IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 25 February 2024, accepted 25 March 2024, date of publication 29 March 2024, date of current version 15 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3383016

RESEARCH ARTICLE

Gravity Dam Deformation Prediction Model Based on I-KShape and ZOA-BiLSTM

MADINIYETI JIEDEERBIEKE[1,2], TONGCHUN LI[2], YANG CHAO[1], HUIJUN QI[1], AND CHAONING LIN[1]

[1]College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210098, China
[2]College of Hydraulic and Civil Engineering, Xinjiang Agricultural University, Ürümqi 830052, China

Corresponding authors: Tongchun Li (ltchhu@163.com) and Yang Chao (chaoyangccc@hhu.edu.cn)

This work was supported in part by the National Key Research and Development Plan of China under Grant 2022YFC3005403, in part by the National Natural Science Foundation of China under Grant 52309151 and Grant 52309157, in part by the Open Research Fund of Key Laboratory of Reservoir and Dam Safety Ministry of Water Resources under Grant YK323007, in part by the Science and Technology Project of Power Construction Corporation of China under Grant DJ-ZDXM-2021-10, and in part by the Excellent Postdoctoral Program of Jiangsu Province under Grant 2022ZB191.

ABSTRACT This research proposes a dam deformation prediction model based on clustering partitioning and Bidirectional Long Short-Term Memory (BiLSTM) networks to address the limitations of traditional monitoring models in characterizing the distribution characteristics of deformation zones in concrete gravity dams. The model takes into account the intrinsic correlations among monitoring points and achieves more comprehensive deformation monitoring by integrating multiple feature information. Firstly, the improved K-Shape algorithm, which takes into account the time series features and spatial coordinate relationships, is used to cluster and partition the spatial measurement points to better capture the spatial distribution characteristics of the deformation region. Following that, the model hyperparameters undergo iterative optimization using the ZOA optimization algorithm to enhance overall model performance. Finally, a ZOA-BiLSTM modelling process incorporating the correlation characteristics of multiple measurement points is proposed. After validation by engineering examples, the clustering results coincide with the spatial distribution characteristics of dam deformation. Meanwhile, the prediction model has high accuracy and robustness, and predicts the dam deformation from the multi-measurement point correlation dimension, which provides a new and effective method to monitor the overall safety state of the dam.

INDEX TERMS Deformation prediction, spatial clustering, bidirectional long-short term memory (BiLSTM), zebra optimization algorithm (ZOA).

I. INTRODUCTION

As the core structures of water conservancy projects, the safety and stability of concrete dams significantly impact the safety of people's lives and properties, economic development, and the ecological environment in downstream areas [1], [2]. Therefore, accurately predicting and monitoring the deformation behavior of concrete dams is particularly important [3], [4]. However, dam deformation patterns are shaped by diverse factors such as water pressure, temperature, and time. These elements collectively contribute to intricate nonlinear characteristics, presenting a technical challenge for precise prediction [5], [6], [7] Current models for predicting dam deformation can be broadly classified into three categories: statistical, deterministic, and hybrid models [8], [9]. A predictive model grounded in statistical principles relies on historical data to infer future deformation trends by analyzing past dam deformation data. While this method is simple and intuitive, its prediction accuracy may be limited when addressing nonlinear and complex dam deformation problems [10], [11] With the rapid advancement of computer technology, sophisticated machine algorithms, including artificial neural networks, support vector machines, random forests, multilayer feed-forward neural networks, genetic algorithms, and other advanced techniques, have increasingly taken a leading role in developing models for

The associate editor coordinating the review of this manuscript and approving it for publication was Qilian Liang.

analyzing and predicting dam deformation. These algorithms offer robust technical support for further improving dam safety monitoring and early warning systems [12], [13]. Hipni et al. [14] successfully predicted the daily water level of a sluice gate through multiple input schemes and effective utilization of SVM algorithm, successfully predicted the daily water level of a sluice gate; Wang et al. [15] introduced a hybrid model, combining backpropagation with a genetic algorithm (GA-BP) and multiple population genetic algorithm (MPGA), building upon the BP model. This integration significantly enhanced both the convergence speed and prediction accuracy of the dam monitoring model; Belmokre et al. [16] based on the Random Forest algorithm proposed a deformation prediction model, which better captures the deformation pattern in the dam monitoring model through multiple different inputs. It has been shown that although the prediction models constructed by machine learning algorithms perform well in simulating static regression relationships, with strong generalizability and high-precision prediction ability, they usually tend to focus on capturing static input-output relationships and ignore the possible time dependence within the deformation data of a single measurement point. Considering that dam deformation may be affected by environmental factors with a lag, ignoring the time dependence may adversely affect the prediction results [7], [17]. This research perspective underscores the importance of a more comprehensive consideration of temporal dynamics in constructing prediction models. In recent years, time series algorithms like Convolutional Neural Network (CNN), Long Short-Term Memory Neural Network (LSTM), and Gated Recurrent Unit (GRU) in the field of deep learning have been widely employed to effectively address the limitations of traditional machine learning algorithms in capturing temporal dependencies [18], [19] The incorporation of time series algorithms markedly enhances the accuracy of dam monitoring models. Leveraging the inherent sensitivity of deep learning models to time series data, they excel in capturing the dynamic effects of environmental factors on dam deformation. This capability enables more precise prediction and monitoring. Qu et al. [20] developed deformation prediction models for the health monitoring of concrete dams based on Rough Set theory (RS) and Long-Short-Term Memory Network (LSTM). They introduced both single-point and multi-point models and suggested a novel evaluation system. This system incorporates quantitative evaluation indices such as model accuracy, robustness, externality, and generalization definition.; Yang et al. [21] aimed at trend analysis of the saturation line of tailings dams, used CNN to identify and learn spatial structures in the time series, and used LSTM cells to detect long-term dependencies, and developed a reliable tailings dam prediction model; Li et al. [22] presented seven contemporary methods for comparison purposes. They additionally put forth a novel combined model designed to predict dam displacement time series. This approach further validated the effectiveness and feasibility of the LSTM model.

The deformation characteristics of a dam are, in fact, a manifestation of the overall structural behavior resulting from the collective work of various dam sections. Previous data analysis methods have primarily focused on individual monitoring points, without delving into the mutual influence of deformations between different parts of the dam. Therefore, it is essential to further and comprehensively explore the collaborative deformation effects among different parts of the dam beyond the deformation prediction model for individual points. This allows for a more in-depth examination of potential connections between monitoring points. Through this approach, we can achieve spatiotemporal coordinated analysis of dam deformations and enhance the predictive accuracy of the model. In the field of data mining, cluster analysis is a common research method. It divides a dataset into several subsets with distinct differences based on a certain similarity [23], [24], [25] Song et al. [26] explored the correlation coefficients among deformation signals from various measurement points. They introduced the concepts of multivariate panel data and K-means clustering theory and devised an analytical method for identifying outliers in dam deformation data; Hu and Ma [27] introduced a zonal deformation prediction model tailored for ultra-high arch dams. This model integrates the hierarchical clustering method with a panel data model, providing a novel approach for predicting deformation in specific zones; Chen et al. [28] embedded a clustering method based on Gaussian mixture model using the minimum-density-entropy-optimized method to achieve reliable identification of spatio-temporal divergence of dam behavior, and established a prediction method based on spatio-temporal clustering and machine learning.The primary inspiration for the partitioning in this study is drawn from the literature [29]. During the review, it was observed that the parameter results obtained from the optimization of different measurement points varied significantly, posing challenges in obtaining consistent input parameters for the entire dam deformation prediction model. Recognizing the importance of deformation similarity, a comprehensive analysis of the correlation among all deformation measurement points of concrete dams becomes crucial.Simultaneously, the division of measurement points and the mitigation of interference caused by differences in deformation patterns among the points can help unveil potential synergistic effects among different regions within the dam. Inspired by the classification of data and time series algorithms, and taking into account the global search capability of the prediction model, this research introduces a novel dam deformation prediction model based on multiple feature information.The model incorporates the Zeroth-Order Optimization Algorithm (ZOA) and Bidirectional Long Short-Term Memory (BiLSTM) algorithms. K-Shape, a method for time series clustering, represents an improved version of the K-Means-based algorithm. The K-Shape method takes into account the vertical stretching and horizontal translation of the data by introducing adaptive alignment and normalization operations of the time series to

capture the shape information of the sequences more accurately [30], [31] BiLSTM, a variant of the recurrent neural network, possesses the capability to memorize time-series data processing, capturing long-term dependencies, a feature inherent in traditional LSTM networks.BiLSTM, on the other hand, by introducing a reverse layer in the network, thus enabling the network to not only process the input sequences forward, but also in reverse [32], [33], [34].ZOA represents a category of optimization algorithms distinguished by its utilization of only zeroth-order information from the objective function in the optimization process, refraining from incorporating additional information such as the gradient or higher-order derivatives. It excels in striking a balance between local and global search [35], [36] In summary, this research improved the K-Shape clustering method by not only considering the similarity of the time series, but also considering the spatial coordinate characteristics of the measurement points in the case of the spatial measurement points of the dam for the clustering partition, followed by the use of ZOA optimization algorithm for the prediction model parameter optimization, and finally established the prediction model of the deformation of concrete dams through BiLSTM, to achieve a more accurate and effective prediction of the deformation behavior of the dam.

