



Article Degradation Trend Prediction of Hydropower Units Based on a Comprehensive Deterioration Index and LSTM

Yunhe Wang ¹, Zhihuai Xiao ¹,*, Dong Liu ^{2,3}, Jinbao Chen ¹, Dong Liu ⁴ and Xiao Hu ⁵

- ¹ School of Power and Mechanical Engineering, Wuhan University, Wuhan 430072, China
- ² College of Energy and Power Engineering, North China University of Water Resources and Electric Power, Zhengzhou 450045, China
 - ³ School of Civil and Hydraulic Engineering, Huazhong University of Science and Technology, Wuhan 430074, China
 - ⁴ China Yangtze Power Co., Ltd., Technical Center, Yichang 443000, China
 - ⁵ Department of Power Electronics Engineering, Hubei Water Resources Technical College, Wuhan 430200, China
 - * Correspondence: xiaozhihuai@126.com

Abstract: Deterioration trend prediction of hydropower units helps to detect abnormal conditions of hydropower units and can prevent early failures. The reliability and accuracy of the prediction results are crucial to ensure the safe operation of the units and promote the stable operation of the power system. In this paper, the long short-term neural network (LSTM) is introduced, a comprehensive deterioration index (CDI) trend prediction model based on the time-frequency domain is proposed, and the prediction accuracy of the situation trend of hydropower units is improved. Firstly, the time-domain health model (THM) is constructed with back-propagation neural network (BPNN) and condition parameters of active power, guide vane opening and blade opening and the timedomain indicators. Subsequently, a frequency-domain health model (FHM) is established based on ensemble empirical mode decomposition (EEMD), approximate entropy (ApEn), and k-means clustering algorithm. Later, the time-domain degradation index (TDI) is developed according to THM, the frequency-domain degradation index (FDI) is constructed according to FHM, and the CDI is calculated as a weighted sum by TDI and FDI. Finally, the prediction model of LSTM is proposed based on the CDI to achieve degradation trend prediction. In order to validate the effectiveness of the CDI and the accuracy of the prediction model, the vibration waveform dataset of a hydropower plant in China is taken as a case study and compared with four different prediction models. The results demonstrate that the proposed model outperforms other comparison models in terms of predicting accuracy and stability.

Keywords: hydropower units; degradation trend prediction; comprehensive deterioration index; long and short-term neural network; ensemble empirical mode decomposition; approximate entropy

1. Introduction

Hydropower units, as the critical equipment for hydropower energy conversion, have always been a focus of attention in the power industry for their safety and stability [1–7]. Along with the continuous development of the unit to the large scale as well as the complex, the degree of integration is getting higher and higher, and the structure is also becoming more and more sophisticated [6]. With the increase in accumulated operation time, hydropower units are prone to abnormal vibration, equipment exhaustion, unit deterioration, and other occurrences [5]. As the deterioration degree increases, the equipment performance of the unit will decline gradually until equipment breakdown happens [7]. Not only will the safe and stable operation of hydropower units and power stations be affected, but this will also bring about economic losses such as the additional cost of maintenance. Consequently, in considering the safety and stability of the hydropower unit and the power



Citation: Wang, Y.; Xiao, Z.; Liu, D.; Chen, J.; Liu, D.; Hu, X. Degradation Trend Prediction of Hydropower Units Based on a Comprehensive Deterioration Index and LSTM. *Energies* **2022**, *15*, 6273. https:// doi.org/10.3390/en15176273

Academic Editor: Chirag Trivedi

Received: 31 July 2022 Accepted: 21 August 2022 Published: 28 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). system, it is helpful to accurately predict the operating status trend of the hydropower unit to prevent early failure by detecting abnormal conditions of the hydropower unit. In this way, scientific and reliable maintenance plans and measures can be planned to optimize the comprehensive benefits of power plant operation. It is, therefore, of major significance to conduct research related to the condition trend prediction of hydropower units [1–7].

At this time, the research on health performance trend prediction of hydropower units is still at the initial stage, and the research experience of rotating machinery is summarized that the equipment quality degradation trend prediction is classified into three steps: (a) establishing a health state model; (b) constructing deterioration index (DI); and (c) achieving the trend prediction of the hydropower unit degradation [1]. The construction of a sensible equipment health state model, the proposal of DI that actually describe the operating state of the unit, and the adoption of an accurate predictive trend prediction model are the essential elements to realize the trend prediction of hydropower unit deterioration. In the existing domestic and international studies, the health model (HM) is constructed by analyzing the correlation between the stability parameters reflecting the operating condition of the unit and the working condition parameters. As abnormal vibration is one of the main causes of unit performance deterioration, the stabilization-related signals such as vibration and oscillation of the unit's shaft system can well describe the operating condition of the unit. Examples of stability parameters used to construct the HM in the current study are the original monitoring data, such as the original measured point values of frame vibration and shaft oscillation, sometimes domain indicator values, such as peak-to-peak and standard deviation (SD), and related working condition parameters, such as working head, active power, and guide vane opening, etc., which can be one-dimensional or multidimensional [1–16]. Shan et al. [1] used back-propagation (BPNN) in their study to establish a health state model with working condition parameters and horizontal vibration values of the Y-direction of the lower bracket. The relative error between the vibration health value and the measured vibration value was used as the DI. And the multi-objective particle swarm algorithm was used to optimize the parameters of the kernel extreme learning machine, and the optimized kernel extreme learning machine model was used to construct the prediction model. Fu et al. [2] applied modal decomposition to the turbine guide Y-direction oscillation monitoring data, aggregated and reconstructed the obtained modal components, and calculated the phase space matrix of each reconstructed modal component, as well as used support vector machines to predict each phase space matrix and summed the predicted values of each component to obtain the final predicted value of the oscillation degree and to achieve the deterioration prediction of the unit operating condition assessment. An et al. [7] developed an HM based on the radial basis function neural network (RBF) for the vibration peaks in the horizontal direction of the upper bracket of the unit and the working condition parameters and calculated the ratio of the health value to the measured value to obtain the unit degradation degree. This time series of the unit degradation degree is decomposed into several intrinsic mode functions (IMFs), and the complexity of the modal components is determined by calculating the approximate entropy (ApEn) of each modal component. When the ApEn is high, the RBF neural network is selected to predict the series. When the ApEn is small, a gray-system model is selected to predict the sequence, and the predictions of the decomposed components are summed to obtain the prediction results of the initial time series. In a similar way, An et al. [8] constructed an HM based on the horizontal vibration of the upper support and the water head and active power. The ratio of the true value to the healthy value under the current working condition parameters was used to evaluate the degree of unit deterioration for the current working condition. A gray-system model and RBF were used to construct the prediction model. Related research work [15,16] similarly developed HMs containing working head and active power for the prediction of the degradation trend of hydropower units.

After analysis of the above studies, it can be seen that: (1) The form of constructing the health status model of hydropower units in the existing studies is relatively easy, and most of the stability parameters for constructing the HM only consider the detection value of the

original single measurement point or a single time-domain index value, and such stability parameters cannot objectively reflect the operating status of the units comprehensively; (2) The form of constructing the DI is also relatively simple, which cannot authentically characterize the state change trend of hydropower units, and the large-scale historical data generated by the condition monitoring system is not effectively applied, and the reliability of its prediction results needs to be advanced [6]. Therefore, to address the shortcomings of constructing unit HM and DI in the current hydropower deterioration trend prediction, it is suggested that a comprehensive deterioration index (CDI) of hydropower units should be constructed with both the time-domain health model (THM) and frequency-domain health model (FHM), so as to comprehensively consider the change characteristics and trends of the operating state of hydropower units in the time-frequency domain to achieve the real-time prediction of hydropower unit deterioration degree. The research objective is to predict the deterioration level of hydropower units in real time.

