

Article

Degradation Trend Prediction of Pumped Storage Unit Based on MIC-LGBM and VMD-GRU Combined Model

Peng Chen ¹, Yumin Deng ², Xuegui Zhang ², Li Ma ², Yaoliang Yan ², Yifan Wu ¹ and Chaoshun Li ^{1,*}¹ School of Civil and Hydraulic Engineering, Huazhong University of Science and Technology, Wuhan 430074, China; hustpengchen@163.com (P.C.); wyfan@hust.edu.cn (Y.W.)² China Yangtze Three Gorges Group Co. Ltd., Wuhan 430010, China; deng_yumin@ctg.com.cn (Y.D.); zhang_xuegui@ctg.com.cn (X.Z.); ma_li2@ctg.com.cn (L.M.); yan_yaoliang@ctg.com.cn (Y.Y.)

* Correspondence: csli@hust.edu.cn; Tel.: +86-158-7180-0142

Abstract: The harsh operating environment aggravates the degradation of pumped storage units (PSUs). Degradation trend prediction (DTP) provides important support for the condition-based maintenance of PSUs. However, the complexity of the performance degradation index (PDI) sequence poses a severe challenge of the reliability of DTP. Additionally, the accuracy of healthy model is often ignored, resulting in an unconvincing PDI. To solve these problems, a combined DTP model that integrates the maximal information coefficient (MIC), light gradient boosting machine (LGBM), variational mode decomposition (VMD) and gated recurrent unit (GRU) is proposed. Firstly, MIC-LGBM is utilized to generate a high-precision healthy model. MIC is applied to select the working parameters with the most relevance, then the LGBM is utilized to construct the healthy model. Afterwards, a performance degradation index (PDI) is generated based on the LGBM healthy model and monitoring data. Finally, the VMD-GRU prediction model is designed to achieve precise DTP under the complex PDI sequence. The proposed model is verified by applying it to a PSU located in Zhejiang province, China. The results reveal that the proposed model achieves the highest precision healthy model and the best prediction performance compared with other comparative models. The absolute average ($|AVG|$) and standard deviation (STD) of fitting errors are reduced to 0.0275 and 0.9245, and the RMSE, MAE, and R^2 are 0.00395, 0.0032, and 0.9226 respectively, on average for two operating conditions.

Keywords: pumped storage unit; degradation trend prediction; maximal information coefficient; light gradient boosting machine; variational mode decomposition; gated recurrent unit



Citation: 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. <https://doi.org/10.3390/en15020605>

Academic Editor: Davide Astolfi

Received: 16 December 2021

Accepted: 12 January 2022

Published: 15 January 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/>).

1. Introduction

Pump storage units (PSUs) store excessive power during light load periods and convert hydro energy into electricity at peak load periods [1]. According to the hydropower status report [2], the installed capacity of PSUs reached 159.5 GW in 2020, accounting for 94% of the capacity of all energy storage facilities. PSUs are playing an increasingly important role in peak–valley reduction and emergency reserves [3–5]. However, the frequent condition conversion and the complex hydro-mechanical–electric coupling aggravate the wear and degradation of PSUs. Degradation trend prediction (DTP) ensures the secure operation of PSUs by evaluating the degradation and predicting the degradation trend of PSUs. Generally, DTP includes the mechanism-analysis approaches and data-driven approaches [6]. The mechanism-analysis approaches describe the degradation process by building a mathematical model based on the failure mechanism. However, complex systems are difficult to describe precisely with mathematical models, which limits their application. With the improvement of monitoring systems, data-driven approaches are attracting increased attention [7]. The data-driven based DTP always consists of two phases: (a) building a healthy model that represents the good running conditions of PSUs and then

constructing the performance degradation index (PDI); (b) establishing a precise prediction model to forecast the future degradation trend of PSUs.

Healthy model building is the process of constructing mapping relationships between working parameters and status data. Its accuracy directly affects the reliability of PDI. In relevant literature, the artificial neural network (ANN) [8], Gaussian process regression (GPR) [9], radial basis function interpolation surface [10,11], Shepard interpolation surface [12], etc. are frequently used. These methods need plenty of computational resources and time, while their performances are always not satisfactory. Recently, the gradient boosting machines (GBMs) [13] have developed rapidly due to their low computational resources, fast training speed, and high fitting accuracy. They are widely used in regression and classification tasks, such as wind speed forecasting [14,15], fault diagnosis [16,17], and anomaly detection [18,19]. The light gradient boosting machine (LGBM) [20] is a novel GBM proposed by Ke et al. in 2017. It has similar performance while requiring far less memory and training time compared with gradient boosting decision tree (GBDT) and has achieved state-of-the-art results in numerous competitions. Considering the outstanding advantages and the potential engineering demand, LGBM would be a great choice to build a healthy model to fit the mapping relationships exactly and save expensive computational resources.

Moreover, the input and the output of the healthy model should be determined. The status data, such as swing, vibration, etc., provide a wealth of information on the operating status of PSUs [21]. It is suitable to use the status data as the output of the healthy model. The working parameters, which describe the operating mode of PSU in detail, should be used as the input of the healthy model [8]. However, some working parameters have a weak correlation with the status data and bring confusion when judging the status of PSUs. To make sure the healthy model entirely learns the characteristics of a PSU under good running conditions, it is necessary to screen the working parameters by correlation. The maximal information coefficient (MIC) [22] can explore not only linear correlation but also nonlinear and nonfunctional correlations between variables, thus achieving remarkable success in data screening. Jiang et al. [23] designed a two-step feature selection method based on MIC to screen the best feature for predicting remaining useful life of the bearing. Ji et al. [24] proposed a novel selection method of software attribute by combining MIC and automatic clustering. Due to its superior performance, MIC is adopted to screen the working parameters in this paper.

The PDI is obtained after building the healthy model, after which the degradation trend of the PSU should be predicted to support decision-making. Classical machine learning methods, such as autoregressive integrated moving average (ARIMA) [25], support vector regression (SVR) [26], ANN [27], etc., are widely used in related works. With the development of machine learning, recurrent neural networks (RNNs) [28] have obtained excellent results in prediction tasks. Park et al. [29] used long short-term memory (LSTM) to predict the remaining useful life of a battery. Xia et al. [30] combined the multi-layer attention and LSTM models to predict the degradation trend of mechanical systems. Wu et al. [31] predicted the remaining useful life of a cooling system by using LSTM and gated recurrent unit (GRU), respectively, finding that GRU performs better than LSTM. Compared with LSTM, GRU [32] only has two gates and fewer parameters, while it often achieves slightly better results [33–35]. However, the degradation trend of PSU is non-periodic, with irregular fluctuation components. Even GRU cannot learn the degradation trend of PSU well. The complexity of PDI sequences brings difficulty for high-precision prediction. One way to solve this issue is to make the PDI sequence simpler. Empirical mode decomposition (EMD) [36] is a classical decomposition method which has been widely used. However, EMD lacks a theoretical foundation and suffers from problems such as mode mixing and boundary effect. To overcome these shortcomings, variational mode decomposition (VMD) [37] has been proposed. It has a sound theoretical foundation and is suitable for dealing with nonlinear and non-stationary series [38]. It decomposes the complex series into a series of approximately orthogonal simple modes and is popular in the fields of signal denoising [39], runoff forecasting [40], wind speed forecasting [41],

etc. Thus, the complex PDI sequences are decomposed into simpler modes by VMD before being fed into GRU to improve the accuracy of prediction.

To achieve precise degradation trend prediction for a PSU, a combined DTP model of a PSU is proposed based on MIC-LGBM and VMD-GRU. Firstly, the working parameters are selected by MIC. Afterwards, the LGBM healthy model is built, and the PDI is obtained by measuring the difference between the benchmark output of the healthy model and the monitoring status data. Finally, the PDI sequence is sent into the VMD-GRU prediction model to obtain a reliable future degradation trend. The main contributions of this work are listed as follows:

- (a) Considering that the relationships between the working condition parameters and the state data are not linear, MIC is utilized to screen the relevant working parameters. The interference of irrelevant working parameters is reduced, and the performance of the healthy model is improved.
- (b) Inspired by the superiority of LGBM, the healthy model is constructed and not only achieves a high-precision fitting result but also consumes fewer computational resources as it has a strongly competitive training speed.
- (c) To address the challenges caused by the complexity of PDI sequences, the VMD-GRU prediction model is designed for reliable prediction. The complex degradation trend is decomposed into a series of simple sequences by VMD, which can be more adequately learned by GRU. An outstanding prediction result is obtained compared with other popular prediction models.

The remainder of this paper is organized as follows. The relevant theoretical background is stated in Section 2. Then, the proposed DTP model is presented in Section 3. In Section 4, model validation, comparative experiments, and analysis are carried out. Conclusions and future work are presented in Section 5.

2. Theoretical Background

2.1. Maximal Information Coefficient

Compared with traditional correlation coefficients, MIC has generality and equitability. It not only captures linear, nonlinear, or even nonfunctional correlations (i.e., generality) but also assigns similar scores to any variables containing equal noise (i.e., equitability) [42]. Supposing a dataset of ordered pairs $D = \{X, Y\} = \{(x_i, y_i), i = 1, \dots, N\}$, where X and Y are variables with length N , the MIC is calculated with the following steps:

Step 1: Divide the D into m -by- n grids G , where $m * n \leq B$, and B is set to $N^{0.6}$ [22] in this paper.

Step 2: Calculate the maximum mutual information (MI) of D under G $MI^*(D|G)$ by Equation (1); then, the (m, n) th term of characteristic matrix $M_{m,n}$ is obtained by normalizing $MI^*(D|G)$ as Equation (2).

$$MI^*(D|G) = \max MI(G) = \max_{m,n} \sum p(x_i, y_i) \log_2 \left(\frac{p(x_i, y_i)}{p(x_i)p(y_i)} \right) \quad (1)$$

$$M_{m,n} = \frac{MI^*(D|G)}{\log_2 \min(m, n)} \quad (2)$$

where $p(x_i, y_i)$ denotes the joint probability density, and $p(x_i)$ and $p(y_i)$ denote the marginal probability densities.

Step 3: Calculate $M_{m,n}$ for all grids that satisfy $m * n \leq B$; then, the MIC of D is the maximum term in the characteristic matrix, namely

$$\text{MIC}(D) = \max_{m*n \leq B} M_{m,n} \quad (3)$$

where MIC ranges from 0 to 1. The larger the MIC, the stronger correlation.