II. IMPROVED KSHAPE CLUSTERING METHOD (I-KSHAPE)

A. K-SHAPE CLUSTERING METHOD

The K-Shape algorithm is an innovative method proposed in 2015 for the clustering problem of time-series data. The method employs the standard inter-correlation distance as the basis of measurement and designs a unique strategy for calculating the center of mass based on the inherent properties of this distance. In order to comprehensively assess the performance of K-Shape clustering, this study introduces two indicators, the contour coefficient and the degree of distortion, for comprehensive consideration, so as to accurately determine the optimal number of clusters k. Through this series of rigorous analytical steps, we were able to draw spatial distribution maps with significant clustering characteristics, which provides strong technical support and theoretical reference for the construction of dam deformation prediction models [30], [31].

Assuming that there are n deformation measurement points in the concrete dam, the number of clusters is n, T denotes the time series of measurement points, and the deformation monitoring dataset is noted as X=($X_1X_2, \cdots, X_k$). Firstly, the K-Shape algorithm's mutual correlation distance method (SBD) is used to calculate the difference between x and y measurement points, that is, the shape information is captured by calculating the shape matrix of the time-series data, in which:

$$SBD\left(\vec{x}, \vec{y}\right) = 1 - \max_{\omega}\left(\frac{CC_{\omega}(\vec{x}, \vec{y})}{\sqrt{R_0(\vec{x}, \vec{x})R_0(\vec{y}, \vec{y})}}\right) \tag{1}$$

where, ω is the possibility of representing all moves; $CC_{\omega}(\vec{x}\vec{y})$ is the number of interrelationships; R_0 is the dot product operation; and the SBD value is between [2, 0], the smaller it is the more similar the timing data are.

Because the center of mass of time-series data is also a time-varying line, there exists the possibility of misaligned peaks and troughs, which does not correctly represent the shape trend. K-Shape determines the center of mass using two approaches. It primarily transforms the center of mass calculation into an optimization problem with an objective function X_i that minimizes the sum of squared distances from the measured values $\vec{x}_i$ of all measurement points to the center of mass $\vec{\mu}_k$. This involves minimizing the sum of squared distances. It is calculated by the formula:

$$\vec{\mu}_k = \underset{\vec{\mu}_k}{\operatorname{argmax}} \sum\nolimits_{\vec{x}_i \in X_k} \left(\frac{\max\limits_{\omega} CC_{\omega}\left(\overrightarrow{x_l}, \overrightarrow{\mu_k}\right)}{\sqrt{R_0\left(\overrightarrow{x_l}, \overrightarrow{x_l}\right) R_0\left(\overrightarrow{\mu_k}, \overrightarrow{\mu_k}\right)}} \right)^2 \tag{2}$$

Expand Formula 4 based on linear algebra methods:

$$\vec{\mu}_k = \underset{\vec{\mu}_k}{\operatorname{argmax}} \frac{\vec{\mu}_k^{T \cdot M \cdot \vec{\mu}_k}}{\vec{\mu}_k^{T \vec{\mu}_k}} \tag{3}$$

Finally, simplify it using the Rayleigh quotient formula:

$$R(M, x) = \frac{x^T M x}{x^T x} \tag{4}$$

In the equation, the maximum value of R is equal to the maximum eigenvalue of matrix M. At this point, transforming the search for the centroid is equivalent to seeking the eigenvector corresponding to the maximum eigenvalue of matrix M.

The optimal cluster number, denoted as k, is comprehensively determined through the elbow method and silhouette coefficient method. In the elbow method, the relationship between the sum of squared errors (SSE) and k results in a graph with an elbow shape. The value of k corresponding to the elbow is considered the optimal cluster number. SSE represents the clustering error for all samples, reflecting the effectiveness of the clustering. The silhouette coefficient method is also employed to assess the quality of clustering, providing a measure of how well-separated clusters are. The formula for the sum of squared errors (SSE) for all measurement points is as follows:

$$SSE = \sum\nolimits_{i=1}^{k} \sum\nolimits_{p \in X_i} |p - \mu_i|^2 \tag{5}$$

where, X_i is the ith cluster; p is the measurement point in X_i; μ_i is the center of mass of X_i. The larger the value of the silhouette coefficient S, the better the clustering effectiveness. The silhouette coefficient for an individual measurement point i is given by:

$$S_i = \frac{b\,(i) - a(i)}{max\,\{a\,(i) - b(i)\}} \tag{6}$$

where, $a(i)$ is the average distance from measurement point i to other measurement points in the same cluster; $b\,(i)$ is the

average distance from measurement point i to all measurement points in other clusters.

The specific steps of the K-Shape algorithm are as follows: 1) Utilize the SBD and MD formulas to compute the distances between measurement points x and y; 2) Based on the newly derived clustering centroids from a comprehensive distance formula, calculate a more reasonable sequence of clustering centroids, denoted as C'; 3) For a dataset X containing n measurement points and a target cluster number k, iterate multiple times until the labels no longer change. In each iteration, recalculate centroids, eventually obtaining updated sequences of clustering centroids for k clusters. The assignment of each sequence to different clusters is determined based on their distances to the new centroids.

B. IMPROVEMENT SECTION

Although the K-Means algorithm usually requires a one-dimensional array as input, it does not mean that it cannot handle high-dimensional data. In practice, one-dimensional arrays consisting of multiple features are usually used as input to the K-Means algorithm. In order to capture more comprehensive time series features and similarities, other metrics such as Manhattan distance can be added to the SBD distance method. This new distance calculation method has better application scenarios. The Manhattan distance formula and the comprehensive distance formula are respectively:

$$MD = d_{ij} = \sum_{h=1}^{dim} \left| x_{ih} - x_{jh} \right| \tag{7}$$

$$d_{ij} = \omega_1 d_{ij}\,(SBD) + \omega_2 d_{ij}\,(MD) \tag{8}$$

where, x is a point in space, dim denotes the dimension, d_{ij} is the Manhattan distance MD between the ith particle and the jth point, and ω is the weight coefficient.

In the process of clustering analysis, considering the relationship of spatial location can capture the actual physical layout more accurately, because the relative positions between points may be closely related to their clustering relationship, and the similarity of the time series cannot be simply considered only. Therefore, in the actual clustering process, incorporating the Euclidean spatial distance formula and defining the threshold condition for spatial location can help to improve the clustering effect. The spatial distance between each pair of measurement points is first calculated using the Euclidean formula (9), which will produce a distance matrix containing the Euclidean distances between all measurement points. When performing clustering, a spatial threshold can be set, for example between 20m and 40m. This threshold will determine which measurement points are considered to be neighbouring. If the Euclidean distance between two measurement points is less than the threshold, they are considered spatially neighbouring points, otherwise they are considered non-neighbouring. Finally the time series data and the spatial distance data are combined into a more complete feature vector that outputs the cluster labels for classification, implying that the combined features of both the time series data and the spatial distance data are taken into account for the cluster analysis. The Euclidean spatial distance between two points is given by the formula:

$$ED = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{9}$$

where x_i and y_i are the two-dimensional spatial coordinates of the measurement point.

III. ZOA-BILSTM DAM PREDICTION MODEL

In order to clearly elaborate the overall framework of the model, the study starts from the initial input variables and introduces each input factor of the HST model, followed by the introduction of the BiLSTM algorithm with powerful time series processing capability to capture the complex dynamic changes in the dam deformation data. Finally, to further improve the prediction accuracy and generalization ability of the model, the ZOA parameter optimization algorithm is integrated in the model to realize the fine tuning of parameters.