State trend measurement is a time series prediction problem where the historical state index values are used to predict the future state index values for the purpose of predicting the future operating state of the unit. Whether a time series can predict future data based on historical data is dependent on the correlation between its future and historical data [17]. The time series reflecting the deterioration trend of the hydropower unit is between the unpredictable white noise time series and the fully predictable periodic signal time series, and therefore, it can be considered to be predictable to a reasonable degree. Due to the strong volatility and nonlinearity of the vibration signal, the calculated deterioration indicator series is strongly non-smooth and contains some noisiness, so it is a challenging topic to achieve an accurate prediction of the indicator series [5]. Experts and scholars have conducted relevant studies and proposed some prediction methods [1–19]. Qin et al. [17] established a long short-term memory neural network (LSTM) prediction model for wind speed prediction on the original wind speed series. Wang et al. [18] proposed a short-term wind speed prediction method based on ensemble empirical mode decomposition (EEMD) and an optimized BPNN. Lu et al. [3] proposed a state trend prediction model for hydropower units based on EEMD and genetic algorithm parameter seeking BPNN. Fu et al. [2] established a state trend prediction model for hydropower units based on aggregated EEMD and SVM theory, which can effectively predict the unit state. Yang et al. [19] proposed a prediction model combining wavelet transform and SVM to achieve short-term prediction of vibration signals. Due to the existence of background noise and electromagnetic interference, the state signals of hydropower units are often non-smooth, which can significantly affect the prediction results. Therefore, the signal needs to be pre-processed before building the prediction model [2]. Through the above literature analysis, the complexity and non-stationarity of the hydropower unit vibration signal lead to the inaccuracy of prediction, and the prediction accuracy of the deterioration trend of hydropower units needs to be enhanced further. LSTM has a unique memory and forgetting pattern, which can effectively deal with the long-term dependence of time series and effectively use the historical input information of time series to achieve an accurate prediction of the deterioration trend of hydropower units [17].

Through the above analysis, a trend prediction model for hydropower units based on CDI and LSTM is proposed in this paper. Firstly, a THM based on BPNN is constructed, and the Sparrow Search Algorithm (SSA) is used to optimize the parameters of BPNN; meanwhile, an FHM based on EEMD combined with approximate entropy and K-mean clustering is constructed; secondly, the time-domain deterioration index (TDI) and frequency-domain deterioration index (FDI) are calculated separately, in particular, the TDI is the variation of the real-time value relative to the health value, and the FDI is the Euclidean distance between the real-time feature vector and the health center vector. After a while, the CDI is obtained through weighted calculation, which can comprehensively and objectively reflect the operating trend and deterioration degree of the unit; a series of smooth modal components are obtained by modal decomposition of the CDI, and the LSTM-based prediction model is constructed for each modal component to make a prediction, and the prediction results of each component are summed to obtain the final prediction result, which reflects the future operating status of the hydropower unit. In this paper, the validity of the CDI to characterize the operating status of hydropower units and the validity of the prediction model to reflect the deterioration trend of the units are examined by the actual operation cases of a hydropower plant in China.

2. Theoretical Background

2.1. SSA Algorithm and BPNN

An artificial neural network is a mathematical model established by simulating the structure of the human brain. BPNN is a feed-forward neural network with the forward transmission of signal and reverse transmission of error. Although it is widely used, there are disadvantages, such as the tendency to fall into local minima and slow convergence speed [20–23]. The setting of initial weights and thresholds of BPNN has a strong influence on the training effect of the network. In this paper, the initial values of BPNN weights and thresholds are optimized using the SSA, and the neural network is trained twice to achieve the global optimum.

Sparrow Search Algorithm (SSA) is a swarm intelligence algorithm proposed in 2020 based on the optimization of social features of a population. The algorithm simulates sparrow foraging and anti-predation behaviors, distinguishes individuals into finders, followers, and vigilantes, and accomplishes the acquisition of resources by continuously updating individual positions, each of which corresponds to a solution, thus obtaining the mathematically optimal solution. Compared with traditional algorithms, the sparrow search algorithm has a simpler structure, is easy to achieve, and has fewer control parameters and better local search capability [24]. If the number of individuals in the population is n, then the population consisting of all individuals can be expressed as shown in Equation (1).

$$X = [X_1, X_2, \cdots X_n]^T \tag{1}$$

where X_i represents an individual in the set $i = 1, 2, \dots, n$.

The respective corresponding fitness function for each individual is shown in Equation (2).

$$F = [f(X_1), f(X_2), \cdots f(X_n)]^T$$
(2)

where $f(X_i)$ represents the fitness of each individual in the set $i = 1, 2, \dots, n$.

Where the discoverer location is updated in the following way, as shown in Equation (3).

$$x_{i,j}^{t+1} = \begin{cases} x_{i,j}^t \cdot \exp\left(\frac{-i}{\alpha \times iter_{\max}}\right), R_2 < ST\\ x_{i,j}^t + Q \cdot L, \qquad R_2 \ge ST \end{cases}$$
(3)

where *t* represents the number of current iterations, $x_{i,j}^t$ represents the position of the *ith* individual in the *tth* generation in the *jth* dimension, α is a random number, $\alpha \in (0, 1)$, *iter*_{max} is the maximum number of iterations, R_2 represents the warning value, $R_2 \in [0, 1]$, *ST* represents the safety threshold, *ST* $\in [0.5, 1]$, *Q* is a random number obeying normal distribution, *L* is the all-1 matrix of $1 \times dim$, and the *dim* represents the dimensionality.

The position of the follower is updated as shown in Equation (4).

$$x_{i,j}^{t+1} = \begin{cases} Q \cdot \exp\left(\frac{x_{worst}^{t} - x_{i,j}^{t}}{i^{2}}\right), & i > \frac{n}{2} \\ x_{p}^{t+1} + \left|x_{i,j}^{t} - x_{p}^{t+1}\right| \cdot A^{+} \cdot L, i \le \frac{n}{2} \end{cases}$$
(4)

where x_{worst}^t denotes the position of the worst adapted individual in the *tth* generation, and x_p^{t+1} denotes the position of the best adapted individual in the *t*+1*th* generation. A denotes the matrix of $1 \times dim$, and each element is randomly preset to -1 or 1, $A^+ = A^T (AA^T)^{-1}$.

The vigilantes' positions are updated, as shown in Equation (5).

$$x_{i,j}^{t+1} = \begin{cases} x_{best}^t + \beta \cdot \left| x_{i,j}^t - x_{best}^t \right|, & f_i \neq f_g \\ x_{best}^t + k \cdot \left(\frac{\left| x_{i,j}^t - x_{best}^t \right|}{|f_i - f_w| + \varepsilon} \right), & f_i = f_g \end{cases}$$
(5)

where x_{best}^t indicates the global optimal position in the *tth* generation, β is the control step, following a normal distribution with a mean 0 and variance 1, $k \in [-1, 1]$, k is a random number, and ε is set as a constant to avoid the denominator being 0. f_i denotes the fitness value of the current individual; f_g and f_w denote the fitness values of the current global optimal and worst individuals, respectively.

2.2. Empirical Modal Decomposition and Approximate Entropy

2.2.1. Ensemble Empirical Mode Decomposition

Empirical mode decomposition (EMD), proposed by Huang et al. [25], is an effective method for adaptive analysis of nonlinear and non-smooth signals. The basic idea of EMD is to perform adaptive smoothing on the original signal and obtain a series of IMFs by decomposing it step by step. Ensemble empirical mode decomposition (EEMD) improves on the traditional EMD decomposition by adding Gaussian white noise to the original data several times to compose a new signal, and the uniform frequency distribution of Gaussian white noise effectively avoids the modal aliasing phenomenon that exists when using EMD for signal decomposition [26]. The EEMD decomposition steps are as Equations (6)–(8).

