

Received September 8, 2020, accepted September 17, 2020, date of publication October 8, 2020, date of current version October 21, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3029562

A Concrete Dam Deformation Prediction Method Based on LSTM With Attention Mechanism

DASHAN YANG, CHONGSHI GU^{ID}, YANTAO ZHU^{ID}, BO DAI, KANG ZHANG^{ID},
ZHIDUAN ZHANG, AND BO LI

College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210098, China

Corresponding authors: Chongshi Gu (csgu@hhu.edu.cn) and Yantao Zhu (zyt50@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 51739003 and Grant 51909173, in part by the Free Exploration Project of Hohai University under Grant B200201058, in part by the Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions under Grant YS11001, and in part by the Basic Research Project Funded of National Key Laboratory under Grant 20195025912.

ABSTRACT Dams are the main water retaining structures in the hydraulic engineering field. Safe operations of dams are important foundations to ensure the hydraulic functionalities of these engineering structures. Deformation, as the most intuitive feature of the dams' operation behaviors, can comprehensively reflect the dam structural states. In this case, the analysis of the dam prototype deformation data and the establishment of a real-time prediction model become frontier research contents in the field of dam safety monitoring. Considering the multi-nonlinear relationships between dam deformation and relative influential factors as well as the time lag effect of these influential factors, this article adopts long-short-term memory (LSTM) network algorithm in deep learning to deal with the long-term dependence existing in dam deformation and explore the deformation law. The method proposed in this work can effectively avoid the gradient disappearance and gradient explosion problems by using the recurrent neural network (RNN). In addition, this work adopts the Attention mechanism to screen the information that has significant influence on deformation, combining the Adam optimization algorithm that has high calculation efficiency and low memory requirement to improve the learning accuracy and speed of the LSTM. The model overfitting is avoided by applying the Dropout mechanism. The effectiveness of this proposed model in studying the long time series deformation prediction of concrete dams is confirmed by case studies, whose MSE (mean square error) and other 4 error indexes can be reduced.

INDEX TERMS Deformation prediction, LSTM network, Adam optimization, Attention mechanism, Dropout mechanism.

I. INTRODUCTION

As the main water retaining structure in the hydraulic engineering field, the safe operation of the concrete dam is an important foundation to achieve many hydraulic functions. In cases of serious operation accidents, dam failures can bring huge loss of life and property. However, dams operating under the complicated boundary conditions of solid-liquid-gas face have many hidden dangers, as the health of the old dams are limited by surveying, construction, management and other technical conditions, while the newly built high dam reservoirs can adopt technologies that lack systematic verification.

The associate editor coordinating the review of this manuscript and approving it for publication was Fuhui Zhou^{ID}.

Deformation monitoring item is the most intuitive and reliable, and is generally regarded as the most important monitoring feature which can show the dam's operation behaviors and structural states, reflecting the dynamic information of dam body operation in real time [1], [2]. Therefore, the analysis of the prototype observation data of dam deformation and the establishment of real-time prediction model become the frontier research contents in the field of dam safety [3], [4]. The prediction model of deformation can be used as the basis of early warning by comparing the predicted values with the measured values, so as to find out the abnormal situation in the operation in time and take corresponding measures quickly to assist the safety management of the dam operation. On the other hand, the operating conditions can also be calculated inversely by controlling the deformation values, which serves as a guiding basis for the reservoir operation and

assist in the decision making of safe operation. Performing the monitoring prediction has intuitive and reliable guiding significance for both identifying and reacting to anomalies in engineering operations.

In traditional deformation prediction models, most of the deformation values are expressed as polynomials of the main influential factors (water pressure, temperature, aging, etc.) through statistical calculation. Such methods lack the ability to express the multiple nonlinear relationship between dam deformation and influential factors, so their prediction results are not good enough [5], [6].

Relying on the development of computers, traditional machine learning (decision tree, random forest, artificial neural network, etc.) has developed rapidly and has great advantages in solving nonlinear problems, which makes machine learning algorithm widely applied in the prediction models of concrete dam deformation and has achieved good results [7]–[10]. Shiyi *et al.* [11] used the random forest method to establish a concrete dam deformation monitoring model and improved the model's robustness of missing data, unbalanced samples, outliers, and other problems, improving the accuracy of prediction. Dai Bo *et al.* [12] effectively overcame the noise interference in the monitoring sequence and improved the prediction accuracy by integrating chaos theory, wavelet theory and radial basis function neural network.

As computing capacity increased significantly with the support of the large data, deep learning [13] (such as CNN, RNN, DBN) becomes one of the main research directions in the field of machine learning. By studying the inherent law of sample data and representation levels, the prediction accuracy of model for complex nonlinear problems and the deformation regularity of mining experience a certain degree of ascension [14]–[16].

The Recurrent Neural network (RNN), has a unique advantage in processing sequential data. Compared with the shallow neural network, RNN can better describe the multicollinearity among variables through the deeper hidden layers. More importantly, different from the traditional FNNs, the weight connection between the same layer neurons of RNN is also established to build the directional information cycle, so as to deal with the problem of the front and back correlations between inputs. This information transmission ability is interpreted visually as memory. Compared with the representation method in the statistical model that reflects the hysteretic nature of the impact factor and includes the prophase term [17], memory is more flexible in searching for effective information periods, with longer mining period.

