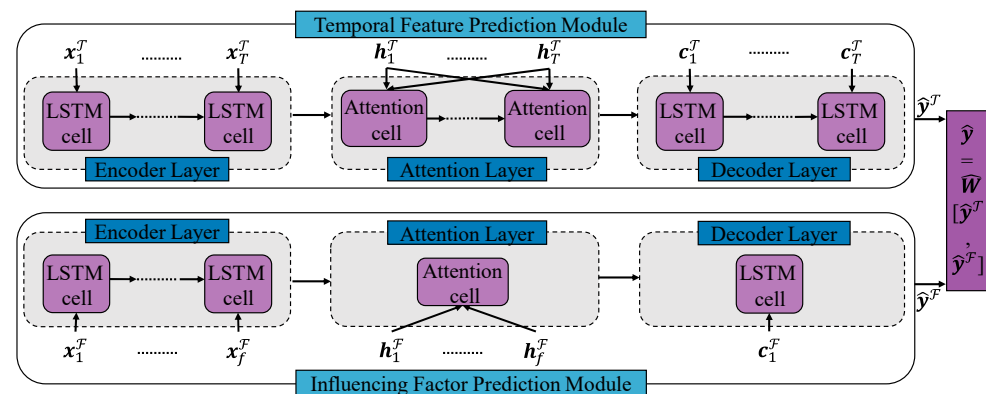


Figure 6. LSTM-seq2seq model with an attention mechanism.

The hidden layer states of the decoders are activated with the sigmoid function, and the final outputs of the temporal prediction layer and the influence factor prediction layer are reshaped into two matrices  $\hat{y}^T$  and  $\hat{y}^F$  of the same dimension  $n \times 1$ . The final output  $\hat{y}$  of the model is

$$\hat{y} = \hat{W} [\hat{y}^T, \hat{y}^F], \tag{27}$$

where  $\hat{W}$  is the weight matrix obtained from the temporal feature prediction layer and the influence factor prediction layer. The dam deformation prediction result is finally obtained after denormalizing  $\hat{y}$ .

The pseudocode of the proposed LSTM-seq2seq model with an attention mechanism and the chaos-based AOA is shown in Algorithm 1.

---

**Algorithm 1: LSTM-seq2seq-AA**

---

Input: original dataset  $\{x, y\}$  with  $x$  influencing factor sequence data and  $y$  dam deformation data, the initial solution  $w_i$  ( $(i = 1, 2, 3 \dots 10)$ ), maximum number of iterations  $M$

Output: Prediction model for dam deformation

- 1: Classify time feature dataset  $\{x^T, y^T\}$  and influencing factor dataset  $\{x^F, y^F\}$
- 2:  $temp = \text{inf}$ ,  $Leader\_Score = \text{inf}$ ,  $Leader\_pos = w_1$  # initialization
- 3: For  $i = 1, 2, \dots, 10$
- 4:     Obtain predicted dam deformation  $\hat{y}_i = \text{LSTM-seq2seq-A}(w_i, x^T, x^F)$
- 5:     Calculate fitness function value  $g_i = g(\hat{y}_i, y)$
- 6:     if ( $g_i < Leader\_Score$ )
- 7:         Update the optimal fitness function value  $Leader\_Score = g_i$  and update the optimal solution  $Leader\_pos = w_i$
- 8:     end if
- 9:   end for
- 10: while  $k < M$
- 11:     if ( $\text{abs}(temp - Leader\_Score) < 10^{-5}$ )
- 12:         update  $Leader\_Score$  and  $Leader\_pos$  with chaotic optimization (14) to (23)
- 13:     end if
- 14:      $temp = Leader\_Score$
- 15:     Update the  $Leader\_Score$  and  $Leader\_pos$  with AOA (10) to (13)
- 16:      $k = k + 1$
- 17: end while
- 18: return  $Leader\_pos$

---

#### 4. Case Study

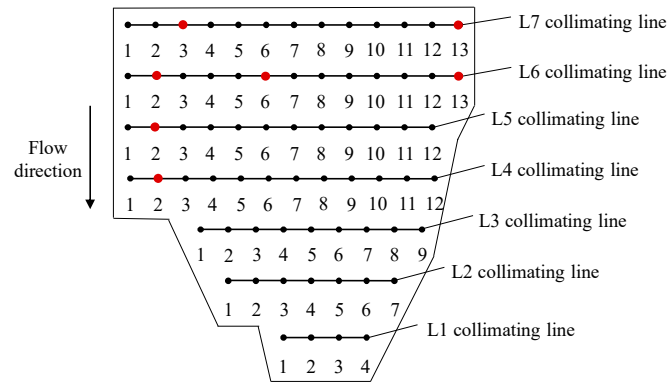
The proposed prediction model was applied to Nuozhadu Dam in southwestern China, which is a gravel-soil core earth-rock dam with a height of 261.5 m, as shown in Figure 7. A monitoring system has been installed to monitor the operation of the dam during the construction of the dam, which records several parameters including deformation, seepage, stress-strain, temperature, water level, etc. In this section, the prediction accuracy of the proposed model is investigated, and the contributions of different factors to the dam deformation are also interpreted with the model.