(1) Add the white noise $n_i(t)$ with a set noise level to the original signal $x_i(t)$ to form the new signal:

$$x_i(t) = x(t) + n_i(t) \tag{6}$$

where $n_i(t)$ denotes the *ith* additive white noise sequence, $x_i(t)$ denotes the additional noise signal of the *ith* trial, $i = 1, 2, \dots, M$, M is the overall average number of times, which is the number of times white noise is added, and its value ranges from 100 to 300.

(2) To decompose the synthesized new signal by EMD, a series of IMFs components $c_{ij}(t)$, and a residual term $r_{ij}(t)$ are obtained:

$$x_i(t) = \sum_{j=1}^{J} c_{ij}(t) + r_{ij}(t)$$
(7)

The number of IMF components is *m*, which $c_{ij}(t)$ is the *jth* component of the decomposition after adding white noise for the *ith* time, *J* is the number of IMFs.

(3) Repeat steps (1) and (2) M times, and average the overall results, and the result is the IMF component of the original signal x(t) obtained by EEMD decomposition:

$$c_j(t) = \frac{1}{M} \sum_{i=1}^{M} c_{ij}(t)$$
(8)

where $c_j(t)$ is the *j*th IMF of the EEMD decomposition, $i = 1, 2, \dots, M$, $j = 1, 2, \dots, J$. Through the EEMD algorithm, the signal is decomposed into a series of IMF components at different time scales, and the fluctuations of each IMF component are smoother compared to the original signal. The advantage of smoothly processing the nonlinear nonstationary indicator series by using EEMD decomposition to avoid the errors due to direct prediction could, in theoretical terms, lead to more accurate prediction results [27].

2.2.2. Approximate Entropy

ApEn characterizes the complexity of a sequence, and the higher the complexity of the sequence, the higher the approximate entropy value [28]. Approximate entropy is widely

used in biomedical signal detection [29] and mechanical equipment fault diagnosis [30], and it has corresponding applications in the field of financial system complexity measurement [31]. Approximate entropy is resistant to strong external interference, does not require a long data length, and can be applied to deterministic signals or noisy signals.

For the data sequence $\{x_1, x_2, \dots, x_N\}$, its ApEn is calculated as follows [28].

(1) Form a set of m – dimensional vectors of $\{x_i\}$ in a continuous order:

$$X(i) = [x(i), x(i+1), \cdots x(i+m-1)]$$
(9)

where X_i represents an individual in the set, $i = 1, 2, \dots, N - m + 1$, N is the number of time series data points and m is the length of the window.

(2) Define the distance d[X(i), X(j)] between X(i) and X(j) to be the one with the largest value of the difference between the two corresponding elements, as follows:

$$d[\mathbf{X}(i), \mathbf{X}(j)] = \max|x(i+k) - x(j+k)|, \ k \in (1, m-1)$$
(10)

where for each value of *i*, calculate the distance $d[\mathbf{X}(i), \mathbf{X}(j)]$ between $\mathbf{X}(i)$ and the remaining corresponding elements of $\mathbf{X}(j)$ ($j = 1, 2, \dots, N - m + 1, j \neq i$).

(3) Iterate through each value, count the number of $d[\mathbf{X}(i), \mathbf{X}(j)] < r$ (*r* is the similarity tolerance, which is a pre-determined threshold, r > 0), and compare the value with the total number of vectors N - m + 1, which denote as $C_i^m(r)$.

$$C_i^m(r) = \frac{1}{N - m + 1} \operatorname{num}\{d[\mathbf{X}(i), \mathbf{X}(j)] < r\}$$
(11)

where $i = 1, 2, \dots, N - m + 1, j = 1, 2, \dots, N - m + 1, j \neq i$.

(4) Firstly, perform the logarithmic operation on $C_i^m(r)$, and then find its average value for all *i*, denoted as $\Phi^m(r)$, as follows:

$$\Phi^{m}(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \ln C_{i}^{m}(r)$$
(12)

(5) Increase the number of dimensions to m + 1 and repeat steps (1) to (4) to obtain $C_i^{m+1}(r)$ and $\Phi^{m+1}(r)$, as follows:

$$C_i^{m+1}(r) = \frac{1}{N-m} \operatorname{num}\{d[\mathbf{X}(i), \mathbf{X}(j)] < r\}$$
(13)

$$\Phi^{m+1}(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} \ln C_i^{m+1}(r)$$
(14)

(6) The ApEn of the sequence is calculated by the following equation:

$$\operatorname{ApEn}(m,r) = \lim_{N \to \infty} [\Phi^m(r) - \Phi^{m+1}(r)]$$
(15)

In practical engineering applications, *N* is a finite value, at this time, the ApEn of the sequence can be calculated by the following equation:

$$ApEn(m, r, N) = \Phi^{m}(r) - \Phi^{m+1}(r)$$
(16)

where *m* is the pattern dimension, given before calculating the approximate entropy; *r* is the similarity tolerance.

It was shown that the value of ApEn(m, r, N) is related to the values of m, r, and N [26]. When m = 2 and $r = (0.1 \sim 0.25)\sigma_x$ (σ_x is the SD of the original data series { x_i }), ApEn(m, r, N) is almost independent of the data length N, as follows:

$$\operatorname{ApEn}(m, r, N) \approx \operatorname{ApEn}(m, r)$$
 (17)

Therefore, in practical calculations, the sequence length is generally between 100–5000, the mode dimension m = 2, and the similarity tolerance $r = (0.1 \sim 0.25)\sigma_x$, which is chosen in this paper as m = 2, $r = 0.2\sigma_x$, σ_x is the SD of original data.

2.3. Long Short-Term Memory Neural Network

LSTM is developed on the foundation of a recurrent neural network (RNN) [32], which combines short-term memory and long-term memory through a special gate. The LSTM is based on the recurrent neural network (RNN) [32], which combines short-term memory and long-term memory with a special gate structure so that the network output has a strong correlation with current and historical inputs, solving the issue that the traditional recurrent neural network only has short-term memory, and can effectively use the time series history information to deal with the long-term correlation of time series.

The basic structure of LSTM is divided into three layers: input layer, hidden layer, and output layer. The hidden layer controls the information transmission by setting the threshold unit (gate structure), which gives it a unique memory pattern, and the structure of LSTM is shown in Figure 1.



Figure 1. The structure of LSTM.

It can be seen that the LSTM hidden layer contains three main gate structures, namely: forgetting gate, input gate, and output gate. Where the forgetting gate is used to filter to retain the information of neuron c_{t-1} history state at moment t - 1, the input gate determines the storage of new input information x_t of neuron at moment t, and the output gate is used to control the information delivery of the output value h_t of the hidden layer. According to Figure 1, the forward propagation algorithm of the LSTM can be derived as shown in Equations (18)–(24).

$$f_t = sig(W_f \cdot [h_{t-1}x_t] + b_f) \tag{18}$$

$$i_t = sig(W_i \cdot [h_{t-1}x_t] + b_i) \tag{19}$$

$$\widetilde{c}_t = \tanh(W_c \cdot [h_{t-1}x_t] + b_c) \tag{20}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \widetilde{c}_t \tag{21}$$

$$o_t = sig(W_o \cdot [h_{t-1}x_t] + b_o)$$
(22)

$$h_t = o_t \cdot \tanh(c_t) \tag{23}$$

$$y_t = sig(W_y \cdot h_t + b_y) \tag{24}$$

where Equation (18) represents the forgetting gate, Equations (19) and (20) represent the input gate, Equation (21) is the neuron state update expression, Equations (22) and (23) are the output gates, and Equation (24) is the calculation output of the neuron at moment t. x_t is the network input information at moment t, h_{t-1} is the network hidden layer state output value at moment t - 1, \tilde{c}_t is the input gate candidate state value, c_{t-1} and c_t is the neuron state at different moments, *sig* is the *sigmoid* activation function, W_f , W_i , W_c , W_o , W_y are the corresponding weight matrix, b_f , b_i , b_c , b_o , b_y are the corresponding threshold vectors, and y_t is the network prediction output at moment t.