The increase of information amount and deeper hidden layers lead to the gradient disappear or gradient explosion phenomenon in the RNN training [18]–[20]. This shortcoming not only limits the possible hidden layer numbers, but also hinders learning laws and relations during the large time interval, making RNN in practice be more preferable to solve short-term dependency problems, because it

is difficult to learn the long-term dependency relationship implicitly from the data. LSTM is a special kind of RNN. It can selectively remove or add information through the "gate" structure of information, the information stored in the form of cell state storage, avoid huge storage and computational costs of long-term dependence problem and have both long-term and short-term dependency relationship, which is more advantageous to mining hidden rules in information. Yang Beibei *et al.* [21] used LSTM to build a model for landslide displacements and found that the model had obvious advantages in step type deformation prediction. Jiaqi [22] combined CNN and LSTM to predict the traffic flow in urban areas, and utilized the advantages of CNN and LSTM in mining the spatial relationship and time dependence of data, finding that the learning efficiency and prediction accuracy were greatly improved. Qu *et al.* [23] combined rough set theory with LSTM theory to improve the accuracy, robustness and externality of model. Wenju *et al.* [24] compared the LSTM-MA (moving average method) model with the LSTM-PCA (principal component analysis) model and found the LSTM-MA model was more suitable for engineering applications due to its convenience. Yangtao *et al.* [25] used STL (Seasonal-Trend decomposition procedure based on Loess) decomposing the deformation into seasonal component, trend component and remainder component, and then predicted seasonal component by extra-trees method while others by stacked LSTM neural network, improving the prediction accuracy and stability of dam deformation prediction by combination of models. The data processing method, through continuous integration of advanced deep learning (neural network), optimization theory and mechanism into the construction of dam deformation prediction model, combined with the existing dam engineering theory and engineering experience, will have important guiding significance for the engineering practice.

II. STATISTICAL MODEL OF CONCRETE DAM DEFORMATION PREDICTION

A. PRINCIPLE OF STATISTICAL MODEL

The influence of load sets on the concrete dam can be found out by the regression calculation on deformation data. Then a mathematical expression can be established to calculate the predictive value (deformation value) of a certain set of load sets [26].

The dam deformation is a displacement vector sum of creep, plastic, and elastic deformation of the concrete dam and bedrock subjected to loads, such as water pressure, uplift pressure, and temperature. Taking the horizontal displacement as an example, the displacement of the dam body under reservoir water pressure is generated, as shown in Figure 1 (a). The displacement of the dam caused by foundation deformation due to internal forces acting on the foundation is shown in Figure 1 (b). The displacement of the dam body caused by foundation rotation due to heavy reservoir water is shown in Figure 1 (c).

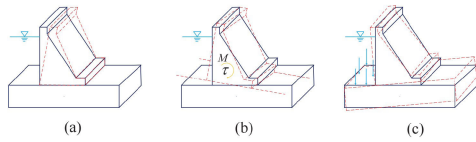


FIGURE 1. Horizontal displacement under the action of reservoir water. The black solid line is the contour line before the displacement of the dam, and the red dotted line is the contour line after the displacement.

B. DEFORMATION COMPONENT

Considering the horizontal displacement, the deformation of a concrete dam can be divided into hydraulic, temperature and aging components as shown below:

$$\delta_x = \delta_H + \delta_T + \delta_\theta \quad (1)$$

where δ_x is horizontal displacement; δ_H is hydraulic component; δ_T is the temperature component; δ_θ is the aging component.

1) THE HYDRAULIC COMPONENT δ_H

The deformation component of the dam body under water pressure and reservoir water weight can be expressed as:

$$\delta_H = \sum_{i=1}^n a_i H^i \quad (2)$$

where, a_i is the regression coefficient of water pressure factor; H is the water depth in front of the dam; For gravity dams, usually $n = 3$, and for arch dams, usually $n = 4$.

2) THE TEMPERATURE COMPONENT δ_T

When considering the mechanics, the temperature component should be the companion temperature measured by thermometers embedded in the concrete and bedrock of the dam body. However, after many years of dam operation, the hydration heat of dam-building concrete has been fully dispersed, and the temperature inside the dam reaches the quasi-stable temperature field. In this case, the dam body temperature is only affected by the boundary temperatures, namely water temperature and air temperature. Assuming that the change of water temperature and air temperature is harmonic, and the deformation is linearly related to the temperature of concrete, the temperature component can be expressed as:

$$\delta_T = \sum_{i=1}^m \left(b_{1i} \sin \frac{2\pi it}{365} + b_{2i} \cos \frac{2\pi it}{365} \right) \quad (3)$$

where, b_{1i}, b_{2i} are the regression coefficients of the temperature factor, generally $m = 1$ or $m = 2$; t is the cumulative number of days from the corresponding monitoring day to the initial monitoring day.

3) THE AGING COMPONENT δ_θ

The cause of the aging component is complex. Generally, the aging component conforms to the change law with a rapid

growth in the early stage and gradual stability in the later stage. Its expression can be written as:

$$\delta_\theta = c_1 \theta + c_2 \ln \theta \quad (4)$$

where, c_1, c_2 are the regression coefficients of aging factor; θ is $t/100$.

C. STATISTICAL MODEL OF DEFORMATION MEASUREMENT POINTS

To sum up, the expressions of the above three components are substituted into Equation (1) to obtain the statistical model of the measured point deformation of the concrete dam. By substituting the corresponding load sets and effect sets into the expression, the problem is transformed into a mathematical problem of the least squares optimization calculation. Finally, the regression coefficients with the minimum error are solved, and the explicit mathematical expression of dam deformation is obtained:

$$\delta_x = \sum_{i=1}^n a_i H^i + \sum_{i=1}^m \left(b_{1i} \sin \frac{2\pi it}{365} + b_{2i} \cos \frac{2\pi it}{365} \right) + c_1 \theta + c_2 \ln \theta + a_0 \quad (5)$$

where, a_0 is the constant term.

