Hydroelectric Generating Unit Fault Diagnosis Using 1-D Convolutional Neural Network and Gated Recurrent Unit in Small Hydro

Guo-Ping Liao, Wei Gao, Geng-Jie Yang, and Mou-Fa Guo, *Member, IEEE*

***Abstract*— Machine learning algorithm based on hand-crafted features from the raw vibration signal has shown promising results in the hydroelectric generating unit (HGU) fault diagnosis in recent years. Such methodologies, nevertheless, can lead to important information loss in representing the vibration signal, which intrinsically relies on engineering experience of diagnostic experts and prior knowledge about feature extraction techniques. Therefore, in this paper, an effective and stable HGU fault diagnosis system using one-dimensional convolutional neural network (1-D CNN) and gated recurrent unit (GRU) based on the sequence data structure is proposed. First, the raw vibration data is reconstructed by data segmentation, which can improve training efficiency. Second, the reconstruction data under the influence of different running conditions and various fault factors can be effectively and adaptively learned by 1-D CNN-GRU and then determine information fault categories via network inference. Finally, four machine learning methods are applied to diagnosis the reconstruction data based on the experimental dataset. The performance of the proposed method is verified by comparing with the results of other machine learning techniques. Furthermore, the fault diagnostic model, which is trained by the practical vibration signal, has successfully applied in engineering practice.**

***Index Terms*— Hydroelectric generating unit (HGU), fault diagnosis, 1-dimension convolutional neural network (1-D CNN), gated recurrent unit (GRU).**

I. Introduction

AS WE know, the generating capacity of small hydro far less than hydroelectricity. However, a recent study shows that there are at least 82,890 small hydros around the world, which is 10 times more than hydroelectricity [1]. For example, according to statistics by the ministry of water resources of China, China had 47,498 small hydro installations at the end of 2017, producing about 247 TWh annually [2]. Therefore, the small hydro is a non-negligible part of the power system. Hydroelectric generating unit is one of the most important devices in the hydropower plant, which greatly affects hydropower station and power system overall performance [3]. In addition, small hydro may be built in remote areas and faced with the unattended operation, which brings difficulties for maintenance and puts higher requirements for the safety of the hydroelectric generating unit (HGU). Therefore, a reliable, precise, and intelligent identification model for HGU faults is practical and significant.

Manuscript received May 18, 2019; revised June 27, 2019; accepted June 27, 2019. Date of publication July 1, 2019; date of current version September 18, 2019. This work was financially supported by the National Natural Science Foundation of China under Grant 51677030. The associate editor coordinating the review of this paper and approving it for publication was Prof. Ruqiang Yan. *(Corresponding authors: Wei Gao; Gengjie Yang.)*

The authors are with the Department of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108, China, and also with the Fujian Smart Electrical Engineering Technology Research Center, Fuzhou University, Fuzhou 350108, China (e-mail: 383861548@qq.com; gaowei0203@fzu.edu.cn; ygj23802@126.com; gmf@fzu.edu.cn).

Digital Object Identifier 10.1109/JSEN.2019.2926095

A. Previous Work

1) HGU Diagnosis: At present, the method fault diagnosis of HGU can be divided into three categories, including model-based, knowledge-based and vibration-signal-based. Saeed *et al.* [4] developed the integration of numerical modeling of Francis turbine by finite element analysis, and then diagnosing the HGU fault based on the proposed model. Simeon *et al.* [5] designed a Web-based HGU fault diagnosis knowledge base system using a hybrid IF-THEN rule presentation method, which can help plant maintainers and operators for solving HGU fault. However, model-based or knowledge-based fault diagnosis methods cannot directly diagnose the cause of the failure, and these methods depend on expert experience. Other researchers adopt the vibration-signal-based method to diagnose HGU faults because the vibration signal is easy to obtain and contains a wealth working status information. For example, Zhu *et al.* [6] utilized kernel independent component analysis (KICA) and principal component analysis (PCA) for HGU fault detection. Lu *et al.* [7] adopted empirical mode decomposition (EMD) to decompose the vibration signal, and then the modified back-propagation neural network (BPNN) is used to HGU fault diagnosis. In [8], an artificial-neural-network was proposed, which immensely improved the accuracy of fault diagnosis for HGU. Yan *et al.* [9] proposed an improved hybrid backtracking search algorithm based on T-S fuzzy model to implement HGU fault diagnosis. A status monitoring and fault diagnostics system for HGU was built in [10], where a support vector machine (SVM) was used for fault diagnostics. Although the above methods are effectively verified in fault diagnosis of HGU scenarios, these works also indicate that it is difficult to analyze the vibration information of HGU owing to their nonlinearity, nonstationary and high-noise-component peculiarities. Based on the research status of HGU fault diagnosis in recent years, it can be observed that signal