2.2. Light Gradient Boosting Machine

Extreme gradient boosting (XGBoost) [43] and LGBM models are popular ensemble learning methods based on GDBT. Compared with XGBoost and GDBT, LGBM consumes fewer computational resources and has a faster training speed while obtaining a similar accuracy [20,44]. The superiority of LGBM is mainly reflected by the following technologies:

- Gradient-based one-side sampling (GOSS). Data with large gradients contribute more to the model and require more attention during training, while those with small gradients are already sufficiently learned by the model. GOSS is applied to make the model focus more on data with large gradients, while avoiding large variations in the distribution of training data. The process of GOSS is described as follows. Firstly, data are sorted according to the gradient by decreasing order. Afterwards, the top $a\%$ of data are retained, and $b\%$ of the remaining data is randomly selected. Finally, the information gain is calculated and the gradient of the selected $b\%$ data is multiplied by $\frac{1-a}{b}$.
- Exclusive feature bundling (EFB). EFB is effective when data are high-dimensional. It bundles mutually exclusive features to reduce the number of features, thus increasing the training speed without reducing the training accuracy.
- Histogram-based algorithm. The continuous features are discretized into K bins, which are utilized to generate the histogram during training, and the optimal segmentation point is found by traversing the discrete value in the histogram, as shown in Figure 1. This approach reduces the memory consumption.
- Leaf-wise growth strategy. The level-wise growth strategy splits plenty of redundant leaves with low gain, which excessively consumes computational resources [45], while the leaf-wise growth strategy achieves higher precision by splitting the leaf with the greatest gain. The comparison of the above growth strategies is shown in Figure 2.

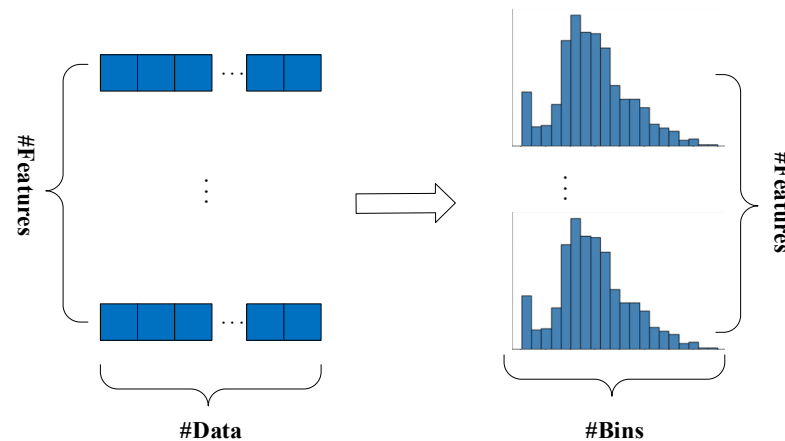


Figure 1. Histogram-based algorithm.

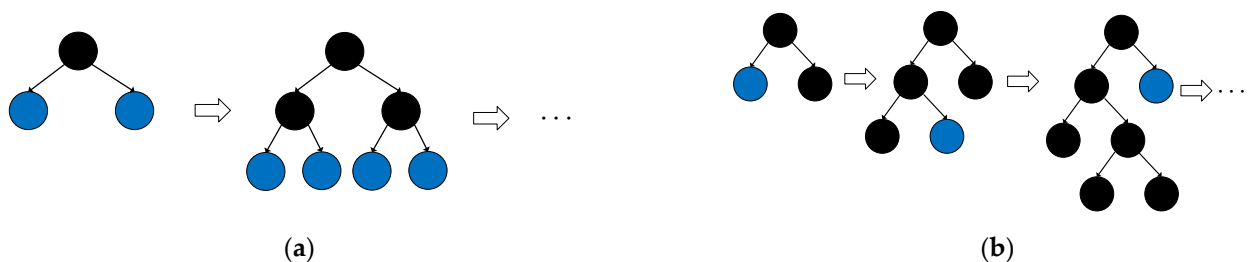


Figure 2. Comparison of (a) level-wise growth strategy and (b) leaf-wise growth strategy.

2.3. Variational Mode Decomposition

VMD is a non-recursive and adaptive signal decomposition method [37]. It obtains a series of approximately orthogonal modes by solving a variational optimization problem. The variational optimization problem is described as follows:

$$\min_{\{u_k\}, \{w_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \quad (4)$$

$$\text{s.t. } \sum_{k=1}^K u_k(t) = x(t) \quad (5)$$

where K denotes the number of modes, $x(t)$ is the original sequence, $\{u_k\} = \{u_1, \dots, u_K\}$ and $\{w_k\} = \{w_1, \dots, w_K\}$ are modes and center frequencies of modes, respectively, $\delta(t)$ is the Dirac distribution, and $*$ denotes convolution operation.

The Lagrange multiplier $\lambda(t)$ and quadratic penalty term are introduced to make the problem unconstrained; the augmented Lagrangian $\mathcal{L}$ is listed as follows:

$$\mathcal{L}(\{u_k\}, \{w_k\}, \lambda) = \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \|x(t) - \sum_{k=1}^K u_k(t)\|_2^2 + \lambda(t), x(t) - \sum_{k=1}^K u_k(t) \quad (6)$$

where α is the balance parameter.

The alternate direction method of multipliers (ADMM) [46] is adopted to solve the augmented Lagrangian $\mathcal{L}$. u_k, w_k are updated as follows:

$$\hat{u}_k^{n+1}(w) = \frac{\hat{x}(w) - \sum_{i < k} \hat{u}_i^{n+1}(w) - \sum_{i > k} \hat{u}_i^n(w) + \frac{\hat{\lambda}^n(w)}{2}}{1 + 2\alpha(\omega - \omega_k^n)^2} \quad (7)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty w |\hat{u}_k^{n+1}(w)|^2 dw}{\int_0^\infty |\hat{u}_k^{n+1}(w)|^2 dw} \quad (8)$$

$$\hat{\lambda}^{n+1}(w) = \hat{\lambda}^n(w) + \tau \left(\hat{x}(w) - \sum_{k=1}^K \hat{u}_k^{n+1}(w) \right) \quad (9)$$

where τ is the iteration factor, and n denotes the number of iterations. $\hat{u}(w)$, $\hat{x}(w)$, and $\hat{\lambda}(w)$ are the Fourier transforms of $u(t)$, $x(t)$, and $\lambda(t)$, respectively.

The iteration is stopped when the following convergence condition is met:

$$\sum_{k=1}^K \frac{\|\hat{u}_k^{n+1}(w) - \hat{u}_k^n(w)\|_2^2}{\|\hat{u}_k^n(w)\|_2^2} < \varepsilon \quad (10)$$

where ε is the convergence threshold.

The number of modes K is key in VMD. On one hand, the modes are still complex if K is too small, which is inconducive to learning degradation trends; thus, the performance of the prediction model cannot be improved effectively. On the other hand, if K is too large, a great amount of computational resources is consumed, even the prediction accuracy is reduced due to the accumulation of prediction errors in sub-models. To determine a suitable K , the ratio of residual energy to the original signal energy R_{res} [47] is used as the criterion of decomposition. The R_{res} is defined as follows:

$$R_{res} = \frac{1}{T} \sum_{t=1}^T \left| \frac{x(t) - \sum_{k=1}^K u_k(t)}{x(t)} \right| \times 100\% \quad (11)$$

where T denotes the length of $x(t)$. The minimum K that satisfies $R_{res} < 1\%$ is the optimal number of modes.

2.4. Gated Recurrent Unit

A traditional recurrent neural network cannot effectively learn long-term dependence [48]. LSTM solves this shortcoming through the gating mechanism, while this structure increases the parameters of the network. GRU [49] is a simplified version of LSTM that only uses two gates, while the prediction accuracy is not reduced. The structure of a GRU cell is shown in Figure 3. Given the current input X_t and the previous hidden state H_{t-1} , the current hidden state H_t is calculated as follows.

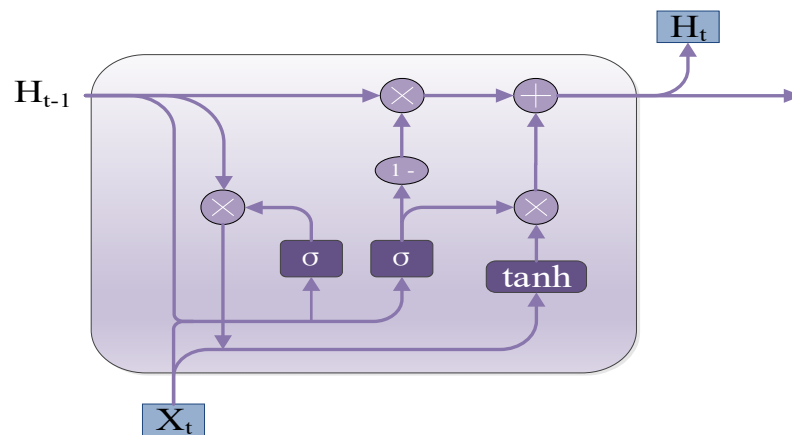


Figure 3. Structure of GRU cell.

Firstly, H_{t-1} and X_t are put into the reset gate R_t and generate the candidate hidden state $\bar{H}_t$:

$$R_t = \text{sigmoid}(W_{rx}X_t + U_{rh}H_{t-1}) \quad (12)$$

$$\bar{H}_t = \tanh(W_{hx}X_t + U_{hh}(R_t \odot H_{t-1})) \quad (13)$$

where $\odot$ denotes the element-wise multiplication. W_{rx} , U_{rh} , W_{hx} , and U_{hh} are the weight matrices.

Afterwards, the update gate Z_t controls how much information in $\bar{H}_t$ is utilized to generate H_t :

$$Z_t = \text{sigmoid}(W_{zx}X_t + U_{zh}H_{t-1}) \quad (14)$$

$$H_t = (1 - Z_t) \odot H_{t-1} + Z_t \odot \bar{H}_t \quad (15)$$

where W_{zx} and U_{zh} are weight matrices.

3. The DTP Model Based on MIC-LGBM and VMD-GRU

To predict the degradation trend of a PSU precisely and provide support for condition-based maintenance, a MIC-LGBM and VMD-GRU-based combined model is proposed. On one hand, it generates a reliable PDI in a short time, only using few computational resources; on the other, its predicted degradation trend is accurate and has a strong correlation with the actual degradation trend. The overall flowchart of the proposed model is shown in Figure 4. Firstly, working parameters are screened by MIC, and the interference information is removed. Secondly, the data in the benchmark state, which represents the good running conditions of PSU, are utilized to generate the LGBM healthy model, and then the PDI sequence is obtained. Lastly, the VMD-GRU prediction model is constructed to predict the degradation trend of PSU.