A. HST DAM DEFORMATION STATISTICAL MODEL

The deformation δ at any point in a concrete gravity dam can be decomposed into three primary components based on its causes. These components include the water pressure component δ_H, the temperature component δ_H, and the aging component δ_θ,among other factors [37], [38].The formula is as follows:

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

The water pressure component δ_H primarily arises from the upstream water load. According to the principles of material mechanics and dam engineering theory verification, it typically demonstrates a linear relationship with the upstream water depth H, H^2, and H^3. The temperature component δ_T can be approximated by employing a periodic function. This assumes a linear relationship between temperature displacement and concrete temperature. Harmonic sine functions can be chosen as factors. The aging component δ_θ provides a comprehensive representation of both the creep and plastic deformation of dam concrete, as well as the compressive deformation of the geological structure of the rock foundation. The typical pattern of aging displacement changes in a normally operating dam is characterized by an initial sharp variation followed by a gradual stabilization in the later stages. The specific calculation formulas for these three components are:

$$\delta_H = \sum_{i=1}^{3} a_i H^i \tag{11}$$

$$\delta_T = \sum_{i=1}^{2} \left(b_i \sin \frac{2\pi it}{365} + c_i \cos \frac{2\pi it}{365} \right) \tag{12}$$

$$\delta_\theta = d_1 \theta + d_2 \ln \theta \tag{13}$$

Here, H represents the upper reservoir level, and a_i, b_i, c_i, d_1, d_2 are regression coefficients, t denotes the number of days from the initial date of dam observation to the current observation data date, and θ is $t/100$.

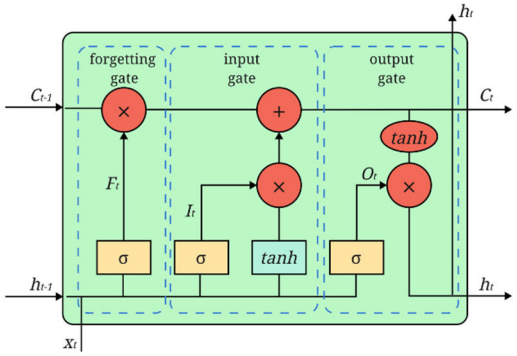


FIGURE 1. Network model of LSTM.

B. BILSTM MODEL

LSTM neural network strengthens its long-term memory capability by introducing control units such as forgetting gate (f_t^l), input gate (i_t^l) and output gate (o_t^l) to maintain and update the cell state [32]. This novel recurrent network structure efficiently addresses the challenges of gradient vanishing and gradient explosion commonly encountered in algorithms like RNN. Leveraging its ability to capture long-term correlations in time-series data, the LSTM network attains faster and more accurate convergence, thereby enhancing the precision of dam deformation prediction [33], [34]. Figure 2 illustrates the LSTM network model featuring three gating structures.

Within the conventional LSTM structure, the responsibility of the forgetting gate lies in determining whether to retain the cell state from the previous time step and selecting relevant information based on a given probability. This mechanism involves processing the hidden layer output from the preceding time step h_{t-1} and the input of the current time step x_t. Subsequently, these two parameters are fed into the sigmoid activation function σ to produce the output of the forgetting gate f_t. f_t and σ are computed as:

$$f_t = \sigma(W_f \cdot (h_{t-1}, x_i) + b_f) \tag{14}$$
$$\sigma\,(x) = 1/(1-e^{-x}) \tag{15}$$

Here, W_f denotes the weight matrix of the forgetting gate, b_f represents the bias term, and σ is the sigmoid activation function. The output f_t of the forgetting gate regulates the extent of forgetting information from the state of the unit in the previous time step. This output takes values in the range of [1, 0], signifying complete retention when $f_t = 1$ and complete forgetting when $f_t = 0$.

The main responsibility of the input gate is to sift through the input information at the current moment to determine what new information should be included in the unit state. This process is accomplished by combining the functions of the sigmoid and tanh layers. the sigmoid layer is responsible for determining the extent to which new information is included, while the tanh layer is responsible for generating possible candidates for new information, also known as $\tilde{a}_t$ the outputs of these two layers together determine the introduction of new information. The relevant formula is:

$$i_t = \sigma(W_i \cdot (h_{t-1}, x_i) + b_i) \tag{16}$$
$$\tilde{a}_t = \tanh(W_c \cdot (h_{t-1}, x_i) + b_c) \tag{17}$$
$$\tanh x = 1-e^{-x})/(1+e^{-x}) \tag{18}$$

Here, W_i and W_c represent the weight matrices for the sigmoid layer and the tanh layer, respectively. Additionally, b_i and b_c are the bias terms corresponding to the sigmoid layer and the tanh layer, respectively. The symbol \(tanh\) refers to the hyperbolic tangent function.

The cell status will be updated after the information is screened by the forget gate and input gate. The new update formula is:

$$C_t = f_t C_{t-1} + i_t a_t \tag{19}$$

The output gate plays a crucial role in extracting valuable information from the current cell state to produce a new hidden layer. This operation initiates by utilizing the sigmoid function to ascertain the portion of the current cell state that should be output. Following this, the tanh function is employed to process the current cell state. Ultimately, the processed information is utilized to generate the new hidden layer h_t. The relevant formulas are as follows:

$$h_t = o_t \tanh C_t \tag{20}$$
$$o_t = \sigma(W_o \cdot (h_{t-1}, x_i) + b_o) \tag{21}$$

where, W_o is the weight matrix of the output gate, and b_o is the bias term.

In summary, the hidden layer output h_t and cell state C_t of the LSTM at the current time step are determined by the hidden layer output h_{t-1} and cell state C_{t-1} of the previous time step, in conjunction with the input x_t at the current time step. While LSTM effectively addresses the issues of gradient vanishing and long-term dependency, it only captures information from the preceding context of the text and cannot leverage information from subsequent portions. Since the semantics of a word are not solely linked to the preceding information but also closely tied to subsequent information, BiLSTM (Bidirectional Long Short-Term Memory) is employed in place of LSTM to incorporate information from both directions. The BiLSTM model comprises two LSTM networks stacked in a bidirectional manner, as illustrated in Figure 3.

As depicted in Fig. 3, within the BiLSTM model, two LSTM gates operating in opposite directions coexist simultaneously at each time step. Here, $\vec{h}_t$ represents the forward output of the LSTM at time step t; $\overleftarrow{h}_t$ represents the reverse output of the LSTM at time step t; h_t signifies the output of the BiLSTM at time step t; and x_t denotes the input at time step t. The state computation for each time step in the BiLSTM model is elucidated in equations (28) and (29). The ultimate output is collectively determined by the states of the LSTMs in both directions. Therefore, the use of BiLSTM instead of the traditional LSTM network can make better use of the contextual information of the text, thus

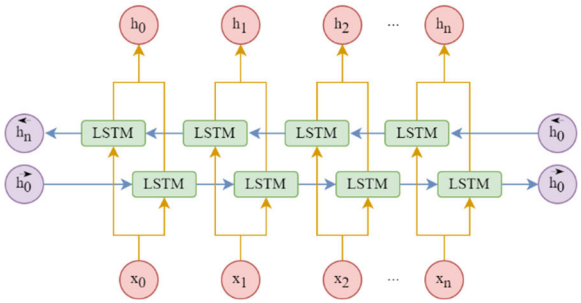


FIGURE 2. Network model of BiLSTM.

improving the model's ability to understand the semantics. By taking into account both the preceding and following information of the text, the BiLSTM model is able to capture the relationships between words more comprehensively, allowing the model to achieve better performance on semantic understanding and related tasks.

$$\vec{h}_t = LSTM(x_t \vec{h}_{t-1}) \tag{22}$$

$$\overleftarrow{h}_t = LSTM(x_t \overleftarrow{h}_{t-1}) \tag{23}$$

$$h_t = \omega_t \vec{h}_t + v_t \overleftarrow{h}_t + b_t \tag{24}$$

C. ZOA OPTIMIZATION ALGORITHM

The Zebra Optimization Algorithm (ZOA) is an emerging optimization algorithm with advantages such as global optimization capability, high efficiency, ease of implementation, wide applicability, low sensitivity to initial values, and good scalability. Therefore, it is considered an effective and reliable optimization algorithm in practical applications. In the ZOA optimization algorithm, zebras act as members of the population and represent potential solutions to the problem. From a mathematical perspective, the position of each zebra in the search space (i.e., the problem domain) directly maps to the values of decision variables. Consequently, vectors can be employed to represent individual zebras, each belonging to the Zebra Optimization Algorithm (ZOA), with the vector elements corresponding to the variable values of the problem. Likewise, the zebra population can be mathematically characterized using a matrix [35], [36]. It's crucial to emphasize that the initial positions of zebras within nZOA population matrix is represented as:

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_t \\ \vdots \\ X_n \end{bmatrix}_{N\times m} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,j} & \cdots & x_{1,m} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ x_{i,1} & \cdots & x_{i,j} & \cdots & x_{i,m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{N,1} & \cdots & x_{N,j} & \cdots & x_{N,m} \end{bmatrix}_{N\times m} \tag{25}$$