After completing the forward propagation of the LSTM, it then enters the backpropagation process; that is, the LSTM is extended into a deep network in time order, and the weights and thresholds are updated iteratively using the back-propagation through time (BPTT) algorithm [33] and the chain rule until the optimal solution is obtained.

3. The Proposed Prediction Model Based on the CDI

3.1. Proposed Model

In this paper, in order to detect the failure signs of hydropower units in advance, realize fault warning, provide sufficient time margin for on-site maintenance and repair work, and thus improve the economic and social benefits of power stations, the EEMD-LSTM prediction model based on the CDI (CDI-EEMD-LSTM) in the time-frequency domain is proposed by making full use of the condition monitoring data of the industrialized information platform of hydropower units. The specific flow of the prediction model is shown in Figure 2, which includes four steps.

Step 1: The HMs of hydropower units are constructed, and this section is divided into two steps. The flowchart of Step 1 is shown in Figure 2.

(1) Constructing a THM of the unit based on SSA-BPNN, with the input being the operating parameters of the unit's historical health state and the output being the time-domain indicators of the unit's historical health state.

(2) Constructing an FHM based on EEMD-APEN and K-mean clustering algorithm, EEMD decomposition is performed on the vibration waveform numbers of the unit's historical health state to obtain the ApEn of each modal component, which constitutes a high-dimensional frequency-domain feature vector, and the health center vector of the hydropower unit is obtained by automatic clustering.

Step 2: The HMs and the health center vector are used to construct the CDI of the hydropower unit, which is divided into three steps. The flowchart of Step 2 is shown in Figure 3.

(1) Input the real-time operating parameters into the THM, obtain the health value under the current operating parameters, and calculate the relative error between the health value and the actual value as the TDI.

(2) Obtain the frequency-domain eigenvectors of the real-time waveform data of the unit vibration based on EEMD-ApEn, and calculate the Euclidean distance between the real-time frequency-domain eigenvectors and the health center vector as the FDI.

(3) Weight and sum the TDI and the FDI to construct the CDI in the time–frequency domain for hydropower units.

Step 3: The CDI-EEMD-LSTM prediction model is constructed.

To further improve the accuracy of trend prediction, EEMD is combined with LSTM. The EEMD of the CDI is performed first, the LSTM prediction model is constructed for each modal component obtained, and the fixed-length data is used as the input of the LSTM prediction model. With the superior accuracy of LSTM in time series prediction, each modal component is predicted, and eventually, the future state trend of the unit health index series is obtained by accumulation. The flowchart of Step 3 is shown in Figure 4.

Step 4: Evaluation and analysis of prediction results.

The prediction results of the CDI-EEMD-LSTM model were evaluated.

The process is shown in Figure 5.



(a)



Figure 2. Flowchart of Step 1: (**a**) is the flowchart of constructing THM; (**b**) is the flowchart of constructing FHM.



Figure 3. Flowchart of Step 2.



Figure 4. Flowchart of Step 3.



Figure 5. Trend prediction of hydropower unit health status based on CDI.

3.2. Evaluation Indicators

So as to evaluate the effectiveness of SSA-BPNN and CDI-EEMD-LSTM, this paper evaluated the prediction results using mean absolute percentage error (MAPE), root Mean

square error (RMSE), and correlation coefficient (CC). MAPE, RMSE, and CC are calculated as shown in Equations (25)–(27), respectively. The lower the RMSE and MAPE, the higher the accuracy of the model prediction. The CC is used to measure the strength of linear correlation between two variables, and a higher value indicates a higher correlation between the two, which also characterizes the more accurate prediction results.

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - y'_i}{y'_i} \right| \times 100\%$$
(25)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - y'_i)^2}$$
(26)

$$R_{xy} = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{N} (y_i - \bar{y})^2}}$$
(27)

4. Experiment Results and Analysis

Vibration signals contain abundant state characteristics, so they can be employed in engineering applications to assess the health status of equipment [34–36]. The condition information of hydropower units is also embedded in the vibration signals, and the actual conditions of unit operation can be obtained by analysis of the vibration signals of hydropower units [37]. In order to validate its effectiveness and engineering application value, this paper uses actual measurement data of hydropower units for experimental analysis.

3.7

A hydroelectric power station unit No. 3 is used as the research object, which is a vertical shaft semi-parachute type. The water turbine type frame is ZZA315-LJ-800, the rated speed is 107.1 r/min, the rated power is 200 MW, the rated head is 47 m, and the generator model is SF200-56/11950. On 28 August 2015, during the start up of Unit No. 3, there were obvious abnormal noises in the upper frame, waterwheel chamber, worm shell and tail pipe, etc. The noise of each department was more intense when the unit was loaded with 200 MW, and there were also obvious vibration and abnormal noises in the cement foundation along the -X direction in the inlet hole of the tail pipe of the hydraulic turbine. On 30 August, it was found that the steel plate of the middle ring of the runner chamber of the No. 3 unit was dislodged, the middle ring and the lower ring had serious cracks, and the blade skirt was seriously damaged.

Through retrospective analysis, the technicians deduced that the lining of the runner chamber fell off between 28 August 2015 at 14:17:03 and 28 August 2015 at 14:37:12. It is caused by the defects in the construction of the runner room, coupled with poor operating conditions, resulting in the fatigue damage of the runner room, the emergence of cracks, and the increase in vibration of the steel structure parts of the runner room, which eventually leads to the combination bolt loosening and shearing, and the rotor blade rubs against the cracked steel plate in the middle ring of the rotor chamber causing the steel plate to be torn and fall off.

A total of 507 sets of data, including point value data and waveform data, were obtained from the power station condition monitoring system before and after the failure of unit No. 3, with a sampling interval of about 20 min. Each set of data contains X and Y pendulum waveforms of the upper guide bearing, X and Y pendulum waveforms of the thrust bearing, X and Y pendulum waveforms of the water guide bearing, and axial vibration A, B, and C waveforms. Each waveform contains 16 key phases, with a total of 4096 points, and the sampling frequency is 458 Hz. Before the failure, unit No. 3 was operating at 63% of guide vane opening, 20% of blade opening, and working head of 50 to 55 m, and the vibration waveform data of axial A direction at the working point near this condition was selected for prediction analysis and verification of unit deterioration trend. All numerical simulation experiments are completed in MATLAB 9.2.

4.1. Constructing the CDI in the Time-Frequency Domain for Hydropower Units

4.1.1. Constructing a Time-Frequency Domain Health Model

Step 1: Constructing a THM for the unit.

The sample data (active power, guide vane opening and blade opening) of the health condition of a hydropower plant hydropower unit No. 3 are used as input, and the SD of the unit's axial A-directional vibration waveform is used as output. Compared with other time-domain indicators, the waveform SD can objectively reflect the operating status of hydropower units, and can identify unit health or fault conditions or even multiple fault categories [38]. The mapping relationship between waveform SD Y(t) and operating parameters X(t) is constructed as Equation (28):

$$Y(t) = f_{SSA-BP}[X_{t1}(t), X_{t2}(t), X_{t3}(t)]$$
(28)

where $[X_{t1}(t), X_{t2}(t), X_{t3}(t)] \in X(t)$.