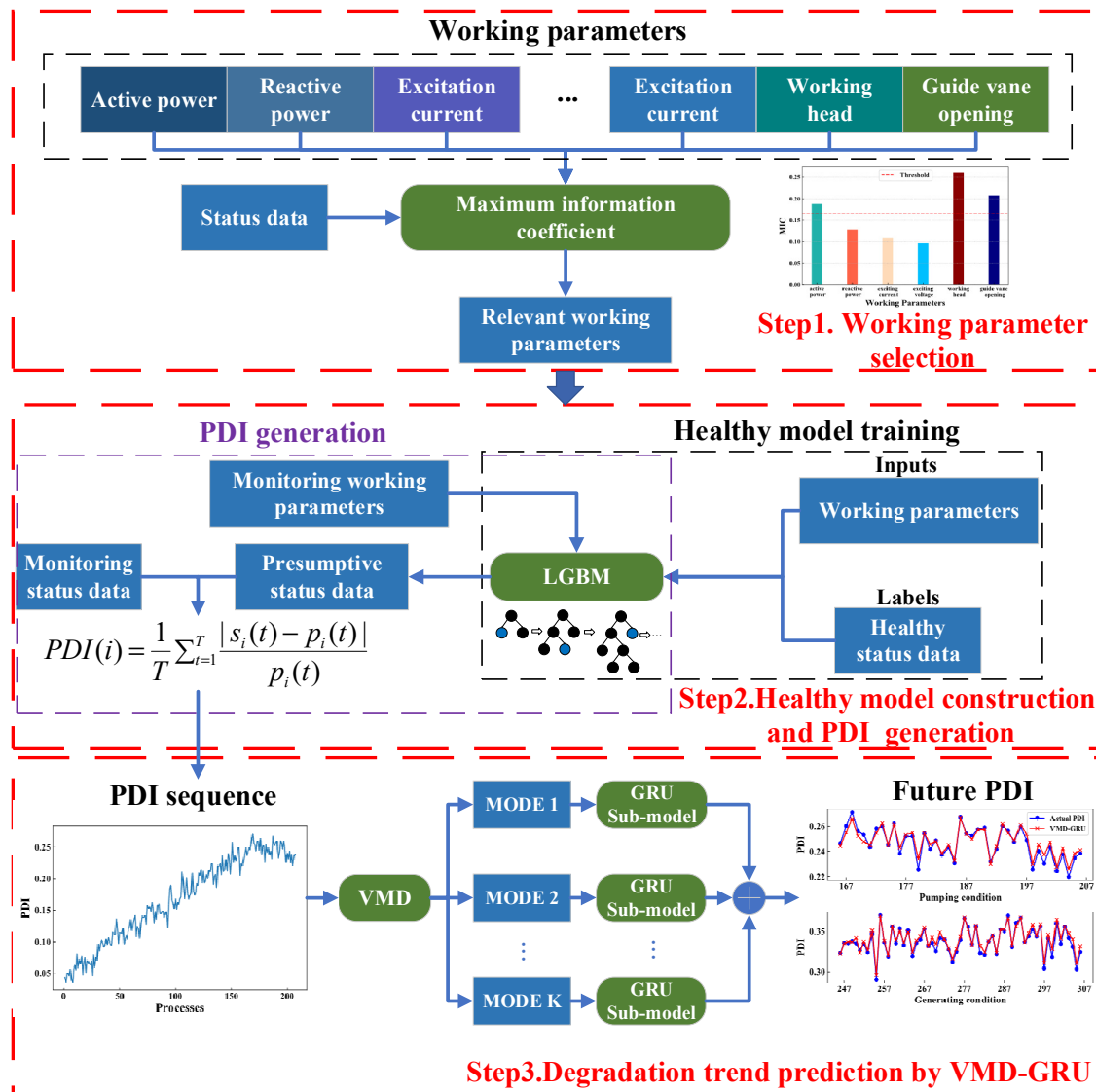


Figure 4. Proposed DTP model.

3.1. Working Parameters Selection by MIC

The status data, such as vibration, swing, etc., reflect the operating status of the PSU directly. Besides, the PSU behaves differently under different working parameters. Working parameters that are poorly correlated with the status data bring interference to the judgment of the PSU's status. Therefore, invalid operating parameters are excluded by MIC to ensure that the inputs to the healthy model are critical to determine the operating status of PSU. Given l working parameters $w^1(t), w^2(t), \dots, w^l(t)$ and the status data $s(t)$ of PSU, the selection of working parameters is carried out as follows:

- (1) Calculate the correlation $c(i), i = 1, \dots, l$ between $w^i(t), i = 1, \dots, l$ and $s(t)$ by the MIC in Section 2.1.
- (2) Obtain the selection threshold δ as follows:

$$\delta = \frac{1}{l} \sum_{i=1}^l c(i) \quad (16)$$

- (3) The working parameter $w^i(t)$ is selected as input of the healthy model if $c(i) \geq \delta$.

3.2. Healthy Model Construction and PDI Generation

3.2.1. Part A: Healthy Model Construction

After selecting the input, the status data are used as the output of the healthy model. Thus, the LGBM healthy model is built as follows:

- (1) The period when the PSU is running well is selected as the benchmark state. The selected working parameters under benchmark state $w^1(t), w^2(t), \dots, w^m(t)$ are used as the input of the LGBM healthy model, and the corresponding healthy status data $h(t)$ are adopted as the output of the LGBM healthy model. Thus, the nonlinear mapping relationship is established as follows:

$$h(t) = f(w^1(t), w^2(t), \dots, w^m(t)) \quad (17)$$

- (2) The trial-and-error method is utilized to determine the optimal parameters of LGBM.

The absolute average ($|AVG|$) and standard deviation (STD) of the fitting errors $E = \{e(i), i = 1, \dots, N\}$ on the test set are introduced to evaluate the effectiveness of the LGBM healthy model. The definition of $|AVG|$ and STD are presented as follows:

$$|AVG| = \left| \frac{1}{N} \sum_{i=1}^N e(i) \right| \quad (18)$$

$$STD = \sum_{i=1}^N \frac{(e(i) - \bar{e})^2}{N - 1} \quad (19)$$

where N is the number of fitting errors E and $\bar{e}$ denotes the average of E . The smaller the $|AVG|$, the smaller fitting error. The smaller the STD , the more stable the healthy model performance.

Moreover, the training time of the healthy model $TIME$ is recorded to illustrate the computational resource consumption.

$$TIME = t_{end} - t_{start} \quad (20)$$

where t_{end} and t_{start} represent the start time and end time of model training, respectively. The smaller the $TIME$, the smaller the cost of computational resources.

3.2.2. Part B: PDI Generation

The performance degradation of the PSU mainly occurs in the pumping condition and generation condition. Therefore, a single pumping or generation process is used as the basis unit of PDI generation in this paper. The PDI $PDI(i)$ of the i th process is calculated as follows:

$$p_i(t) = f(w_i^1(t), w_i^2(t), \dots, w_i^m(t)) \quad (21)$$

$$PDI(i) = \frac{1}{T} \sum_{t=1}^T \frac{|s_i(t) - p_i(t)|}{p_i(t)} \quad (22)$$

where $s_i(t)$ is the monitoring status data, $w_i^1(t), w_i^2(t), \dots, w_i^m(t)$ are the monitoring working parameters selected by the MIC, f denotes the mapping relationship learned by the LGBM healthy model, $p_i(t)$ implies the presumptive status data under corresponding working parameters when the PSU is running well, and T denotes the number of points in the i th process.

3.3. Degradation Trend Prediction with VMD-GRU

The VMD-GRU prediction model is constructed after obtaining PDI. As shown in the bottom of Figure 4, the PDI sequence is decomposed into a series of modes at first. Then, the GRU sub-models are built for each mode separately. Finally, the predicted values of all modes are added to obtain the future PDI. The structure of the GRU sub-model is

shown in Figure 5. The long-term dependence of the PDI sequence is extracted by the GRU layer; then, the output of the final GRU cell is sent to the full connected layers to obtain the predicted value of mode.

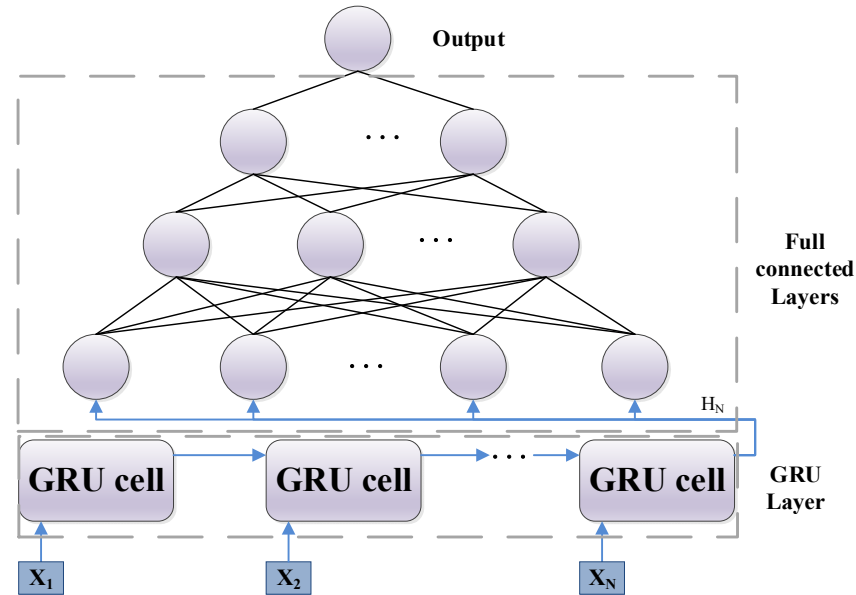


Figure 5. Structure of the GRU sub-model.

RMSE, MAE, and R^2 are selected as metrics for evaluating the performance of the prediction model.

(a) RMSE:

$$\text{RMSE} = \sqrt{\sum_{i=1}^N \frac{(y_i - \tilde{y}_i)^2}{N}} \quad (23)$$

(b) MAE:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \tilde{y}_i| \quad (24)$$

(c) R^2 :

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \tilde{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (25)$$

where y_i and $\tilde{y}_i$ denote the actual PDI and predicted PDI, respectively. N is the length of the actual PDI, and $\bar{y}$ is the average of y_i .

