J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 8919--8937

Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Construction of a Soil and Water Loss Probability Model and Landslide Risk Prediction for Hybrid Pumped Storage Engineering Based on an Online Monitoring Database

Lei Lei[1], Xiaolong Wei[1], Liang Wang[1], Qingyun Chen[1,✉], Lv Wang[1], Duozhi Kang[2]

[1] *State Grid (Xi'an) Environmental Protection Technology Center Co., Ltd., Xi'an, Shaanxi, 710100, China*
[2] *Electric Power Research Institute of State Grid Shaanxi Electric Power Co., Ltd., Xi'an, Shaanxi, 710100, China*

ABSTRACT
In the context of building a new type of energy system, pumped storage projects have been widely adopted as a form of energy storage with the most mature technology and the most economical investment. In this paper, a hybrid pumped storage project online monitoring system based on multi-sensors is proposed, and an online monitoring database is designed and constructed. Based on the data in the online monitoring database, the soil erosion of the hybrid pumped storage project is calculated and analyzed by combining the CSLE model. Then the attention mechanism is combined with BiLSTM model to construct the landslide risk prediction model of hybrid pumped storage project. The soil erosion during the construction of the hybrid pumped storage project is mainly distributed on the construction land, garden land, grassland and cropland, among which the construction land has the largest area of soil erosion (132.19 km^2), followed by the area of soil erosion of cropland (29.24 km^2). The MAPE is between 0.002% and 0.005% when predicting landslide risk deformation of hybrid pumped storage project using CNN-BiLSTM-ATT model. And using the model in this paper can minimize the error of rainfall on the prediction of landslide risk deformation and realize the safe and stable construction and operation of hybrid pumped storage projects.

Keywords: online monitoring database, CSLE model, attention mechanism, hybrid pumped storage project, soil erosion, landslide risk prediction

1. Introduction

China is one of the countries in the world where soil erosion is more serious, and soil erosion has become a major environmental problem in China. With the sustained and rapid development of China's economy and the increase of investment in infrastructure construction, the development and

✉ Corresponding author.
E-mail address: ep_Environmental@163.com (Q. Chen).
Received 12 January 2024; Revised 25 March 2024; Accepted 25 December 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-490

construction projects have increased year by year, so the soil erosion caused by them has shown a rapidly rising trend and has become one of the important sources of soil erosion [1-2].

In recent years, the construction pace of China's pumped storage project shows an accelerating trend. Pumped storage project is characterized by large scale of project, long construction period, large amount of soil and ballast disposal, serious soil erosion hazards, large potential hazards in the scope of influence of the project, and complex types of soil erosion [3-6]. It is generally constructed in the mountains and mountains, the existence of the project upper and lower reservoir area determines the complexity of the geological environment conditions such as topography and geomorphology, geological structure, engineering geology, hydrogeological conditions and the degree of development of geologic hazards in the construction project area to be between moderate and complex, and the total geological environment conditions are moderate or complex [7-10]. Therefore, to do a good job of soil and water conservation monitoring of pumped storage projects can maximize the control of anthropogenic soil and water erosion caused by project construction, ensure the safety of the ecological environment in the project construction area and the normal production and life of the people, which is a major task facing the process of project construction [11-14]. At present, there is no set of relatively perfect technical standard system for soil and water conservation monitoring in pumped storage projects, which affects the accuracy and uniformity of the soil and water conservation monitoring results to a certain extent [15-16]. The establishment of soil erosion probability model and landslide risk prediction model suitable for pumped storage power station project based on online monitoring database is an important part of soil and water conservation monitoring work of pumped storage power station at this stage, which is of great practical significance for soil and water conservation monitoring work [17-19].

Pumped storage projects are currently an important means of supporting the development of new energy sources due to their advantages of flexible operation, mature technology, economic and environmental protection, and the ability to provide rotational inertia to the power system. Starting from designing an online monitoring system for hybrid pumped storage projects, the article constructs a cluster database for storing large-scale hydrogeological data related to hybrid pumped storage projects. Using the integrated data, the CSLE model is used to calculate and determine the erosion factors and erosion intensity of the hybrid pumped storage project. The CNN-BiLSTM-ATT model for landslide risk prediction of hybrid pumped storage projects was constructed by combining the time series processing capability of BiLSTM and the feature extraction capability of the attention mechanism. The hybrid pumped storage power plant project in the reservoir area of the Yellow River MED power station is taken as an example to calculate its soil erosion and landslide risk prediction effect.

2. Online monitoring system for hybrid pumped storage projects

Led by the dual-carbon target, the hybrid pumped storage power station, as a more economical, effective and mature energy storage technology, plays a more obvious role in the process of building a new energy system. Due to the special characteristics of hybrid pumped storage power station, which pumps water for energy storage when electricity consumption is low and discharges water for power generation when electricity consumption is high, the basic requirements for engineering geology are different from those of large-scale conventional hydropower projects, and the engineering geological investigation work is facing a great challenge. Therefore, how to use the existing technology for hybrid pumped storage project soil and water geological information of real-time monitoring, become a hybrid pumped storage project to enhance the power station must be resolved, so as to better power system operation to provide peak shifting, frequency regulation and other auxiliary services, enhance the flexibility of the power system.

2.1. Hybrid pumped storage engineering geological survey

2.1.1. Hybrid pumped storage projects. Hybrid pumped storage power station is a kind of power generation main body that integrates two types of hydropower forms, namely, traditional pumpless hydropower station and pumped storage power station, with the functions of pumped storage and runoff power generation, and it can be used as a high-quality large-scale flexible regulating power source. In the key period of energy green low-carbon transformation and development, vigorously develop hybrid pumped storage power station group with strong energy storage, regulation and supply capacity is to realize the power supply, power structure optimization, and build a new renewable energy-dominated power system as the inevitable choice [20]. Hybrid pumped storage project power station construction form shown in Figure 1, mainly contains two ways, one is a built conventional power station reservoir as the upper reservoir, downstream another lower reservoir. The second is to expand the pumped storage units in the upper reservoir of the already-built stepped hydropower station, or to directly convert some conventional hydropower units into pumped storage units to act as peak shifters.

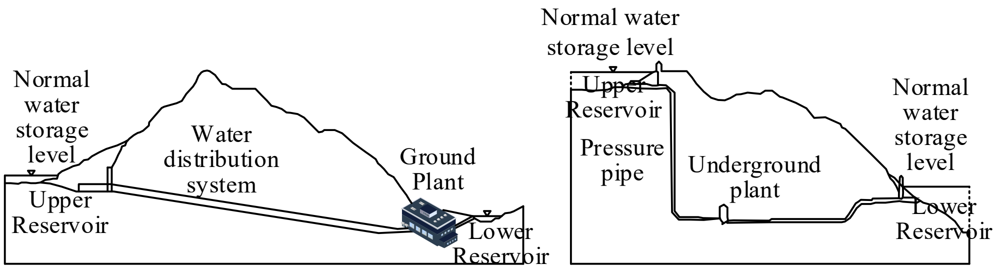


Fig. 1. Hybrid pumped storage power station

Using the joint operation of different units to complete pumping and power generation not only effectively improves the original power plant operation mode, increases the ability of peak shifting and valley filling, accident standby, etc., realizes the full use of water resources, and effectively improves the economy and market competitiveness of the power plant. To a certain extent, it also relieves the pressure of thermal power peaking, ensures the power quality and maintains the safe and stable operation of the system, and at the same time reduces the environmental protection investment caused by the large amount of exhaust gas emission.

2.1.2. Key points of soil and water geological investigations. The current regulations require that the construction of hybrid pumped storage power station using built reservoirs, the engineering geological investigation should focus on collecting engineering geological data during the design, construction and operation of built reservoirs, on the basis of reviewing the reservoir leakage, bank stability, foundation stability of hub buildings and other engineering water and soil geological problems related to the construction and operation of pumped storage power stations, and focusing on evaluating the impact of water level changes on the stability of the bank of the reservoir. Focusing on the investigation of water transmission and power generation system engineering water and soil geological conditions, the renovation and expansion project should also carry out specialized engineering geological investigation.