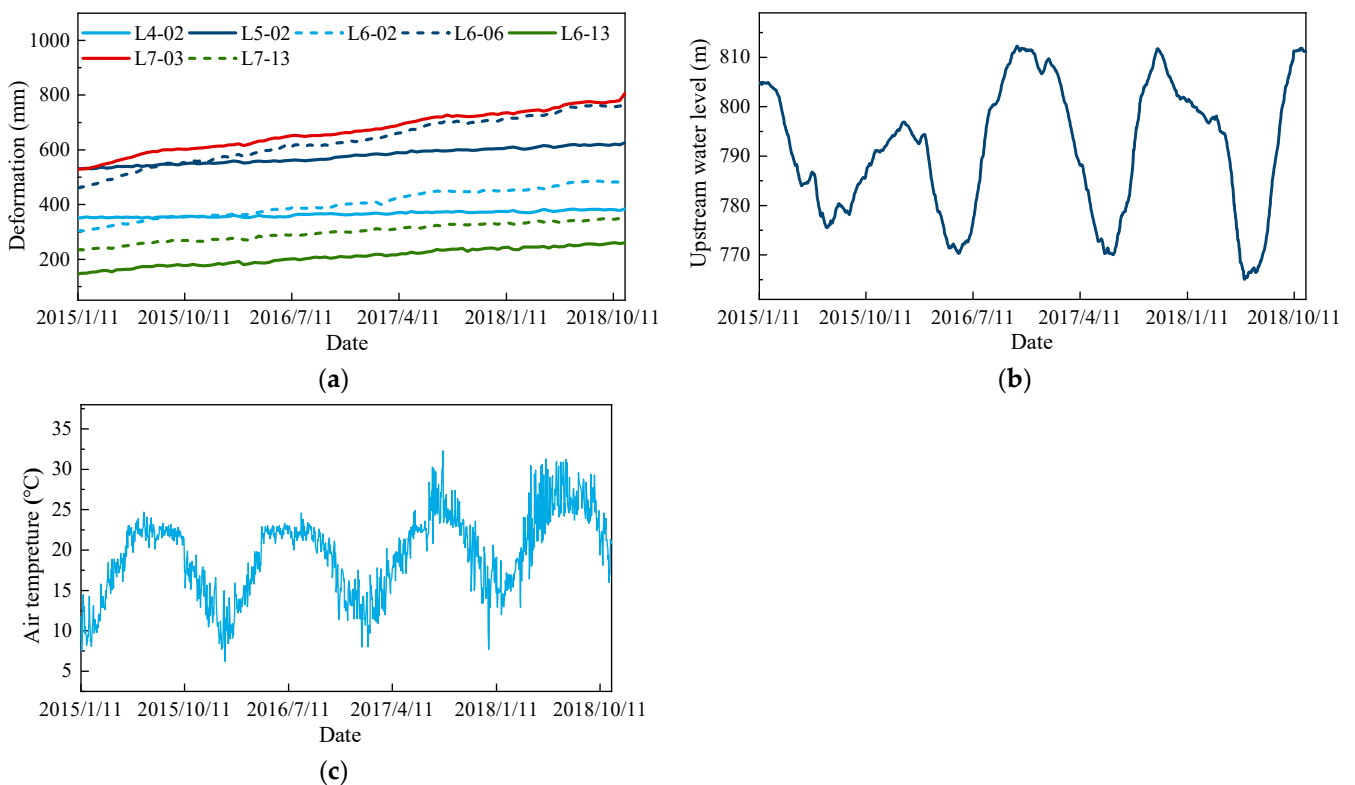
**Figure 7.** Scene of Nuozhadu Dam.

#### 4.1. Data Collection and Preprocessing

Figure 8 shows the layout of the deformation monitoring points on the dam. Without loss of generality, we chose seven monitoring points for model validation (i.e., L4-02, L5-02, L6-02, L6-06, L6-13, L7-03, and L7-13). In this study, the monitoring data were collected from 11 January 2015 to 10 November 2018 (1400 days) including deformation, water level, and air temperature, as shown in Figure 9.



**Figure 8.** Layout of monitoring points on collimating lines, in which seven points (in red) were selected for the model validation.



**Figure 9.** Monitoring data of the dam over time. (a) Deformation at the monitoring points. (b) Upstream water level. (c) Air temperature.

According to Equation (5), the input vector of the prediction model is composed of three types of influencing factors including the hydrostatic pressure component,  $\{p, p^2, p^3\}$ , the temperature component,  $\{\cos(\frac{2\pi d}{365}), \cos(\frac{4\pi d}{365})\}$ , and the time effect component,  $\{\eta, \ln \eta\}$ . Due to the spatial correlation among multiple monitoring points, the deformation data of

multiple points were directly integrated into the monitoring model. Therefore, the set of influencing factors  $\mathcal{F}$  can be considered as

$$\mathcal{F} = \left\{ p, p^2, p^3, \cos\left(\frac{2\pi d}{365}\right), \cos\left(\frac{4\pi d}{365}\right), \eta, \ln \eta, D1, D2, D3, D4, D5, D6 \right\}, \quad (28)$$

where  $D1$  to  $D6$  are the other six monitoring points. The LSTM-seq2seq-AA model utilizes these HST-derived inputs as sequential features, allowing it to learn complex nonlinear relationships between hydrostatic pressure, temperature, aging, and deformation.

The data are further normalized with

$$\bar{x} = \frac{x - x_{min}}{x_{max} - x_{min}}, \quad (29)$$

where  $x$  represents the original data,  $\bar{x}$  represents the normalized value, and  $x_{min}$  and  $x_{max}$  are the minimum and maximum values in the original data, respectively.

#### 4.2. Hyperparameters of the Prediction Model

The model was built with Keras 2.2.4. The chaos-based AOA (c-AOA) was applied to optimize the hyperparameters of the model. After optimization, we obtained the optimal hyperparameters of the temporal feature prediction module and the influence factor prediction module. In the temporal feature prediction module, the number of time steps  $T = 20$ , the learning rate  $L_r = 0.008$ , the number of neurons in an LSTM unit of the encoder  $N_u^e = 64$ , the number of neurons in an LSTM unit of the decoder was  $N_u^d = 64$ . In the influencing factor prediction module, the hyperparameters were  $T = 20$ ,  $L_r = 0.0077$ ,  $N_u^e = 64$ ,  $N_u^d = 64$ . The number of epochs  $N$  was set to 150.

We selected six other learning models (i.e., LSTM, LSTM with attention mechanism (LSTM-A), LSTM-seq2seq, LSTM-seq2seq-A, support vector machine (SVM), and multilayer perceptron (MLP)) to compare their prediction performance with our proposed model for the deformation at point L6-13. The chaos-based AOA and fivefold cross validation were used to determine the hyperparameters of all six models, which are given in Table 1.  $N_u$  in the LSTM-based models represents the number of LSTM units. The kernel function of SVM is the Gaussian kernel, with parameters  $q$  and  $\kappa$  the penalty and the kernel parameter, respectively.  $N_h$  in MLP represents the number of hidden layers, and  $N_l$  is the number of hidden neurons.