III. LSTM CONCRETE DAM DEFORMATION PREDICTION MODEL BASED ON ATTENTION MECHANISM

A. LONG-SHORT-TERM MEMORY NETWORK

In view of the association between pre and post deformation of a concrete dam, and the shortcoming of RNN learning that the gradient disappearance and gradient explosion are prone to occur, LSTM network is introduced to construct a concrete dam deformation prediction model of measuring points. LSTM as an improvement of RNN, can mine the long-term and short-term correlation laws of deformation sequences and realize the synchronous prediction of the deformation at measuring points.

Figure 2 shows a node in the LSTM structure [27]. The core function of the LSTM lies in its ability to remember information, also known as a cell. The cell state of the previous node C_{t-1} , is passed along a separate chain, and the LSTM updates the cell state through a ‘‘gate’’ structure. At the same time, the hidden state of the output of the previous node h_{t-1} is

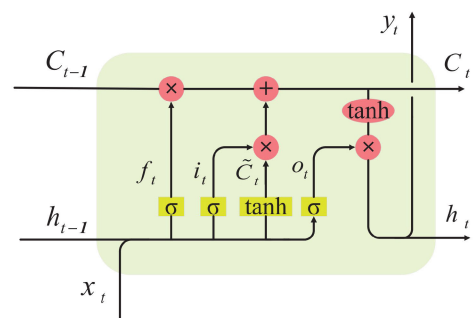


FIGURE 2. The schematic diagram of a LSTM node.

passed to the current node along another chain, which forms the input layer of the node with the input x_t of the current node.

In Figure 2, the process line f_t represents the forgetting gate, which is used to determine the discarding of information in the cell. The equation is as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (6)$$

where, W_f is the weight matrix; b_f is the bias term of the forgetting gate; By means of the *Sigmoid* function, the value is in the interval (0, 1), where “0” means complete abandonment, and “1” means complete retention.

The process lines of i_t and \tilde{C}_t combine to form a renewal gate which is used to update the information stored in the cell.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (7)$$

where, W_i is the weight matrix, and b_i is the bias term.

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (8)$$

where, W_C is the weight matrix, and b_C is the bias term.

At this point, the cell state has been updated by the formula.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (9)$$

The output h_t of the current node can be calculated from the cell state of the current node.

$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o) \quad (10)$$

$$h_t = o_t * \tanh(C_t) \quad (11)$$

where, o_t is the intermediate computation; W_o is the weight matrix, and b_o is the bias term.

The output h_t here will be the input to the next node as the hidden state, and the final output (deformation fitting value) typically needs to put h_t into a linear layer and then classified using *Softmax* to get the required data.

Like other neural network algorithms, LSTM can learn the complex nonlinear relationship between effect sets and load sets through training. More importantly, its advantage lies in the existence of cell states that enable the LSTM to selectively retain the previous deformation and the corresponding load when calculating the deformation under the current load. Consequently, the neural network has the “memorability” and the ability to learn the laws between the deformation and the previous load set.

B. ADAM OPTIMIZATION ALGORITHM

Adam is a first-order optimization algorithm put forward by Kingma in 2015 [28], which combines the exponential moving average of the Momentum algorithm and the gradient update rules of the RMSprop algorithm. By estimating and correcting the first and second moments of the gradient, the learning rate of each parameter is dynamically adjusted so that the learning rate of each iteration has a certain range and the parameters variation is relatively stable.

Adam algorithm requires less memory and therefore becomes especially suitable to solve problems involved

excessive data and parameter sizes in deep learning. Meanwhile, it is also very suitable for the solution of unsteady targets, problems involving high noises or sparse gradients.

After parameter initialization and random objective function $f(\theta)$ determination, the gradient of objective function $f_i(\theta)$ with respect to θ at step length t can be obtained from the formula. This gradient is the partial derivative of $f_i(\theta)$ with respect to θ under the step t .

$$g_t = \nabla_{\theta} f_i(\theta) \quad (12)$$

According to the obtained gradient, the first order and second order momentum matrices are estimated and updated preliminarily.

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t \quad (13)$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2 \quad (14)$$

The moving mean is estimated using the first moment (mean) and the second original moment (with a partial variance) of the gradient. However, because these moving averages are initialized as 0 vectors, the moment estimated value deviates toward 0, especially in the initial time step and in the case where the decay rate is very small (that is, when the β is close to 1). So, bias correction is required.

$$m'_t = \frac{m_t}{1 - \beta_1^t} \quad (15)$$

$$v'_t = \frac{m_t}{1 - \beta_2^t} \quad (16)$$

Finally, the parameter θ is updated.

$$\theta_t = \theta_{t-1} - \frac{\alpha \cdot m'_t}{\sqrt{v'_t} + \varepsilon} \quad (17)$$

The Adam algorithm is a gradient update optimization algorithm that is often used to replace stochastic gradient descent and other optimization algorithms in the deep learning model. It has strong robustness and relatively simple parameter adjustment process, so the default parameters can handle most of the problem with high computational efficiency.

C. ATTENTION MECHANISM

The Attention mechanism is inspired by the selective attention of human vision, in which the eyes scan the global image to find the focus area and pay more attention, while suppressing the acquisition of useless information. The application of Attention mechanism in deep learning can quickly screen the key conditions that help to make decisions from big data, thus reducing the computation costs, saving the storage space, and improving learning efficiency and accuracy. The soft attention refers to the mechanism of weighted averaging of the input information attention [29].