decomposition, feature extraction, and pattern recognition are three main steps in fault diagnosis based on vibration signal. In particular, the quality of the extracted features is crucial to diagnosis accuracy. Therefore, many forms of time-frequency analysis methods are gradually proposed. For example, wavelet transform (WT) [11], PCA [12], EMD [13] and its improved approaches [14], [15], *etc*. However, there are exists various disadvantage. The scalable analysis window in WT results unsatisfied in time-frequency representation concentration [16], PCA based on second-order statistics and linearity assumption unsuitable to deal non-linear vibration signal directly [17], and the issue of mode mixing and end effect exists in EMD [13]. Furthermore, traditional intelligent fault diagnosis methods design and perform feature extraction and classification separately, which will affect the final diagnosis performance [18]. Furthermore, unsupervised learning methods are also used in HGU fault diagnosis. Ahmed *et al.* [19] proposed a minimum spanning tree (MST)-based hydropower turbines anomaly detection method, which helps distinguish the anomalous data from the regular dataset. In [19], although the fault detection ability of this method was remarkable, it didn't mention how to classify different fault types. In the application scenarios of this study, the vibration signal of the hydroelectric generating unit has large noise, nonlinearity, and nonstationary. The noise component includes ambient noise and abnormal sound that is emitted when the HGU faults happen. Moreover, the frequency of these abnormal sounds is difficultly determined. Therefore, traditional intelligent diagnosis methods may neglect these fault features. Furthermore, incorrectly collected vibration signal or dispersion in the dataset might cause over-fitting in existing intelligent methods for HGU. Therefore, it is necessary to build deep architectures, which combines automatic and effective feature extraction and intelligent diagnosis.

2) Deep-Learning-Based Approaches: Deep learning is an algorithm based on learning data representations in machine learning, which can attempt to extract high-level representations of raw data via multiple stacked layers of nonlinear information processing units in a hierarchical architecture [20]. Therefore, deep learning approaches can overcome the disadvantage of the traditional intelligent methods. Many researchers adopted deep learning for fault identification in rotating machinery scenarios. For examples, Pan *et al.* [21] proposed a model combing the convolutional neural network (CNN) and long-short-term memory recurrent neural network to input the raw vibration signal for fault diagnosis. Liu *et al.* [22] developed a novel method for fault diagnosis of bearing using GRU non-linear predictive denoising autoencoders, which provided a new way to process multi-dimension vibration signal. In [23], a novel method, which is utilized to automatically feature learning and fault diagnosis in the rolling bearing, is based on improved convolutional deep belief network with compressed sensing. As studied by Ince *et al.* [24] and Li *et al.* [25], 1-D CNN or GRU is a novelty deep learning method to address the complex characteristics of vibration signal in real-world industrial application and improved the performance of fault diagnosis considerably. The above research works are effectively verified in their respective application scenarios. Unfortunately, in recent years, the deep learning which does not require any feature extraction has not been used in HGU fault diagnosis. Therefore, this study attempts to solve the problem of fault diagnosis for hydro units by using the existing deep learning method.

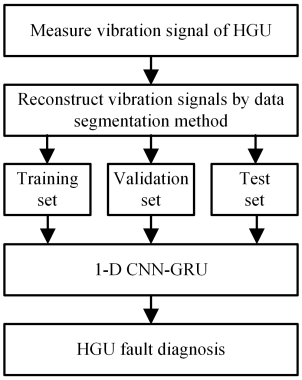


Fig. 1. Flow chart of proposed HGU fault diagnosis.

B. Contribution

Based on the above analysis, this study aims to build a novel method for HGU fault diagnosis system using 1-D CNN-GRU, which can achieve an elegant fault diagnosis performance. The flow chart of the proposed algorithm is depicted in Fig. 1. In the real industrial environment, it is difficult to collect a large number of various fault samples for training. Therefore, fault samples are collected through the numerical simulation model. Moreover, four standard machine learning classifiers, namely 1-D CNN, GRU, BPNN, SVM, are constructed to compare the results by using measured and simulation vibration signal, respectively. The primary contributions of this study can be summarized as follows:

1) In order to improve the ability of adaptively extracting features of the proposed method and the efficiency of its training process, the vibration signals are reconstructed in chronological. By comparing with the raw data, it can be observed the time steps of the reconstruction data decreased and its features with respect to each time step increased. As a result, it will take full advantage of the sequential property of the raw vibration signal in automatic extraction feature stage.
2) The reconstruction data includes various important vibration information. Thus, in order to enhance the capacity for capturing complementary and rich features directly from the raw vibration signals, 1-D CNN combined GRU is adopted. The reason is that 1-D CNN-GRU can combine the speed and lightness of convnets with the order-sensitivity of GRU.
3) A novel approach for HGU fault diagnosis using 1-D CNN and GRU can prominently improve the accuracy of the binary classification and multiclass faults diagnosis of HGU based on the experimental results in our study. Specifically, the diagnostic system built in this study is successfully applied in engineering practice, which has shown promising robustness and outstanding diagnosis

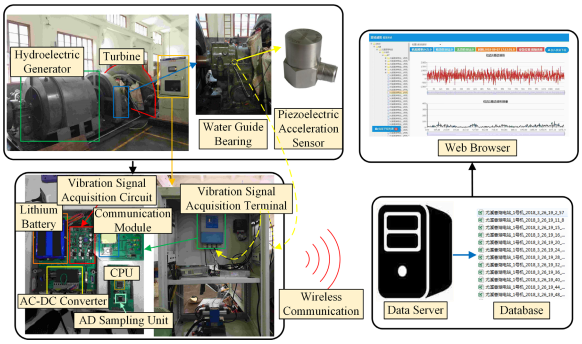


Fig. 2. Online vibration signal acquisition system for HGU.