**Table 1.** Hyperparameters of different models for comparison.

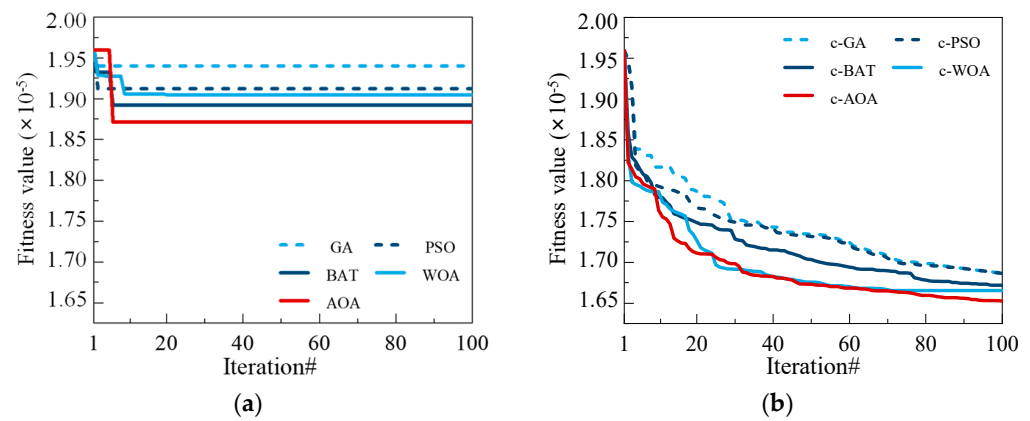
Model	Hyperparameters
LSTM	$N = 150, T = 14, L_r = 0.005, N_u = 49$
LSTM-A	$N = 150, T = 14, L_r = 0.005, N_u = 73$
LSTM-seq2seq	$N = 150, T = 20, L_r = 0.012, N_u^e = 73, N_u^d = 68$
LSTM-seq2seq-A	$N = 150, T = 20, L_r = 0.007, N_u^e = 63, N_u^d = 59$
SVM	$q = 7.21, \kappa = 5.49$
MLP	$N_h = 1, N_l = 15$

## 5. Result

### 5.1. Validation of Meta-Heuristic Training of the Model

We first demonstrated the advantage of using the AOA as the meta-heuristic training method. Taking the deformation prediction at monitoring point L6-13 as an example, we first compared the AOA with other classical meta-heuristic algorithms including the genetic algorithm (GA), the particle swarm optimization (PSO), the bat-inspired optimization algorithm (BAT), and the wolf optimization algorithm (WOA) to optimize the learnable parameters of the model. The convergence curves of different algorithms are shown in Figure 10a. It can be observed that all the algorithms converged quickly after 20 iterations to sub-optimal values, and using the AOA converged to the smallest value, reducing the

fitness function value from  $1.96 \times 10^{-5}$  to  $1.87 \times 10^{-5}$ . Afterward, we considered their implementations enhanced with the double chaotic mappings, as introduced in Section 3.2.2 (i.e., c-GA, c-PSO, c-BAT, c-WOA, and c-AOA), to optimize the learnable parameters of the model, and the convergence curves are shown in Figure 10b. Again, the AOA gave the smallest value, reducing the fitness function value from  $1.96 \times 10^{-5}$  to  $1.65 \times 10^{-5}$ . The chaos-based enhancements overcome this limitation by introducing chaotic search strategies that initiate independent parallel searches around the current best solution when convergence stalls. As can be seen, with the same 100 iterations, the double chaotic mappings significantly improved the optimization results, while the original meta-heuristic algorithms fell prematurely into the local optima. We present the deformation prediction results of each chaotic meta-heuristic algorithm for monitoring point L6-02 in Table 2. It can be seen that the differences in prediction evaluation metrics were minimal, indicating an almost similar performance. Additionally, as shown in Tables 3 and 4, AOA and c-AOA achieved comparable optimization goals with reduced time consumption compared to other meta-heuristic algorithms, enhancing the model's practicality for real-time dam monitoring applications.



**Figure 10.** Convergence of five meta-heuristic optimization algorithms to optimize learnable parameters. (a) Original optimization algorithms. (b) Chaos-based optimization algorithms.

**Table 2.** Comparison of the time consumption of all chaotic meta-heuristic algorithms.

	c-GA	c-PSO	c-BAT	c-WOA	c-AOA
MAPE (%)	0.244	0.242	0.242	0.243	0.242
RMSE (mm)	0.676	0.677	0.675	0.677	0.675
MAE (mm)	0.491	0.490	0.490	0.489	0.489

**Table 3.** Comparison of the time consumption of all classical meta-heuristic algorithms.

	GA	PSO	BAT	WOA	AOA
Time consumption (hours)	0.28	0.31	0.24	0.26	0.21

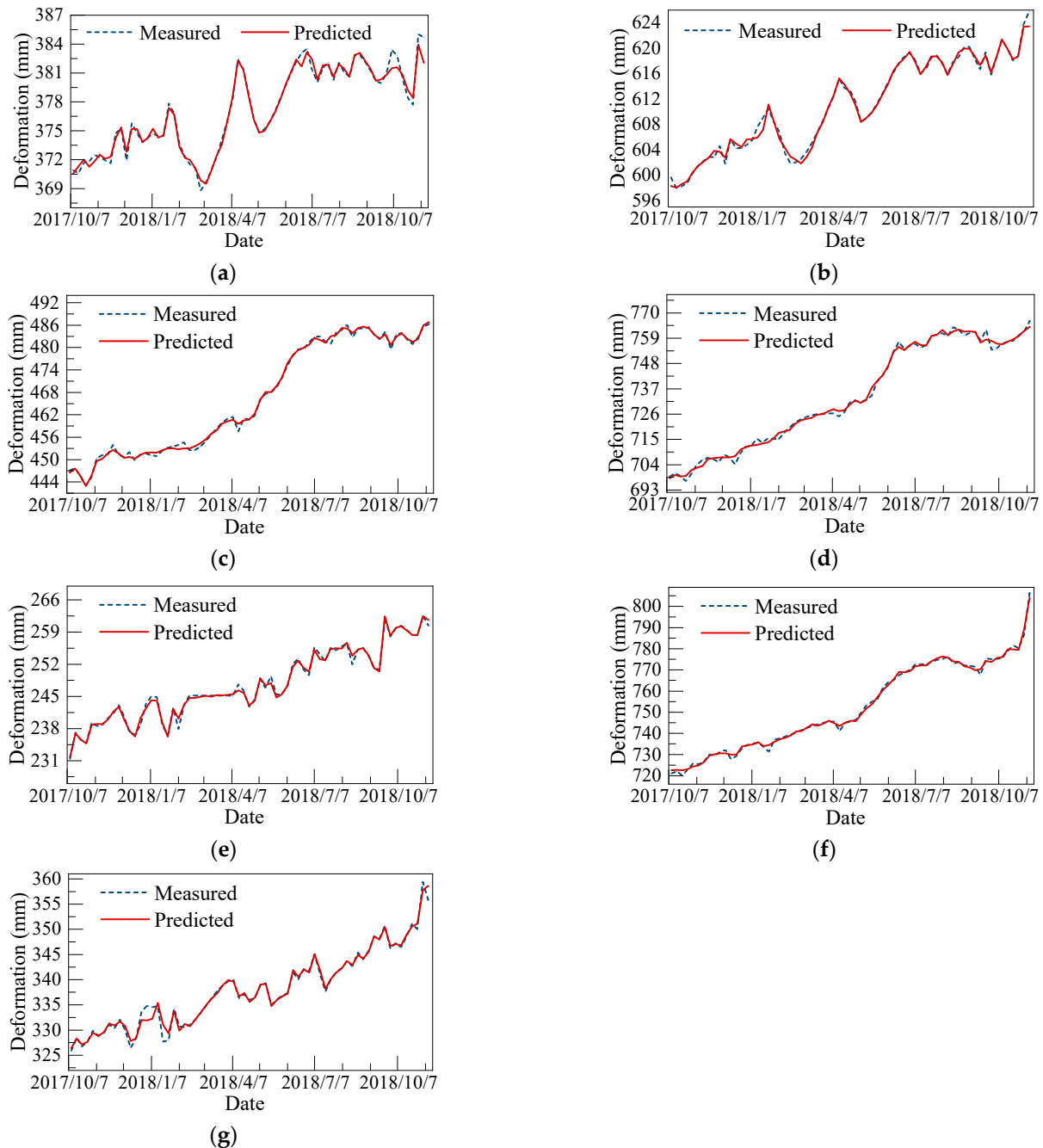
**Table 4.** Comparison of the time consumption of all chaotic algorithms.