Hybrid pumped storage power station engineering water and soil geological investigation should be combined with the key engineering geological problems of the hub building already built and planned and designed to carry out the corresponding work, should be above and below the reservoir inlet/outlet location selection, underground plant cavern group perimeter rock stability and the permeability characteristics of the outer mountains should be the focus of the engineering geological investigation.

Affected by the construction conditions, technical and economic rationality, some power stations need to choose the appropriate topographic and geological conditions to arrange the ground plant, such as some pumped storage power stations have chosen the semi-underground plant program, the pit stability and the engineering geological conditions of the slopes behind the plant should be taken as the key point of the investigation for this kind of project. Through effective soil and water geological investigation, to ensure that the hybrid pumped storage project in the construction process will not occur large-scale soil erosion or landslide risk occurs, to better ensure the safe and stable operation of the hybrid pumped storage project.

2.2. *Online monitoring system and database framework*

2.2.1. Pumped storage online monitoring systems. Aiming at the construction characteristics and management difficulties of hybrid pumped storage project power station construction, this paper proposes a hybrid pumped storage project online monitoring system based on multi-sensors, which is mainly composed of application layer, platform layer, network layer and perception layer, and its main system architecture is shown in Figure 2 [21]. The system realizes a comprehensive display of the overall situation of the hybrid pumped storage project construction, and provides online supervision functions for the project construction and management personnel, including information on the completion of the construction progress, the distribution of the construction operations, the soil and water quality data and early warning. The quality data of the on-site construction process is unified in the system display, which comprehensively demonstrates the construction situation, effectively solves the soil and water geological survey supervision problems in the process of engineering construction, and provides a means of supervision for the soil and water geological control in the process of hybrid pumped storage project construction.

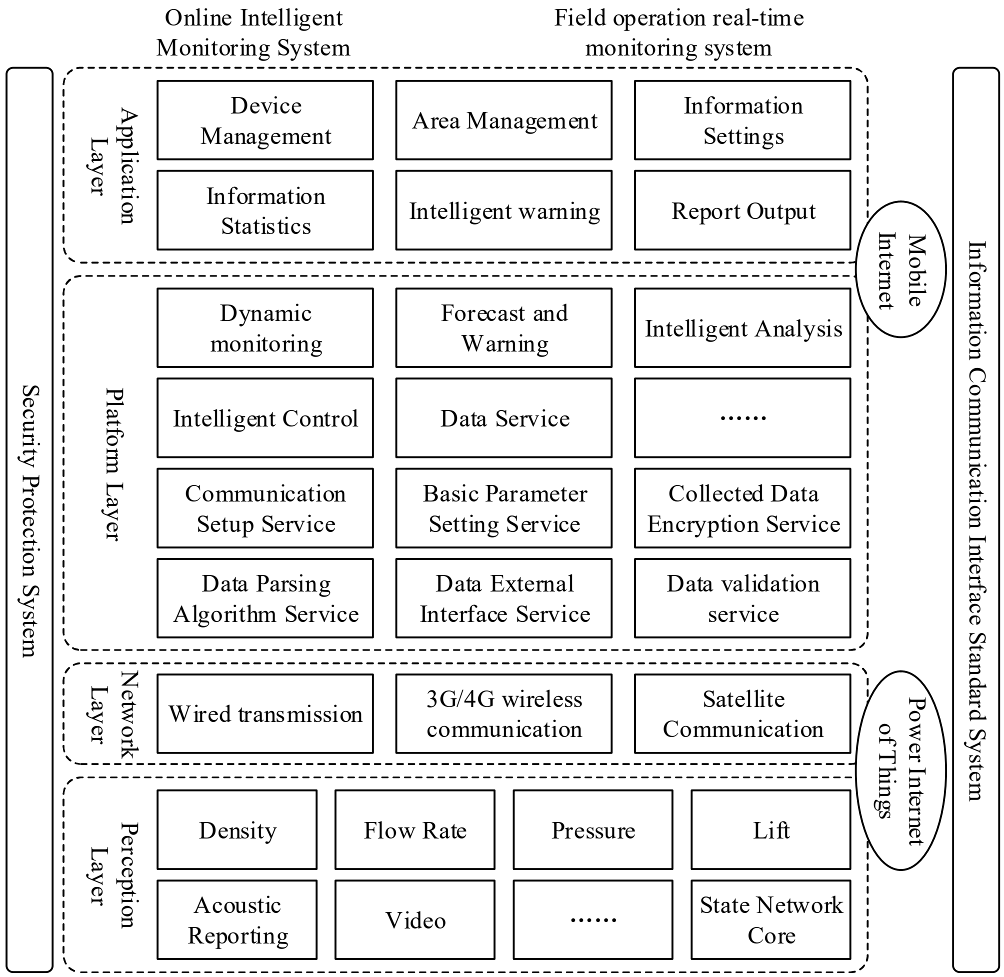


Fig. 2. Pumped storage can monitor the system online

1) Integrated design. Hybrid pumped storage project power station site environmental conditions are complex, and the construction site work surface is large, frequent movement of equipment, the traditional recorder equipment in the field construction monitoring process of its equipment dispersed layout installed in the work surface, the equipment is more messy, civilized construction is poor. The use of non-contact intelligent monitoring technology not only to solve the soil and water quality data collection accuracy, data quality reliability problems, the integrated design of the device to take into account the complexity of the site construction environment, operational mobility, installation, maintenance convenience and other needs.

2) No network data sharing. Non-contact online monitoring system not only realizes real-time data collection and transmission in a network environment, but also on-site non-contact recording and collecting device with independent data processing unit and data protection, authentication mechanism, and on-site raw data recording. In the absence of network environment, the on-site management mobile terminal can be used to authenticate the on-site non-contact record collection device, and the relevant data records can be collected offline, and the offline data collected by the mobile terminal in the place with network environment can be automatically transferred to the online platform, thus realizing the sharing of data collection in the absence of network environment.

2.2.2. Online monitoring database design. The online real-time monitoring of hybrid pumped storage project will collect the real-time change data of soil and water geology in the process of engineering construction through sensors, and it is necessary to store the relevant data through the database to provide data support for timely analysis of soil erosion and landslide risk.

Real-time monitoring data has the characteristics of fast change frequency, large data volume and continuous generation, according to the characteristics of real-time data management, in the face of the proliferation of measurement point scale, the scale of the second level of deployment of measurement points can reach millions, such as the use of a single real-time database is difficult to meet the demand for data access, query. Distributed strategy, the application of joint deployment of multiple real-time database program to achieve the so-called "cluster database". Based on this, this paper designs an online monitoring database as shown in Figure 3, the cluster database is divided into storage carrier layer and measurement point location transparent layer. The storage carrier layer is mainly used to store specific measurement point data, which consists of several real-time/historical databases and a measurement point relationship library. On top of the storage carrier layer, the measurement point location transparency layer realizes the transparency of the corresponding storage location, and the whole cluster database can be understood as a logical whole. The point location transparency layer maintains point location metadata, which indicates in which database each point is stored in the storage carrier layer.