At present, the Attention mechanism is mainly used in combination with Encoder-Decoder framework. The input (Source) will be expressed in the form of (K, V). K is the key value of the element, which is its position, and V is

the value of the element, which is its assigned attention. Each element of the output (Target) is represented by q , which is query. To calculate attention, it is necessary to first calculate the attention scores s_i of each input quantity:

$$s_i = F(q, k_i) \tag{18}$$

where, there are many kinds of attention rating functions, such as: Additive model $F(q, k_i) = v^T \tanh(W \cdot k_i + U \cdot q)$; The dot product model $F(q, k_i) = k_i^T \cdot q$; Scale dot product model $F(q, k_i) = k_i^T \cdot q / \sqrt{d}$; Bilinear model $F(q, k_i) = k_i^T \cdot W \cdot q$. Among these models, W , U and v are network parameters that can be learned and trained, and d is the dimension of input information.

Then, the attention scores are converted numerically. Generally, there are two ways of conversion. One can use the probability distribution of normalization, which makes the sum of the weight coefficients of attention scores equal to 1; the other can use *SoftMax* function to highlight the weight of important elements.

$$\alpha_i = \text{SoftMax}(s_i) = \frac{\exp(s_i)}{\sum_{j=1}^N \exp(s_j)} \tag{19}$$

where, α_i is the weight coefficient corresponding to the element in Source. N is the amount of input information.

Finally, the weighted average is used to get the attention value of the input information.

$$\text{Attention}(K, V, q) = \sum_{i=1}^N \alpha_i v_i \tag{20}$$

y_i in the Figure 3 is the output of the Attention mechanism.

The function of the Attention mechanism is to consider the weight of the input elements, so as to pay more attention to the sequence fragments with high similarity to the current input, weaken the attention to the fragments with large differences in the sequence, and consider the role of global and local relations. For the prediction model of the concrete dam deformation, more attention can be paid to the rule learning and the factor significance selection under the conditions with similar deformation. Weights can be assigned according to the significance of factors.

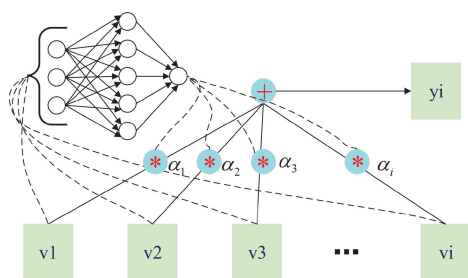


FIGURE 3. Schematic diagram of attention mechanism.

D. DROPOUT ALGORITHM

Dropout [30] is an optimization method proposed by Hinton in 2012 to improve the generalization ability of neural networks. It has good effects on preventing model overfitting and alleviating the time-consuming problem of deep learning in model integration training.

According to the set abandonment rate, the Dropout mechanism can randomly block some neurons in the hidden layer temporarily, so that only the remaining neurons can keep normal training.

This random discard mechanism makes it impossible for two neurons to work at the same time every iteration, and prevented the common action of feature detector. Therefore, it avoids the dependence of weight update on the fixed relation of neurons, and reduces the influence of the correlation between feature vectors on the prediction results, so this mechanism enhances the robustness of the neural network losing specific neuronal connections.

At the same time, by randomly discarding some neurons, many different subnetworks are extracted from the complete model. By averaging the results of different sub-neural networks, different overfitting of different networks can be offset to achieve a balanced effect.

Neural network operation without joining Dropout:

$$h_i^{l+1} = f(W_i^{l+1}h^l + b_i^{l+1}) \tag{21}$$

where, h is the input or output of a neuron; l is the update times of the matrix; and i is the number of the neurons. Here h^l is the output of a neuron of the upper layer as the input of the Dropout layer, while h^{l+1} is the output of the Dropout layer as the input of the next neuron.

Neural network operation after joining Dropout:

$$r_j^l \sim \text{Bernoulli}(p) \tag{22}$$

$$\tilde{h}^l = r^l * h^l \tag{23}$$

$$h_i^{l+1} = f(W_i^{l+1}\tilde{h}^l + b_i^{l+1}) \tag{24}$$

where, Bernoulli (p) is the discrete probability distribution function to generate probability vector, that is, a randomly generated 0 and 1 vector.

By inserting the Dropout layer between the LSTM layers, the model overfitting is prevented and the robustness of the model to special conditions is enhanced.

IV. CONSTRUCTION OF THE PREDICTION MODEL

On the premise of knowing the measured deformation data and environmental impact factors of concrete dam measurement points, the LSTM network is used to extract the long-term and short-term features contained in the data; the Adam conducting optimization, Attention mechanism highlight the influence of important features on prediction; and the Dropout mechanism prevents model from overfitting. The synchronous prediction model of measured point deformation of concrete dam is established by utilizing all the above algorithms.

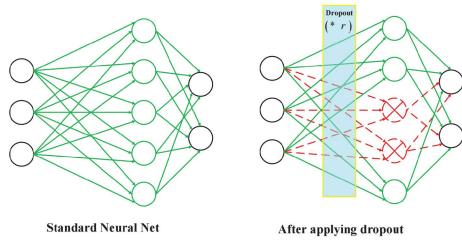


FIGURE 4. Schematic diagram of Dropout mechanism. Green indicates that the neurons are active and can normally carry out input, output and matrix update, while red indicates that the neurons are temporarily dormant, blocking all its work.