	c-GA	c-PSO	c-BAT	c-WOA	c-AOA
Time consumption (hours)	7.54	7.99	6.87	6.84	6.13

## 5.2. Prediction Performance

The dam deformations at the seven points L4-02, L5-02, L6-02, L6-06, L6-13, L7-03, and L7-13 were predicted with the proposed model. The data of 1000 days from 11 January 2015

to 6 October 2017 formed the training dataset, while the remaining data of 400 days from 7 October 2017 to 10 November 2018 were used as the test dataset. For all of the predictions, the MAE loss function converged within 100 iterations. The comparison illustrated in Figure 11 demonstrates a good agreement between the deformation prediction results and the monitoring data. The performance evaluation metrics are given in Table 5. The MAPE, MAE, and RMSE ranged from 0.042% to 0.242%, 0.257 mm to 0.937 mm, and 0.483 mm to 1.019 mm, respectively. The average MAPE, RMSE, and MAE of the seven points were 0.125%, 0.739 mm, and 0.493 mm, respectively.



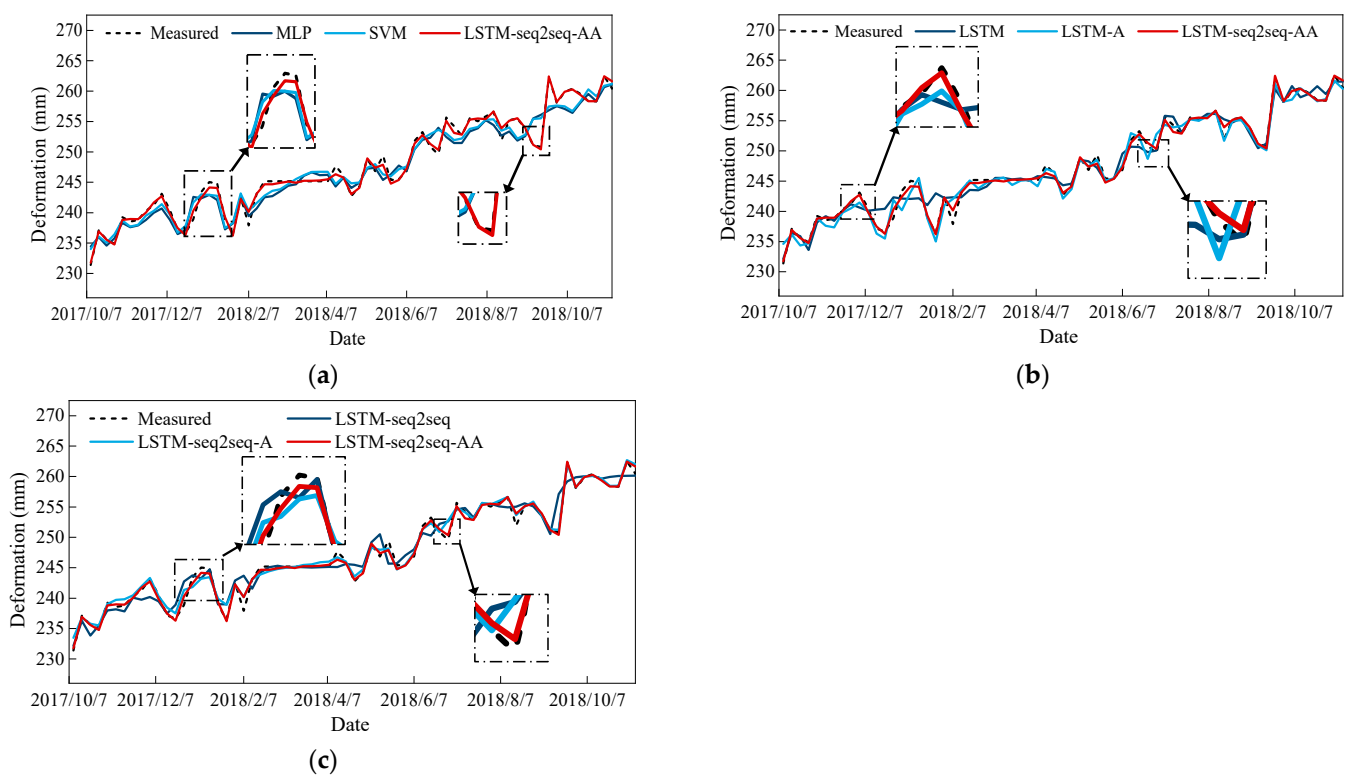
**Figure 11.** Prediction results at the seven monitoring points: (a) L4-02; (b) L5-02; (c) L6-02; (d) L6-06; (e) L6-13; (f) L7-03; (g) L7-13.

**Table 5.** Evaluation metrics of the proposed model at the seven monitoring points.

	L4-02	L5-02	L6-02	L6-06	L6-13	L7-03	L7-13	AVG
MAPE (%)	0.086	0.042	0.242	0.137	0.156	0.096	0.116	0.125
RMSE (mm)	0.483	0.694	0.675	1.019	0.695	0.983	0.626	0.739
MAE (mm)	0.257	0.376	0.489	0.937	0.301	0.702	0.391	0.493

### 5.3. Comparison

To further validate the model, we compared it with several commonly used traditional machine learning and deep learning models: support vector machine (SVM), multilayer perceptron (MLP), long short-term memory (LSTM), LSTM sequence-to-sequence (LSTM-seq2seq), attention-based LSTM (LSTM-A), and attention-based LSTM-seq2seq (LSTM-seq2seq-A). The comparison of prediction results for point L6-13 using these different models is shown in Figure 12, with the performance evaluation metrics listed in Table 6.