The SD of the axial A-directional vibration waveform Y(t) and the corresponding operating conditions (active power $X_{t1}(t)$, guide vane opening $X_{t2}(t)$ and blade opening $X_{t3}(t)$), of 170 sets were selected when the unit was in the initial healthy state in the middle and early August. Moreover, 110 groups were randomly selected as the training set, 30 groups as the validation set, and 30 groups as the test set for the THM. The original values of the SD of the waveform of the axial A-direction vibration of the unit are shown in Figure 6, where the timepoint is the sample points in chronological order.



Figure 6. SD of the original axial A-directional vibration waveform.

When using SSA for the initial weights and initial thresholds of BPNN for the optimization search, to avoid over-learning and under-learning, both the initial weights a and thresholds b are set to [-30, 30], and MAPE is selected as the fitness function of SSA. The settings and modeling results of BPNN are shown in Table 1 and Figure 7.

Table 1. The parameters of BPNN.

Epochs	Training Function	Goal	Learning Rate
5000	traingda	$1e^{-7}$	0.1

From Table 1 and Figure 7, it is observed a good fit between the fitting output calculated by the THM and the real SD. The RMSE of the THM is 0.1806 and MAPE is 30.23%, which shows that the error values are low and the model is fitted with high precision. It can accurately reflect the relationship between the operating parameters and the time-domain



characteristics under the unit's health condition and provides a basis for constructing a TDI sequence for the unit.

Figure 7. Fitting results and fitting errors of SSA-BP HM.

Step 2: Construct the FHM and a frequency-domain health center feature vector.

While time-domain features can characterize the unit deterioration condition to a certain degree, in practice, the frequency-domain features of the vibration signal can reflect more information about the unit status, and the nonlinearity and non-stability of the unit deterioration trend will also be reflected by the signal frequency-domain features. Therefore, it is necessary to consider the frequency-domain characteristics of the unit to construct the CDI.

In this paper, to construct an FHI based on EEMD decomposition with ApEn and K-mean clustering, the unit frequency-domain feature vectors are extracted based on EEMD-ApEn, and the central feature vector in the health state is obtained using K-mean clustering. The specific steps implemented in this paper are shown below.

(1) 170 sets of axial A-directional vibration waveform data corresponding to the THM are selected, and EEMD is performed for each set of waveform data with noise level k = 0.2, overall mean M = 100, and the number of decompositions is 6. The eigenmodal components $IMF1 \sim IMF6$ are obtained after decomposition.

(2) Calculate the ApEn of each modal component separately to obtain the frequencydomain feature vector L in the health state.

$$L = [ApEn1, ApEn2, \cdots, ApEn6]$$
⁽²⁹⁾

(3) The feature vectors of the unit health samples are automatically clustered using the K-means method to obtain the cluster center Ω in the health condition, and the cluster center Ω is the health center vector of the FHM. This health center vector value characterizes the frequency-domain feature vector possessed by the unit during normal operation, so it can be used as the health vector to evaluate the unit deterioration from the frequency-domain perspective.

4.1.2. Constructing the Comprehensive Degradation Index

This section is divided into three steps.

Step 1: Obtain a sequence of TDI.

A sequence of TDI is constructed from the sample data of the unit in the process of developing from the healthy state to the fault state. A number of 230 sets of working parameters of the unit are selected in chronological order, and the working condition data during the development of the unit fault are input into Equation (28) to obtain the health value Y(t) under the current corresponding working condition, and the relative error

between the actual value T(t) and the health value Y(t) is adopted as the TDI of the unit, and the mathematical expression is Equation (30).

$$R_1(t) = \frac{Y(t) - T(t)}{T(t)} \times 100\%$$
(30)

where $R_1(t)$ is the sequence of vibration signal time-domain deterioration indicators as shown in Figure 8, which represents the deviation degree and deviation direction of the time-domain characteristics of the unit relative to the normal value in the physical sense.



Figure 8. Unit time-domain degradation index sequence $R_1(t)$.

The actual values of SD of axial vibration waveforms are compared with the healthy values, as shown in Figure 9. From Figure 9, it can be observed that the difference between the actual value series and the healthy value series is not significant at the beginning of operation, and the unit is well in health at this time. With an increasing operating time of the unit, the gap between the actual value series and the healthy value series gradually increases, the operating performance of the unit gradually deteriorates, and the TDI deviates toward the trend of more than the healthy value and the deviation degree increases sharply.



Figure 9. Comparison between T(t) and Y(t).

As shown in Figures 5 and 6, the TDI series has a strong non-stationary nature with fluctuating changes. The degree of deviation of the healthy value generally has an increasing trend, indicating that the condition of the unit gradually deteriorates with increasing operation time.

Step 2: Obtain a sequence of FDI.

The 230 sets of vibration waveform data corresponding to the CDI are selected, and the frequency-domain features of the unit are extracted based on EEMD-ApEn to obtain the frequency-domain feature vector collection L(t), and the Euclidean distance between the L(t) and the health center vector Ω is calculated to obtain the sequence of FDI $R_2(t)$. $R_2(t)$ is calculated as Equation (31).

$$\begin{cases} R_2(t) = \frac{\|L(t) - \Omega\|}{\|\Omega\|} \times 100\%, R_1(t) > 0\\ R_2(t) = -\frac{\|L(t) - \Omega\|}{\|\Omega\|} \times 100\%, R_1(t) < 0 \end{cases}$$
(31)

where $R_2(t)$ is sequence of the FDI shown in Figure 10, which physically indicates the deviation of the unit frequency-domain characteristics relative to the normal value, and constant positive because of the Euclidean distance. L(t) is the approximate entropy eigenvector of the actual measured vibration signal at time t. In order to ensure the homogeneity between the frequency-domain and time-domain degradation directions to avoid the phenomenon of mutual cancellation, $R_2(t)$ and $R_1(t)$ need to have the same sign, which means the TDI and FDI deviate from the normal value at the moment of the same time to maintain the same direction.



Figure 10. The sequence of FDI $R_2(t)$.

As shown in Figure 10, the $R_2(t)$ series has a strong non-smoothness and fluctuates with an overall increasing trend, which also indicates that the FDI can reflect the changing characteristics of the unit state gradually deviating from the healthy state with increasing operation time.

Step 3: Obtain the CDI in the time–frequency domain.

The TDI $R_1(t)$ and the FDI $R_2(t)$ are weighted and summed to obtain the time series of CDI in the time–frequency domain of hydropower units, as shown in Equation (32).

$$R(t) = \omega_1 \cdot R_1(t) + \omega_2 \cdot R_2(t) \tag{32}$$

where ω_1 and ω_2 are the weighting coefficients. To enhance the sensitivity of the degradation index to the abnormal data, the weights are taken as shown in Equations (33) and (34).

$$\omega_1 = \frac{|R_2(t)|}{|R_1(t)| + |R_2(t)|} \tag{33}$$

$$\omega_2 = \frac{|R_1(t)|}{|R_1(t)| + |R_2(t)|} \tag{34}$$

As seen in Figure 11, the R(t) series indicates that the unit health deteriorates gradually with increasing operating time.



Figure 11. The sequence of CDI in the time–frequency domain of the unit R(t).

4.2. Prediction Model of Unit Deterioration Trend Based on CDI-EEMD-LSTM

Based on the unit CDI obtained in Section 4.1, combined with the signal processing capability of EEMD and the time series prediction capability of LSTM, a forecasting investigation on the future trend of unit health status is carried out. To be consistent with engineering practice, the first 3/4 of the series in chronological order is used as the training set X_{Train} and the remaining part as the test set X_{Pred} . In the CDI-EEMD-LSTM, the CDI sequence can be decomposed into different frequency modal components by EEMD, and the fixed-length data in each modal component is used as the input of the LSTM, and the powerful nonlinear fitting ability of the LSTM is used to make single-step prediction for each modal component, and the prediction results of each modal component are accumulated to finally output the single-step prediction results of the CDI sequence. RMSE, MAPE, and similarity coefficient CC are employed as prediction effectiveness evaluation metrics, which are calculated as shown in Equations (25)–(27). The division of the training set and the prediction test set of the integrated degradation index is shown in Figure 12.