Here, X represents the zebra population, where X_i denotes the ith zebra. $x_{i,j}$ represents the value of the jth decision variable suggested by the ith zebra. N is the number of population members (zebras), and m is the number of decision variables. The objective function's value can be obtained by assessing the proposed values of the decision variables for each zebra. The obtained values of the objective function are specified as a vector, expressed by the formula:

$$F = \begin{bmatrix} F_1 \\ \vdots \\ F_i \\ \vdots \\ F_N \end{bmatrix}_{N\times 1} = \begin{bmatrix} F(X_1) \\ \vdots \\ F(X_i) \\ \vdots \\ F(X_N) \end{bmatrix}_{N\times 1} \tag{26}$$

Here, F represents the vector of objective function values, with F_i indicating the objective function value of the ith zebra stripe. Given that both the positions of zebras and the values of the objective function undergo updates in each iteration, the identification of the best candidate solution becomes crucial at each iteration. To achieve this, two natural behaviors observed in zebras in the wild, namely foraging and defense strategies against predators, are leveraged to update members of the Zebra Optimization Algorithm (ZOA) population. Consequently, in each iteration, ZOA members undergo updates in two distinct stages.

During the foraging stage, updates to the population members are performed by simulating zebra foraging behavior. In ZOA, the population's most exceptional member is designated as the zebra pioneer, guiding other members to move closer to its position within the search space. Consequently, the mathematical modeling of zebras' position updates during the foraging stage can be expressed using equations (16) and (17).

$$x_{i,j}^{new,P1} = x_{i,j} + r\cdot(PZ_j - I \cdot x_{i,j}) \tag{27}$$

$$X_i = \begin{cases} X_i^{new,P1}, & F_i^{new,P1} < F_i \\ X_i, & else \end{cases} \tag{28}$$

Here, $X_i^{new,P1}$ represents the updated state of the ith zebra in the first stage, $x_{i,j}^{new,P1}$ is its value in the jth dimension,and $F_i^{new,P1}$ is its objective function value, PZ denotes the pioneer zebra, PZ_j is its value in the jth dimension, r is a random number within the interval [1, 0], and I= round(1 + rand), where rand is a random number within the interval [1, 0]. Therefore, $I \in [1]$, [39], and if the parameter $I = 2$, the variation in population movement will be larger.

In the defense stage, the defense strategy of zebras against predator attacks is simulated to update the positions of ZOA population members in the search space. In the design of the ZOA algorithm, the probability of a zebra being attacked by a lion and other predators is assumed to be the same. When faced with a lion attack, zebras seek to evade the lion by moving closer, as simulated by Equation (18) denoted as S1. In the case of attacks from other predators, the zebra herd converges near the attacked zebra and forms a defensive structure, as modeled by Equation (18) denoted as S2. During the update of zebra positions, only new positions with higher

objective function values are accepted, and the updating condition is expressed in Equation (19).

$$x_{i,j}^{new,P2} = \begin{cases} S1 : x_{i,j} + R \cdot (2r-1) \cdot \left(1 - \frac{t}{T}\right) \cdot x_{i,j} P_s \leq 0.5 \\ S2 : x_{i,j} + r \cdot (AZ_j - I \cdot x_{i,j}), else \end{cases} \tag{29}$$

$$X_i = \begin{cases} X_i^{new,P2} & , F_i^{new,P2} \leq F_i \\ X_i, & else \end{cases} \tag{30}$$

Here, $X_i^{new,P2}$ signifies the updated state of the ith zebra in the second stage, $x_{i,j}^{new,P2}$ denotes its value in the jth dimension, $F_i^{new,P2}$ is its objective function value, t represents the iteration round, T is the maximum iteration count, R is a constant equal to 0.01, P_s is the probability of choosing one of two randomly generated strategies within the interval [1, 0], AZ is the state of the zebra being attacked, and AZ_j is its value in the jth dimension.

During each iteration of ZOA, the population members are updated based on foraging and defense strategies. The continuous update process of the algorithm population follows equations (16) to (19) until the entire algorithm execution is complete. The best candidate solution is continuously updated and retained throughout the consecutive iterations. The specific implementation process is illustrated in Figure 1.

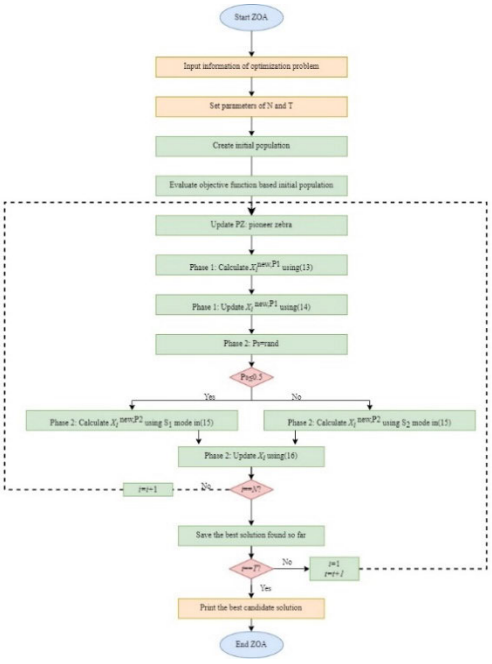


FIGURE 3. ZOA Flowchart.

D. IMPLEMENTATION PROCESS

According to the HST dam deformation statistical model, the initial input variables for the BiLSTM model can be summarized as nine, namely H, H^2, H^3, $\sin\frac{2\pi it}{365}$, $\cos\frac{2\pi it}{365}$, $\sin\frac{4\pi it}{365}$, $\cos\frac{4\pi it}{365}$, θ, and $\ln\theta$. In the BiLSTM network, key indicators affecting prediction accuracy include parameters such as learning rate, training epoch, batch size, and the number of neurons in each layer, which directly determine the LSTM model's network structure. Upon obtaining partition results through the K-Shape clustering algorithm and selecting representative measurement points, the ZOA optimization algorithm is employed to automatically optimize the five parameters of the LSTM model. Subsequently, based on the optimization results, the model undergoes training and prediction. To evaluate the impact of hyperparameter optimization on model performance, a comparative analysis is conducted with non-optimized BiLSTM and LSTM models. The proposed ZOA-BiLSTM dam deformation prediction model framework, based on K-Shape clustering partition, is depicted in Figure 3, with the specific process outlined as follows:

1) DATA PREPROCESSING

In order to improve the predictive ability and clustering effect of the model, it is necessary to clean and preprocess the original observational data first, including completing the missing values, removing the outliers and normalizing the processed data, in order to unify the benchmarks and distributions of different features.

2) CLUSTERING PARTITIONING

The improved K-shape clustering algorithm was used to effectively partition the dam deformation measurement points, and the cohesive clustering of time series and spatial relationship was carried out for all the measurement points based on the distance matrix and the center of mass sequence of each measurement point to obtain the spatial partitioning results. After that, the typical measuring points of different partitions were selected according to the cluster category of each measuring point.

3) PARAMETER OPTIMIZATION

The deformation time series of typical measuring points are divided into training set and testing set according to a certain proportion, and the training set is partially utilized to perform the hyper-parameter optimization using Eqs. (7) to (9) mentioned in Section II-B, and ultimately the stopping conditions are satisfied and the globally optimal parameter combinations are output to obtain the optimal parameters of the LSTM structure.

4) MODEL PREDICTION

In order to build the prediction model and verify its accuracy, the optimal parameters from the optimisation search are substituted into the BiLSTM model and an error measure is agreed upon to verify the prediction accuracy of the BiLSTM model by comparing the predicted values of the test set with the true values, thus illustrating the model's monitoring capability.

The specific realization framework is shown in Figure 4.

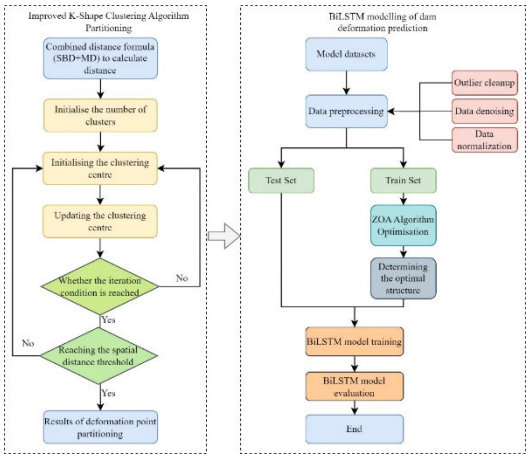


FIGURE 4. ZOA-BiLSTM model prediction process.