The construction process of the model is shown in Figure 5, and the analysis steps are as follows:

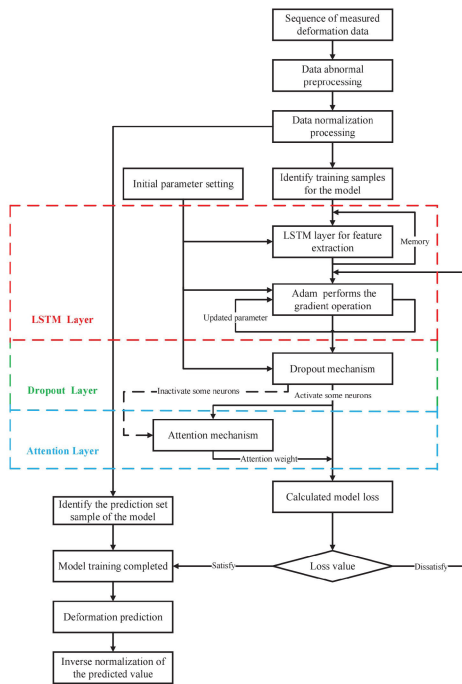


FIGURE 5. Schematic diagram of Dropout mechanism.

Step1: Preprocess the deformation monitoring data. For the monitoring data series with strong regularity, the Pauta Criterion can be used to judge whether the data are abnormal or not.

$$y_i \leq \hat{y}_i \pm 2.58\sigma \quad (25)$$

where, y_i is the measured value; \hat{y}_i is the corresponding predicted value in the prediction model established through the analysis of the historical data sequence y_i ; σ is the standard deviation Δ of the residual sequence. If Equation (25) is satisfied, the measured value is considered to be normal; otherwise, the measured value is considered abnormal and has to be removed.

Step2: Normalize the data and determine the training samples. After the measured value sequence is checked and if there are no obvious abnormalities, in order to align the

variables to the same orders of magnitude (the presence of a large number of orders of magnitude variables can make the model neglect the influence of small orders of magnitude variables and lose the information contained in these kinds of variables), variables shall be normalized. The training sample and prediction sample of the model are determined after data normalization.

$$x'_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (26)$$

where, x_i is the original variable; x'_i is the variable obtained after min-max standardization; x_{\max} and x_{\min} are the maximum and minimum values in the original data sequence, respectively.

Step3: Initialize the parameters. Initialize Adam parameters. The learning rate (step factor) is usually set to $\alpha = 0.001$, to control the updating ratio of weights. The exponential decay rate of the first-order moment estimation is set as $\beta_1 = 0.9$. The exponential decay rate of the second-moment estimation is set as $\beta_2 = 0.999$. $\epsilon = 1E - 0.8$, ϵ is set to a very small number to prevent the possible error of dividing by zero in the calculation. The Dropout level drop probability p is set to be 0.3.

Step4: The training set data go through model LSTM layer, Dropout layer and Attention layer in turn to conduct model training. During model training, the cell states in the LSTM are constantly updated, so that the structure has the “memory function”. In the Adam optimization algorithm, parameters such as step size will be updated in real time as the gradient drops.

Step5: Determine whether the model needs to continue training according to the loss value. If the requirements are met, the predicted samples can be imported into the trained model to predict the deformation of the measured point of the concrete dam.

Where, in Step4, the specific construction of the model is shown in Figure 6.

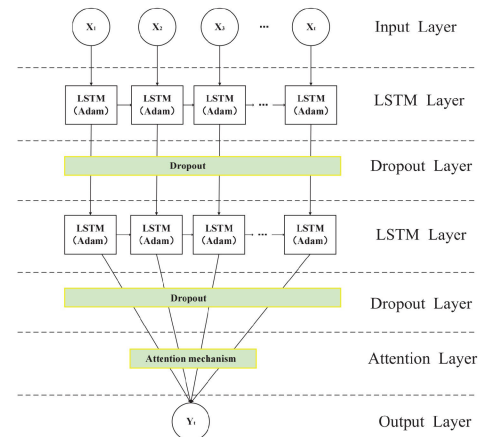


FIGURE 6. LSTM structure of the attention mechanism.

After the input vector enters the hidden layer of the LSTM network through the input layer, the Adam optimization algorithm is used to solve the weight matrix, so as to improve

the solving efficiency, reduce the memory consumption and ensure the output accuracy. The output of the LSTM layer will be used as the input of the Dropout layer to prevent the overfitting problem during model training by masking part of the eigenvalues. The output of the first two layers of LSTM network and the two layers of Dropout are encoded as inputs to the Attention mechanism.

According to the importance of each characteristic, the Attention module assigns corresponding weight to each characteristic, and obtains the characteristic representation of the deformation value as output according to the weight and the updated cell state.

After the LSTM network updating the cell state and the Attention mechanism adjusting the weight, the calculated eigenvalue is decoded by the activation function at the full connection layer to obtain the final output – deformation value.

V. CASE STUDIES

A. PROJECT OVERVIEW

In Sichuan province, China, there is a concrete double-curved arch dam with a height of more than 200m and is divided into 39 dam sections. The power station junction consists of barrage, flood discharge, energy dissipation and water conveyance structures. Figure 7 is the engineering plane layout (local), and Figure 8 is the upstream elevation layout of the horizontal displacement measurement points. The vertical lines are arranged on the arch dam section 4#, 11#, 21#, 33# and 37# of the crest, the horizontal corridor and foundation corridor inside the dam. Among them, the red ones are the vertical points, and the magenta ones are the inverted vertical points.