**Figure 12.** Comparison of the proposed model and other learning models for L6-13. (a) Results of MLP, SVM, and LSTM-seq2seq-AA. (b) Results of LSTM, LSTM-A, and LSTM-seq2seq-AA. (c) Results of LSTM-seq2seq, LSTM-seq2seq-A, and LSTM-seq2seq-AA.

**Table 6.** Hyperparameters of different models for comparison.

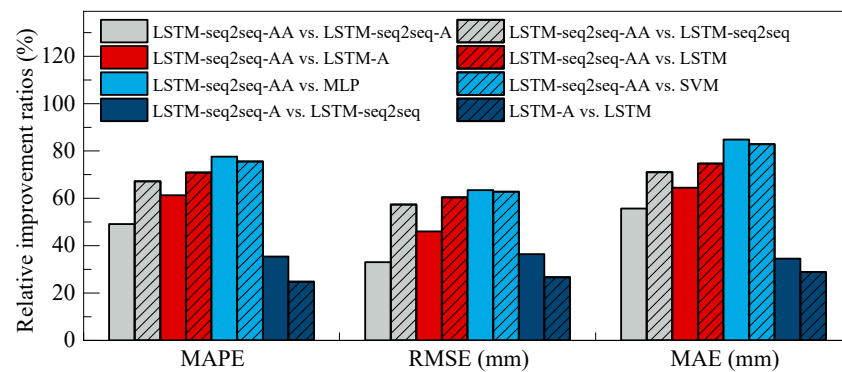
	LSTM-seq2seq-AA	LSTM-seq2seq-A	LSTM-A	LSTM-seq2seq	LSTM	MLP	SVM
MAPE (%)	0.156	0.307	0.403	0.475	0.536	0.697	0.637
RMSE (mm)	0.695	1.038	1.287	1.632	1.755	1.902	1.868
MAE (mm)	0.301	0.679	0.846	1.037	1.189	1.986	1.756

As illustrated in Figure 12, the proposed model consistently aligned most closely with the actual data, particularly in regions with high deformation fluctuations. A detailed view of point L6-13 highlights these differences, where the proposed method showed the lowest deviation from the observed values compared to other models. LSTM-seq2seq and



LSTM-A provided better fits than the standard LSTM, and LSTM-seq2seq-A aligned more closely with the observed values than LSTM-seq2seq. The proposed method achieved the best overall fit, confirming its effectiveness in capturing dynamic deformation patterns. Table 6 further supports these findings, showing that the proposed LSTM-seq2seq-AA model outperformed all of other models for deformation prediction, with the lowest MAPE, MAE, and RMSE values across the seven monitoring points.

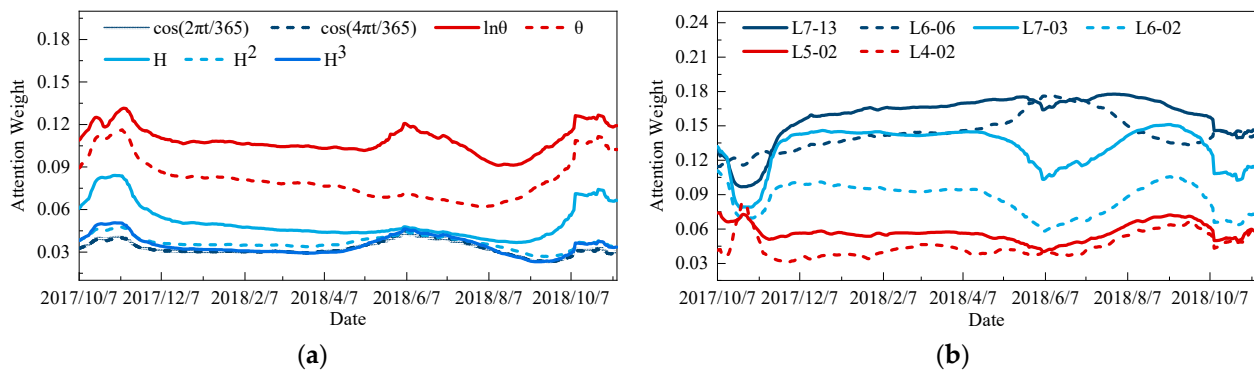
The relative improvement ratios between different models are plotted in Figure 13. The performance of the proposed model was distinguishingly better than the traditional machine learning models as it addressed the issue of long-term dependence. The relative improvement over the other deep learning models indicates that the attention mechanism improves the prediction accuracy with the adaptive weight assignment.



**Figure 13.** Relative improvement ratios between the deformation prediction models.

#### 5.4. Contributions of Influencing Factors

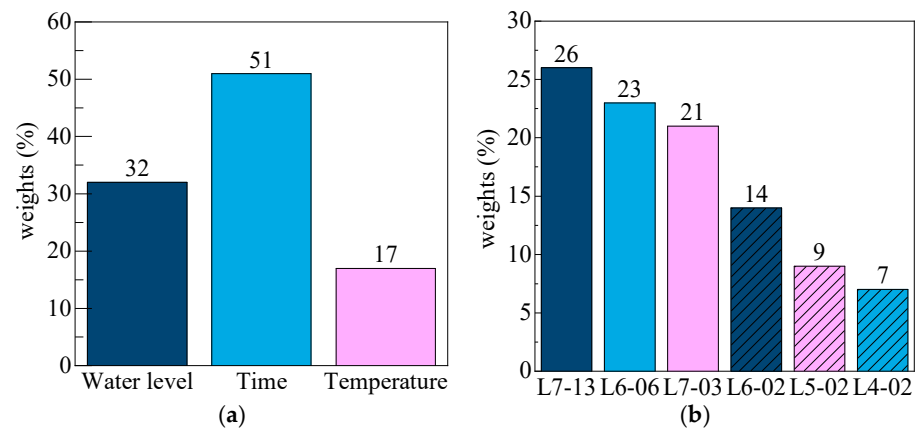
In the proposed model, the contribution of each influencing factor to deformation prediction can also be dynamically quantified. Figure 14a,b displays the time-varying attention weights of seven influencing factors and other spatially correlated monitoring points for point L6-13, respectively. Figure 15a,b shows the average attention weights of the seven environmental factors and six other monitoring points.



**Figure 14.** Attention weights of different influencing factors and other spatially correlated monitoring points of L6-13. (a) Attention weights of influencing factors. (b) Attention weights of other spatially correlated monitoring points.