E. MODEL EVALUATION INDICATORS

To comprehensively assess the predictive accuracy of the model, various metrics, including Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE), are commonly utilized. The specific formulas for calculating these metrics are presented in Equations (31) to (33).

$$MAE(y_i\hat{y}) = \frac{1}{m}\sum_1^m |y_i - \hat{y}| \tag{31}$$

$$MAPE(y_i\hat{y}) = \frac{100\%}{m}\sum_1^m \frac{|y_i - \hat{y}|}{|y_i|} \tag{32}$$

$$RMSE(y_i\hat{y}) = \sqrt{\frac{1}{m}\sum_1^m (y_i - \hat{y})^2} \tag{33}$$

where, y_i represents the true values of denoised load data; $\hat{y}$ denotes the predicted values; and m is the number of data points to be predicted. MAE reflects the actual situation of prediction errors, MAPE assesses the model accuracy, RMSE evaluates the prediction precision. Smaller values for these three parameters indicate higher model accuracy and better prediction performance.

IV. CASE STUDY

A. PROJECT OVERVIEW

This study focuses on a roller-compacted concrete gravity dam with a height of 112m. The dam's crest elevation is 153.00m, with a width of 6m, and a total installed capacity of 193.2MW. The design flood level for this reservoir is 151.88m, normal storage level is 150.00m, dead storage level is 130.00m, and the total reservoir capacity is 717.3 million cubic meters. The dam consists of 10 dam segments and the main construction started in September 2007, reaching completion in November 2011. To monitor the along-river deformation of the dam, tensile wires were installed at elevations of 153m, 120m, and 88m within the corridors on the dam crest. Additionally, 8 inclinometer points were buried at elevations of 153m, 120m, 88m, 60m, and 43m on both the left and right banks of the dam. In this study, stable tensile wire points were selected as objects for clustering. All tensile wires were partitioned, and typical points were then selected based on the partition results. Due to some tensile wires being unable to collect data or producing significantly abnormal data, the number of valid and available measurement points was reduced to 19. The specific information for each measurement point is shown in Table 1. Each tensile wire measurement point has approximately 180 sets of actual raw data.

TABLE 1. Distribution of measurement points.

dam height / Dam section	2#	3#	4#	4#	5#	5#	6#	7#	8#	9#
153(m)	EX 1-2	EX 1-3	EX 1-4	EX 1-5	EX 1-6	EX 1-7	EX 1-8	EX 1-9	EX 1-10	EX 1-11
120(m)		EX 2-2	EX 2-3		EX 2-4		EX 2-5	EX 2-6	EX 2-7	
88(m)			EX 3-2		EX 3-3		EX 3-4			

B. DATA PREPROCESSING

Due to installation errors or aging issues, some tensile wires exhibited significantly inaccurate monitoring data, leading to the preliminary removal of such measurement points, as in the case of the EX1-1 point near the left bank. Subsequently, to further effectively eliminate outliers, the remaining points underwent denoising processes. This time, two algorithms, Gaussian filtering and Symlet wavelet filtering, were used to denoise the raw data in an overlapping manner, and through the complementary effect of the two algorithms, the data features and noise types were comprehensively processed, and significant outlier processing effects were obtained. For missing data in the middle, linear interpolation was employed for data supplementation, further enhancing the quality and accuracy of the data. Finally, to eliminate the influence of different dimensions and numerical magnitudes among features, the data underwent normalization. The denoising results for the EX2-4 point, compared with the original data, are shown in Figure 5.

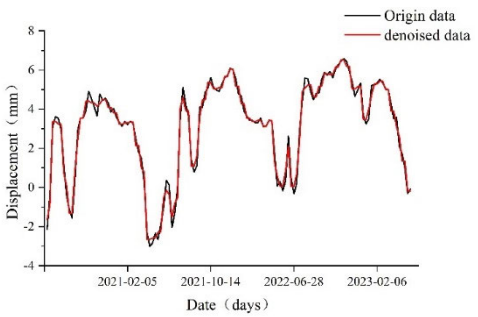
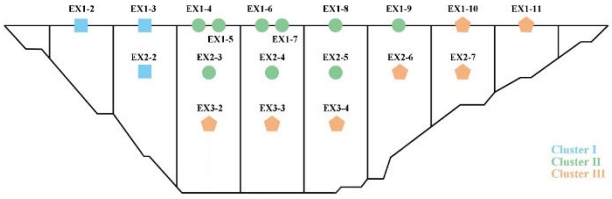


FIGURE 5. De-noising comparison chart.

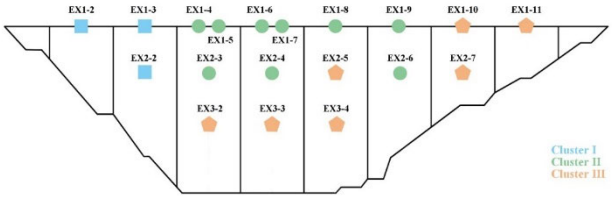
C. CLUSTER ANALYSIS BASED ON I-KSHAPE

In order to gain a more precise understanding of the dam's behavior and variability in different regions, this study

employs the improved K-Shape clustering algorithm detailed in Section II-B to refine the partitioning of the dam's deformation measurement points. By applying the distance calculation method described in Eqs. (7) to (9), the deformation differences among the measuring points were quantified under the same weights. On this basis, the center of mass of each region was determined and the distances from all the measurement points to their corresponding centers of mass were further calculated. After calculating the combined distances based on SBD and MD, the spatial distance conditions between the measurement points were further verified through the Euclidean distance formula to confirm whether the clustering results were in accordance with the actual physical spatial structure distribution. Through several iterations of this process, the position of the center of mass was continuously updated and optimized until the center of the clusters reached a stable state. Eventually, three independent clusters, i.e., the three deformation regions of the dam, were derived, as shown in Figure 6(a). Figure 6(b) shows the clustering results obtained on the basis of the simple SBD distance measurement method.



(a)Improved K-Shape Clustering Algorithm Results



(b)Ordinary K-Shape Clustering Algorithm Results

FIGURE 6. Results of K-Shape clustering partitioning of dam Gauge points.

As can be seen from Figure 6, the improved clustering algorithm produces results that are more in line with the normal spatial and temporal distribution characteristics. Before the improvement, the EX2-5 measurement point is in partition III, and the EX2-6 measurement point is in partition II. Although the EX2-5 measurement point has some similarity with partition III in the time series, the actual spatial distribution of the left, right, and upper parts are all in partition III, and the partition results are obviously unreasonable. Therefore, on this basis, the integrated spatio-temporal distance metric is introduced, the MD and ED distance formulas are added, and the spatial distance threshold is defined, which effectively corrects the defects of the clustering results. By setting the threshold value to 30 m and performing the clustering analysis based on the improved integrated distance formula, it is found that the clustering results of the EX2-5 measurement points are classified as partition II, while the clustering results of the EX2-6 measurement points are classified as partition III. This improvement not only enhances the interpretability of the spatio-temporal distribution of the measurement points, but also further verifies the accuracy of the clustering results after adding the spatial distance condition.

In Figure 6(a), Partition I is near the left bank of the dam, Partition II measuring points are concentrated in the middle and upper part of the dam, and Partition III measuring points are concentrated in the right bank and middle and lower part of the dam. According to the dam monitoring practice and engineering experience, in general, the deformation of the measurement point at the bottom position is usually relatively small, while the deformation of the measurement point at the high position is relatively large. This is due to the uneven water pressure and foundation constraints on the dam structure, resulting in different stresses on the upper and lower parts of the dam, which causes the deformation difference between the measuring points at different locations. In this example the location of the three survey points at 88m elevation is at the lowest elevation and therefore the deformation is relatively small and is a zoning type.At the same time, due to the relatively long length of the right bank section, and by the influence of the inverted plumb line, the deformation of the measuring points near the bank slope is also similar to the deformation of the measuring points at low elevation is not obvious, and therefore also incorporated into the same group in the partitioning process. Comparatively speaking, the upper and middle parts of the dam body are the main load bearing area, so the deformation in this part will be more obvious, and is divided into a separate partition. The left bank measurement points are relatively small and are also divided into one partition. Comprehensively, the clustering partitioning results have a certain correlation with the structural analysis partitioning of concrete dams, and are more in line with the radial displacement distribution characteristics of the dam, which proves that the partitioning results are more reasonable. After clustering, the change rule of deformation of measurement points in the same partition has high similarity, which can comprehensively reflect the regional distribution characteristics of dam deformation in spatial dimension. Figure 7 shows the clustering results. The silhouette coefficient is a metric that combines cluster tightness and separation to evaluate the clustering results, and generally the clusters where the points with the largest silhouette coefficients are located have a higher average tightness and a better degree of separation. The cluster point with the highest contour coefficient in Figure 7 is 3.Distortion is the sum of the squares of the distances from each sample point to the centroid of the cluster to which it belongs, and generally according to the elbow rule, the turning point with a larger slope is considered as the best cluster result. The position of cluster point 3 in Figure 7 has a larger slope relatively speaking. Based on the observation of the two metrics, cluster point 3 indicates a high degree of tightness and good separation.