4. Case Study

The proposed DTP model was verified on a PSU located in China, and comparison experiments were conducted to illustrate the superiority of components of the proposed model. All experiments were carried out in the Python 3.6.4 environment running on a computer with R7 5800h CPU, GTX3060ti GPU.

4.1. Data Source

The structure of the PSU is shown in Figure 6. It has a single-stage mixed-flow pump-turbine unit with a capacity of 375 MW. The single-shaft vertical-stage pump-turbine is concatenated with the power generating motor, which has a rated speed of 375 r/min, through the main shaft. The monitoring system has worked since 19 May 2017. According to the operation reports, no abnormalities or accidents occurred from 15 January 2018 to 15 February 2018, which means the PSU ran well. This period was chosen as the benchmark state. The PSU had poor performance between 1 March 2019 to 1 October 2019, and this

period was used for generating the PDI. Besides, the PSU has different characteristics under different operating conditions, and DTP was carried out on the pumping condition and generating conditions, respectively. There were 207 pumping processes and 307 generating processes during PDI generation.

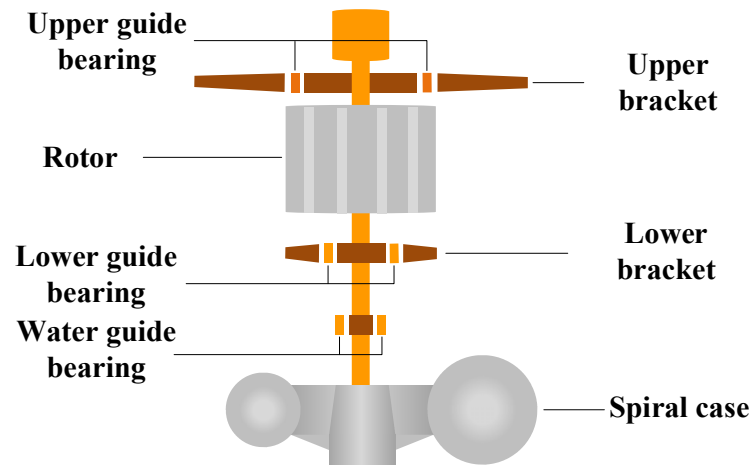


Figure 6. The structure of the PSU.

4.2. Working Parameter Selection Based on MIC

Among the status data, swing and vibration reflect the operating status of the PDU significantly [9]. The swing of the upper guide bearing was chosen to reflect the status of the PSU in this paper. Therefore, the output of the healthy model was determined as swing. The working parameters, such as active power, reactive power, excitation recurrent, excitation voltage, working head, guide vane opening, etc., determine the operation mode of the PSU. The working parameters have different effects on the swing; however, they also have different correlations with the swing. The relationships between the swing and the working parameters under the benchmark state are shown in Figure 7.

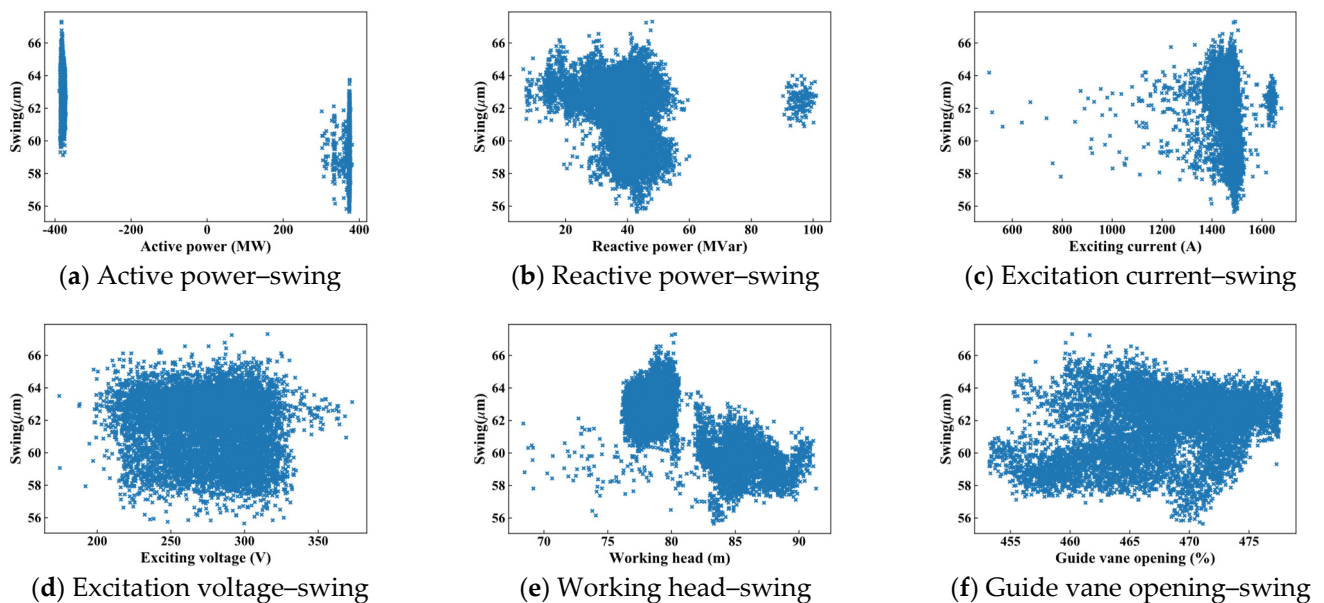


Figure 7. Relationships between working parameters and swing.

From Figure 7a, it can be seen that swing was distributed from 59.12–67.31 μm and 55.64–63.75 μm under the pumping condition and generating condition, respectively. This

indicates that PSU has different vibration characteristics under these operating conditions, and it is necessary to generate a PDI for all operating conditions separately to ensure consistency. Figure 7b shows that the reactive power of PSU is mostly distributed between 6.33 MVar and 60 Mvar, while a small amount is distributed from 91.2 Mvar to 102.3 MVar. Moreover, the relationships between the swing and the above working parameters are not linear and are difficult to discern directly; the same conclusions can be drawn in Figure 7c–f. Thus, MIC was utilized to extract the complex relationships between the swing and the working parameters; then, the selection threshold δ was calculated by Equation (16) to exclude the working parameters that have weak correlations with the swing. The MIC and δ are shown in Figure 8, where $\delta = 0.553$. The active power, working head, and guide vane opening were selected as the inputs of the healthy model since their MICs are greater than δ .

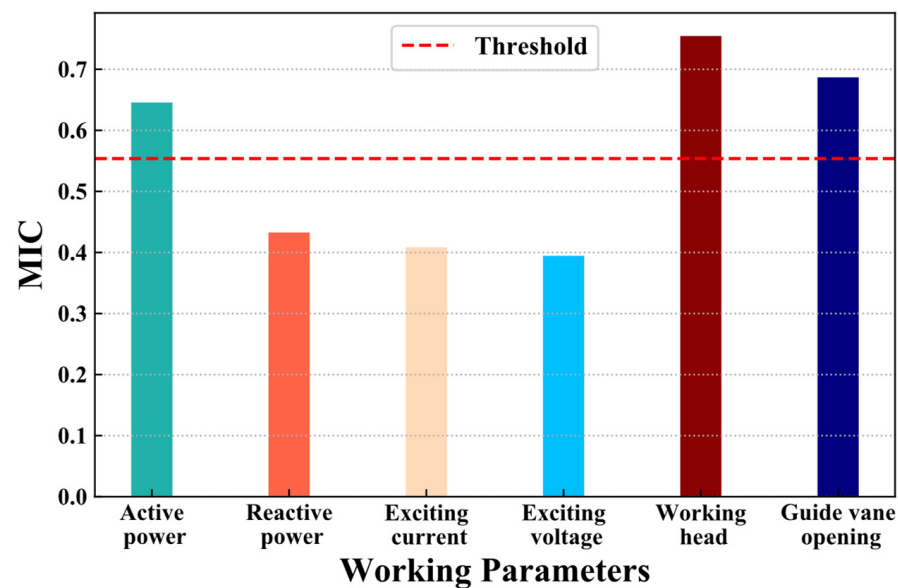


Figure 8. MIC between swing and working parameters.

4.3. Healthy Model Establishment and PDI Construction

4.3.1. Comparative Healthy Models and Parameter Settings

After determining the input and output of the healthy model, the LGBM healthy models were built under two operating conditions, respectively. In the benchmark state, there were 5009 samples under the pumping condition and 3175 samples under the generating condition. In total, 90% of the benchmark state was used for training, and the remaining 10% was applied for testing. The GPR, Classification and Regression Tree (CART) [50], and XGBoost are small and effective; thus, they were adopted for the comparison with the LGBM. The optimal parameters of healthy models were obtained through the trial-and-error method, as listed in Table 1. Moreover, ablation experiments were conducted to illustrate the importance of working parameter selection; i.e., all working parameters were taken as the input of the healthy model.

Table 1. Parameter setting of healthy models.

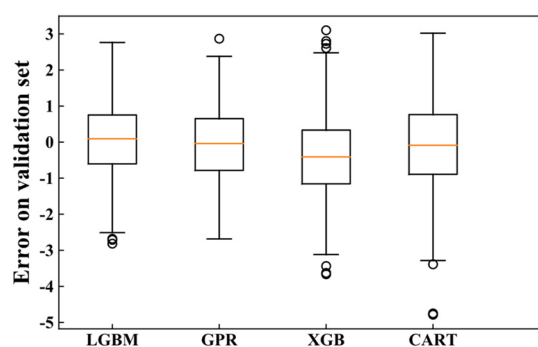
Model		Parameter Settings
GPR		Kernel = 'RBF', alpha = '1e-9'.
CART		Criterion = 'MSE', Min_samples_split = 2, Min_samples_leaf = 1.
XGBoost		Booster = 'gbtree', eta = 0.1, Max_depth = 7, Min_child_weight = 1, Sub_sample = 0.82.
LGBM		Max_depth = 8, Num_leaves = 19, Min_child_samples = 30, Sub_sample = 0.85.