From Figure 14a, it can be observed that in the time-varying attention weight plot for environmental factors, the two curves for the time-effect component  $\{\eta, \ln \eta\}$  consistently held the highest positions, and the hydrostatic pressure component  $\{p, p^2, p^3\}$  were mostly larger than the temperature component  $\{\cos \frac{2\pi d}{365}, \cos \frac{4\pi d}{365}\}$ , indicating that among the environmental factors, the time-effect components had the greatest impact on deformation. Figure 14b shows that in the time-varying attention weight plot for monitoring points, the curve for point L7-13 remained in a relatively high position, suggesting that deformation

at this point plays a dominant role in influencing the deformation monitoring data for point L6-13.



**Figure 15.** Average attention weights of influencing factors and other points. (a) Weights of influencing factors for L6-13. (b) Average attention weights of influencing factors and other points.

Figure 15a shows that among the environmental factors, the aging factor had the highest attention weight, or contribution rate, to the deformation of the monitoring point, at 51%, followed by the water level and temperature, with contribution rates of 32% and 17%, respectively. Among the deformation monitoring points, point L7-13 contributed the most to point L6-13, at 26%, followed by L6-06, L7-03, and L6-02, while points L5-02 and L4-02 had the smallest contribution rates.

Based on long-term studies of earth-rock dam deformation, the time-effect components were the most significant, the hydrostatic pressure component was less significant, and the temperature component was negligible, which aligns with the attention weight results calculated in this study [33]. Additionally, from the perspective of spatial distance from point L6-13, L7-13 was the closest, followed by L6-06 and L7-03, while L6-02 was slightly farther, and L5-02 and L4-02 were the farthest. Generally, the closer the points are to each other, the greater the influence of deformation between them, which corresponds closely to the results in Figures 14 and 15, thus proving the reliability of the model in this study.

## 6. Discussion

The proposed LSTM-seq2seq-AA model demonstrates an innovative approach to improving dam deformation prediction by leveraging a chaos-based arithmetic optimization algorithm (c-AOA) and an attention mechanism. These components address core challenges in complex time-series modeling by optimizing predictive accuracy and enhancing model interpretability. The attention mechanism plays a critical role in refining predictions by dynamically focusing on the most relevant input factors for each time step. This allows the model to adaptively emphasize essential temporal patterns in the data, capturing the nuanced influences of various environmental and structural factors. Such adaptability is particularly beneficial in managing the variable conditions surrounding dam deformation, as it enables the model to assess factor significance in real-time. While spatial correlations among monitoring points were not the focus of this study, future work could consider integrating a spatial attention layer, which could further improve the accuracy by capturing dependencies across different monitoring locations. The chaos-based AOA offers an effective solution for hyperparameter optimization, surpassing conventional algorithms in its ability to navigate complex solution spaces and avoid local minima. By employing chaotic mapping, AOA introduces diversity into the search process, enhancing model convergence and stability. This optimization approach contributes to robust predictive performance, ensuring reliable deformation predictions under varying conditions.

Despite the model's demonstrated efficacy, there are limitations to its applicability across different dam types. Concrete gravity dams, for instance, exhibit more stable mate-

rial characteristics and deformation driven largely by creep and thermal stress, potentially requiring model adjustments to capture these specific dynamics accurately. Similarly, the complex deformation patterns in arch dams might challenge the model's current configuration. Tailoring the model for these structural variations could improve its adaptability and broaden its application scope within the field.

This study addressed key challenges in building sciences, particularly in the prediction of structural deformation and the health monitoring of large infrastructure systems. The proposed model is particularly valuable for structural health monitoring in civil engineering and construction and enhances the ability to predict deformation trends and assess risk, enabling timely maintenance and intervention strategies. By improving deformation prediction accuracy, this model supports the development of early warning systems for critical infrastructure including dams, bridges, and high-rise buildings. Such systems can significantly reduce the risk of catastrophic failures and increase the overall resilience of infrastructure.

## 7. Conclusions

In this study, an LSTM sequence-to-sequence model integrated with an attention mechanism and a chaos-based arithmetic optimization algorithm (AOA) was proposed for dam deformation prediction. The conclusions drawn from this research are as follows:

(1) The proposed LSTM-seq2seq-AA model significantly enhanced the prediction accuracy for dam deformation. Quantitatively, the model achieved average RMSE, MAE, and MAPE values of 0.739 mm, 0.493 mm, and 0.125%, respectively, across seven monitoring points. This performance markedly surpassed that of the traditional machine learning models and other LSTM-based models. Qualitatively, the integration of the attention mechanism improved the interpretability of the model by dynamically assigning weights to influencing factors, while the chaos-based AOA effectively optimized the learnable parameters, avoiding convergence to the local optima.

(2) Given its high accuracy and robustness, the proposed model is recommended for implementation in dam safety monitoring systems. It is particularly suitable for real-time deformation prediction and the risk assessment of earth-rock dams. Engineers and practitioners can utilize this model to enhance early warning systems and inform maintenance decisions, thereby improving the overall safety management of dam infrastructure.

(3) Future studies should explore the application of the LSTM-seq2seq-AA model to different types of dams and structural health monitoring scenarios to validate its generalizability. Investigating the incorporation of additional influencing factors, such as material properties and environmental conditions, could further refine the model. Moreover, integrating other advanced optimization algorithms might enhance the prediction performance, opening new avenues for research in the predictive modeling of complex engineering structures.

(4) While the model demonstrated superior performance, it was tested on data from a single earth-rock dam. To ensure broader applicability, future work should involve testing the model on various dam types and incorporating larger datasets. Additionally, extending the model to predict other structural behaviors, such as stress distribution and seepage patterns, could provide a more comprehensive tool for structural health monitoring.

**Author Contributions:** Conceptualization, J.W.; Methodology, L.W.; Validation, D.T.; Writing—original draft preparation, L.W.; Writing—review and editing, J.W.; Supervision, X.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest.