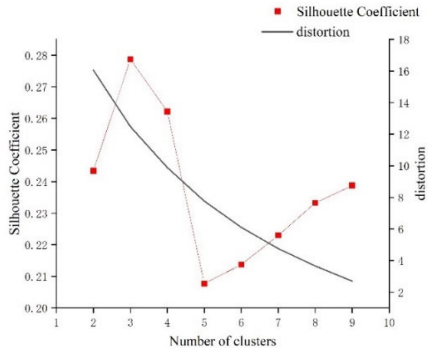


FIGURE 7. Plot of distortion vs. contour coefficient.

D. MODEL PARAMETER OPTIMIZATION

In this research, the optimization of the hyperparameters of the prediction model is carried out for each of the three regional measurement points of clustering, and the main hyperparameters of optimization include the training rate, the number of training times, the number of batches and the number of neurons in two layers, which are combined with the model framework established in the previous paper, and the ZOA algorithm is initialized, and meanwhile, several experiments and optimization processes are carried out in the range of parameter values. The maximum number of iterations in the ZOA algorithm is 1000, the population size is 100, the lower and upper bounds of the search space are from 1 to 100, and the objective function is the MAE function. In the optimization process, it is obviously found that the optimization results of the same type of region have a very high degree of similarity, which means that similar regions have similar data distribution and noise characteristics. On the other hand, there are some differences in the optimization results of different regions, implying that the diversity of data in different geographical locations or environmental conditions may be the main reason for this phenomenon. Table 2 shows the optimization results for each region.

E. CONTRAST ANALYSIS

Since the measuring points in the same region have certain similarity in the change rule and spatial distribution, this paper selected the measuring points in the II region for model hyperparameter optimisation and training, and the selected measuring points are EX1-4, EX1-8, EX2-3, EX2-4, and the selected time period is from 7th May 2019 to 25th September 2023, respectively. After the optimal parameters were obtained through ZOA algorithm optimization, the BiLSTM model was retrained and predicted. Meanwhile, the research initially established a default BiLSTM model featuring 50 neurons in layer 1, 100 neurons in layer 2, 20 training iterations, a batch size of 10, and a learning rate of 0.002. This model, along with an equivalent LSTM model using the same parameters, was then compared to the optimized BiLSTM model. Through this comparison, the paper allows for an intuitive observation of the predictions made by both

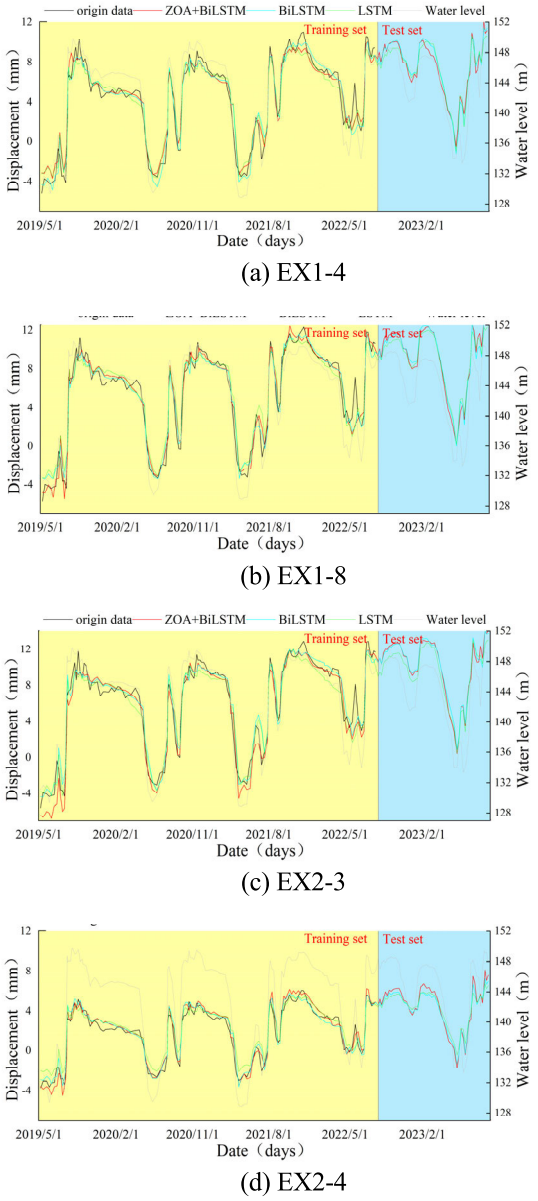
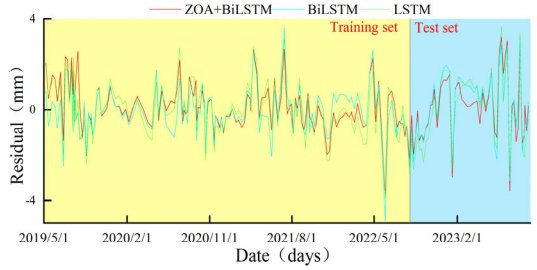
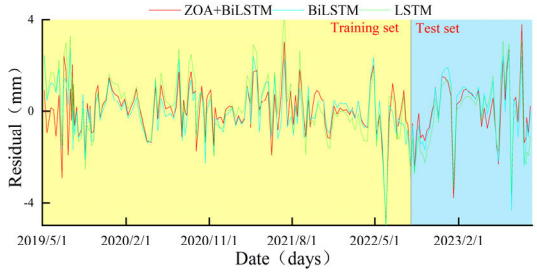


FIGURE 8. Forecast result chart.

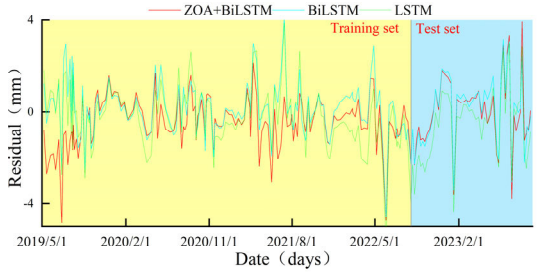
the optimized and unoptimized models. According to the selected four monitoring points, training and predictions were conducted for three models, all using the "Adam" optimizer. Table 3 presents a comparative analysis of performance indicators for the four points after training. It is observed that the ZOA-BiLSTM model exhibits decreased EMAE, RMSE, and MAPE values compared to the other two models. This suggests that the predictive accuracy of the ZOA-BiLSTM model surpasses that of the BiLSTM and LSTM models. The ZOA-BiLSTM model has a MAPE ranging from 0.044% to 3%, while the BiLSTM model's MAPE ranges from 0.322% to 5.559%. The error improvement in predictive accuracy for the ZOA-BiLSTM model is between 0.278% and 2.559%, while the LSTM model's MAPE ranges from 2.709% to



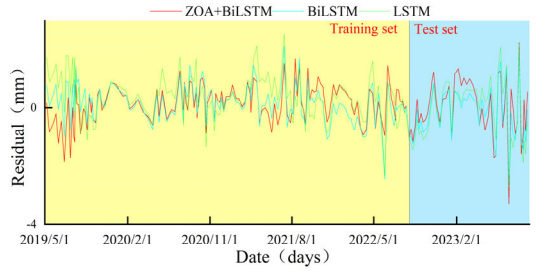
(a) EX1-4



(b) EX1-8



(c)EX2-3



(d)EX2-4

FIGURE 9. Residual plots of different model.