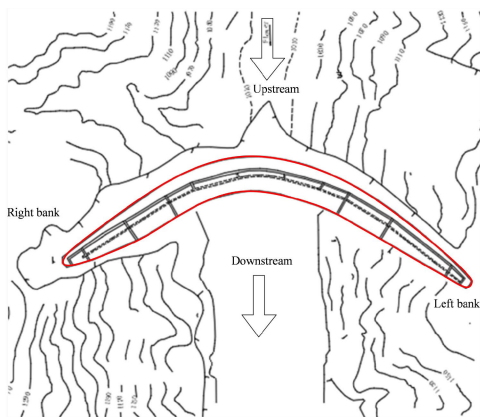


FIGURE 7. Engineering layout plan (local).

Considering the low frequency of manual observation, the measurement points are selected from the automatic vertical measurement points for the comparative analysis. Among them, TCN15 vertical measuring point of the dam crest on section 33# has high reliability and few instrument faults such as missing or jumping, which can provide relatively accurate deformation monitoring data with longer data series.

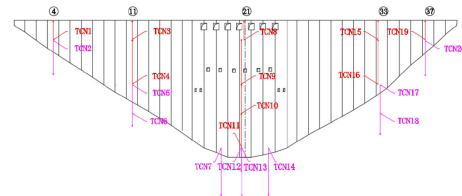


FIGURE 8. Elevation drawing of upstream vertical layout of arch dam.

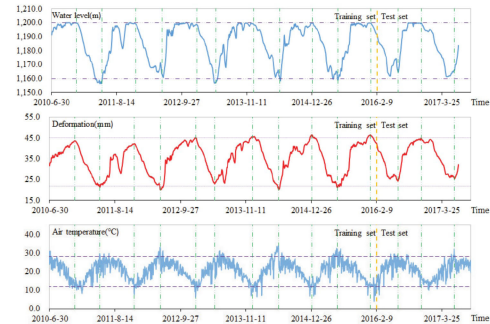


FIGURE 9. Measured process lines of the upstream water level, air temperature and deformation. The left side of the orange dotted line is the training set sequence, while the right side is the test set sequence. The green dotted line is roughly the date of the extreme displacement value in each deformation period, and the purple line is used to mark the approximate change interval of each monitored quantity.

Stepwise regression model, RNN model and LSTM model with attention mechanism were respectively established for the comparative analysis of the monitoring data in a modeling period of 2562 days from July 1st, 2010 to July 5th, 2017. The 2050 days data from July 1st, 2010 to February 9th, 2016 were selected as the training set, and the remaining data were used as the test set.

B. SELECTION AND PROCESSING OF THE PREDICTION MODEL FACTORS

Select the influence subsets of the model input, $H - H_0$, $(H - H_0)^2$, $(H - H_0)^3$, $(H - H_0)^4$, $\sin(2\pi t/365) - \sin(2\pi t_0/365)$, $\cos(2\pi t/365) - \cos(2\pi t_0/365)$, $\cos(4\pi t/365) - \cos(4\pi t_0/365)$, $\sin(4\pi t/365) - \sin(4\pi t_0/365)$, $\theta - \theta_0$, $\ln \theta - \ln \theta_0$. The above 10 impact factors are commonly used in the statistical model of arch dams according to theoretical knowledge and engineering experience of experts.

The abnormal data processing was carried out for the influence factor subset and deformation data, and the interpolation processing was carried out for the missing data, if necessary, to ensure the continuity of the data. To enhance the model’s learning efficiency, the normalization processing is carried out for the processed data.

C. CONSTRUCTION OF THE DEFORMATION PREDICTION MODEL

The normalized influence factor subset was taken as the input of the model, and the normalized deformation value was taken as the output of the model. The loss function is selected as

follow:

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (27)$$

where y_i is the measured displacement value; \hat{y}_i is the calculated value of the model; m is the number of the data series.

The training set data were used to make the model fully learn the deformation law of the dam, and the trained prediction model was used to predict the deformation of the prediction set data. In order to verify the validity of the model, the results were compared with those generated by the statistical model, the LSTM model and the RNN model with the Attention mechanism. In the neural network model, the Adam algorithm is used for parameter optimization. In order to prevented overfitting, two Dropout layers were inserted into the model. The prediction results of the four models are shown in Figure 11, and the prediction effects are shown in the Table 1. The remaining error evaluation indexes and formulas are as follows:

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (28)$$

$$SMAPE = \frac{100\%}{m} \sum_{i=1}^m \frac{|\hat{y}_i - y_i|}{(|\hat{y}_i| + |y_i|) / 2} \quad (29)$$

$$MAPE = \frac{100\%}{m} \sum_{i=1}^m \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (30)$$

$$MAE = \frac{1}{m} \sum_{i=1}^m |\hat{y}_i - y_i| \quad (31)$$

where RMSE is root mean square error; MAE is the mean absolute error; MAPE is mean absolute percentage error; and SMAPE is symmetric mean absolute percentage error.

From Figure 10 and Figure 11, each model has achieved relatively good imitative and forecast effect. The calculated values of the model fit well with the measured values in most periods, while the deviation is obvious in the period near the extreme value of deformation every year, especially around the annual maximum. As we can see from Figure 9, most of the radial displacement maxima are located in the maximum range of the water level in the data series (corresponds to the dotted green line in the Figure 9). Meanwhile, the extreme values of the water level over the years are close to each other (the purple horizontal line above the water level in the Figure 9). However, The temperature fluctuates frequently during the period of high water level, which is not completely consistent with the assumption that the temperature changes harmonically, resulting in the prediction deviation of the model.