(1) Determining the input step length

The length of input data in LSTM affects the effect of prediction; too long input length will lead to information redundancy and affect the efficiency of the prediction model; too short will affect the accuracy of the prediction model. Firstly, the input length of the CDI prediction model is calculated by computing the autocorrelation coefficient of the index series. The autocorrelation of the CDI time series is shown in Figure 13, and the autocorrelation coefficient of each index is still above the 95% confidence level when the delay step is <10. In order to calculate the prediction effect of time series with different input step lengths within delay Step 10, RMSE, MAPE, and similarity coefficient R were used as evaluation indicators. The test was repeated 20 times, and the results are shown

in Table 2. When the input length is 8, the RMSE and MAPE metrics of the prediction results are the smallest, and the similarity coefficient CC is the largest, so the input length of the model is determined to be 8 and the prediction step is 1, that is, LSTM is an 8-input, 1-output network.



Figure 12. Division of training set and prediction test set of CDI.



Figure 13. Sequential autocorrelation analysis of CDI.

Table 2. Evaluation of prediction effects with different input lengths.

Input Length	RMSE	MAPE/%	CC
2	0.0320	22.66	0.7563
4	0.0306	23.64	0.7404
6	0.0310	24.86	0.7360
8	0.0292	21.67	0.9030
10	0.0334	26.48	0.6754

(2) Signal decomposition

Firstly, a series of indicators h is decomposed by EEMD to obtain each modal component, as shown in Figure 14. In terms of the number of sets and decomposition noise intensity selection, with the increase of the number of sets, the effect of Gaussian white noise added in the EEMD process on the decomposition effect gradually decreases and stabilizes, and different studies have shown that the effect of noise intensity on the result error is also relatively mild [39,40], so when using EEMD to process the time series of the degradation index, the number of sets is set to 100, and the auxiliary noise intensity is 0.2.



Figure 14. Modal components of the CDI.

(3) Analysis of prediction results

The prediction results of the proposed CDI-EEMD-LSTM are shown in Figure 15, and the evaluation indexes are shown in Table 3. Based on the signal processing capability of EEMD and the nonlinear fitting capability of LSTM, the prediction value of the model is almost consistent with the trend of the real value and can well reflect the fluctuation of the series R(t). The RMSE is 0.019, the MAPE is 15.1%, and the CC is 0.903, which can accurately predict the changing trend of the unit health status.



Figure 15. Prediction results of CDI-EEMD-LSTMNN.

Table 3. Index of prediction evaluation of CDI-EEMD-LSTM prediction model.

Prediction Model	MAPE/%	RMSE	CC
CDI-EEMD-LSTM	15.1	0.019	0.903

Figure 16 shows the prediction results of LSTM for each modal component of the CDI, and it can be seen that the prediction results of each component basically match the actual values. Further analysis of the prediction errors of each component shows that the smoother the modal component, the better the prediction outcome. Among the prediction results for all six modal components, the IMF1 component has the maximum prediction error and is the main source of deviation between the predicted and actual values of the index.



Figure 16. Prediction results of each modal component of the CDI. (**a**) is the prediction result of IFM1; (**b**) is the prediction result of IFM2; (**c**) is the prediction result of IFM3; (**d**) is the prediction result of IFM4; (**e**) is the prediction result of IFM5; (**f**) is the prediction result of IFM6.

4.3. Multi-Model Comparison Validation

4.3.1. Comparison of Indexes

The unit characteristics indexes can reflect the unit state changes, and it can be known from the literature [6] that the SD, peak-to-peak value, skewness, and kurtosis of the signal waveform increase constantly with time, reflecting the trend of unit deterioration in the same way as the CDI of the unit. The trends of the CDI proposed in this paper are consistent with the trends of the DI proposed in the literature [6]. In order to measure the sensitivity of the CDI proposed in this paper to reflect the state change of the unit, the indicator gradient rate (IGR) of the indexes is applied. Calculate the IGR of each index, that is, the sharpness of changes, to measure the sensitivity of the indexes to the change of unit status. The four points of H, I, II, and III, where the change of indexes amplitude is obvious, are selected in Figure 17, and the indexes are known to be healthy at the point of H. The change of the indexes' amplitude at the point I, II, and III relative to the amplitude of the indexes at the point of H is calculated as the IGR of the indexes, separately. The calculation is shown in the following Equations (35)–(37).

$$k_1 = \left| \frac{r(\mathbf{I}) - r(H)}{r(H)} \right| \tag{35}$$

$$k_2 = \left| \frac{r(\mathrm{II}) - r(H)}{r(H)} \right| \tag{36}$$

$$k_3 = \left| \frac{r(\mathrm{III}) - r(H)}{r(H)} \right| \tag{37}$$

where r(I), r(II), and r(III) are the amplitudes of signal SD, kurtosis, skewness, and peakto-peak values at moments I, II, and III, respectively; k_1 , k_2 , and k_3 are the IGR of signal SD, kurtosis, skewness, and peak-to-peak values at moments I, II, and III.



Figure 17. IGR of each index.

The IGR of each index is shown in Table 4 and Figure 17. In Table 4, where A, B, C, D, and E respectively represent the indicators: SD, kurtosis, skewness, peak-to-peak, and CDI. It can be seen that the IGR of the CDI is much better than the other indexes, which indicates that the index is more sensitive to the changes in the unit's operation status and can be more representative of the deterioration process from normal to failure than the traditional time and frequency-domain indexes.

Table 4. IGR of each indicator.

IGR	Α	В	С	D	Ε
k_1	0.022	0.095	0.317	0.055	0.700
k_2	0.146	0.237	0.627	0.467	3.715
k_3	0.134	0.063	0.219	0.104	4.463

4.3.2. Comparison of Predicted Results

To further verify the advantages of the proposed model prediction model in predicting the trend of failure sign indicators of hydropower units, four control groups were designed to verify the accuracy of the prediction model, where the four control groups are EEMD-GA-BP (the first control group), original sequence-LSTM (the second control group), EMD-LSTM (the third control group), and EEMD-SVM (the fourth control group). The selection of these models is based on the basic model and similar model to the proposed model, and the first control group is proposed in the 42nd literature [41], and the fourth control group is proposed in the second literature [2]. The proposed model is compared with four control groups and the original sequence. In the control group experiments, the training set and test set divisions and all parameters of the LSTM network are kept the same as those of the experimental group, and in using EEMD-GA-BPNN for deterioration indicator prediction, the BPNN structure is a three-layer structure of input layer-hidden layer-output layer, and the initial weights of the neural network and the initial threshold value of the neural network are optimized by genetic algorithm. The final prediction results obtained for each comparison model and the experimental model after training are shown in Figure 18. The evaluation indexes of the 5 prediction models are compared, as shown in Table 5.



Figure 18. Comparison of prediction results of different prediction models.

Predicted Models	RMSE	MAPE	СС
FEMD-GA-BPNN	0.313	0.047	0.576
EEMD-SVM	0.232	0.191	0.831
Original Sequence-LSTM	0.285	0.037	0.621
EMD-LSTM	0.152	0.021	0.884
Proposed model	0.151	0.019	0.903

Table 5. Evaluation indicators of prediction results of different forecasting models.