7.942%, resulting in an error improvement of 2.665% to 4.942% in comparison. Considering the overall fitting performance, especially for the EX1-4 monitoring point, the ZOA-BiLSTM model shows a superior predictive effect. Compared to the BiLSTM model, the ZOA-BiLSTM model exhibits a 1.782% improvement in MAPE, and in comparison with the LSTM model, there is a 2.665% enhancement, showcasing a strong predictive performance. The MAE of EX1-4 measurement points is reduced by 1.9 × 10.3mm and 6.22 × 10.3mm compared with the other two models, which effectively reduces the prediction error. The RMSE

TABLE 2. Hyperparameter optimization results.

clustering region	Parameters	Range
I	Training rate(lr)	0.001
	Epoch	256
	Hidden_1	68
	Hidden_2	130
	batch_size	18
II	Training rate(lr)	0.001
	Epoch	150
	Hidden_1	92
	Hidden_2	180
	batch_size	15
III	Training rate(lr)	0.001
	Epoch	220
	Hidden_1	68
	Hidden_2	132
	batch_size	18

of EX1-4 measurement points is reduced by 0.302mm and 0.688mm compared with the other two models, which indicates that the ZOA-BiLSTM model achieves a remarkable enhancement in prediction accuracy and its accuracy indexes are all significantly better than the control model. Through the comparative analysis of the three indicators, it can be seen that the proposed prediction model shows satisfactory stability in the overall performance, and the BiLSTM model optimised by ZOA strengthens the global optimisation seeking ability of the model, which makes the prediction results of the ZOA-BiLSTM model more reliable.

TABLE 3. Comparison of predicted indicators for measuring points.

Monitoring point	Model	MAE(mm)	RMSE (mm)	MAPE(%)
EX1-4	ZOA-BiLSTM	1.86×10^{-3}	0.582	0.044
	BiLSTM	3.76×10^{-3}	0.884	1.826
	LSTM	8.08×10^{-3}	1.270	2.709
EX1-8	ZOA-BiLSTM	3.16×10^{-4}	0.245	3.00
	BiLSTM	1.04×10^{-3}	0.423	4.695
	LSTM	7.90×10^{-3}	0.956	7.523
EX2-3	ZOA-BiLSTM	1.10×10^{-3}	0.384	3.074
	BiLSTM	4.79×10^{-3}	0.545	5.559
	LSTM	6.36×10^{-3}	0.938	7.942
EX2-4	ZOA-BiLSTM	1.71×10^{-3}	0.476	0.110
	BiLSTM	4.77×10^{-3}	0.881	0.322
	LSTM	5.31×10^{-3}	0.949	3.218

In order to deeply verify the model accuracy, this study demonstrates the training and prediction results of the three models for four measurement points. Figure 8 shows the fitting results of the training and prediction models, while Figure 9 presents the corresponding residual plots. From the figures, it can be observed that the fitting results for all four observation points are relatively ideal, with the

ZOA-BiLSTM model exhibiting a better fitting degree, and its residuals are noticeably smaller compared to other models. In the prediction curves for the EX1-4 and EX1-8 observation points, the ZOA-BiLSTM model nearly coincides with the true curves, indicating its effectiveness in predicting the complex nonlinear variations between short-term dam deformations and influencing factors. This precise understanding of changing patterns further demonstrates the high efficiency of the ZOA algorithm in optimizing predictive performance. Based on this validation, it can be concluded that dam monitoring data analysis using artificial intelligence algorithms exhibits good predictive performance, making it suitable for post-monitoring data analysis in dam engineering.

V. CONCLUSION

This research employs the improved K-Shape algorithm for dam clustering and utilizes three models, namely ZOA-BiLSTM, BiLSTM, and LSTM, to establish deformation prediction models for four monitoring points of a gravity dam. Through comparative analysis of prediction and residual effects, the following conclusions are drawn:

(1) The improved K-Shape clustering partitioning results based on SBD, MD and ED distance metrics are in line with the general deformation law of the dam body and can reflect the spatial distribution characteristics of the dam body deformation.

(2) The use of multiple denoising algorithms superimposed on each other can improve the denoising effect of the measured data, complementary monitoring of random outliers in the data at the same time, and further enhance the quality and reliability of the data.

(3) By employing the ZOA optimization algorithm, optimal hyperparameter combinations for the LSTM model can be identified, enhancing the predictive performance and accuracy of the model. This approach aids in a better understanding and utilization of information within dam deformation data, thereby improving the prediction of dam deformation trends.

(4) The ZOA-BiLSTM model constructed based on the clustering partitioning results is able to explore the intrinsic correlation of the measurement points, which is more accurate and reasonable than the BiLSTM and LSTM models, and accurately reflects the overall security state of the dam, which provides a new technique for predicting the deformation of the dam with high accuracy. The method is simple and efficient, and can be modified and applied to the prediction and analysis of other monitoring effects of dams.

It is anticipated that through this approach, the monitoring capabilities for dam safety can be enhanced, thereby reducing potential risks and ensuring the safety of people's lives and property. This, in turn, supports sustained economic development and safeguards the ecological environment. Simultaneously, we hope that this research can provide novel perspectives and methodologies for the monitoring and prediction techniques in concrete dam surveillance.

REFERENCES

[1] X. Fu, C. Gu, and D. Qin, "Deformation features of a super-high arch dam structural system," *Optik*, vol. 130, pp. 681–695, Feb. 2017.

[2] Z. Wu and H. Su, "Dam health diagnosis and evaluation," *Smart Mater. Struct.*, vol. 14, no. 3, pp. 130–136, Jun. 2005.

[3] H. Su, Z. Wen, X. Sun, and X. Yan, "Multisource information fusion-based approach diagnosing structural behavior of dam engineering," *Struct. Control Health Monitor.*, vol. 25, no. 2, p. e2073, Feb. 2018.

[4] X. F. Huang, D. J. Zheng, M. Yang, H. Gu, H. Z. Su, X. B. Cui, and W. H. Cao, "Displacement aging component-based stability analysis for the concrete dam," *Geomech. Eng.*, vol. 14, no. 3, pp. 241–246, Feb. 2018.

[5] S. Chen, C. Gu, C. Lin, K. Zhang, and Y. Zhu, "Multi-kernel optimized relevance vector machine for probabilistic prediction of concrete dam displacement," *Eng. Comput.*, vol. 37, no. 3, pp. 1943–1959, Jan. 2020.

[6] W. Ge, X. Wang, Z. Li, H. Zhang, X. Guo, T. Wang, W. Gao, C. Lin, and P. van Gelder, "Interval analysis of the loss of life caused by dam failure," *J. Water Resour. Planning Manage.*, vol. 147, no. 1, Jan. 2021.

[7] C. Lin, X. Wang, Y. Su, T. Zhang, and C. Lin, "Deformation forecasting of pulp-masonry arch dams via a hybrid model based on CEEMDAN considering the lag of influencing factors," *J. Structural Eng.*, vol. 148, no. 7, Jul. 2022.

[8] X. Cui, H. Gu, C. Gu, W. Cao, and J. Wang, "A novel imputation model for missing concrete dam monitoring data," *Mathematics*, vol. 11, no. 9, p. 2178, May 2023.

[9] Y. Chen, K. Wang, M. Zhao, Y. Xiong, C. Li, and J. Liu, "Identification and reconstruction of anomalous data in dam monitoring considering temporal correlation," *Smart Mater. Struct.*, vol. 32, no. 11, Nov. 2023, Art. no. 115009.

[10] J. Mata, A. T. de Castro, and J. S. da Costa, "Constructing statistical models for arch dam deformation," *Structural Control Health Monitor.*, vol. 21, no. 3, pp. 423–437, Mar. 2014.

[11] B. Dai, C. Gu, E. Zhao, and X. Qin, "Statistical model optimized random forest regression model for concrete dam deformation monitoring," *Struct. Control Health Monitor.*, vol. 25, no. 6, p. e2170, Jun. 2018.

[12] Y. Xing, Y. Chen, S. Huang, P. Wang, and Y. Xiang, "Research on dam deformation prediction model based on optimized SVM," *Processes*, vol. 10, no. 9, p. 1842, Sep. 2022.

[13] X. Li, Z. Wen, and H. Su, "An approach using random forest intelligent algorithm to construct a monitoring model for dam safety," *Eng. Comput.*, vol. 37, no. 1, pp. 39–56, Jan. 2021.

[14] A. Hipni, A. El-shafie, A. Najah, O. A. Karim, A. Hussain, and M. Mukhlisin, "Daily forecasting of dam water levels: Comparing a support vector machine (SVM) model with adaptive neuro fuzzy inference system (ANFIS)," *Water Resour. Manage.*, vol. 27, no. 10, pp. 3803–3823, Aug. 2013.

[15] X. Wang, K. Yang, and C. Shen, "Study on MPGA-BP of gravity dam deformation prediction," *Math. Problems Eng.*, vol. 2017, pp. 1–13, Jan. 2017.

[16] A. Belmokre, M. K. Mihoubi, and D. Santillán, "Analysis of dam behavior by statistical models: Application of the random forest approach," *KSCE J. Civil Eng.*, vol. 23, no. 11, pp. 4800–4811, Nov. 2019.