From Figure 11, compared with other models, the LSTM model with attention mechanism is closer to the measured deformation values, especially in May and June each year when the radial displacement is relatively small, under the condition of relatively high temperatures and low water levels. Among them, the calculated values of statistical model

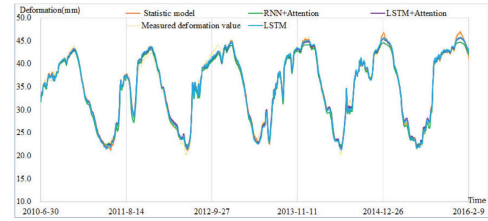


FIGURE 10. Results comparison of different models.

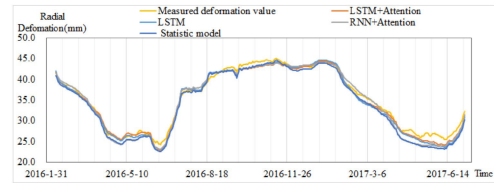


FIGURE 11. Prediction results comparison of different models.

TABLE 1. Statistical table of prediction models' accuracy.

	MSE	RMSE	MAE	MAPE	SMAPE
Statistic model	1.77	1.33	1.09	3.52%	3.63%
RNN+Attention	1.30	1.14	0.91	2.96%	3.03%
LSTM	1.06	1.03	0.81	2.56%	2.62%
LSTM+Attention	0.69	0.83	0.67	2.06%	2.09%

and the extremum of radial displacement in a few years belong to training set are very similar. But in other years, the prediction deviation of statistical model for extremum often bigger than other models. This may be due to the poor ability of the statistical model to express the nonlinear relationship between variables, and the inability to describe the complexity of the deformation under the combination of special conditions.

As can be seen from Table 1, the MSE of LSTM model with attention mechanism model decreases from 1.77mm to 0.69mm compared with the statistical model, decreasing by 61%. Compared with the RNN model, the MAE of LSTM model with Attention also decreased by about 26%, which shows the advantage of LSTM over RNN in long sequence to some extent. Meanwhile, by adding attention mechanism, the errors of LSTM model are reduced by about 20%, which suggests that the Attention mechanism is helpful to improve the prediction accuracy through the weight distribution among the influencing factors.

VI. CONCLUSION

In this article, the construction method of a concrete dam deformation prediction model is studied. Here are the summary of this work:

1) The proposed model utilizes LSTM in deep learning as the carrier to give full play to its "memorability" advantage in the time series data mining;

2) The Adam algorithm is used in this model as a gradient descent algorithm to optimize the parameters in a time-saving and efficient way;

3) The Dropout layer is inserted between the LSTM layers to prevent the overfitting phenomenon during training and increase the model's robustness by randomly shielding the work of some neurons;

4) The addition of the Attention mechanism in front of the output layer enables the model to give corresponding weight to the characteristics of influencing deformation when making predictions. Meanwhile, the addition of the Attention mechanism equips the model with the ability to consider the role of global and local connections between deformation and long series of influencing factors;

5) The validity of the method is verified by an engineering example. Compared with other prediction models, this proposed model achieves smaller errors in various aspects.