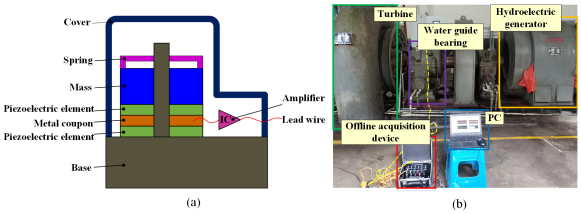


Fig. 3. (a) Model of IEPE sensor. (b) Offline vibration signal acquisition device for HGU.

performance. As far as I know, this is the pioneering work applied for HGU fault diagnosis, which can implement automatically features extraction from vibration signal and fault diagnosis, simultaneously.

C. Work Organization

The remainder of this paper is organized as follows: In Section II, introduces the case rig briefly. Then, the main frequency of common faults, data description, the data segmentation method, and fault diagnosis framework are described in Section III. Section IV presents different experimental results, and practical application of the binary diagnostic model. Finally, Section V summarizes the proposed method.

II. Case Rig

In this study, the small HGU operating in real-environment is used as case object, which includes a turbine with the type HLD54-WJ-71 and a hydroelectric generator with the type SFW1250-6/1180. Then, the vibration signal of HGU under different loads is collected by online acquisition system consisting of vibration signal acquisition terminal, data server, fault analysis module, and web browser. Among them, vibration signal acquisition terminal includes an acceleration sensor with the type LC0166C, AC-DC converter, AD sampling unit, vibration signal acquisition circuit, communication module, and CPU. This case rig is shown in Fig. 2. Moreover, the type of acceleration sensor LC0166C is integrated electronics piezoelectric (IEPE). The model of the sensor is shown in Fig. 3(a). It can be observed from Fig. 3 (a) that IEPE makes integrating the sensitive electronics as close as possible to the transducer to ensure better noise immunity. In other words, it is suitable that the IEPE sensor is used to collect data in a real operating environment of small hydro. The measuring range, sensitivity, resolution of the sensor are ± 5 g (where $g = 9.8 ms^{-2}$), $1000 mV g^{-1}$, 0.00002g, respectively. The sampling frequency and sampling period of online acquisition system are 3000Hz, and 0.4s, respectively.

The general process of collecting vibration signals by this online acquisition system is as follows: when the HGU runs, the vibration value of HGU is converted to the vibration acceleration value by piezoelectric accelerometer, and then the vibration signal acquisition circuit converts these varying acceleration value into voltage signals, which represent the vibration information of HGU in the sampling period. Then, the vibration signal acquisition terminal periodically connects to the socket port of data server through the communication module, thereby transmitting and storing the data packets to the data server. Finally, these vibration signal waveforms and their corresponding diagnostic information will be displayed on the web server. In Fig. 2, the yellow dotted lines represent the acquisition flow of vibration signals.

In this study, the principles of sensor selection are as follows: Firstly, ensure that the converted electrical signal is not saturated and not too small. Secondly, the price of the sensor cannot be too expensive given the demand for large-scale applications. Finally, the acceleration sensor can be easily amounted on the water guide bearing shell, which belongs to micro-vibration. Before selecting the installation position of the sensor, the vibration signals of HGU from multiple stations collected through the offline vibration signal acquisition device were stored in a PC as shown in Fig. 3 (b). Then, these vibration signals were analyzed by signal processing methods. The analysis results indicated that the vibration signals located at the horizontal position of the water guide bearing have the strongest vibration intensity and the most obvious vibration frequency characteristics. Therefore, the ideal installation position of the acceleration sensor is finally determined when the analysis result and the key role of bearing in HGU are comprehensively considered. The final installation site is shown in Fig. 2.

III. Methodology

A. Faults Frequency in HGU

HGU belongs to a kind of rotating machinery, which mostly has a vibration phenomenon [26]. When faults occur in HGU, the fault features are mainly expressed in the vibration signal. The excitation sources of vibration signal mainly include the factors of hydraulic, mechanical, and electrical. According to [27], the main frequency of typical fault in HGU is counted in Table I.

It is clearly shown from Table I that all the malfunctions, except polar frequency vibration and Karman vortices, are related to rotation frequency f_0. Among them, the main frequency of HGU faults is low frequency, which includes misalignments, unbalance, bend of rotor, unequal gap of water seal, rotor winding inter turn loosening, and eccentric draft tube surges.

TABLE I
MAINLY INTERMEDIATE FREQUENCY OF TYPICAL FAULT IN HGU

Excitation Sources	Type of Fault	Main Frequency (Hz)
Hydraulic factors	Eccentric draft tube surges	$f = (1/6\sim1/2)\, f_0$
	Karman vortices	$f = Cw/\delta$
	Unequal gap of water seal	$f = f_0$
	Hydraulic disequilibrium	$f = Z_0 f_0$
Mechanical factors	Unbalance	$f = f_0$
	Bend of rotor	
	Rotor winding inter turn loosening	
	Bearing or shaft misalignments	$f = 2f_0$
	Thrust bearing unevenness	$f = mf_0$
Electrical factors	Polar frequency vibration	$f = 50n$ (n=1, 2, 3, …)
	Rotation frequency vibration	$f = kf_0$ (k=1, 2, 3, …)

where f_0=n_N/60, n_N is the rated speed of the hydro unit, C is the coefficient associated with the Reynolds number, generally is equal to 0.2, w is the blade exit edge velocity, δ is the tip edge thickness, Z_0 is the number of runner blades, and m is the number of thrust bearing.