[17] B. Chen, M. He, Z. Huang, and Z. Wu, "Long-tern field test and numerical simulation of foamed polyurethane insulation on concrete dam in severely cold region," *Construct. Building Mater.*, vol. 212, pp. 618–634, Jul. 2019.

[18] S. Yang, X. Han, C. Kuang, W. Fang, J. Zhang, and T. Yu, "Comparative study on deformation prediction models of wuqiangxi concrete gravity dam based on monitoring data," *Comput. Model. Eng. Sci.*, vol. 131, no. 1, pp. 49–72, 2022.

[19] W. Liu, J. Pan, Y. Ren, Z. Wu, and J. Wang, "Coupling prediction model for long-term displacements of arch dams based on long short-term memory network," *Struct. Control Health Monitor.*, vol. 27, no. 7, Jul. 2020.

[20] X. Qu, J. Yang, and M. Chang, "A deep learning model for concrete dam deformation prediction based on RS-LSTM," *J. Sensors*, vol. 2019, pp. 1–14, Oct. 2019.

[21] J. Yang, J. Qu, Q. Mi, and Q. Li, "A CNN-LSTM model for tailings dam risk prediction," *IEEE Access*, vol. 8, pp. 206491–206502, 2020.

[22] Y. Li, T. Bao, J. Gong, X. Shu, and K. Zhang, "The prediction of dam displacement time series using STL, extra-trees, and stacked LSTM neural network," *IEEE Access*, vol. 8, pp. 94440–94452, 2020.

[23] I. Khan, Z. Luo, J. Z. Huang, and W. Shahzad, "Variable weighting in fuzzy k-means clustering to determine the number of clusters," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 9, pp. 1838–1853, Sep. 2020.

[24] I. Khan, Z. Luo, A. K. Shaikh, and R. Hedjam, "Ensemble clustering using extended fuzzy k-means for cancer data analysis," *Expert Syst. Appl.*, vol. 172, Jun. 2021, Art. no. 114622.
[25] I. Khan, J. Z. Huang, and K. Ivanov, "Incremental density-based ensemble clustering over evolving data streams," *Neurocomputing*, vol. 191, pp. 34–43, May 2016.
[26] J. Song, S. Zhang, F. Tong, J. Yang, Z. Zeng, and S. Yuan, "Outlier detection based on multivariable panel data and k-means clustering for dam deformation monitoring data," *Adv. Civil Eng.*, vol. 2021, pp. 1–11, Dec. 2021.
[27] J. Hu and F. Ma, "Zoned deformation prediction model for super high arch dams using hierarchical clustering and panel data," *Eng. Computations*, vol. 37, no. 9, pp. 2999–3021, Apr. 2020.
[28] W. Chen, X. Wang, Z. Cai, C. Liu, Y. Zhu, and W. Lin, "DP-GMM clustering-based ensemble learning prediction methodology for dam deformation considering spatiotemporal differentiation," *Knowl.-Based Syst.*, vol. 222, Jun. 2021, Art. no. 106964.
[29] J. Madiniyeti, Y. Chao, T. Li, H. Qi, and F. Wang, "Concrete dam deformation prediction model research based on SSA-LSTM," *Appl. Sci.*, vol. 13, no. 13, p. 7375, Jun. 2023.
[30] J. Paparrizos and L. Gravano, "K-shape: Efficient and accurate clustering of time series," *ACM SIGMOD Rec.*, vol. 45, no. 1, pp. 69–76, Jun. 2016.
[31] Y. Zhang, Y. Liu, Z. Yu, W. Xiong, L. Wang, Q. Ai, Z. Li, K. Huang, R. Hao, and Z. Jiang, "Improving aggregated load forecasting using evidence accumulation k-shape clustering," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, Aug. 2020, pp. 1–5.
[32] G. Kang, Y. Xiao, J. Liu, Y. Cao, B. Cao, X. Zhang, and L. Ding, "Tatt-BiLSTM: Web service classification with topical attention-based BiLSTM," *Concurrency Comput., Pract. Exper.*, vol. 33, no. 16, Aug. 2021.
[33] D. Xia, N. Yang, S. Jian, Y. Hu, and H. Li, "SW-BiLSTM: A spark-based weighted BiLSTM model for traffic flow forecasting," *Multimedia Tools Appl.*, vol. 81, no. 17, pp. 23589–23614, Jul. 2022.
[34] S. Hochreiter and J. Schmidhuber, "LSTM can solve hard long time lag problems," in *Proc. Adv. Neural Inf. Process. Syst.*, 1997, pp. 473–479.
[35] S. Mohapatra and P. Mohapatra, "American zebra optimization algorithm for global optimization problems," *Sci. Rep.*, vol. 13, no. 1, Mar. 2023.
[36] E. Trojovská, M. Dehghani, and P. Trojovský, "Zebra optimization algorithm: A new bio-inspired optimization algorithm for solving optimization algorithm," *IEEE Access*, vol. 10, pp. 49445–49473, 2022.
[37] D. Yang, C. Gu, Y. Zhu, B. Dai, K. Zhang, Z. Zhang, and B. Li, "A concrete dam deformation prediction method based on LSTM with attention mechanism," *IEEE Access*, vol. 8, pp. 185177–185186, 2020.
[38] Z. Wu, C. Gu, H. Su, and B. Chen, "Review and prospect of calculation analysis methods in hydro-structure engineering," *J. Hohai Univ. Natural Sci.*, vol. 43, no. 5, pp. 395–405, 2015.
[39] C.-S. Gu, J.-X. Song, and H.-T. Fang, "Analysis model on gradual change principle of effect zones of layer face for rolled control concrete dam," *Appl. Math. Mech.*, vol. 27, no. 11, pp. 1523–1529, Nov. 2006.

MADINIYETI JIEDEERBIEKE was born in Fuyun County, Xinjiang, in April 1992. He received the B.S. degree in water conservancy and hydropower engineering and the M.S. degree in hydraulic structures from Hohai University, in 2017 and 2019, where he is currently pursuing the Ph.D. degree. He is also a full-time Faculty Member with the College of Water Resources and Civil Engineering, Xinjiang Agricultural University, with research interests in the analysis of water conservancy engineering visualization simulation technology. He is presiding more than two longitudinal tasks and has published more than ten articles. His research interests include visual simulation technology analysis of water conservancy engineering and dam safety monitoring.

TONGCHUN LI was born in Taizhou, Jiangsu, in June 1963. He is currently Hohai second-class a Professor and a Ph.D. Supervisor. He published more than 60 articles. Three-Dimensional stochastic finite element analysis reliability technology applied to the Laxiwa, Three Gorges, Ertan, and Xiaowan projects. He proposed the method of solving the internal force of complex structures directly from finite element analysis results, such as arch dam stress analysis and reinforced concrete anchored pile internal force analysis method. He also proposed the finite element iterative solution method for slope stability analysis when the sliding surface is known, which was used in the stability analysis of high slopes in Hongjiatu, Three Gorges, and Baise projects; and with Spanish scholars, he proposed the equi-interpolated stabilized finite element solution method for the interaction between the soil body and the pore fluid dynamics, and the interfacial element method for the dynamics interaction between the dam and the foundation. He won the second prize of the National Scientific and Technological Progress and the Ministry of Electric Power Second Prize for Scientific and Technological Progress.

YANG CHAO was born in Suqian, Jiangsu, in November 1994. He received the B.S. degree in water conservancy and hydropower engineering from the North China University of Water Conservancy and Electric Power, in 2017, and the M.S. degree in hydraulic structures from Hohai University (HHU), in 2020, where he is currently pursuing the Ph.D. degree, with the research interests in safety monitoring and health diagnosis of water conservancy projects, two articles published, and four software copyrights.

HUIJUN QI was born in Handan, Hebei, in 1988. He received the Ph.D. degree in hydraulic structures from Hohai University. He was a Postdoctoral Research in software engineering. He is currently an Associate Researcher and a Master's Tutor with Hohai University. He has successfully presided more than a youth project funded by the National Natural Science Foundation of China, a project under the central university basic operating expenses, and has published more than 20 articles. His primary research interests include the three-dimensional design of water conservancy engineering, optimization, and intelligent analysis of hydraulic structures.

CHAONING LIN was born in Ningbo, Zhejiang, in 1994. She received the Ph.D. degree from Hohai University, in 2021, specializing in hydraulic structures. She is currently a Postdoctoral Scholar with Hohai University. She has published more than 20 articles, has been granted four patents for inventions, has presided more than a project of the National Natural Foundation of China, and has been a Reviewer for several journals, with focusing on the structural health diagnosis, risk assessment, and reliability analysis of dams.