It can be visually seen from Figure 18 that among the several prediction models compared, the constructed EEMD-LSTM model prediction results are more closely matched to the actual change trend of the deterioration indicators. In contrast, when the LSTM is used to predict the original indicator sequence directly, the non-smoothness of the sign indicators during the normal-to-fault evolution leads to a serious deviation of the prediction results from the actual changes of the indicators.

The prediction result of the EEMD-GA-BPNN model is similar to that of the EEMD-SVM model, but the prediction result of the LSTMNN model is better than that of the BPNN and SVM models. From Table 5, it is observed that EMD-LSTM and EEMD-LSTM

outperformed EEMD-GA-BP and EEMD-SVM in all evaluation metrics. In the EEMD-LSTM model compared with the EEMD-GA-BP model, the RMSE and MAPE decreased by 0.162 and 0.028, respectively, and the CC improved by 0.327, while compared with the EEMD-LSSVM model, the RMSE and MAPE decreased by 0.081 and 0.172, respectively, and the CC improved by 0.072. The superiority of the LSTM model in nonlinear fitting is demonstrated. It can be shown that the LSTM has an outstanding prediction effect and higher accuracy for time-series indicators during the unit state fading process compared with the traditional BPNN and SVM and has a great advantage in the self-learning of time series. It can thus be shown that the LSTM is feasible for the prediction model of the time series of deterioration indicators.

Meanwhile, according to each evaluation index in Table 5, it can be seen that EEMD-LSTM compared with the original sequence-LSTM model, RMSE and MAPE decreased by 0.134 and 0.018, respectively, and CC improved by 0.282, which greatly improved the prediction accuracy. Compared with the EMD-LSTM model, RMSE and MAPE were reduced by 0.001 and 0.002, respectively, and CC was improved by 0.019. It can be seen that the EEMD-LSTM model has the best evaluation indicators and the highest prediction accuracy. The EEMD-LSTM predicts the degraded indicators more effectively than the EMD-LSTM, indicating the advantage of EEMD over EMD in signal smoothness decomposition, which shows that decomposing the mutated signals into smoothed component signals and reducing the non-smoothness and non-linearity of the indicator sequences can enhance the accuracy of prediction.

In the case study of this paper, the default values given by the toolbox are used for most of the parameters of the LSTM model, which may lead to a slight increase in the evaluation metrics. Overall, the overall prediction performance of the proposed EEMD-LSTM model is better than several of the remaining comparative models and can be used to predict the deterioration trend of hydropower units.

5. Conclusions

In order to improve the measurement accuracy of non-stationary and non-linear state trends of hydropower units, a trend prediction model (EEMD-LSTM) based on CDI is proposed in this paper. A THM is established by considering the mapping relationship between operating parameters such as active power, guide vane opening and blade opening, and the time-domain indicators, and an FHM is constructed based on EEMD-ApEn and the K-mean clustering algorithm. Based on the above health models, TDI and FDI were constructed, respectively, and the CDI was formed by weighted fusion. The main conclusions of this paper are as follows:

- 1. Autocorrelation analysis was performed on the deteriorated indicator series to obtain the appropriate correlation length. Too long or too short correlation length of the indicator series can cause excessive prediction errors. The analysis showed that the prediction input step length of the deterioration indicator series could be 8 steps when using historical data for prediction.
- 2. The EEMD-LSTM model compared with the EEMD-GA-BPNN model, RMSE and MAPE decreased by 0.162 and 0.028, respectively, and CC enhanced by 0.327, while compared with the EEMD-SVM model, RMSE and MAPE decreased by 0.081 and 0.172, respectively, and CC enhanced by 0.072. It can be obtained that the LSTM is more effective in predicting the time-series indicators in the asymptotic process, which is outstanding relative to the traditional prediction model in terms of time series self-learning. Thus, it can be shown that LSTM is feasible for predicting the deterioration trend of hydropower units.
- 3. EEMD-LSTM compared with the original sequence-LSTM model, RMSE and MAPE were reduced by 0.134 and 0.018, respectively, and CC was improved by 0.282, and the prediction accuracy was greatly improved. Compared with the EMD-LSTM model, RMSE and MAPE are reduced by 0.001 and 0.002, respectively, and CC is improved by 0.019. The EEMD-LSTM model has the highest evaluation prediction

accuracy. EEMD-LSTM outperforms EMD-LSTM in predicting degraded indicators, which shows the advantage of EEMD over EMD in signal smoothness decomposition, thus showing that decomposing mutated signals into smooth component signals and reducing the non-smoothness and non-linearity of indicator sequences can enhance the prediction accuracy.

In the case study of this paper, most of the parameters of the LSTM model use the default values given by the toolbox, which may lead to a slight increase in the evaluation metrics. Overall, the overall prediction performance of the proposed EEMD-LSTM model is better than the remaining several comparative models and can be used to predict the deterioration trend of hydropower units.

Author Contributions: Conceptualization, Y.W.; methodology, Y.W. and D.L. (Dong Liu, liudong-whu@126.com); software, J.C.; validation, Y.W. and X.H.; formal analysis, D.L. (Dong Liu, lewist-whu@163.com); investigation, Y.W. and D.L. (Dong Liu, liudongwhu@126.com); resources, Z.X.; data curation, Y.W.; writing—original draft preparation, Y.W.; writing—review and editing, D.L. (Dong Liu, liudongwhu@126.com); visualization, Y.W. and Z.X.; supervision, Z.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (NSFC) (No. 51979204) and China Postdoctoral Science Foundation (Grant No. 2020M682416).

Data Availability Statement: The data used to support the findings of this study are not available because it involves the parameters and full characteristic data of the actual hydropower station.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Shan, Y.; Liu, J.; Xu, Y.; Zhou, J. A combined multi-objective optimization model for degradation trend prediction of pumped storage unit. *Measurement* **2021**, *169*, 108373. [CrossRef]
- Fu, W.; Zhou, J.; Zhang, Y.; Zhu, W.; Xue, X.; Xu, Y. A state tendency measurement for a hydro-turbine generating unit based on aggregated EEMD and SVR. *Meas. Sci. Technol.* 2015, 26, 125008. [CrossRef]
- Hu, X.; Li, C.; Tang, G. A hybrid model for predicting the degradation trend of hydropower units based on deep learning. In Proceedings of the 2019 Prognostics and System Health Management Conference (PHM-Qingdao), Paris, France, 2–5 May 2019; pp. 1–5.
- 4. Liu, D.; Lai, X.; Xiao, Z.; Hu, X.; Zhang, P. Fault diagnosis of rotating machinery based on convolutional neural network and singular value decomposition. *Shock. Vib.* **2020**, 2020, 1–13. [CrossRef]
- Li, C.; Tang, G.; Xue, X.; Saeed, A.; Hu, X. Short-term wind speed interval prediction based on ensemble GRU model. *IEEE Trans.* Sustain. Energy 2019, 11, 1370–1380. [CrossRef]
- Liu, D.; Lai, X.; Hu, X.; Xiao, Z. Research on online assessment method of condition deterioration of hydropower units based on vibration signal. J. Water Resour. 2021, 52, 461–473.
- 7. An, X.; Pan, L. Characteristic parameter degradation prediction of hydropower unit based on radial basis function surface and empirical mode decomposition. *J. Vib. Control* **2013**, *21*, 2200–2211. [CrossRef]
- An, X.; Pan, L.; Yang, L. Condition parameter degradation assessment and prediction for hydropower units using Shepard surface and ITD. *Trans. Inst. Meas. Control* 2014, 36, 1074–1082. [CrossRef]
- 9. Guo, L.; Lei, Y.; Li, N.; Yan, T.; Li, N. Machinery health indicator construction based on convolutional neural networks considering trend burr. *Neurocomputing* **2018**, *292*, 142–150. [CrossRef]
- 10. Li, J.; Yao, X.; Wang, X.; Yu, Q.; Zhang, Y. Multiscale local features learning based on BP neural network for rolling bearing intelligent fault diagnosis. *Measurement* **2019**, *153*, 107419. [CrossRef]
- 11. Wu, J.; Hu, K.; Cheng, Y.; Zhu, H.; Shao, X.; Wang, Y. Data-driven remaining useful life prediction via multiple sensor signals and deep long short-term memory neural network. *ISA Trans.* **2019**, *97*, 241–250. [CrossRef]
- 12. Lu, S.; Li, Q.; Bai, L.; Wang, R. Performance predictions of ground source heat pump system based on random forest and back propagation neural network models. *Energy Convers. Manag.* **2019**, *197*, 111864. [CrossRef]
- 13. Zhang, X.; Jiang, Y.; Li, C.; Zhang, J. Health status assessment and prediction for pumped storage units using a novel health degradation index. *Mech. Syst. Signal Process.* **2022**, *171*, 108910. [CrossRef]
- 14. Liu, D.; Zeng, H.; Xiao, Z.; Peng, L.; Malik, O.P. Fault diagnosis of rotor using EMD thresholding-based de-noising combined with probabilistic neural network. *J. Vibroeng.* **2017**, *19*, 5920–5931.
- 15. Chen, P.; Deng, Y.; Zhang, X.; Ma, L.; Yan, Y.; Wu, Y.; Li, C. Degradation Trend Prediction of Pumped Storage Unit Based on MIC-LGBM and VMD-GRU Combined Model. *Energies* **2022**, *15*, 605. [CrossRef]