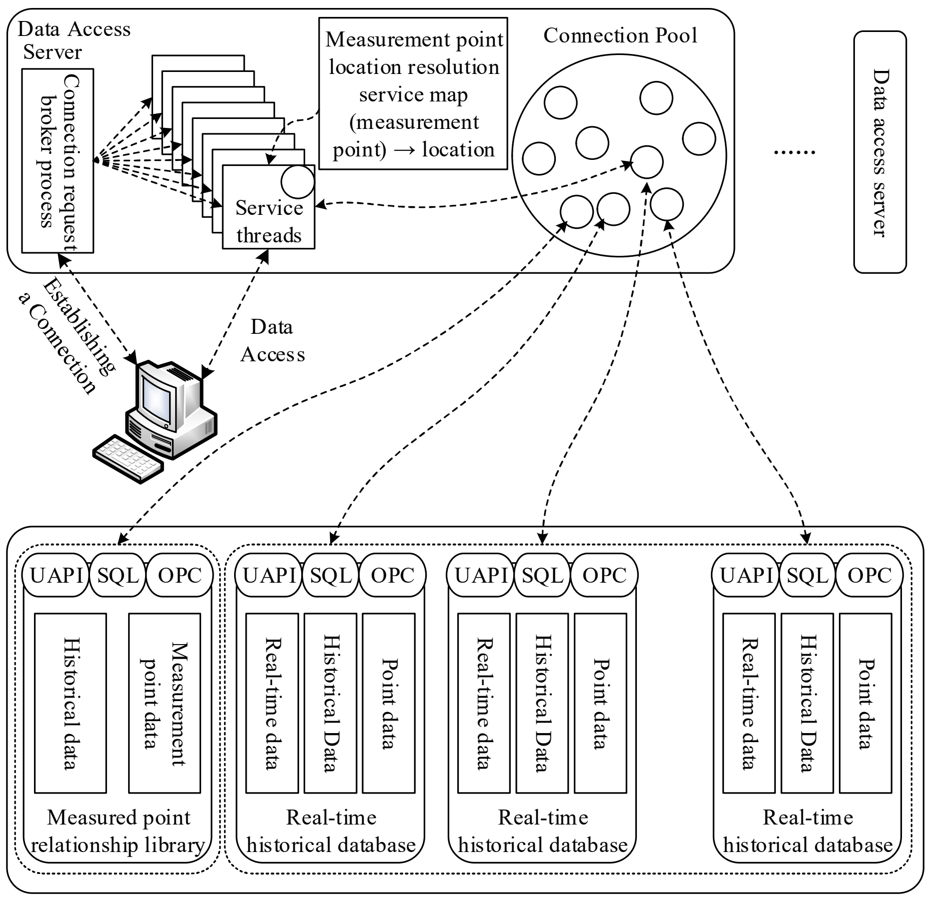


Fig. 3. Water storage project online monitoring database

3. Probabilistic erosion and landslide risk prediction models

With the construction of new power systems mainly based on new energy sources, the role of pumped storage as a system regulating resources is becoming more and more prominent, and the rapid development of hybrid pumped storage power plant projects is still the main theme, which will play an important role in the "double carbon" background. The hybrid pumped storage project in the construction process and after the completion of the construction, it may have a corresponding impact on the surrounding soil and water geology, how to realize the hybrid pumped storage project erosion analysis and landslide risk prediction, is to ensure the stable operation of the hybrid pumped storage project key content.

3.1. CSLE model and erosion intensity determination

3.1.1. Calculation of erosion factors for soil and water erosion. The CSLE model integrates seven key factors to predict the amount of soil erosion by multiplying them together. These factors include rainfall erosivity (R), soil erodibility (K), slope length (L), slope gradient (S), vegetation cover and biological measures (B), engineering measures (E) and tillage measures (T) [22]. Each factor has a direct influence on the soil erosion process and is combined to calculate the soil erosion modulus (A), the basic form of the equation is:

$$A = R \cdot K \cdot L \cdot S \cdot B \cdot E \cdot T \tag{1}$$

The rainfall erosivity factor R was calculated based on the daily rainfall data of the study area by dividing the annual data into 24 half-monthly cycles and calculating the mean values, accumulating these mean values to obtain the annual R values, and then generating the corresponding raster data in geographic information software using Kriging geostatistical spatial interpolation method. I.e.:

$$\overline{R} = \sum_{k=1}^{24} \overline{R}_{Half\ Moon\ k} \tag{2}$$

$$\overline{R}_{Half\ Moon\ k} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=0}^{m} \left(\alpha \cdot P_{i,j,k}^{1.7265} \right) \tag{3}$$

$$\overline{WR}_{Half\ Moon\ k} = \frac{\overline{R}_{Half\ Moon\ k}}{\overline{R}} \tag{4}$$

where $\overline{R}$ and $\overline{R}_{Half\ Moon\ k}$ denote the multi-year average rainfall erosivity and k half-monthly rainfall erosivity, $k = 1, 2 \cdots\cdots 24$, refers to the division of the year into 24 half-months, $i = 1, 2 \cdots\cdots N$, refers to the yearly sequence, and m refers to the number of days with rainfall greater than 10 mm in the kth half-month of the ith year. $P_{i,j,k}$ is the erosive daily rainfall in the kth half-month j of the ith year, and α is the parameter, from May to September $\alpha = 0.3937$ and from October to April $\alpha = 0.3101$ of the following year.

The topographic factors affecting soil erosion mainly include slope length factor and gradient factor. In this study, the slope length factor is mainly calculated using the segmented slope length factor, and the slope length factor is calculated by the formula:

$$L_i = \frac{\lambda_i^{m+1} - \lambda_{i-1}^{m+1}}{\left(\lambda_i - \lambda_{i-1}\right) \cdot (22.13)^m} \quad m = \begin{cases} 0.2\theta \le 1^\circ \\ 0.31^\circ < \theta \le 3^\circ \\ 0.43^\circ < \theta \le 5^\circ \\ 0.5\theta > 5^\circ \end{cases} \tag{5}$$

where L_i is the ind grid slope length factor, λ_i and λ_{i-1} are the slope lengths (m) of the ith and $i-1$th slope sections, and m is the slope length index. Its slope factor is calculated as:

$$S = \begin{cases} 10.8 \sin\theta + 0.03 & \theta \le 5^\circ \\ 16.8 \sin\theta - 0.5 & 5^\circ < \theta \le 10^\circ \\ 21.9 \sin\theta - 0.96 & \theta \ge 10^\circ \end{cases} \tag{6}$$

where, S is the slope factor and θ is the slope.

After obtaining the vegetation cover data, the vegetation cover data was used to calculate the vegetation cover and biological measure factor with the 24 half-month rainfall ratio data (RBL) and the cover curve data of the garden land. The engineering measure factor was defined as the ratio of soil loss with the implementation of soil and water conservation engineering measures to the soil loss without the implementation of measures to quantify the effectiveness of soil and water conservation measures. Tillage factor is an important factor in the CSLE model, and the smaller its value, the better the soil and water conservation benefits brought by the implementation of such conservation tillage measures.

3.1.2. Determination of erosion intensity of soil erosion. Based on the soil erosion classification and grading standards, combined with the calculated correlation modulus of soil erosion factors, in order to evaluate the erosion intensity of each grid. Table 1 shows the grading of soil erosion intensity, which is mainly divided into six grades: slight, mild, moderate, strong, very strong and intense.

Table 1. Soil erosion intensity classification

Level	Average erosion modulus [t/(km^2·a)]	Average erosion modulus (mm/a)
Micro- Ⅰ	<200, <500, <1000	<0.15, <0.37, <0.74
Mild- Ⅱ	200, 500, 1000~2500	0.15, 0.37, 0.74~1.9
Moderate-Ⅲ	2500~5000	1.9~3.7
Strong-Ⅳ	5000~8000	3.7~59.
Very Strong- Ⅴ	8000~15000	5.9~11.1
Intense-Ⅵ	>15000	>11.1

3.2. Landslide risk prediction for pumped storage projects

3.2.1. Bidirectional long and short-term memory networks. Long Short-Term Memory (LSTM) network is a variant of recurrent neural network, a neural network model capable of processing time series data and remembering long term dependencies. Bidirectional Long Short-Term Memory Network (BiLSTM) is a variant of LSTM that not only has the ability to learn long-distance sequence relationships, but also learns the association of flow sequence data in both positive and negative directions, taking into account both past and future information to better capture the contextual information of the input sequence. In this paper, by combining CNN with BiLSTM, we can fully utilize the local feature extraction capability of CNN and the global contextual information capturing capability of BiLSTM, and then realize the extraction of landslide feature data from hybrid pumped storage projects, which can provide accurate data for landslide risk prediction.