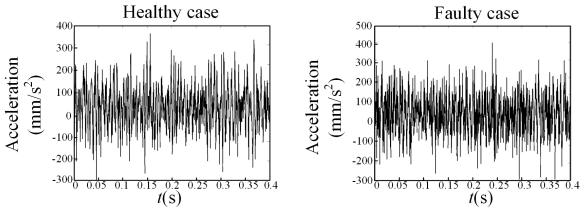


Fig. 4. Different type of the vibration signal.

B. Data Description

1) Measured Vibration Signal: The measured vibration signal, which is used to verify the effectiveness of the proposed method in binary classification, is collected from Xianghu hydropower station, Fujian, China. On March 19, 2018, the one of HGU in Xianghu Hydropower Station had shut down to maintenance due to excessive bearing temperature. By calculating the strength of the main shaft and the force of bearing, it was concluded that the fault was caused by the excessive clearance of bearing bush. The maintenance of HGU had completed on March 26. In addition, the development of these faults is a gradual process [28]. Through the waveform analysis, combining with engineering experience of diagnostic experts and prior knowledge, the vibration signals of March 13 to 19 and March 26 to 30 were picked as the faulty and healthy data, respectively. In fact, the measured signal contains the case of the hydropower unit running at different loads which include light-load, half-load, rated-load, and heavy-load, *etc*. These vibration signals and their waveforms are shown in Table II and Fig. 4, respectively.

Although the case object is the big unit, the sensor is fixed at water guide bearing shell as shown in Fig. 2. Moreover, in this paper, it can be observed from Fig. 4 that the acceleration value of the measured vibration signal ranges from −0.3g to 0.5g (where g = 9.8ms^{-2}). Therefore, the 5g measuring range of sensor is sufficient to meet the requirements of this research.

TABLE II
SUMMARY OF MEASURED VIBRATION SIGNAL

Health Condition	Description of States	Category Label	Sample Number
Healthy	Normal case	Class 1	540
Faulty	Excessive bush gap	Class 2	540

2) Simulation Fault Vibration Signal: In order to verify the multi-faults classification performance of the proposed methods, a large number of HGU fault samples are required. However, in real hydropower unit operating environments, it is difficult to acquire a large number of fault samples for classifier training [29]. Therefore, in this paper, the numerical simulation model is used to simulate the various faults of HGU at guide bearing. As stated in [30], Equation (1) is used to simulate the vibration signal of HGU at guide bearing.

$$s(t) = \sum_{i=1}^{6} A_i \sin 2\pi f_i t \tag{1}$$

where i is the number of mainly intermediate-frequency, A_i is the corresponding amplitude under different main frequency that varies within a certain range, f_i is the mainly intermediate-frequency of mechanical and hydraulic excitation.

Furthermore, in [31], by comparing the change of the spectrogram before and after the rotor winding inter-turn loosening, it can be observed that the rotation frequency is significantly increased and the remaining multiples rotation frequency is almost unchanged. Based above analysis, the principle of Equation (2) is that the sine wave signal, which contains a mainly intermediate-frequency of specific fault, is superimposed on the healthy vibration signal. In addition, the various fault mainly intermediate frequencies are obtained by substituting the rated parameters of the HGU (the rated speed $n_N = 1000$r/min, the number of runner blades $Z_0 = 15$, and the number of thrust bearing $m = 4$) into Table I. The numerical simulation model is expressed as.

$$s_j(t) = s_0 + \sum_{i=j}^{j+n-1} A_i \sin 2\pi f_i t \tag{2}$$

where s_0 is the measured vibration signal in healthy state, n is the number of fault critical frequencies with respect to a specific fault, A_i is the corresponding amplitude under different faults that varies within a certain range, f_i is the corresponding frequency under different faults, j is the number of fault category.

Five different fault samples (i.e., eccentric draft tube surges, thrust bearing unevenness, hydraulic disequilibrium, unbalance, and mixed fault) were collected in Fig. 5. It is clearly seen that a variety of fault information reflected in Table I is included within the frequency range of 500Hz. Therefore, the sampling frequency of 3000Hz is sufficient to meet the requirements of the research according to Nyquist-Shannon sampling theorem [32].

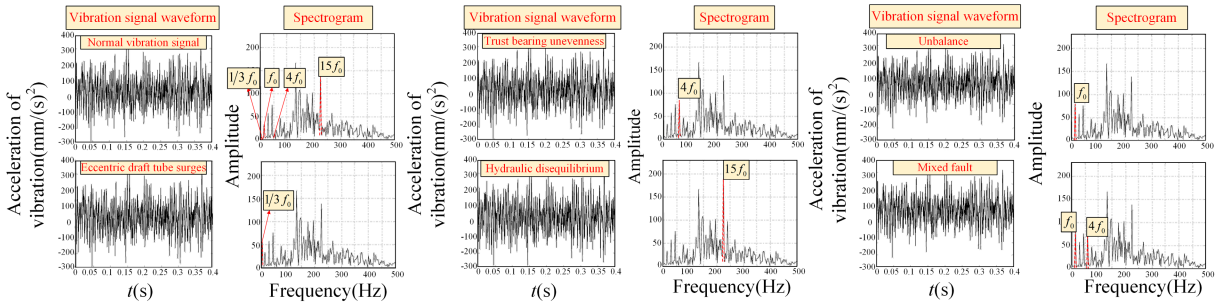


Fig. 5. Simulation fault vibration signal and its spectrogram.