4.3.2. Performance Analysis and Discussion of Healthy Models

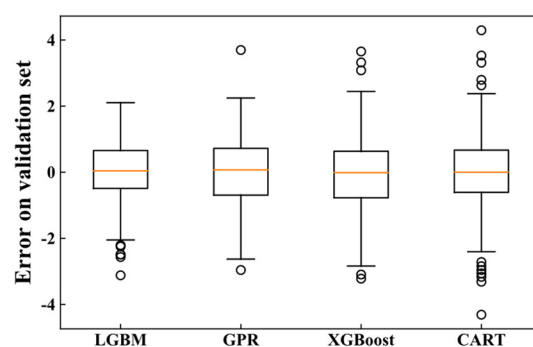
The evaluation metrics of different healthy models on the testing set are listed in Table 2. The bolded values represent the best metrics. The box plots of fitting errors on the testing set are shown in Figure 9, and the mean values of evaluation metrics of two operating conditions are presented in Figure 10.

Table 2. Fitting errors of healthy models.

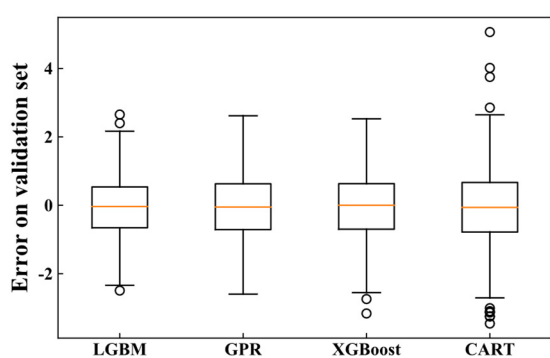
Parameter Selection	Healthy Model	Pumping Condition			Generating Condition		
		AVG	STD	TIME	AVG	STD	TIME
NO MIC	GPR	0.065	1.033	2.317	0.014	1.037	1.700
	CART	0.093	1.246	0.182	0.036	1.127	0.096
	XGBoost	0.384	1.174	1.539	0.030	1.128	0.937
	LGBM	0.082	1.055	0.755	0.028	0.994	0.652
MIC	GPR	0.049	0.985	2.261	0.029	1.071	1.553
	CART	0.056	1.169	0.023	0.023	1.063	0.036
	XGBoost	0.040	0.994	0.314	0.023	0.970	0.612
	LGBM	0.036	0.911	0.302	0.019	0.938	0.563



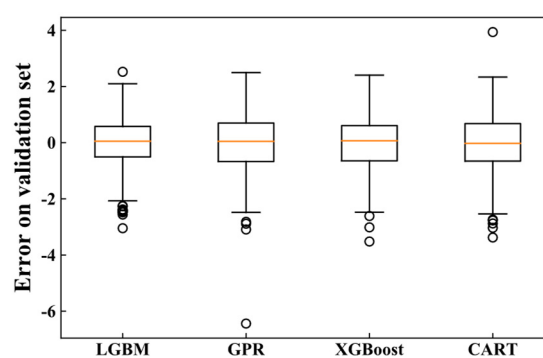
(a) Fitting errors under pumping condition (NO MIC)



(b) Fitting errors under generating condition (NO MIC)



(c) Fitting errors under pumping condition (MIC)



(d) Fitting errors under generating condition (MIC)

Figure 9. The box plots of fitting errors on the testing set.

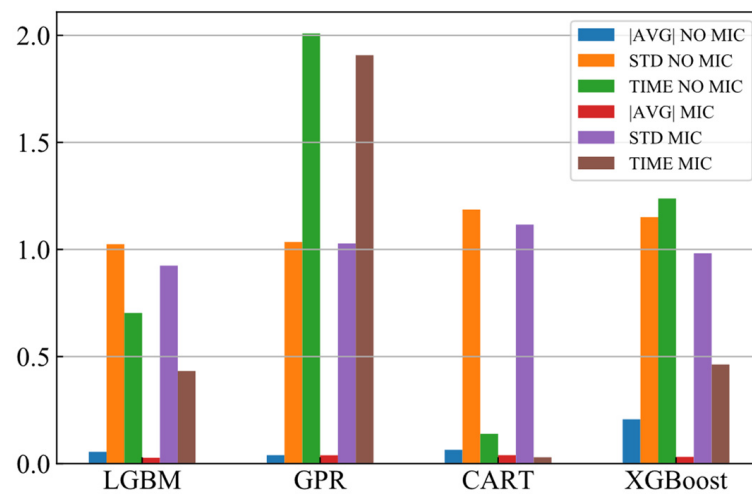


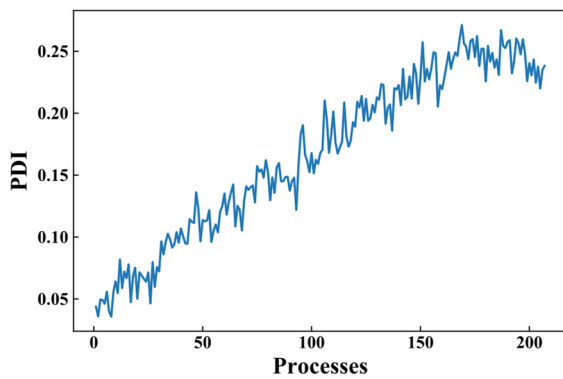
Figure 10. Average evaluation metrics under two conditions.

When the working parameters were selected by MIC, LGBM achieved the minimum $|AVG|$ and STD among all healthy models under both the pumping condition and generating condition. This reveals that LGBM has the smallest fitting error and the most stable performance. The $|AVG|$ and $TIME$ of XGBoost are close to those of LGBM, while STD is 6.2% higher than LGBM on average for two operating conditions. This indicates that XGBoost is more likely to produce outliers. CART takes the shortest training time among healthy models, while $|AVG|$ and STD are larger than XGBoost and LGBM. Although it trains quickly, its fitting error is not satisfactory. GPR took the longest time for training and consumed the most computational resources but performed the worst under the generating condition. The above conclusions are also verified in Figure 9c,d.

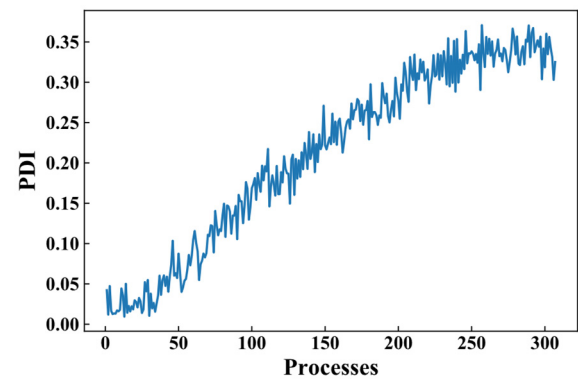
As shown in Table 2, when working parameters were not selected by MIC (NO MIC), most of healthy models had a larger $|AVG|$ and STD and required a longer training time. This shows that the redundant components in the working parameters not only decrease the accuracy and stability of healthy models but also cost more in terms of computational resources. Interestingly, the working parameter selection improves the performance of GPR under the pumping condition while its capability reduces under the generating condition. The working parameter selection makes the distributions of fitting errors on the testing set more concentrated, as shown in Figure 9. From Figure 10, it can be seen that parameter selection improved the average performance of GPR, CART, XGBoost, and LGBM under two operating conditions, where $|AVG|$ improved by 1.2%, 38.9%, 84.7%, and 50%, STD improved by 0.6%, 5.9%, 14.6%, and 9.7%, and $TIME$ improved by 5.1%, 78.8%, 62.6%, and 38.5%. MIC greatly improves the capability of XGBoost and LGBM.

4.3.3. PDI Construction with LGBM Healthy Model

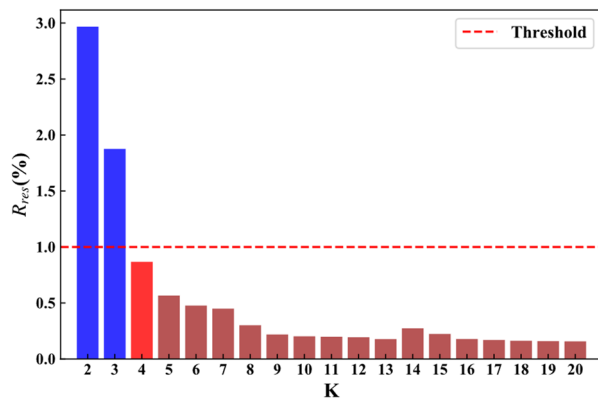
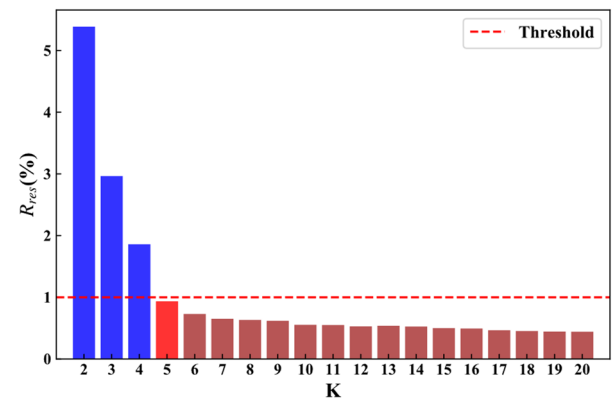
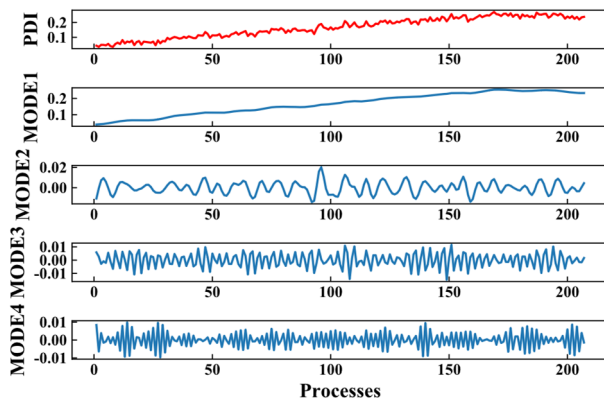
Based on the reliable LGBM healthy model, the effective PDI sequences are generated. For the i th process, the presumptive status data $p_i(t)$ are obtained by Equation (21); then, the $PDI(i)$ is calculated by Equation (22). The PDI sequences of two operating conditions are shown in Figure 11. The PDI sequences of two operating conditions have similar overall increasing trends. This indicates that the degradation of PSU gradually increases with operation time, which is consistent with the records in the operating reports. In addition, the PDI sequences are so complex that there are plenty of recursive components and nonlinear components. These components are clearly demonstrated in the PDI sequence under the generating condition, as shown in Figure 12b. They seriously affect the performance of the prediction model.