The BiLSTM network consists of an input layer, a forward hidden layer, a backward hidden layer, and an output layer. The forward hidden layer processes the input sequences in forward chronological order, while the reverse hidden layer processes the input sequences in reverse chronological order, which enables the network to acquire bidirectional features by learning the forward and reverse information of the input sequences at each time step [23].

The input sequence of the BiLSTM unit is denoted as $x=\left(x_1,x_2,\cdots,x_n\right)$, where n denotes the length of the sequence. $\overrightarrow{h_t}=\left(\overrightarrow{h_1},\overrightarrow{h_2},\cdots,\overrightarrow{h_n}\right)$ denotes the forward hidden sequence, $\overleftarrow{h_t}=\left(\overleftarrow{h_1},\overleftarrow{h_2},\cdots,\overleftarrow{h_n}\right)$ denotes the reverse hidden sequence, and $y_t=\left(y_1,y_2,\cdots,y_n\right)$ denotes the output sequence. The final output vector is the combined effect of the forward and reverse information streams, i.e., $y_t=f\left(\overrightarrow{h_t},\overleftarrow{h_t}\right)$. The formulation of the BiLSTM network structure is shown below:

$$\overrightarrow{h_t}=\sigma\left(w_{\vec{h}x}x_t+w_{\vec{h}\vec{h}}h_t+b_{\vec{h}}\right) \tag{7}$$

$$\overleftarrow{h}_t = \sigma\left(w_{\overleftarrow{h}x} x_t + w_{\overleftarrow{h}\overleftarrow{h}} h_t + b_{\overleftarrow{h}}\right) \tag{8}$$

$$y_t = w_{y\overrightarrow{h}} \overrightarrow{h}_t + w_{y\overleftarrow{h}} \overleftarrow{h}_t + b_y \tag{9}$$

where $h_{\overrightarrow{h}}$, $h_{\overleftarrow{h}}$, and denote the corresponding bias values for forward, reverse, and output, respectively, and $w_{\overrightarrow{h}x}$, $w_{\overleftarrow{h}x}$, $w_{\overrightarrow{h}\overrightarrow{h}}$, $w_{\overleftarrow{h}\overleftarrow{h}}$, $w_{y\overrightarrow{h}}$, and $w_{y\overleftarrow{h}}$ denote the weight matrices for input-to-forward, input-to-reverse, forward-to-forward, reverse-to-reverse, forward-to-output, and reverse-to-output, respectively.

3.2.2. Principles of Attention Mechanisms. The basic idea of the attention mechanism is to select from the input information that is more critical to the current task goal. It captures more valuable information by assigning corresponding weights to each input information according to its importance. Self-attention mechanism is an improvement of the attention mechanism, which reduces the dependence on external information and is better at capturing the internal relevance of data [24].

In the self-attention mechanism, h_t denotes the hidden state vector at the t nd moment, e_t denotes the similarity weight of h_t, α_t denotes the normalized similarity weight of h_t, and R denotes the self-attention output matrix obtained by weighted summation. The computational process can be divided into three steps as follows:

1) The similarity weights are computed using a multilayer perceptual machine (MLP) for each moment of data, i.e.:

$$e_t = W_b \tanh\left(W_a \left[h_t; H\right]\right) \tag{10}$$

where W_a denotes the connection weights between the input and hidden layers of the MLP, W_b denotes the connection weights between the hidden and output layers of the MLP, and $H = \left(h_{t-1}, h_t, \ldots, h_{t+i}\right)$ denotes the hidden state matrix.

2) Normalize these similarity weights using SoftMax function, i.e.:

$$a_t = soft\max\left(e_t\right) = \frac{\exp\left(e_t\right)}{\sum_{j=t-1}^{t+i}\left(e_j\right)} \tag{11}$$

where $\exp()$ denotes an exponential function with a natural constant e as the base.

3) The normalized similarity weights and the corresponding data are weighted and summed to obtain the self-attention output matrix R, i.e:

$$R = \sum_{j=t-1}^{d} \alpha_j \overline{h}_j \tag{12}$$

where h_j denotes the hidden state vector at moment j and α_j denotes the similarity weight after h_j normalization.

In the landslide risk prediction of hybrid pumped storage projects, the self-attention mechanism can utilize the MLP to calculate the similarity weights of the hydro-soil geological data at each moment. The self-attention mechanism can capture more valuable information by assigning corresponding weights according to the importance of the hydro-soil geologic data at each moment to the prediction results. Therefore, in order to make BiLSTM fully utilize the differences of the hydrogeological data, this paper combines BiLSTM with the self-attention mechanism to predict landslide risk from the hydrogeological data.

3.2.3. Landslide risk prediction models. In this paper, the combined CNN-BiLSTM-Attention model is utilized to predict the landslide risk of a hybrid pumped storage project, and its structure is schematically shown in Fig. 4. It mainly contains input layer, CNN layer, BiLSTM layer, attention layer and output layer.

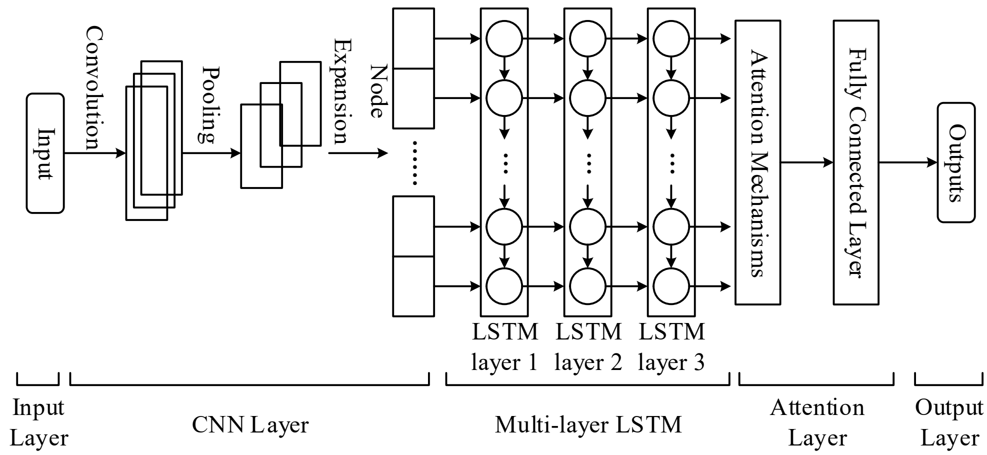


Fig. 4. Schematic diagram of the CNN-BiLSTM-Attention prediction model

1) Input layer. Specify the format of the input data, including batch size, time step, and feature dimension, defaulting the batch size to 1, the time step to t, and the feature dimension to n. A sample can be represented as a matrix of real sequences $R_t \times n$, and denote x_i as the vector representation of the data at the i th time step in $R_t \times n$.

2) CNN layer: The CNN layer can extract the spatial connection between different feature values in the data, thus making up for the shortcomings of LSTM that cannot capture the spatial components of the data, and at the same time, the features it extracts still have temporal order. Sample data into the CNN layer will be sequentially convolution, pooling and node expansion (dimensionality reduction) operations. For sequence data, this model takes one-dimensional convolution, and the convolution kernel is only convolved according to a single time-domain direction.

3) BiLSTM layer. This model adopts BiLSTM neural network, in which the output of the previous LSTM layer is the input of the next layer, passing to the next layer, and the output of the last LSTM hidden layer will enter the attention layer for further processing.