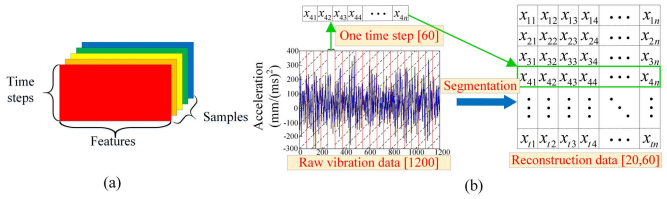


Fig. 6. (a) 3-D tensor for vibration datasets. (b) Data reconstruction progress.

C. Data Segmentation

The vibration signals are a kind of time series data, which should be stored in a 3-dimension (3-D) tensor with an explicit time axis. Besides, each vibration signal can be encoded as a sequence of a 2-dimension (2-D) tensor. The 3-D tensor of the vibration signal datasets is shown in Fig. 6 (a), where different colors represent different categories of samples.

As already stated in Section II, the vibration signals are collected from the water guide bearing. In this paper, considering that the hydroelectric generator grid connected frequency is 50 Hz and the sampling period is set to 0.4s. Moreover, the proposed method can extract features more effectively by input the samples of complete periodicity. Therefore, the sampling number is 1200, which suggest that the time step of each data is 1200 and each time step contains only one feature value. Once these data trained by proposed method could reduce the network ability of adaptive feature extraction and increase its training time. Therefore, the raw vibration data should be reconstructed to reduce time steps and increase the features of each time step. The specific process is as follows:

The sampling frequency is 3000 Hz and the rated rotating speed of HGU is 1000r/min, which means about 180 data points will be collected for one revolution (one revolution is 3/50 s). Therefore, the vibration signal with a length of 1200 is divided into 20 segments (i.e., 20 time steps) in sequence and 60 points per segment (i.e., 1/3 revolution). This segmentation process is shown in Fig. 6(b).

The data segmentation method is used to reconstruct the raw vibration data, which laid the foundation for the neural network to run faster, better extract features and higher fault identification accuracy.

D. 1-D CNN-GRU Architecture

The overall structure of 1-D CNN-GRU consists of three parts: the input layer, the hidden layer, and the output layer. In addition, the hidden layer contains the convolution layer, the pooling layer, the GRU layer, and the fully-connected layer. Therefore, 1-D CNN-GRU-based architecture contains 6 layers as shown in Fig. 7.

CNN is a multi-layered neural network, where each layer contains multiple feature planes consisting of multiple independent neurons [33]. 1-D CNN is often used in time-sequence models [24], which convolutional output are one-dimensional. Fig. 8 shows that the basic difference of 1-D CNN and 2-D CNN is the sliding direction of filters. As stated in section III-C, the time-axis is important for the vibration signal, which is stored in a 3-D tensor. In addition, the filters of 1-D CNN are slides along time-axis and the input shape of 1-D CNN is 3-D tensor. Therefore, 1-D CNN is adopted in this study.

In this study, rectified linear unit (ReLU) [34] is selected as activation function in 1-D CNN. He *et al.* [35] confirmed that he_normal which is the implementation of weight initialization based on Gaussian distribution would make the data input to the ReLU having good constant variance. Therefore, the he_normal initializer is chosen to initialize weight in convolution layer, which will draw samples from a truncated normal distribution centered on 0 with Standard deviation formula as follow.

$$std = \sqrt{2/fan} \tag{3}$$

where *fan* is the number of input units in the weight tensor.

Suppose the input sequence matrix is D, and the i-th output sequence matrix is S_i, then $S_0 = D$.

When $i = 1$, the current layer is convolution layer. The convolutional layer parameters consist of a set of learnable filters, which is small spatially ($c \times d$, where c is the size of the kernel and d is the dimension of input data). The output feature matrix S_i can be express as:

$$S_i = f(S_{i-1} \otimes w_i + b_i) \tag{4}$$

where w is the weight matrix and b is the bias. In this study, the nonlinear activation function uses ReLU. The zero padding is adopted in convolution layers so that the size of S_1 unified into $m \times n$(m is the time steps of input data and n is the number of filters).