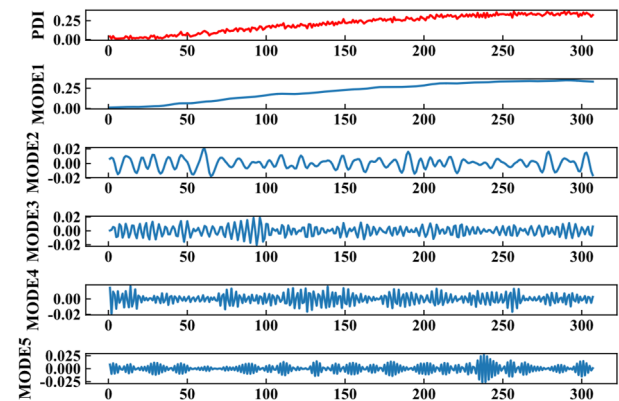
(a) PDI sequence under pumping condition



(b) PDI sequence under generating condition

Figure 11. PDI sequences under (a) pumping and (b) generating conditions.(a) R_{res} of different K under pumping condition(b) R_{res} of different K under generating condition

(c) Decomposition result under pumping condition



(d) Decomposition result under generating condition

Figure 12. Decomposition results of PDI sequences by VMD.

4.4. Degradation Trend Prediction of PSU

4.4.1. Comparative Prediction Models and Parameter Settings

The complex PDI sequence brings challenges for predicting degradation trends accurately. To solve this problem, the VMD-GRU prediction model was constructed in this paper. VMD was utilized to decompose the PDI sequence into a series of simple modes; then, GRU sub-models were built for each mode separately, and the predicted values of the sub-models were summed to obtain the future PDI at last. The following comparative experiments were conducted to confirm the superiority of VMD-GRU. Firstly, the popular prediction models were used to illustrate the challenge brought by the complexity of

the PDI sequence, including ANN [27], SVR [26], LSTM [29], and GRU [33]. Afterwards, VMD-ANN, VMD-LSTM, and VMD-GRU were compared with ANN, LSTM, and GRU, respectively, to demonstrate the performance improvement resulting from decomposition; then, the EMD-ANN, EMD-LSTM, and EMD-GRU prediction models were set up to demonstrate the effectiveness of VMD. Lastly, the validity of GRU was proved by comparing GRU, EMD-GRU, and VMD-GRU with corresponding models. The optimal structures and parameter settings of prediction models were determined by the trial-and-error method, as shown in Table 3.

Table 3. Parameter setting of prediction models.

Model	Parameter Settings
ANN	Four full connected layers, number of neurons: 256, 64, 4, 1.
SVR	C = 0.8, Kernel = 'RBF', Epsilon = 0.001, Tol = 0.001.
LSTM	(a) LSTM Layer, 128 units. (b) Three full connected layers, number of neurons: 256, 64, 1.
GRU	(a) GRU Layer, 128 units. (b) Three full connected layers, number of neurons: 256, 32, 1.
EMD-ANN	(a) EMD: Decompose until meeting the stopping condition in [51]. (b) ANN sub-model: it has same structure as ANN prediction model.
EMD-LSTM	(a) EMD: Decompose until meeting the stopping condition in [51]. (b) LSTM sub-model: The structure is same as LSTM prediction model.
EMD-GRU	(a) EMD: Decompose until meeting the stopping condition in [51]. (b) GRU sub-model: The structure is same as GRU prediction model.
VMD-ANN	(a) VMD : K is set to 4 under pumping condition, while K = 5 under generating condition. (b) ANN sub-model: it has same structure as ANN prediction model.
VMD-LSTM	(a) VMD : K is set to 4 under pumping condition, while K = 5 under generating condition. (b) LSTM sub-model: The structure is same as LSTM prediction model.
VMD-GRU	(a) VMD : K is set to 4 underpumping condition, while K = 5 under generating condition. (b) GRU sub-model: The structure is same as GRU prediction model.

The time step was set to 5 in all prediction models; i.e., $PDI(i-5), PDI(i-4), \dots, PDI(i-1)$ were used to predict $PDI(i)$. The first 80% of the PDI sequence was employed for training and the remaining 20% was utilized for testing. To eliminate randomness, all prediction results were the averages of 10 repeated experiments.

To determine the optimal number of modes for VMD, the R_{res} of different numbers of modes K was calculated under two conditions, respectively. The R_{res} is shown in Figure 12a,b. It can be seen that $R_{res} < 1\%$ when $K \geq 4$ under the pumping condition and $K \geq 5$ under the generating condition. Therefore, K was set to 4 and 5 under the pumping condition and generating condition, respectively. The decomposition results under an optimal K are presented in Figure 12c,d. These demonstrate that the smooth and simple modes are obtained by VMD, and they store different information of the PDI sequence. Additionally, the decomposition results of EMD are illustrated in Figure 13. EMD suffers from the mode mixing severely that multiple frequency components appear in the same mode. This reveals that the modes generated by EMD are more complex than those of VMD. Thus, we can conclude that VMD makes PDI sequences simpler, and it provides a better decomposition compared with EMD.

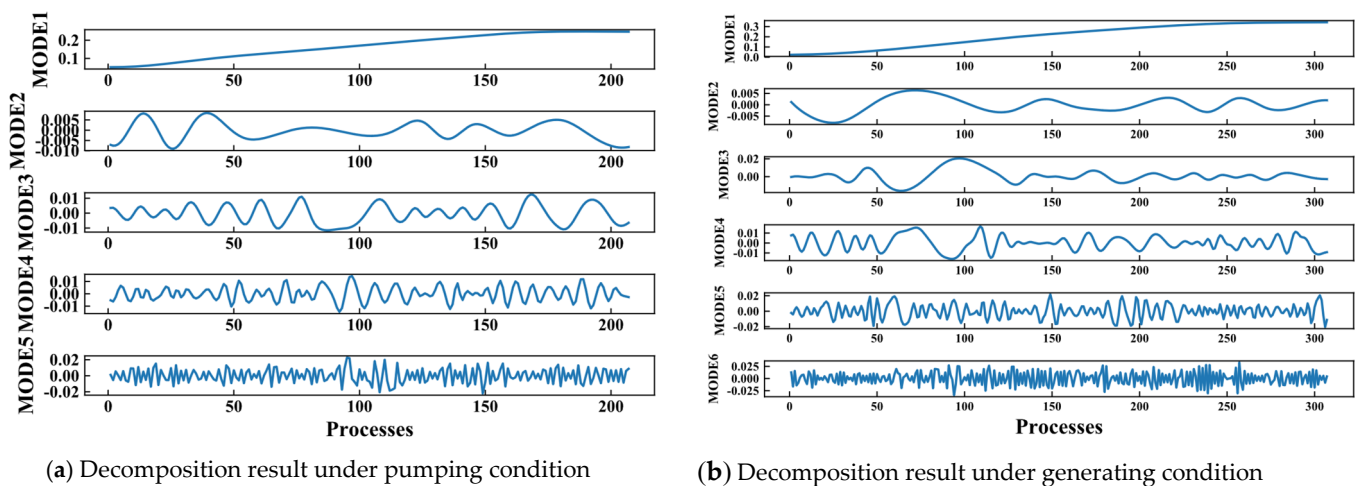


Figure 13. Decomposition result of PDI sequences by EMD.

4.4.2. Performance Analysis and Discussion of Prediction Models

The prediction results of VMD-GRU on the testing sets are shown in Figure 14, and those of comparative experiments are shown in Figure 15. The RMSE, MAE, and R^2 are listed in Table 4. The bolded values indicate the optimal metrics. The analysis of the results and discussion are presented below.

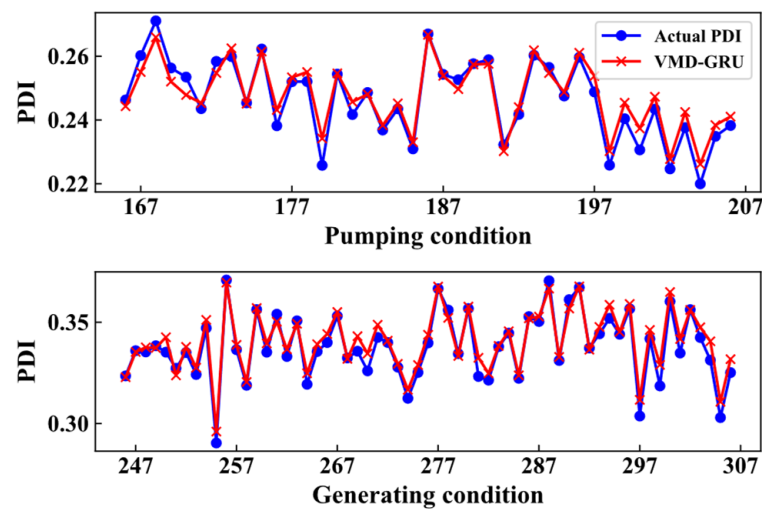


Figure 14. Performance of VMD-GRU.