4) Attention layer. Attention can improve the role of important time steps in LSTM, thus further reducing the model prediction error. Attention is essentially to find the weighted average sum of the output vectors of the last LSTM layer.The output vectors of the LSTM hidden layer are used as inputs to the attention layer, which is trained by a fully connected layer, and then the outputs of the fully connected layer are normalized using the SoftMax function, resulting in the assigned weights of each hidden layer vector, and the magnitude of the weights indicates the importance of the hidden state of each time step for the prediction of the results. .

5) The output layer is the last layer of the CNN-BiLSTM-ATT model, and its role is to output the results of the prediction model's processing of the input data.

4. Pumped storage project erosion and landslide risk prediction

Landslides are a kind of geological disaster widely distributed in nature, and their occurrence often causes serious casualties and economic losses. During the construction process of hybrid pumped storage projects, the degree of disturbance to the surrounding environment by engineering activities is increasing, which, together with the influence of hydrological conditions and geological activities, will greatly trigger the risk of landslides. Therefore, how to realize the soil erosion calculation and landslide risk prediction of hybrid pumped storage project through effective data monitoring has become an important guarantee for the stable construction of hybrid pumped storage project.

This paper takes the hybrid pumped storage power plant project in the reservoir area of the Yellow River MED power station as an example, and aims to provide an example reference for the stable operation of the hybrid pumped storage project through soil erosion calculation and landslide risk prediction.

4.1. Calculation of soil erosion in pumped storage projects

4.1.1. Research data sources and processing:

1) Remote sensing image data. High-resolution remote sensing images with 2.5m spatial resolution from January to September 2022 in the study area were used for the interpretation of land use and soil and water conservation measures. MODIS Normalized Vegetation Index (NVI) products for the study area from 2020 to 2023 were collected at a temporal resolution of one period every 15d, 24 periods per year, and a spatial resolution of 200m.

2) Rainfall data. Day-by-day rainfall data from 50 stations in the study area from 1995 to 2016 were obtained from the National Meteorological Science Data Center and used to calculate rainfall erosivity.

3) Soil and topographic data. The relevant topographic and soil data of the study area obtained from the online monitoring database established in this paper.

4) Soil and water conservation key project data. Types of national key soil and water conservation projects, project implementation areas, and distribution, quantity or area of soil and water conservation measures, with the types of measures mainly based on afforestation, grass planting, sealing, ecological restoration, terracing and horizontal steps.

5) Anthropogenic disturbance data. Information on the type, scale, and spatial distribution of production and construction projects is obtained from the online monitoring system designed in this paper.

4.1.2. Soil erosion in different land-use types:

1) Area and intensity of soil erosion of different land use types. Based on the calculation method of soil erosion and soil erosion factor given in the previous section, combined with the relevant soil and water and geological data of the study area, it is brought into the CSLE model to obtain the soil erosion of different land use types as shown in Table 2.

The total land area of the study area is 2574.6km^2, and according to the results of remote sensing interpretation in 2023, more than 75% of the entire area is covered by watersheds and water conservancy facilities, building land and cultivated land, with a total area of 1951.7km^2, which accounts for 75.81% of the land area of the study area, and it is mainly distributed in the left and right sides of the river. This is followed by forest and grassland, with a combined area of 533.05km^2, or 20.7%, while parkland, transportation land and other land account for 0.91%, 2.47% and 0.1% of the total area, respectively. In terms of soil erosion on the land use level category, soil erosion in the study area is mainly distributed on construction land, garden land, grassland and cultivated land. Construction land has the largest area of soil erosion, which is 132.19km^2, accounting for 69.15% of the total area of soil erosion in the study area. This was followed by arable land with an erosion area of 29.24km^2, which accounted for 15.3% of the total erosion area of the study area, and forest land, grassland and garden land with an erosion area of 17.87km^2, 7.06km^2, and 4.47km^2, respectively, which altogether accounted for 15.38% of the total erosion area of the study area. The area of soil erosion on other land and transportation land is small, only 0.18km^2 and 0.16km^2. In addition, construction land and garden land have the highest percentage of soil erosion in their land area, 28.97% and 19.01%, respectively, followed by cropland, other land and forest land with 6.84%, 6.82% and 5.11%, respectively, and transportation land with a relatively low percentage of 0.25%. The area of strong and above erosion in the whole study area amounted to 29.05km^2, mainly distributed in cultivated land (16.88km^2) and construction land (11.18km^2).

Table 2. Soil erosion area of different land use types

Land type	Land area (km^2)	Soil erosion area (km^2)						Ratio (%)
		I	II	III	IV	V	VI	
Ploughing	427.38	1.24	2.08	9.04	4.51	5.43	6.94	6.84
Garden	23.52	0.01	0.06	3.72	0.63	0.05	0.00	19.01
Woodland	349.64	4.08	8.24	5.43	0.12	0.00	0.00	5.11
Grass	183.41	3.69	2.05	1.26	0.06	0.00	0.00	3.85
Construction	456.27	85.37	9.46	26.18	9.05	2.13	0.00	28.97
Transportation	63.69	0.12	0.00	0.00	0.00	0.00	0.04	0.25
Water conservancy facilities	1068.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Other	2.64	0.06	0.02	0.01	0.01	0.02	0.06	6.82

2) Main control factors of soil erosion in different land use types. Using the soil erosion raster data superimposed on the land use vector data, calculate the average erosion modulus of arable land, garden land, forest land, grassland and construction land, classify the results into two categories of no soil erosion and soil erosion in accordance with soil erosion classification and grading standards, statistically analyze the correlation between soil erosion categories and factors under different land use categories, and quantitatively evaluate the main control factors affecting the occurrence of soil erosion in different land use categories, the specific results are shown in Table 3. The specific results are shown in Table 3. The specific results are shown in Table 3. In the table, *,** denote at 5% and 1% confidence level respectively.

The results showed that the S factor and slope were the main factors affecting the occurrence of soil erosion on cropland and construction land, both of which reached more than 0.7 at the 1% confidence level. Since the S factor is calculated based on slope, it can be assumed that the main controlling factor for cropland and construction land is slope. Garden land, forest land and grassland are mainly influenced by the B factor, i.e., vegetation cover is the main factor influencing soil erosion in garden land, forest land and grassland. In addition, slope is also one of the main factors affecting soil erosion in garden, woodland and grassland. Rainfall erosivity factor R and soil erodibility factor K, etc. have relatively small effects on soil erosion of each land use type in hybrid pumped storage projects. Therefore, in the construction process of the hybrid pumped storage project, we should pay attention to the changes in slope and vegetation cover, and fully ensure that the probability of soil erosion is minimized on the basis of ensuring the construction of the project, which is more in line with the concept of sustainable development.

Table 3. Analysis of dominant control factor in soil and soil drift area

Type	Ploughing	Garden	Woodland	Grass	Construction
R	-0.095**	0.026*	0.023	0.131**	0.093**
K	-0.182**	0.265**	0.245**	0.146**	0.152
L	-0.123**	0.467**	0.446**	0.472**	0.604**
S	0.756**	0.431**	0.431**	0.416**	0.725**
B	0.054**	0.492**	0.482**	0.505**	0.206

E	0.432**	-	0.183**	-	-
Slope	0.705**	0.395**	0.402**	0.351**	0.715**

4.1.3. Changes in erosion dynamics. In order to further analyze the impact of the construction of the hybrid pumped storage project on soil erosion, this paper divides the erosion types into three categories: natural (T1), sloping cultivated land (T2) and production and construction (T3). The soil and water geological data in 2015 are further utilized to calculate the erosion area and compare it with the erosion area in 2023, so as to clarify the dynamic change of soil and water erosion during the construction of the hybrid pumped storage project. Table 4 shows the dynamic change of soil erosion in the study area from 2015 to 2023, where R1, R^2 and R3 are the changes in the proportion of the study area, the proportion of the eroded area and the proportion of the study area, respectively.