## References

1. Su, H.; Wen, Z.; Wang, F.; Wei, B.; Hu, J. Multifractal Scaling Behavior Analysis for Existing Dams. *Expert Syst. Appl.* **2013**, *40*, 4922–4933. [[CrossRef](#)]
2. Li, B.; Yang, J.; Hu, D. Dam Monitoring Data Analysis Methods: A Literature Review. *Struct. Control Health Monit.* **2020**, *27*, e2501. [[CrossRef](#)]
3. Huang, B.; Kang, F.; Li, J.; Wang, F. Displacement Prediction Model for High Arch Dams Using Long Short-Term Memory Based Encoder-Decoder with Dual-Stage Attention Considering Measured Dam Temperature. *Eng. Struct.* **2023**, *280*, 115686. [[CrossRef](#)]
4. Wen, Z.; Zhou, R.; Su, H. MR and Stacked GRUs Neural Network Combined Model and Its Application for Deformation Prediction of Concrete Dam. *Expert Syst. Appl.* **2022**, *201*, 117272. [[CrossRef](#)]
5. Hu, J.; Jiang, H.; Li, X. An Optimized Zonal Deformation Prediction Model for Super-High Arch Dams. *Structures* **2023**, *50*, 758–774. [[CrossRef](#)]
6. Ren, Q.; Li, M.; Li, H.; Shen, Y. A Novel Deep Learning Prediction Model for Concrete Dam Displacements Using Interpretable Mixed Attention Mechanism. *Adv. Eng. Inform.* **2021**, *50*, 101407. [[CrossRef](#)]
7. Prakash, G.; Dugalam, R.; Barbosh, M.; Sadhu, A. Recent Advancement of Concrete Dam Health Monitoring Technology: A Systematic Literature Review. *Structures* **2022**, *44*, 766–784. [[CrossRef](#)]
8. Gamse, S.; Zhou, W.-H.; Tan, F.; Yuen, K.-V.; Oberguggenberger, M. Hydrostatic-Season-Time Model Updating Using Bayesian Model Class Selection. *Reliab. Eng. Syst. Saf.* **2018**, *169*, 40–50. [[CrossRef](#)]
9. Gamse, S.; Henriques, M.J.; Oberguggenberger, M.; Mata, J.T. Analysis of Periodicities in Long-term Displacement Time Series in Concrete Dams. *Struct. Control Health Monit.* **2020**, *27*, e2477. [[CrossRef](#)]
10. Li, F.; Wang, Z.; Liu, G.; Fu, C.; Wang, J. Hydrostatic Seasonal State Model for Monitoring Data Analysis of Concrete Dams. *Struct. Infrastruct. Eng.* **2015**, *11*, 1616–1631. [[CrossRef](#)]
11. Kao, C.-Y.; Loh, C.-H. Monitoring of Long-Term Static Deformation Data of Fei-Tsui Arch Dam Using Artificial Neural Network-Based Approaches: Long-Term Static Deformation Data of Fei-Tsui Arch Dam. *Struct. Control Health Monit.* **2013**, *20*, 282–303. [[CrossRef](#)]
12. Wang, X.; Yang, K.; Shen, C. Study on MPGA-BP of Gravity Dam Deformation Prediction. *Math. Probl. Eng.* **2017**, *2017*, 2586107. [[CrossRef](#)]
13. Ranković, V.; Grujović, N.; Divac, D.; Milivojević, N. Development of Support Vector Regression Identification Model for Prediction of Dam Structural Behaviour. *Struct. Saf.* **2014**, *48*, 33–39. [[CrossRef](#)]
14. Su, H.; Chen, Z.; Wen, Z. Performance Improvement Method of Support Vector Machine-Based Model Monitoring Dam Safety: Performance Improvement Method of Monitoring Model of Dam Safety. *Struct. Control Health Monit.* **2016**, *23*, 252–266. [[CrossRef](#)]
15. Su, H.; Li, X.; Yang, B.; Wen, Z. Wavelet Support Vector Machine-Based Prediction Model of Dam Deformation. *Mech. Syst. Signal Process.* **2018**, *110*, 412–427. [[CrossRef](#)]
16. Kang, F.; Li, J.; Dai, J. Prediction of Long-Term Temperature Effect in Structural Health Monitoring of Concrete Dams Using Support Vector Machines with Jaya Optimizer and Salp Swarm Algorithms. *Adv. Eng. Softw.* **2019**, *131*, 60–76. [[CrossRef](#)]
17. Kang, F.; Liu, X.; Li, J. Temperature Effect Modeling in Structural Health Monitoring of Concrete Dams Using Kernel Extreme Learning Machines. *Struct. Health Monit.* **2020**, *19*, 987–1002. [[CrossRef](#)]
18. Liu, X.; Kang, F.; Ma, C.; Li, H. Concrete Arch Dam Behavior Prediction Using Kernel-Extreme Learning Machines Considering Thermal Effect. *J. Civ. Struct. Health Monit.* **2021**, *11*, 283–299. [[CrossRef](#)]
19. Chen, W.; Wang, X.; Tong, D.; Cai, Z.; Zhu, Y.; Liu, C. Dynamic Early-Warning Model of Dam Deformation Based on Deep Learning and Fusion of Spatiotemporal Features. *Knowl.-Based Syst.* **2021**, *233*, 107537. [[CrossRef](#)]
20. Panjapornpon, C.; Bardeeniz, S.; Hussain, M.A. Deep Learning Approach for Energy Efficiency Prediction with Signal Monitoring Reliability for a Vinyl Chloride Monomer Process. *Reliab. Eng. Syst. Saf.* **2023**, *231*, 109008. [[CrossRef](#)]
21. Lu, S.; Yang, J.; Yang, B.; Li, X.; Yin, Z.; Yin, L.; Zheng, W. Surgical Instrument Posture Estimation and Tracking Based on LSTM. *ICT Express* **2024**, *10*, 465–471. [[CrossRef](#)]
22. Krones, F.; Marikkar, U.; Parsons, G.; Szmul, A.; Mahdi, A. Review of Multimodal Machine Learning Approaches in Healthcare. *Inf. Fusion* **2025**, *114*, 102690. [[CrossRef](#)]
23. Fahad Mon, B.; Wasfi, A.; Hayajneh, M.; Slim, A.; Abu Ali, N. Reinforcement Learning in Education: A Literature Review. *Informatics* **2023**, *10*, 74. [[CrossRef](#)]
24. Zheng, Y.; Xu, Z.; Xiao, A. Deep Learning in Economics: A Systematic and Critical Review. *Artif. Intell. Rev.* **2023**, *56*, 9497–9539. [[CrossRef](#)] [[PubMed](#)]
25. Joseph, L.P.; Deo, R.C.; Prasad, R.; Salcedo-Sanz, S.; Raj, N.; Soar, J. Near Real-Time Wind Speed Forecast Model with Bidirectional LSTM Networks. *Renew. Energy* **2023**, *204*, 39–58. [[CrossRef](#)]
26. Xu, H.; Wu, L.; Xiong, S.; Li, W.; Garg, A.; Gao, L. An Improved CNN-LSTM Model-Based State-of-Health Estimation Approach for Lithium-Ion Batteries. *Energy* **2023**, *276*, 127585. [[CrossRef](#)]
27. Wang, Z.; Liu, N.; Chen, C.; Guo, Y. Adaptive Self-Attention LSTM for RUL Prediction of Lithium-Ion Batteries. *Inf. Sci.* **2023**, *635*, 398–413. [[CrossRef](#)]
28. Hu, Z.; Gao, Y.; Ji, S.; Mae, M.; Imaizumi, T. Improved Multistep Ahead Photovoltaic Power Prediction Model Based on LSTM and Self-Attention with Weather Forecast Data. *Appl. Energy* **2024**, *359*, 122709. [[CrossRef](#)]