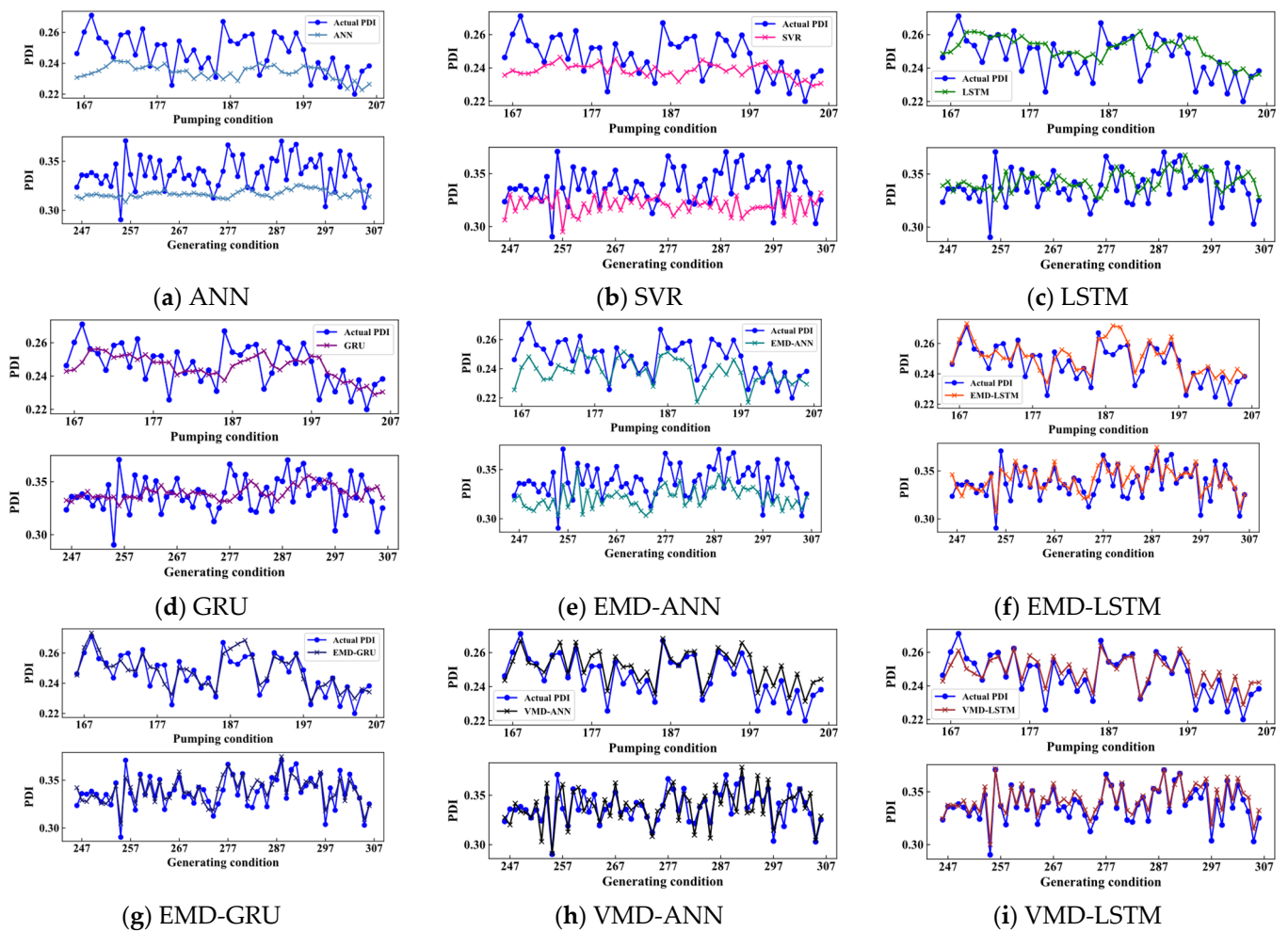


Figure 15. Performance of the comparative prediction models.

Table 4. Performance of prediction models.

Prediction Models	Pumping Condition			Generating Condition		
	RMSE	MAE	R^2	RMSE	MAE	R^2
ANN	0.0173	0.0149	−0.9761	0.0271	0.0232	−1.7304
SVR	0.0147	0.0125	−0.4360	0.0249	0.0208	−1.3031
LSTM	0.0129	0.0095	−0.1060	0.0190	0.0146	−0.3337
GRU	0.0118	0.0096	0.0677	0.0178	0.0136	−0.1780
EMD-ANN	0.0124	0.0106	−0.0187	0.0206	0.0177	−0.5713
EMD-LSTM	0.0076	0.0061	0.6160	0.0099	0.0078	0.6305
EMD-GRU	0.0062	0.0050	0.7457	0.0085	0.0068	0.7328
VMD-ANN	0.0065	0.0056	0.7234	0.0123	0.0100	0.4435
VMD-LSTM	0.0053	0.0045	0.8131	0.0066	0.0053	0.8405
VMD-GRU	0.0035	0.0029	0.9171	0.0044	0.0035	0.9281

(1) Challenges brought by the complex PDI sequence

From Table 4, the R^2 values of ANN, SVR, LSTM, and GRU are less than 0.07 under two operating conditions. This shows that the predicted PDI is poorly correlated with the actual PDI; thus, the developing trend of PDI is not effectively learned by these prediction models. The RMSE and MAE of four models are large; additionally, their prediction results can only fall roughly between the upper and lower envelopes of the actual PDI, as shown in Figure 15a–d. These show that the prediction models have large deviations. As shown in Figure 11 and Table 4, the PDI sequence is more complex under generating condition, and the prediction models perform worse under the generating condition than pumping

condition. These factors reveal that the performance of prediction models is inversely proportional to the complexity of the PDI sequence.

Thus, we can conclude that the complexity of the PDI sequence increases the difficulty of accurate prediction. In this case, even popular prediction models are not effective. The more complex the PDI sequence, the worse the performance.

(2) Comparison of VMD-based models with other models

VMD-ANN, VMD-LSTM, and VMD-GRU are compared with ANN, LSTM, and GRU, respectively, to verify the prediction performance improvement due to PDI sequence simplification by VMD. As listed in Table 4, the VMD-based models show a significant performance improvement. Compared with GRU, the RMSE and MAE of VMD-GRU are improved by 237.1% and 231% under the pumping condition and 304.5% and 288.5% under the generating condition. VMD-LSTM improves RMSE by 143.3% and 187.8% and MAE by 111.1% and 175.4% under the two conditions compared with LSTM. Similar results can be obtained by comparing VMD-ANN with ANN. These indicate that simple modes are more conducive to learning PDI sequences and reducing the prediction bias. In addition, the R^2 values of VMD-based models are greatly increased compared to the corresponding models. VMD-GRU reaches the optimal R^2 , reaching 0.917 under the pumping condition and 0.928 under the generating condition, while the R^2 values of most benchmark prediction models are less than 0. Similar results can be acquired by comparing EMD-based models with ANN, LSTM, and GRU. These results show that decomposition is helpful to learning the long-term dependence in a PDI sequence.

As listed in Table 4, the VMD-based models have smaller errors and learn more long-term dependence in the PDI sequence compared with the EMD-based models. For example, the RMSE, MAE, and R^2 of the VMD-GRU are improved by 85.2%, 83.3%, and 19.9% compared with those of the EMD-GRU on average for the two operating conditions. This indicates that the VMD-based prediction models perform better, so the modes obtained by VMD are more effective.

Therefore, it can be concluded that the simple modes obtained by decomposition are helpful to learning the trend of PDI sequences and reducing the prediction error. Besides, the modes generated by VMD are more efficient than those of EMD.

(3) Comparison of GRU-based models with other models

Compared with ANN, SVR, and LSTM, GRU achieves the smallest RMSE and MAE and the largest R^2 . Under the two operating conditions, the RMSE of GRU improved by 49.4% and 62.9% on average and MAE improved by 32.2% and 41.6% on average compared to ANN and SVR. This indicates that GRU has a smaller prediction bias and better learning ability compared with these models. The performance of LSTM is much better than ANN and SVR, yet the RMSE and MAE of LSTM are still larger than GRU by 7.42% and 2.89%. Besides, LSTM has more parameters and requires more time and computational resources. Therefore, GRU has the best prediction performance when the PDI sequence is not decomposed.

Under the two operating conditions, the EMD-GRU achieves the optimal evaluation metrics and performs best among EMD-based models. Compared with EMD-ANN and EMD-LSTM, the RMSE is improved by 121.1% and 19.5% on average, and the MAE is improved by 136.1% and 18.4% on average. Moreover, the R^2 of EMD-GRU is also the best among EMD-based models, reaching 0.7457 and 0.7328 under the pumping condition and generating condition, respectively. Comparing the VMD-based models, it can be seen that the RMSE and MAE of VMD-GRU are improved by 137.9% and 143.7% on average over VMD-ANN and 50.6% and 53.1% over VMD-LSTM under the two operating conditions. The R^2 of VMD-GRU is also the best among VMD-based models. These factors show that the GRU-based models learn the degradation trend most fully among the corresponding models when the PDI sequence is decomposed.

VMD-GRU achieves the largest R^2 among all prediction models, which are greater than 0.9 under the two operation conditions. This shows that the predicted values of VMD-

GRU have a strong correlation with the actual PDI. Besides, VMD-GRU has the smallest RMSE and MAE among all prediction models. These factors indicate that VMD-GRU has the smallest prediction error and learns the information in PDI sequence most adequately.

Based on the above analysis, we can conclude that VMD-GRU learns the long-term dependence of PDI sequence best, achieves the highest prediction accuracy, and every component of it is indispensable.

5. Conclusions

To predict the degradation trend of PSU reliably, a novel combined model based on MIC-LGBM and VMD-GRU is proposed in this paper. Firstly, MIC is utilized to eliminate the working parameters that have weak correlations with the status of PSU. Secondly, the LGBM healthy model is built to establish the mapping relationship between the selected working parameters and status data under good running conditions; then, the PDI is obtained on the basis of the LGBM healthy model and monitoring data. Lastly, the VMD-GRU prediction model is designed to predict the degradation trend reliably. Experimental validation and comparative analysis show that the proposed model requires less computational resources while establishing the most accurate and stable healthy model and predicts the degradation trend most reliably.

However, all hyperparameters are adjusted by the trial-and-error method in this paper, which consumes a great deal of time. The intelligent optimization algorithms perform outstandingly in hyperparameter tuning and will be used in our future work. Besides, the more reliable the degradation trend prediction, the greater the reference for decision-making. Building a more effective prediction model will also be a research focus in the future.

Author Contributions: Conceptualization, methodology, software, writing—original draft preparation, P.C.; validation, P.C., Y.D. and X.Z.; formal analysis, Y.W.; investigation, L.M.; resources, Y.Y.; writing—review and editing and funding acquisition, C.L. 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 (No. 51879111) and the Hubei Provincial Natural Science Foundation of China (No. 2019CFA068).

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Wu, G.; Shao, X.; Jiang, H.; Chen, S.; Zhou, Y.; Xu, H. Control strategy of the pumped storage unit to deal with the fluctuation of wind and photovoltaic power in microgrid. *Energies* **2020**, *13*, 415. [CrossRef]
2. Hydropower Status Report 2021. Available online: <https://www.hydropower.org/status-report> (accessed on 3 January 2022).
3. Lai, X.; Li, C.; Zhou, J.; Zhang, N. Multi-objective optimization of the closure law of guide vanes for pumped storage units. *Renew. Energy* **2019**, *139*, 302–312. [CrossRef]
4. Zhao, Z.; Yang, J.; Yang, W.; Hu, J.; Chen, M. A coordinated optimization framework for flexible operation of pumped storage hydropower system: Nonlinear modeling, strategy optimization and decision making. *Energy Convers. Manag.* **2019**, *194*, 75–93. [CrossRef]
5. Zheng, Y.; Chen, Q.; Yan, D.; Liu, W. A two-stage numerical simulation framework for pumped-storage energy system. *Energy Convers. Manag.* **2020**, *210*, 112676. [CrossRef]
6. Lei, Y.; Li, N.; Guo, L.; Li, N.; Yan, T.; Lin, J. Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mech. Syst. Signal Process.* **2018**, *104*, 799–834. [CrossRef]
7. 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]
8. Hu, X.; Li, C.; Tang, G. In 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), Qingdao, China, 25–27 October 2019; pp. 1–5.