In terms of change in erosion area, the erosion area of the study area in 2023 decreased by 85.26km^2 or 17.14% compared to 2015. In terms of erosion type, the erosion of sloping cropland has the largest decrease of 23.36%, and its slight and mild erosion area decreases as high as 79.26% and 40.40%, respectively. It indicates that in the process of constructing the hybrid pumped storage project, the corresponding soil erosion control measures are also being implemented for the sloping cultivated land, and good results have been achieved, making the erosion of the sloping cultivated land under better control, and the degree of erosion has been reduced. The natural erosion area also had a relatively large reduction compared with 2015, with a reduction ratio of 21.57%, among which, the erosion reduction area of the slight type was the largest of 32.35km^2, accounting for 71.18% of the total area of natural erosion reduction. From the spatial distribution of erosion patches, the erosion area of natural and sloping cultivated land decreased significantly, but various types of soil erosion of various intensities are still distributed in the east and north of the study area, which is also the key management area to improve the soil erosion around the hybrid pumped storage project in the future.

During the period from 2015 to 2023, the soil erosion control of forest grassland, sloping cultivated land and production and construction projects, etc. implemented for the study area has achieved certain results, and the area and degree of erosion have been improved to a certain extent; however, soil erosion in the watershed is still widespread, and different prevention and control strategies should be implemented through differentiation in order to further improve the situation of water and soil erosion in the sub-watersheds. For natural erosion, measures such as natural sealing, banning cultivation and grazing can be taken to promote forest vegetation restoration, thus reducing regional natural soil erosion. Sloping cultivated land erosion management requires selective construction of high-standard farmland, horizontal terraces and slope terraces, and implementation of biological measures such as plant hedges and strip-tillage, depending on the slopes on which the area is situated. At the same time, man-made soil and water erosion caused by production and construction activities is effectively prevented by carrying out remote sensing supervision of soil and water conservation, joint law enforcement actions, and increasing on-site inspection and informationization supervision.

Table 4. The dynamic change of soil and soil loss in 2015~2023 study area

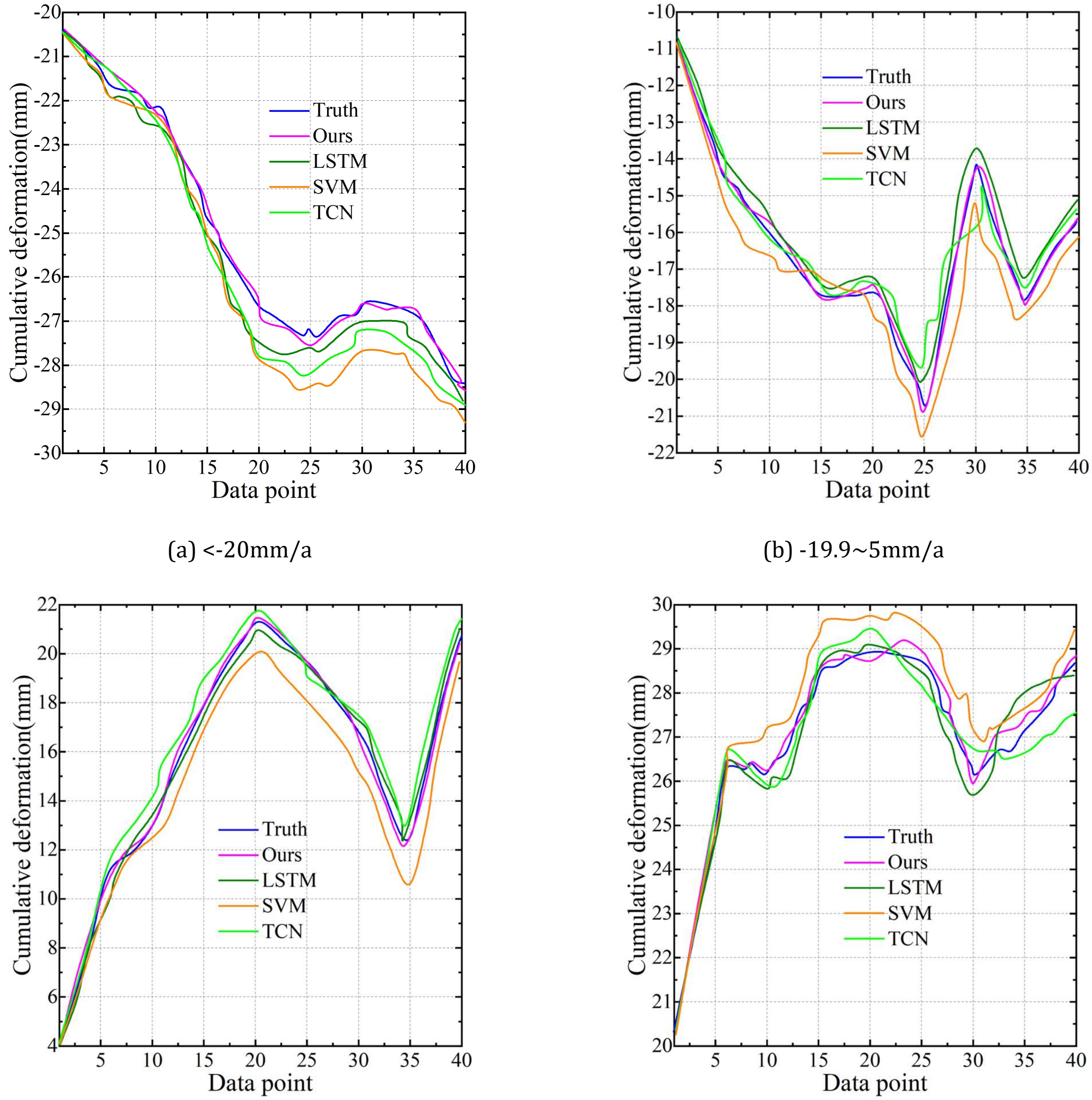
Type	2015			2023			Change	
	Area (km^2)	R1(%)	R^2(%)	Area (km^2)	R1(%)	R^2(%)	Area (km^2)	R3(%)
T1-Ⅰ	143.59	5.58	29.1	111.24	4.32	26.99	-32.35	-1.26
T1-Ⅱ	52.06	2.02	10.55	42.08	1.63	10.21	-9.98	-0.39
T1-Ⅲ	11.21	0.44	2.27	9.04	0.35	2.19	-2.17	-0.09
T1-Ⅳ	3.06	0.12	0.62	2.51	0.10	0.61	-0.55	-0.02
T1-Ⅴ	0.78	0.03	0.16	0.39	0.02	0.09	-0.39	-0.01
T1-Ⅵ	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
T1-total	210.7	8.19	42.7	165.26	6.42	40.09	-45.44	-1.77
T2-Ⅰ	5.98	0.23	1.21	1.24	0.05	0.31	-4.74	-0.18

T2- Ⅱ	3.49	0.14	0.71	2.08	0.08	0.51	-1.41	-0.06
T2-Ⅲ	10.37	0.40	2.10	9.04	0.35	2.19	-1.33	-0.05
T2-Ⅳ	5.32	0.21	1.08	4.51	0.18	1.09	-0.81	-0.03
T2- Ⅴ	5.46	0.21	1.11	5.43	0.21	1.32	-0.03	0
T2- Ⅵ	7.53	0.29	1.53	6.94	0.27	1.68	-0.59	-0.02
T2-total	38.15	1.48	7.74	29.24	1.14	7.10	-8.91	-0.34
T3- Ⅰ	94.61	3.67	19.17	85.37	3.32	20.72	-9.24	-0.35
T3- Ⅱ	10.98	0.43	2.22	9.46	0.37	2.30	-1.52	-0.06
T3-Ⅲ	30.24	1.17	6.13	26.18	1.02	6.35	-4.06	-0.15
T3-Ⅳ	10.18	0.40	2.06	9.05	0.35	2.20	-1.13	-0.05
T3- Ⅴ	5.09	0.20	1.03	2.13	0.08	0.52	-2.96	-0.12
T3- Ⅵ	97.37	3.78	19.73	85.37	3.32	20.72	-12	-0.46
T3-total	248.47	9.65	50.34	217.56	8.46	52.81	-30.91	-1.19
Total	497.32	19.32	100	412.06	16.02	100	-85.26	-83.98