29. Yazdinejad, A.; Kazemi, M.; Parizi, R.M.; Dehghantanha, A.; Karimipour, H. An Ensemble Deep Learning Model for Cyber Threat Hunting in Industrial Internet of Things. *Digit. Commun. Netw.* **2023**, *9*, 101–110. [[CrossRef](#)]
30. Mohammad-Alikhani, A.; Nahid-Mobarakeh, B.; Hsieh, M.-F. One-Dimensional LSTM-Regulated Deep Residual Network for Data-Driven Fault Detection in Electric Machines. *IEEE Trans. Ind. Electron.* **2024**, *71*, 3083–3092. [[CrossRef](#)]
31. Kheddar, H.; Himeur, Y.; Awad, A.I. Deep Transfer Learning for Intrusion Detection in Industrial Control Networks: A Comprehensive Review. *J. Netw. Comput. Appl.* **2023**, *220*, 103760. [[CrossRef](#)]
32. Chen, X.; Chen, Z.; Hu, S.; Gu, C.; Guo, J.; Qin, X. A Feature Decomposition-Based Deep Transfer Learning Framework for Concrete Dam Deformation Prediction with Observational Insufficiency. *Adv. Eng. Inform.* **2023**, *58*, 102175. [[CrossRef](#)]
33. Zhou, Y. Multi-Expert Attention Network for Long-Term Dam Displacement Prediction. *Adv. Eng. Inform.* **2023**, *57*, 102060. [[CrossRef](#)]
34. Shu, X.; Bao, T.; Li, Y.; Gong, J.; Zhang, K. VAE-TALSTM: A Temporal Attention and Variational Autoencoder-Based Long Short-Term Memory Framework for Dam Displacement Prediction. *Eng. Comput.* **2022**, *38*, 3497–3512. [[CrossRef](#)]
35. Zhu, T.; Chen, Z.; Zhou, D.; Xia, T.; Pan, E. Adaptive Staged Remaining Useful Life Prediction of Roller in a Hot Strip Mill Based on Multi-Scale LSTM with Multi-Head Attention. *Reliab. Eng. Syst. Saf.* **2024**, *248*, 110161. [[CrossRef](#)]
36. Shi, J.; Zhong, J.; Zhang, Y.; Xiao, B.; Xiao, L.; Zheng, Y. A Dual Attention LSTM Lightweight Model Based on Exponential Smoothing for Remaining Useful Life Prediction. *Reliab. Eng. Syst. Saf.* **2024**, *243*, 109821. [[CrossRef](#)]
37. Li, Y.; Bao, T.; Gong, J.; Shu, X.; Zhang, K. The Prediction of Dam Displacement Time Series Using STL, Extra-Trees, and Stacked LSTM Neural Network. *IEEE Access* **2020**, *8*, 94440–94452. [[CrossRef](#)]
38. Liu, W.; Pan, J.; Ren, Y.; Wu, Z.; Wang, J. Coupling Prediction Model for Long-term Displacements of Arch Dams Based on Long Short-term Memory Network. *Struct. Control Health Monit.* **2020**, *27*, e2548. [[CrossRef](#)]
39. Yang, D.; Gu, C.; Zhu, Y.; Dai, B.; Zhang, K.; Zhang, Z.; Li, B. A Concrete Dam Deformation Prediction Method Based on LSTM With Attention Mechanism. *IEEE Access* **2020**, *8*, 185177–185186. [[CrossRef](#)]
40. Li, X.; Krivtsov, V.; Arora, K. Attention-Based Deep Survival Model for Time Series Data. *Reliab. Eng. Syst. Saf.* **2022**, *217*, 108033. [[CrossRef](#)]
41. Gundu, V.; Simon, S.P. PSO–LSTM for Short Term Forecast of Heterogeneous Time Series Electricity Price Signals. *J. Ambient Intell. Humaniz. Comput.* **2021**, *12*, 2375–2385. [[CrossRef](#)]
42. Chen, Z.; Yang, C.; Qiao, J. The Optimal Design and Application of LSTM Neural Network Based on the Hybrid Coding PSO Algorithm. *J. Supercomput.* **2022**, *78*, 7227–7259. [[CrossRef](#)]
43. Abualigah, L.; Diabat, A.; Mirjalili, S.; Abd Elaziz, M.; Gandomi, A.H. The Arithmetic Optimization Algorithm. *Comput. Methods Appl. Mech. Eng.* **2021**, *376*, 113609. [[CrossRef](#)]
44. Li, C.; Mei, X. Application of SVR Models Built with AOA and Chaos Mapping for Predicting Tunnel Crown Displacement Induced by Blasting Excavation. *Appl. Soft Comput.* **2023**, *147*, 110808. [[CrossRef](#)]
45. Xu, T.; Gao, Z.; Zhuang, Y. Fault Prediction of Control Clusters Based on an Improved Arithmetic Optimization Algorithm and BP Neural Network. *Mathematics* **2023**, *11*, 2891. [[CrossRef](#)]
46. Wei, B.; Luo, S.; Yuan, D. Optimized Combined Forecasting Model for Hybrid Signals in the Displacement Monitoring Data of Concrete Dams. *Structures* **2023**, *48*, 1989–2002. [[CrossRef](#)]
47. Wu, Z.; Ruan, H. Inverse analysis of safety monitoring data from concrete dams. *J. Hohai Univ. (Nat. Sci.)* **1989**, *2*, 10–18. (In Chinese)
48. Wu, Z.; Chen, J. Methods and models for the analysis of dam safety monitoring. *Adv. Sci. Technol. Water Resour.* **1989**, *9*, 48–52+54–64. (In Chinese)
49. Feng, J.; Zhang, J.; Zhu, X.; Lian, W. A Novel Chaos Optimization Algorithm. *Multimed. Tools Appl.* **2017**, *76*, 17405–17436. [[CrossRef](#)]
50. Luo, Y.; Yu, J.; Lai, W.; Liu, L. A Novel Chaotic Image Encryption Algorithm Based on Improved Baker Map and Logistic Map. *Multimed. Tools Appl.* **2019**, *78*, 22023–22043. [[CrossRef](#)]
51. Kiani, F.; Nematzadeh, S.; Anka, F.A.; Findikli, M.A. Chaotic Sand Cat Swarm Optimization. *Mathematics* **2023**, *11*, 2340. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.