When $i = 2$, the current layer is pooling layer, and its role is to reduce the time steps of the sequence features while maintaining the scale invariance of features. In this study,

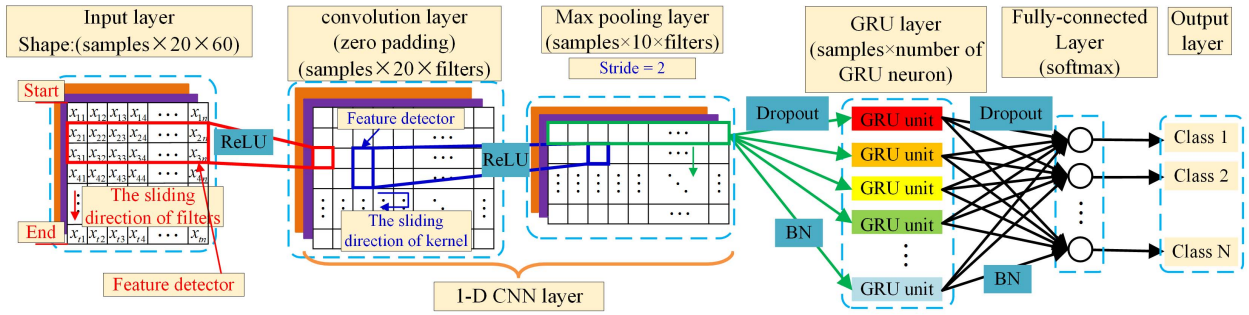


Fig. 7. Framework of 1-D CNN-GRU.

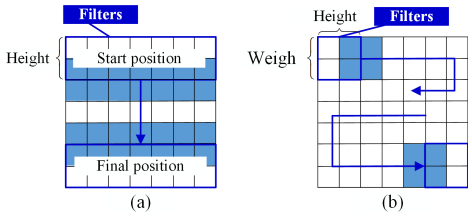


Fig. 8. Sliding direction of filters. (a) 1-D CNN. (b) 2-D CNN.

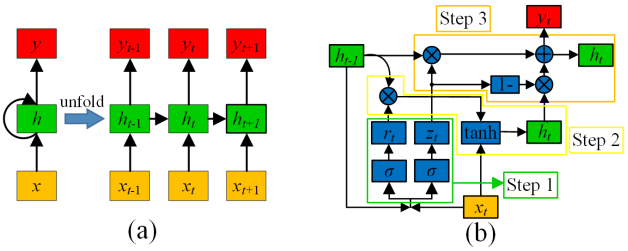


Fig. 9. Architecture. (a) Basic architecture of RNN. (b) Basic architecture of GRU.

the Max pooling method is used in the pooling layer, whose output is the maximum of the previous feature matrix. The output feature matrix S_i is defined as follows:

$$S_i = Y(S_{i-1}) \tag{5}$$

where Y is the pooling function. The size of S_2 is $m/z \times n$, where z is the pooling layer scale value of the current layer.

When $i = 3$, the current layer is GRU layer. The architecture of standard RNN is shown in Fig. 9 (a).

GRU is a popular RNN architecture designed for store the previous state information better than traditional RNN, which can address the problem of gradient vanishing and gradient explosion in traditional RNN training process [36]. The main difference between the GRU and standard RNN is that the hidden units' structure of the standard RNN was replaced by GRU cells. The architecture of GRU hidden unit is shown in Fig. 9(b).

The first step is to get states of the reset gate and the update gate by the last transmitted h_{t-1} and the current input x_t, whose output could be expressed as:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \tag{6}$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \tag{7}$$

where σ is a sigmoid function, W and U are weight matrix, and b is the bias.

The second step is to determine what new information is stored in the current node $\hat{h}_t$ through the reset gate, which could be calculated as:

$$\hat{h}_t = \tanh(W_h x_t + U_h(r_t \circ h_{t-1}) + b_h) \tag{8}$$

Finally, the update gate is used to perform both forgetting and remembering steps, and its output could be calculated as:

$$h_t = z_t \circ h_{t-1} + (1 - z_t) \circ \hat{h}_t \tag{9}$$

where $z_t \circ h_{t-1}$ indicates selective "forgetting" of the original hidden state, $(1\text{-}z_t) \circ \hat{h}_t$ represents selective "memory" of current node information.

When $i = 4$, the current layer is the fully-connected layer, and the softmax function is selected as the activation function. In other words, the fully-connected layer performs the final classification. In this layer, the probability of currently each sample corresponds to each category is calculated. Then, a new feature expression ($y_{predict}$) form is obtained, which could be calculated as:

$$y_{predict}(i) = f(L = l_i | S_3; (W, b)) \tag{10}$$

where f is the softmax activation function, S_3 is the features obtained from GRU, l_i is the calculation result of the input data of the i-th class.

The cross-entropy is usually the best choice when dealing with models of output probabilities. In addition, L2 regularization can be viewed as an effective way of compromising between finding small weights and minimizing the cost function [37]. Therefore, the cross-entropy and L2 regularization are adopted to prevent overfitting. The training goal of the model is to minimize the loss function calculation, which could be calculated as:

$$Loss(W, b) = -\sum_{i=1}^{n}\sum_{t=1}^{c}(y^{(t)}(i) * \log(y^{(t)}_{predict}(i)) + \frac{\lambda}{2}\sum_{i=1}^{n} W_i^2 \tag{11}$$

where $y^{(t)}(i)$ is an indicator variable (0 or 1), if the t-th category is the same as the category of the i-th sample, it is 1, otherwise, it is 0; $y^{(t)}_{predict}(i)$ is the probability of the i-th sample corresponds to t-th category; n is the number of samples; c is the number of class; λ is the coefficient of L2 regularization.

The Adam optimizer which can adaptively adjust learning rates for different parameters [38], is used to the process of minimizing the loss function. The backpropagation algorithm is used to find the gradient of the loss function with respect to the parameter (*W, b*), and the value of the parameter (*W, b*) is updated according to the optimal gradient complete the training process.