4.2. Landslide risk prediction for pumped storage projects

4.2.1. Landslide risk deformation prediction. This experiment involved 95,748 slope points, each of which served as a set of experimental data. These data record the deformation data collected every 15 days from April 15, 2023 to February 28, 2024, a total of 40 data points. What needs to be accomplished in this experiment is the accurate prediction of the deformation data to solve the lag problem that exists in the deformation data. In order to verify the effectiveness and superiority of the model in this paper, this paper selects four models that are good at dealing with time series data, LSTM model, TCN model and SVR model to carry out controlled experiments. Fig. 5 shows the deformation prediction results of different models, where Fig. 5(a)~(d) shows the cumulative deformation values for the average loss thickness less than -20 mm/a, -19.9~5 mm/a, 5~19.9 mm/a and more than 20 mm/a, respectively.

Based on the results presented in the figure, it can be seen that thanks to BiLSTM's powerful ability to process time series, both the BiLSTM model based on the attention mechanism selected in this paper and the traditional LSTM model performed well in this prediction task. Due to the introduction of the attention mechanism, the selected model in this paper is more capable of capturing the time series information, so the BiLSTM model based on the attention mechanism performs the best in the task of landslide risk deformation prediction for hybrid pumped storage projects. In contrast, the TCN model performs between this paper's model and the LSTM model due to its good robustness, while the SVR model performs less well in this prediction task. From the figure, it can be clearly seen that the BiLSTM neural network deformation prediction model based on the attention mechanism adopted in this paper performs well in all deformation rate intervals and different deformation trends, and even for the points with large fluctuations in deformation, the model does not have obvious deviations, showing high stability and reliability, which verifies the effectiveness of the model adopted in landslide deformation prediction. Ten data points were randomly selected to calculate the average absolute percentage error between the predicted value and the true value, and the accuracy of the model in the 10 data points fell between 0.002% and 0.005%, and the fluctuation was small. Although there are cases where the accuracy of the LSTM model and TCN model exceeds that of this paper's model for a single data point, their overall accuracy is lower than that of this paper's model and their performance is unstable. Thus, based on the hydrogeological data in the online monitoring database, the landslide risk prediction model of the hybrid pumped storage project based on the attention mechanism and the BiLSTM model can realize the accurate prediction of landslide risk deformation, and provide data guarantee for the safe construction of the hybrid pumped storage project.



(a) <-20mm/a

(b) -19.9~5mm/a

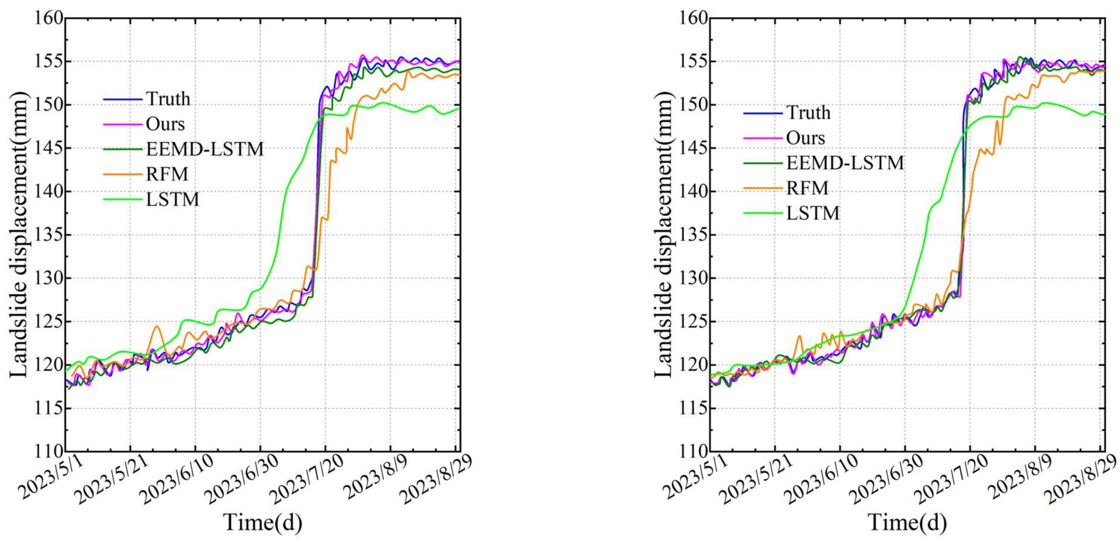
(c) 5~19.9mm/a

(d) >20mm/a

Fig. 5. The deformation prediction of different models

4.2.2. Landslide risk displacement prediction. To validate the performance of CNN-BiLSTM-ATT model for landslide risk displacement prediction in hybrid pumped storage projects, data obtained from other monitoring points of landslides in the study area are used as an auxiliary validation set. In previous landslide displacement prediction, many classical machine learning and deep learning models have achieved good results, such as the EEMD-LSTM model, the LSTM model, and the random forest model (RFM). Therefore, in this paper, the prediction results of these 3 models are compared with those of the CNN-BiLSTM-ATT model, and 2 scenarios with and without rainfall are considered. The results of the comparison between the predicted and actual values of landslide displacements obtained by using different learning models based on the monitoring point data are shown in Fig. 6, where Fig. 6(a)~(b) shows the cases with and without rainfall, respectively. Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Coefficient of Determination (R^2) were also selected as model evaluation indexes, and the landslide risk displacement prediction results of different models were obtained as shown in Table 5.

Due to the relatively short time of the selected landslide displacement data, a single LSTM model will produce a relatively large error when dealing with a small amount of data, while the LSTM model after adding the EEMD method for data decomposition has a great improvement in the accuracy of short-term landslide risk displacement prediction. The random forest method also has a good effect on the displacement prediction of landslide risk in hybrid pumped storage projects, but the change of its accuracy index is significantly higher than that of the EEMD-LSTM model and the CNN-LSTM-ATT model after adding rainfall data. The EEMD-LSTM and CNN-LSTM-ATT models outperform the random forest approach in all accuracy metrics in the presence and absence of rainfall.The CNN-LSTM-ATT model outperforms the traditional random forest model, the LSTM model, and the EEMD-LSTM model in prediction. According to the data from the landslide monitoring sites of the hybrid pumped storage project, the predictive evaluation indexes of the model in the absence of rainfall showed RMSE and MAE of 0.4213 mm and 0.3151 mm, respectively, MAPE of 0.2374%, and R^2 of 0.9998. The model still maintains high accuracy when there is rainfall, and the model also achieves good results in predicting landslide risk displacements at the monitoring sites of hybrid pumped storage projects, showing excellent performance in comparison with other models. In addition, the accuracy of the CNN-LSTM-ATT model is slightly better than that of the EEMD-LSTM model, and the systematic error that occurs in the decomposition of landslide risk displacements into trend and period term data is significantly optimized in the CNN-LSTM-ATT model compared to the EEMD-LSTM model, which makes the CNN-LSTM-ATT model a better research tool in the hybrid pumped-storage project for landslide risk displacement prediction has better research significance.