9. Zhou, J.; Shan, Y.; Liu, J.; Xu, Y.; Zheng, Y. Degradation tendency prediction for pumped storage unit based on integrated degradation index construction and hybrid CNN-LSTM model. *Sensors* **2020**, *20*, 4277. [\[CrossRef\]](#)
10. An, X.; Pan, L. Characteristic parameter degradation prediction of hydropower unit based on radial basis function surface and empirical mode decomposition. *J. Vib. Control*. **2015**, *21*, 2200–2211. [\[CrossRef\]](#)
11. An, X.; Yang, L.; Pan, L. Nonlinear prediction of condition parameter degradation trend for hydropower unit based on radial basis function interpolation and wavelet transform. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2015**, *229*, 3515–3525. [\[CrossRef\]](#)
12. 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. [\[CrossRef\]](#)
13. Sagi, O.; Rokach, L. Ensemble learning: A survey. *WIREs Data Mining Knowl Discov.* **2018**, *8*, e1249. [\[CrossRef\]](#)
14. Cai, R.; Xie, S.; Wang, B.; Yang, R.; Xu, D.; He, Y. Wind speed forecasting based on extreme gradient boosting. *IEEE Access* **2020**, *8*, 175063–175069. [\[CrossRef\]](#)
15. Zheng, H.; Wu, Y. A XGBoost model with weather similarity analysis and feature engineering for short-term wind power forecasting. *Appl. Sci.* **2019**, *9*, 3019. [\[CrossRef\]](#)
16. Tao, T.; Liu, Y.; Qiao, Y.; Gao, L.; Lu, J.; Zhang, C.; Wang, Y. Wind turbine blade icing diagnosis using hybrid features and Stacked-XGBoost algorithm. *Renew. Energy* **2021**, *180*, 1004–1013. [\[CrossRef\]](#)
17. Trizoglou, P.; Liu, X.; Lin, Z. Fault detection by an ensemble framework of extreme gradient boosting (XGBoost) in the operation of offshore wind turbines. *Renew. Energy* **2021**, *179*, 945–962. [\[CrossRef\]](#)
18. Alsaleh, A.; Binsaeedan, W. The influence of salp swarm algorithm-based feature selection on network anomaly intrusion detection. *IEEE Access* **2021**, *9*, 112466–112477. [\[CrossRef\]](#)
19. Truong, D.; Tran, D.; Nguyen, L.; Mac, H.; Tran, H.A.; Bui, T. Detecting web attacks using stacked denoising autoencoder and ensemble learning methods. In Proceedings of the Tenth International Symposium on Information and Communication Technology, Association for Computing Machinery, Hanoi, Vietnam, 4–6 December 2019; pp. 267–272.
20. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A highly efficient gradient boosting decision tree. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Curran Associates Inc., Long Beach, CA, USA, 4–6 December 2017; pp. 3149–3157.
21. Cerrada, M.; Sánchez, R.-V.; Li, C.; Pacheco, F.; Cabrera, D.; Valente de Oliveira, J.; Vásquez, R.E. A review on data-driven fault severity assessment in rolling bearings. *Mech. Syst. Signal Process.* **2018**, *99*, 169–196. [\[CrossRef\]](#)
22. Reshef David, N.; Reshef Yakir, A.; Finucane Hilary, K.; Grossman Sharon, R.; McVean, G.; Turnbaugh Peter, J.; Lander Eric, S.; Mitzenmacher, M.; Sabeti Pardis, C. Detecting novel Associations in large data sets. *Science* **2011**, *334*, 1518–1524. [\[CrossRef\]](#)
23. Jiang, Y.; Li, C.; Yang, Z.; Zhao, Y.; Wang, X. Remaining useful life estimation combining two-step maximal information coefficient and temporal convolutional network with attention mechanism. *IEEE Access* **2021**, *9*, 16323–16336. [\[CrossRef\]](#)
24. Ji, H.; Huang, S.; Wu, Y.; Hui, Z.; Lv, X. In A New attribute selection method based on maximal information coefficient and automatic clustering. In Proceedings of the 2017 International Conference on Dependable Systems and Their Applications (DSA), Beijing, China, 31 October–2 November 2017; pp. 22–28.
25. Zhao, L.; Zhang, Y.; Li, J. Research on constructing a degradation index and predicting the remaining useful life for rolling element bearings of complex equipment. *J. Mech. Sci. Technol.* **2021**, *35*, 4313–4327. [\[CrossRef\]](#)
26. Liu, F.; Li, L.; Liu, Y.; Cao, Z.; Yang, H.; Lu, S. HKF-SVR optimized by Krill Herd algorithm for coaxial bearings performance degradation prediction. *Sensors* **2020**, *20*, 660. [\[CrossRef\]](#)
27. Liu, Z.; Guo, Y. In A neural network approach for prediction of bearing performance degradation tendency. In Proceedings of the 2017 9th International Conference on Modelling, Identification and Control (ICMIC), Kunming, China, 10–12 July 2017; pp. 204–208.
28. Bao, H.; Yan, Z.; Ji, S.; Wang, J.; Jia, S.; Zhang, G.; Han, B. An enhanced sparse filtering method for transfer fault diagnosis using maximum classifier discrepancy. *Meas. Sci. Technol.* **2021**, *32*, 085105. [\[CrossRef\]](#)
29. Park, K.; Choi, Y.; Choi, W.J.; Ryu, H.; Kim, H. LSTM-based battery remaining useful life prediction with multi-channel charging profiles. *IEEE Access* **2020**, *8*, 20786–20798. [\[CrossRef\]](#)
30. Xia, J.; Feng, Y.; Lu, C.; Fei, C.; Xue, X. LSTM-based multi-layer self-attention method for remaining useful life estimation of mechanical systems. *Eng. Fail. Anal.* **2021**, *125*, 105385. [\[CrossRef\]](#)
31. Wu, S.; Jiang, Y.; Luo, H.; Yin, S. Remaining useful life prediction for ion etching machine cooling system using deep recurrent neural network-based approaches. *Control Eng. Pract.* **2021**, *109*, 104748. [\[CrossRef\]](#)
32. Cho, K.; Merriënboer, B.V.; Gulcehre, C.; Ba Hdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y.J.C.S. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2019; pp. 1724–1734.
33. Chen, J.; Jing, H.; Chang, Y.; Liu, Q. Gated recurrent unit based recurrent neural network for remaining useful life prediction of nonlinear deterioration process. *Reliab. Eng. Syst. Saf.* **2019**, *185*, 372–382. [\[CrossRef\]](#)
34. Du, B.; He, Y.; An, B.; Zhang, C. Remaining useful performance estimation for complex analog circuit based on maximal information coefficient and bidirectional gate recurrent unit. *IEEE Access* **2020**, *8*, 102449–102466. [\[CrossRef\]](#)
35. Lu, Y.-W.; Hsu, C.-Y.; Huang, K.-C. An autoencoder gated recurrent unit for remaining useful life prediction. *Processes* **2020**, *8*, 1155. [\[CrossRef\]](#)

36. 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.J.P.M.P.; Sciences, E. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. A* **1998**, *454*, 903–995. [\[CrossRef\]](#)
37. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [\[CrossRef\]](#)
38. He, X.; Luo, J.; Zuo, G.; Xie, J. Daily runoff Forecasting using a hybrid model based on variational mode decomposition and deep neural networks. *Water Resour. Manag.* **2019**, *33*, 1571–1590. [\[CrossRef\]](#)
39. Xiao, F.; Yang, D.; Guo, X.; Wang, Y. VMD-based denoising methods for surface electromyography signals. *J. Neural Eng.* **2019**, *16*, 056017. [\[CrossRef\]](#) [\[PubMed\]](#)
40. He, X.; Luo, J.; Li, P.; Zuo, G.; Xie, J. A hybrid model based on variational mode decomposition and gradient boosting regression tree for monthly runoff forecasting. *Water Resour. Manag.* **2020**, *34*, 865–884. [\[CrossRef\]](#)
41. Ali, M.; Khan, A.; Rehman, N.U. Hybrid multiscale wind speed forecasting based on variational mode decomposition. *Int. Trans. Electr. Energy Syst.* **2018**, *28*, e2466. [\[CrossRef\]](#)
42. Liang, T.; Zhang, Q.; Liu, X.; Lou, C.; Liu, X.; Wang, H. Time-frequency maximal information coefficient method and its application to functional corticomuscular coupling. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2020**, *28*, 2515–2524. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Chen, T.; Guestrin, C. In XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference, Association for Computing Machinery, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
44. Wang, Y.; Wang, T. Application of improved lightGBM model in blood glucose prediction. *Appl. Sci.* **2020**, *10*, 3227. [\[CrossRef\]](#)
45. Zhang, Y.; Zhu, C.; Wang, Q. LightGBM-based model for metro passenger volume forecasting. *IET Intell. Transp. Syst.* **2021**, *14*, 1815–1823. [\[CrossRef\]](#)
46. Bertsekas, D.P. *Constrained Optimization and Lagrange Multiplier Methods*. *Constrained Optimization and Lagrange Multiplier Methods*; Academic Press: Cambridge, MA, USA, 1982.
47. Liu, Y.; Yang, C.; Huang, K.; Gui, W. Non-ferrous metals price forecasting based on variational mode decomposition and LSTM network. *Knowl. Based Syst.* **2020**, *188*, 105006. [\[CrossRef\]](#)
48. Wang, Y.; Zou, R.; Liu, F.; Zhang, L.; Liu, Q. A review of wind speed and wind power forecasting with deep neural networks. *Appl. Energy* **2021**, *304*, 117766. [\[CrossRef\]](#)
49. 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* **2020**, *11*, 1370–1380.
50. Yu, J.; Jang, J.; Yoo, J.; Park, J.H.; Kim, S. A fault isolation method via classification and regression tree-based variable ranking for drum-type steam boiler in thermal power plant. *Energies* **2018**, *11*, 1142. [\[CrossRef\]](#)
51. Lei, Y.; Lin, J.; He, Z.; Zuo, M.J. A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mech. Syst. Signal Process.* **2013**, *35*, 108–126. [\[CrossRef\]](#)