In particular, the working condition of HGU is varied hugely. The measured vibration signal exists large ambient noise and strong dispersibility. In other words, this measured data might potentially cause neural network overfitting. Some necessary measures must be adopted to address this problem. Batch normalization, which is demonstrated as an effective method to avoid overfitting of the network, can dramatically accelerate the training of networks [39]. Dropout can prevent neuron units from co-adapting too much, and then significantly reduces overfitting [40]. Moreover, the L2 regularization method is added to the cost function. Through the regularization methods of processing, the proposed method can avoid overfitting and internal covariate shift.

The principle of 1-D CNN-GRU is as follows: Firstly, 1-D CNN is used to adaptive extraction feature from measured vibration signal. Secondly, GRU can automatically learn the relation between different time steps in the rich features of 1-D CNN output. Finally, the softmax is used to classification for GRU output. Therefore, the proposed method can simultaneously realize vibration signal automatically features extraction and fault diagnosis.

E. General Procedure of Diagnosis Process

The flow chart of our diagnosis process is shown in Fig. 10 and the general steps are summarized as follow:

Step 1: Collect the vibration signals of HGU. In this study, the vibration signal is collected from two ways, one is to obtain various types of fault signals through numerical model simulation, and the other is to collect the measured vibration signals at the water guide bearing by the online acquisition system. The operating conditions of the measured signals include light-load, half-load, rated-load, and heavy-load, etc.

Step 2: The samples of known types are labeled and a database containing various types of faults for HGU is established.

Step 3: Reconstruct the vibration signal into a matrix by data segmentation and normalization.

Step 4: Split dataset into training, validation, and test set.

Step 5: A network combining 1-D CNN and GRU algorithm is trained. Methods such as batch normalization, L2 regularization, and dropout are adopted to optimize network parameters in order to avoid over-fitting and internal covariate shift. Ultimately, the trained network parameters are used to build the fault diagnosis model of HGU.

Step 6: The predicted labels are obtained, after test samples are input into the fault diagnosis model. For the test set of known labels, the performance of the diagnostic model can be evaluated based on the predicted labels and true labels.

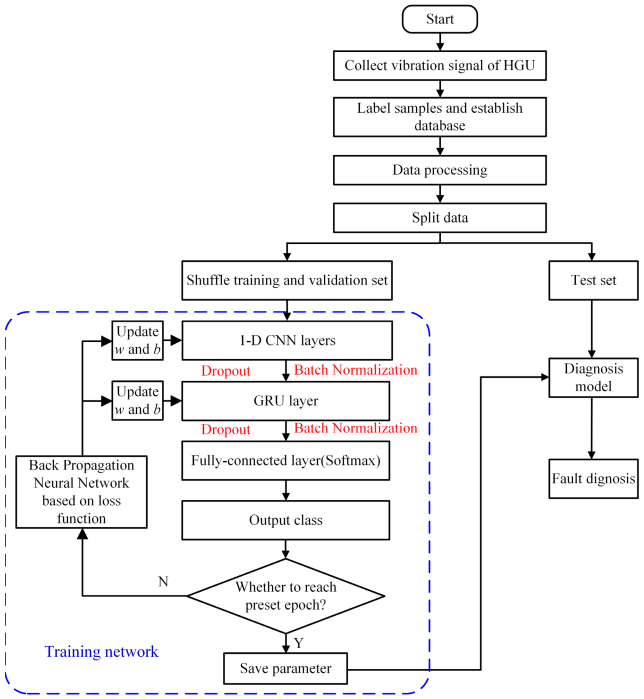


Fig. 10. Flow chart of diagnosis process.

IV. Experiments Verification

In this study, all the experiments were performed on Ubuntu system platform, and executed on a computer with two NVIDIA GTX-1080Ti. Certainly, the model is trained offline on the above stated platform, and then this model is used for online fault diagnosis on a general computer.

A. Performance Metrics

1) Binary Classification: The measured vibration signals, which are described detailedly in section III-B, are used to validate the binary classification performance of the proposed method. The training set, the validation set, and the test set are randomly selected from 540 measured signals, and the ratio is 8:1:1. Then, the training, validation, and test set are reconstructed by data segmentation method in section III-C. Before training, each data is normalized between -1 and 1. The network parameters of the model are continuously adjusted to meet high test accuracy and training speed, as shown in Section VI-B experiment1. The trained network parameters are listed in Table III.

It can be shown from Fig. 11, the fast convergence speed and no overfitting phenomenon during the training and validation process, indicating that the regularization method adopted in this study is effective. The blue and the orange curve express the training accuracy and validation accuracy in Fig. 11 (a), respectively. In Fig. 11 (b), the blue and the orange curve express the training loss and the validation loss, respectively.