(a) With rainfall (b) Without rainfall

Fig. 6. Prediction results of different model

Table 5. Performance evaluation of prediction results (with/without rainfall)

Index	Rainfall	Ours	EEMD-LSTM	LSTM	RFM
RMSE (mm)	With	0.4357	0.6251	1.9852	0.8364
	Without	0.4213	0.6185	3.4261	1.0402
MAE (mm)	With	0.2926	0.4253	1.3359	0.6362
	Without	0.3151	0.4192	2.7354	0.7317
MAPE (%)	With	0.2462	0.2638	0.7898	0.5036
	Without	0.2374	0.2574	1.5764	0.5783
R^2	With	0.9995	0.9985	0.9816	0.9891

	Without	0.9998	0.9985	0.9451	0.9882

5. Conclusion

Based on the hybrid pumped storage online monitoring system, the article calculates the soil erosion during the construction of the hybrid pumped storage project through the online monitoring data combined with the CSLE model, and predicts and analyzes the deformation and displacement of the landslide risk by combining the CNN-BiLSTM-ATT model.

1) During the construction of the hybrid pumped storage project, its soil erosion is mainly distributed on the construction land, garden land, grassland and cultivated land. The erosion area of construction land is the largest (132.19 km^2), followed by the erosion area of cultivated land (29.24 km^2), both of which account for 69.15% and 15.3% of the total erosion area in the study area, respectively. This indicates that the hybrid pumped storage project will affect the building land and cultivated land to some extent.

2) When the CNN-BiLSTM-ATT model is utilized for landslide risk deformation prediction of hybrid pumped storage projects, its average absolute percentage error falls between [0.002%, 0.005%] and fluctuates less. When performing landslide risk displacement prediction, each evaluation index with and without rainfall is significantly better than the comparison model. This indicates that it is feasible to utilize the CNN-BiLSTM-ATT model for landslide risk prediction of hybrid pumped storage projects, which can be better predicted by online monitoring data to fully ensure the safe and stable operation of hybrid pumped storage projects.

Funding

This work was supported by Science and Technology Project of State Grid Shaanxi Electric Power Co., Ltd. (5226KY23001W).

References

[1] HOU, L. T. (2022). Research on soil erosion monitoring of production and construction projects based on RUSLE. HUBEI AGRICULTURAL SCIENCES, 61(4), 66.

[2] Ruiz, R. A., De Vilder, L. H., Prasasti, E. B., Aouad, M., De Luca, A., Geisseler, B., ... & Goseberg, N. (2022). Low-head pumped hydro storage: A review on civil structure designs, legal and environmental aspects to make its realization feasible in seawater. Renewable and Sustainable Energy Reviews, 160, 112281.

[3] Li, S., Hu, Y., & Zhang, L. (2022). Coupling coordination relationship of pumped storage power station and eco-environment system. Journal of Energy Storage, 52, 105029.

[4] Hunt, J. D., Byers, E., Riahi, K., & Langan, S. (2018). Comparison between seasonal pumped-storage and conventional reservoir dams from the water, energy and land nexus perspective. Energy conversion and management, 166, 385-401.

[5] Lu, B., Stocks, M., Blakers, A., & Anderson, K. (2018). Geographic information system algorithms to locate prospective sites for pumped hydro energy storage. Applied energy, 222, 300-312.

[6] Kim, H. G., & Park, C. Y. (2021). Landslide susceptibility analysis of photovoltaic power stations in Gangwon-do, Republic of Korea. Geomatics, Natural Hazards and Risk, 12(1), 2328-2351.

[7] Liu, F., Yang, K., Yang, T., Gao, Y., Li, J., Liu, Q., & Fu, Q. (2022). Pumped storage hydropower in an abandoned open-pit coal mine: Slope stability analysis under different water levels. Frontiers in Earth Science, 10, 941119.

[8] Alnaqbi, S. A., Alasad, S., Aljaghoub, H., Alami, A. H., Abdelkareem, M. A., & Olabi, A. G. (2022). Applicability of hydropower generation and pumped hydro energy storage in the Middle East and North Africa. Energies, 15(7), 2412.

[9] Li, Y., Chen, J., Li, Z., Han, X., Zhai, S., Li, Y., & Zhang, Y. (2021). A case study of debris flow risk assessment and hazard range prediction based on a neural network algorithm and finite volume shallow water flow model. Environmental Earth Sciences, 80, 1-19.

[10] Luo, P., Han, X., & Chen, N. (2022). Study on complexity planning model of pumped storage in abandoned open-pit mine: a case study of Fushun open-pit mine. Frontiers in Earth Science, 10, 808371.

[11] Choi, E., Ha, J. G., Hahm, D., & Kim, M. K. (2021). A review of multihazard risk assessment: Progress, potential, and challenges in the application to nuclear power plants. International Journal of Disaster Risk Reduction, 53, 101933.

[12] Mikroutsikos, A., Theocharis, A. I., Koukouzas, N. C., & Zevgolis, I. E. (2024). Slope stability of reclaimed coal mines through a new water filling index. Journal of Rock Mechanics and Geotechnical Engineering, 16(3), 828-839.

[13] Miao, F., Wu, Y., Li, L., Tang, H., & Li, Y. (2018). Centrifuge model test on the retrogressive landslide subjected to reservoir water level fluctuation. Engineering geology, 245, 169-179.

[14] Hu, J., Wang, C., & Ma, X. (2024, February). Countermeasures for Ecological Environment Restoration of Pumped Storage Power Stations in Northern Stone Mountain Areas. In IDEE 2023: Proceedings of the International Conference on Industrial Design and Environmental Engineering, IDEE 2023, November 24–26, 2023, Zhengzhou, China (p. 441). European Alliance for Innovation.

[15] Park, H. J., Shin, H. J., Kim, D. H., & Lee, S. O. (2024). A Score-Based Evaluation Model for Rehabilitation of Existing Pumped Storage Hydropower Plant Construction. International Journal of Engineering & Technology Innovation, 14(2).

[16] Miao, K., Bai, Z., Huang, Y., Huang, Y., & Su, Y. (2022). Research on seepage control of Jurong pumped storage hydroelectric power station. Water, 14(2), 141.

[17] Huang, Y., Zhou, Z., Wang, J., & Chen, Z. (2018). Investigation of the hydraulic connection between the upper storage reservoir and the underground powerhouse in the Hongping Pumped Storage Hydroelectric Powerstation. Environmental earth sciences, 77, 1-11.

[18] Wang, Z., Liu, J., Zhong, S., Qiao, L., Li, W., & Guo, J. (2023). Hydrogeological model for underground oil storage in rock caverns. Tunnelling and Underground Space Technology, 132, 104880.

[19] Wang, E., Zhong, J., Zhao, Y., & Mao, W. (2015). Analysis of seepage and seepage control measures in the rock masses of the Huilong pumped-storage power station. Bulletin of Engineering Geology and the Environment, 74, 1453-1462.

[20] Datta Debidas & Biswas Subhasankar Debsarma. (2022). Pumped Storage Project Initiative of West Bengal to meet Peakpower Demand and Grid Stabilization. Water and Energy International(8),35-37.

[21] Benneth Oyinna,Lois E. David,Kenneth E. Okedu,Ogheneruona E. Diemuodeke,Akhtar Kalam & Ilhami Colak. (2024). SMS based reservoir condition monitoring sensor system for pumped hydro storage in mini-grids. Measurement: Sensors101242-.

[22] Yi Zou,Yimei Wang,Yanhu He,Lirong Zhu,Shiyu Xue,Xu Liang & Changqing Ye. (2024). Soil Erosion Characteristics in Tropical Island Watersheds Based on CSLE Model: Discussion of Driving Mechanisms. Land(3).

[23] Yue Xu,Quan Pan,Zengfu Wang & Baoquan Hu. (2024). A Novel Trajectory Prediction Method Based on CNN, BiLSTM, and Multi-Head Attention Mechanism. Aerospace(10),822-822.

[24] Di Wang,Sha Li & Xiaojin Fu. (2024). Short-Term Power Load Forecasting Based on Secondary Cleaning and CNN-BILSTM-Attention. Energies(16),4142-4142.