REFERENCES

- [1] W. Bowen, Y. Dongyan, C. Lei, W. Yongbing, and X. Zhenkai, "Multi-scale deformation combination forecast model for concrete dam based on BP-ARIMA," *Hydro-Sci. Eng.*, vol. 168, no. 2, pp. 55–63, 2018.
- [2] W. Bangbin, N. Jingtai, S. Huaizhi, Y. Meng, W. Zhongru, and C. Xinbo, "An approach for deformation modulus mechanism of super-high arch dams," *Struct. Eng. Mech.*, vol. 69, no. 5, pp. 557–566, 2019.
- [3] C. Peijiang, G. Chongshi, and G. Yanchang, "Maximum entropy method for determining dam safety monitoring indices," *Geomatics Inf. Sci. Wuhan Univ.*, vol. 33, no. 11, pp. 1126–1129, 2008.
- [4] C. Maosen, Q. Xiumei, and X. Ning, "A chaos-optimized neural network model for dam safety monitoring," *Rock Soil Mech.*, vol. 27, no. 8, pp. 1348–1444, 2006.
- [5] G. Chongshi, S. Huaizhi, and W. Shaowei, "Research progress on calculation model and monitoring method of long-term deformation characteristics of high concrete dam," *J. Hydroelectr. Eng.*, vol. 35, no. 5, pp. 1–14, 2016.
- [6] L. Yiyong, Z. Hao, and Z. Liting, "GA-SVM based multidimensional grey deformation prediction model research," *J. Zhejiang Univ. Technol.*, vol. 38, no. 1, pp. 79–83, 2010.
- [7] G. Runlong, "Discussion on machine learning algorithm under big data," *Telecom World*, vol. , no. 5, pp. 279–280, 2019.
- [8] Z. Yunkai, "Introduction to machine learning and its related algorithms," *Public Commun. Sci. Technol.*, vol. 11, no. 5, pp. 153–154, 2019.
- [9] E. Z. Naeini, K. Prindle, and W. Zhongde, "Translated world seismology," *Translated World Seismol.*, vol. 50, no. 5, pp. 442–452, 2019.
- [10] L. Haopeng, "Research on intelligent robot based on machine learning method," *Telecom World*, vol. 26, no. 4, pp. 247–248, 2019.
- [11] C. Shiyi, Y. Jie, C. Lin, Z. dongjian, and H. Chenrui, "Concrete dam deformation monitoring model based on Copula-RF," *J. China Three Gorges Univ. Natural Sci.*, vol. 42, no. 2, pp. 18–23, 2020.
- [12] D. Bo and C. Bo, "Chaos-based dam monitoring sequence wavelet RBF neural network prediction model," *Water Resour. Hydropower Eng.*, vol. 47, no. 2, pp. 80–85, 2016.
- [13] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.
- [14] C. Xianchang, *Research on Deep Learning Algorithm and Application Based on Convolutional Neural Network*. Hangzhou, China: Zhejiang Gongshang Univ., 2014.
- [15] N. Zhenhai and J. Dan, "An overview of deep learning research," *Sci. Technol. Innov. Herald*, vol. 12, no. 30, p. 224, 2015.
- [16] Q. Y. A. L. Zhaoce, "The realization and development of deep learning—From neural network to machine learning," *Electron. Technol. Softw. Eng.*, no. 11, pp. 30–31, 2017.
- [17] J. Zhang, X. Cao, J. Xie, and P. Kou, "An improved long short-term memory model for dam displacement prediction," *Math. Problems Eng.*, vol. 2019, pp. 1–14, Apr. 2019.
- [18] S. Hochreiter, *Untersuchungen Zu Dynamischen Neuronalen Netzen*. Munich, Germany, Technische Universität München, 1991.
- [19] J. Schmidhuber, *Sepp Hochreiter's Fundamental Deep Learning Problem*. Manno, Switzerland: Scuola Universitaria Della Svizzera Italiana, 1991.
- [20] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.
- [21] Y. Beibei, Y. Kunlong, and D. Juan, "Chinese journal of rock mechanics and engineering," *Chin. J. Rock Mech. Eng.*, vol. 37, no. 10, pp. 2334–2343, 2018.
- [22] W. Jiaqi, *Urban Area Traffic Flow Prediction Based on CNN and LSTM*. Dalian, China: Dalian Univ. Technology, 2019.
- [23] X. Qu, J. Yang, and M. Chang, "A deep learning model for concrete dam deformation prediction based on RS-LSTM," *J. Sensors*, vol. 2019, no. 1, pp. 1–14, 2019.
- [24] L. Wenju, P. Jianwen, R. Yisha, W. Zhigang, and W. Jinting, "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, 2020, Art. no. e2548.
- [25] 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.
- [26] W. Zhongru, "Theory and method of dam and dam foundation safety monitoring and its application," *Jiangsu Sci.-Tech. Inf.*, no. 12, pp. 1–6, 2006.
- [27] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [28] D. Kingma and B. Jimmy, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: <https://arxiv.org/abs/1412.6980>
- [29] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*. [Online]. Available: <https://arxiv.org/abs/1409.0473>
- [30] G. Hinton, N. Srivastava, A. Krizhevsky, R. R. Salakhutdinov, and I. Sutskever, "Improving neural networks by preventing co-adaptation of feature detectors," *Comput. Sci.*, vol. 3, no. 4, pp. 212–223, Jul. 2012.



DASHAN YANG received the B.E. degree in water conservancy and hydropower engineering from Hohai University, Nanjing, China, where he is currently pursuing the M.D. degree with the College of Water Conservancy and Hydropower Engineering. His research interests include risk assessment of dam and structural health monitoring (SHM).

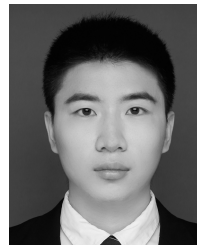


CHONGSHI GU was born in Qidong, Jiangsu, China, in 1962. He received the B.Eng. and Ph.D. degrees in water conservancy and hydropower engineering from Hohai University, Nanjing, China, in 1985 and 1997, respectively.

Since 2004, he has been a Professor with the Water Conservancy and Hydropower Engineering College, Hohai University, and was the Dean of the college, from 2010 to 2020. He is the author of six books and more than 150 articles. His research interests include early diagnosis for hydraulic structures, dam safety evaluation, and structural mechanics. He is the Chief Editor of *Advances in Science and Technology of Water Resources*. He was a recipient of the Award of Scientific and Technological Progress of the State, in 2005, 2007, and 2015.



YANTAO ZHU received the B.E. degree in water conservancy and hydropower engineering from Hohai University, Nanjing, China, where he is currently pursuing the Ph.D. degree with the College of Water Conservancy and Hydropower Engineering. He is currently a Visiting Scholar with the University of California at Los Angeles, Los Angeles. His research interests include risk assessment of dam and structural health monitoring (SHM).



ZHIDUAN ZHANG received the B.E. degree from the College of Water Conservancy and Hydropower Engineering, Taiyuan University of Technology, Taiyuan, China. He is currently pursuing the M.D. degree with the College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing, China. His research interests include machine learning and hydraulic structural health monitoring.



BO DAI received the B.E. degree from the College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing, China, where he is currently pursuing the Ph.D. degree with the College of Water Conservancy and Hydropower Engineering. He is currently a Visiting Scholar with the National University of Singapore. His research interests include machine learning, structural health monitoring, and hydraulic engineering.



KANG ZHANG received the B.S. degree in water resources and hydropower engineering from the Hefei University of Technology, Xuancheng, China, in 2014. He is currently pursuing the Ph.D. degree in hydraulic structural engineering with Hohai University. His research interests include dam safety monitoring theory and dam safety status assessment.



BO LI received the bachelor's degree majoring in agricultural water conservancy project from Jiangxi Agricultural University. He is currently pursuing the master's degree with the College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing, China. His research interests include machine learning, structural health monitoring, and roller compacted concrete dam.

...