- 16. Tuerxun, W.; Chang, X.; Guo, H.Y.; Jin, Z.J.; Zhou, H.J. Fault diagnosis of wind turbines based on a support vector machine optimized by the sparrow search algorithm. *IEEE Access* **2021**, *9*, 69307–69315. [CrossRef]
- 17. Qin, Q.; Lai, X.; Zou, J. Direct Multistep Wind Speed Forecasting Using LSTM Neural Network Combining EEMD and Fuzzy Entropy. *Appl. Sci.* **2019**, *9*, 126. [CrossRef]
- Wang, S.X.; Zhang, N.; Wu, L.; Wang, Y.M. Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method. *Renew. Energy* 2016, 94, 629–636. [CrossRef]
- 19. Zou, M.; Zhou, J.; Liu, Z.; Zhan, L. A hybrid model for hydroturbine generating unit trend analysis. In Proceedings of the Third International Conference on Natural Computation (ICNC 2007), Hainan, China, 24–27 August 2007; Volume 2, pp. 570–574.
- 20. Jin, T.; Cheng, Q.; Chen, H.; Wang, S.; Guo, J.; Chen, C. Fault diagnosis of rotating machines based on EEMD-MPE and GA-BP. *Int. J. Adv. Manuf. Technol.* **2021**, 1–12. [CrossRef]
- 21. Zichun, Y. The BP Artificial Neural Network Model on Expressway Construction Phase Risk. Syst. Eng. Procedia 2012, 4, 409–415.
- Liu, H.; Mi, X.; Li, Y. Comparison of two new intelligent wind speed forecasting approaches based on wavelet packet decomposition, complete ensemble empirical mode decomposition with adaptive noise and artificial neural networks. *Energy Convers. Manag.* 2018, 155, 188–200. [CrossRef]
- 23. Zhou, K.B.; Zhang, J.Y.; Shan, Y.; Ge, M.-F.; Ge, Z.-Y.; Cao, G.-N. A hybrid multi-objective optimization model for vibration tendency prediction of hydropower generators. *Sensors* **2019**, *19*, 2055. [CrossRef] [PubMed]
- Xue, J.; Shen, B. A novel swarm intelligence optimization approach, sparrow search algorithm. Syst. Sci. Control Eng. 2020, 8, 22–34. [CrossRef]
- Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.-C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. A Math. Phys. Eng. Sci.* 1998, 454, 903–995. [CrossRef]
- 26. Ma, F.; Chen, X.; Du, J. Bearing Fault Diagnosis Based on Improved Hilbert-Huang Transform. In Proceedings of the 5th International Conference on Civil Engineering and Transportation, Guangzhou, China, 28–29 November 2015; pp. 551–555.
- 27. Huang, Y.; Yang, L.; Liu, S.; Wang, G. Multi-step wind speed forecasting based on ensemble empirical mode decomposition, long short term memory network and error correction strategy. *Energies* **2019**, *12*, 1822. [CrossRef]
- 28. Pincus, S.M. Approximate entropy as a measure of system complexity. Proc. Natl. Acad. Sci. USA 1991, 88, 2297–2301. [CrossRef]
- 29. Zhang, Z.A.; Zhou, Y.A.; Chen, Z.Y.B.; Tian, X.; Du, S.; Huang, R. Approximate entropy and support vector machines for electro-encephalogram signal classification. *Neural Regen. Res.* **2013**, *88*, 1844–1852.
- 30. Unal, M.; Onat, M.; Demetgul, M.; Kucuk, H. Fault diagnosis of rolling bearings using a genetic algorithm optimized neural network. *Measurement* 2014, *58*, 187–196. [CrossRef]
- Teng, Y.; Shang, P.; He, J. Multiscale fractional-order approximate entropy analysis of financial time series based on the cu-mulative distribution matrix. *Nonlinear Dyn.* 2019, 97, 1067–1085. [CrossRef]
- 32. Werbos, P.J. Backpropagation Through Time—What It Does And How To Do It. Proc. IEEE 1990, 78, 1550–1560. [CrossRef]
- Singh, A.; Singh, L.P.; Singh, S.; Singh, H.; Chhuneja, N.K.; Singh, M. Evaluation and analysis of occupational ride comfort in rotary soil tillage operation. *Measurement* 2019, 131, 19–27. [CrossRef]
- 34. Rezaeianjouybari, B.; Shang, Y. Deep learning for prognostics and health management, State of the art, challenges, and opportunities. *Measurement* **2020**, *163*, 107929. [CrossRef]
- 35. Soualhi, M.; Nguyen, K.T.P.; Soualhi, A.; Medjaher, K.; Hemsasc, K.E. Health monitoring of bearing and gear faults by using a new health indicator extracted from current signals. *Measurement* **2019**, *141*, 37–51. [CrossRef]
- Li, H.; Wang, Y.; Yang, X.P.; Jia, R.; Luo, X.Q. Vibration fault diagnosis for hydropower generating unit based on multiwavelet and PSO-RBF neural network. J. Northwest A F Univ. Nat. Sci. Ed. 2017, 45, 227–234.
- Li, L.M.; Wang, Z.S. Method of redundant features eliminating based on k-means clustering. *Appl. Mech. Mater.* 2014, 488, 1023–1026. [CrossRef]
- Lu, N.; Xiao, Z.; Malik, O.P. Feature extraction using adaptive multiwavelets and synthetic detection index for rotor fault diagnosis of rotating machinery. *Mech. Syst. Signal Process.* 2015, 52, 393–415. [CrossRef]
- Wang, W.; Chen, Q.; Yan, D.; Geng, D. A novel comprehensive evaluation method of the draft tube pressure pulsation of Francis turbine based on EEMD and information entropy. *Mech. Syst. Signal Process.* 2019, 116, 772–786. [CrossRef]
- Liu, D.; Xiao, Z.; Hu, X.; Zhang, C.; Malik, O. Feature extraction of rotor fault based on EEMD and curve code. *Measurement* 2019, 135, 712–724. [CrossRef]
- Lu, D.; Xiao, Z.H.; Liu, D.; Hu, X.; Deng, T. EEMD-GA-BP-based state trend prediction of hydropower units. *China Rural Water Conserv. Hydropower* 2021, 186–194. [CrossRef]