The binary-class confusion matrix records the testing classification results in a detailed manner as shown in Fig. 12. The value of the intersection of the last row and the last column of the confusion matrix represents the classification accuracy. The element value of the confusion matrix represents

TABLE III
PARAMETERS USED IN BINARY CLASSIFICATION EXPERIMENT

Description	Value
Numbers of filters in convolution layer	128
Size of kernel in convolution layer	8
Strides in maxpooling layer	2
Pool size in maxpooling layer	2
Numbers of neurons in GRU layer	128
Numbers of neurons in Dense layer	2
Learning rate of Adam	0.0001
Dropout	0.2
Batch size	32
Maximum epochs	60

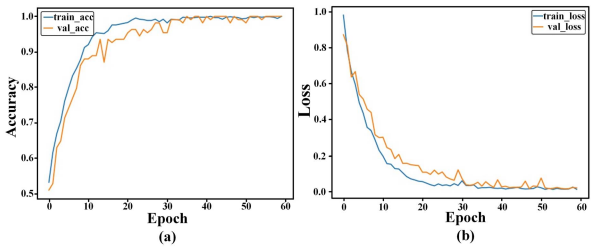


Fig. 11. Curve of training and validation progress.

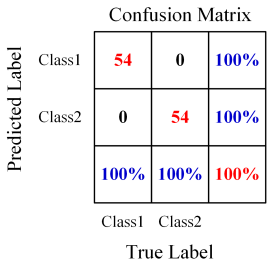


Fig. 12. Diagnosis results for binary-category.

the number of matches or mismatches sample between the true label and the predicted label. In addition, the last row and the last column of the confusion matrix represent the precision and the recall of each label, respectively. In this figure, it can be observed that the precision is 100% and also recall is 100%, which occurs in each class. Furthermore, for all test samples, the overall classification accuracy of the confusion matrix is 100%, which suggests that the proposed method has a promising performance for measured data classification.

2) Multiclass Classification: Section III-B introduces the simulation process of five common fault vibration signal in detail. Then, these fault signals are reconstructed into new structured data by data segmentation. In particular, 800 healthy measured data were randomly extracted from database. With these data, samples of five fault types were simulated, with 800 data for each type, for establishing fault dataset. The multiclass dataset includes healthy and faulty dataset listed in Table IV.

Samples from each type of signal were randomly selected to constitute the training set, validation set, and test set, with a ratio of 8:1:1. Before training, each data is normalized between −1 and 1. The parameters used in multiclass classification experiment are listed in Table V.

TABLE IV
SUMMARY OF SIMULATED VIBRATION SIGNAL

Health Condition	Description of States	Category Label	Sample Number
Healthy	Normal case	Class 1	800
Faulty	Eccentric draft tube surges	Class 2	800
	Unbalance	Class 3	800
	Guide and Runner blades fracture	Class 4	800
	Thrust bearing unevenness	Class 5	800
	Mixed fault	Class 6	800

TABLE V
PARAMETERS USED IN MULTICLASS CLASSIFICATION EXPERIMENT

Description	Value
Numbers of filters in convolution layer	128
Size of kernel in convolution layer	8
Strides in maxpooling layer	2
Pool size in maxpooling layer	2
Numbers of neurons in GRU layer	256
Numbers of neurons in Dense layer	6
Learning rate of Adam	0.0001
Dropout	0.2
Batch size	64
Maximum epochs	200

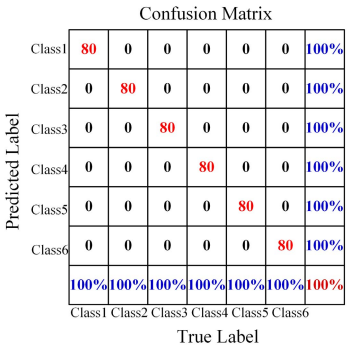


Fig. 13. Diagnosis results for multiclass classification.

As can be seen from Fig. 13, the confusion matrix of the multiclass classification records the classification results under all conditions in detail. It can be observed that the overall classification accuracy of multiclass classification is 100% indicating that the proposed method has superior performance for the multiclass classification.

The running time of different training and test dataset for the fault diagnosis algorithm based on deep learning framework Keras, is recorded in Table VI. In this table, the training time of all epoch and the test time of one sample are calculated, respectively. Particularly, all diagnosis models are trained offline and then used for testing vibration signals. Therefore, our proposed 1-D CNN-GRU approach is suitable for real-time diagnosis.

B. Discussion

1) Experiment 1: Parameters of Network: Experiment 1 is used to determine the parameters of binary classification. Table VII shows the parameters of the proposed model used for

TABLE VI
COST TIME FOR 1-D CNN-GRU FOR DIFFERENT DATASET

Datasets	Training Time (s)	Test Time (s)
Measured datasets	61.0623	0.0524
Simulation datasets	84.7805	0.0436

TABLE VII
PARAMETERS LISTED WITH THE TRAINING TIME AND ACCURACY STATISTICS OF MODEL PERFORMANCE. THE ACRONYMS OF EACH PARAMETERS: NUMBERS OF FILTERS IN CONVOLUTION LAYER = NFC, SIZE OF KERNEL IN CONVOLUTION LAYER = SFC, NUMBERS OF NEURONS IN GRU LAYER = NNG

Model number	NFC	SKC	NNG	Training time(s)	Accuracy (%)
M1	64	4	64	58.6302	93.52
M2	64	4	128	59.6611	94.44
M3	64	4	256	60.1613	92.59
M4	64	8	64	57.6803	94.44
M5	64	8	128	58.7929	96.29
M6	64	8	256	59.8027	95.37
M7	64	16	64	55.9902	93.52
M8	64	16	128	57.0542	95.37
M9	64	16	256	58.9934	94.44
M10	128	4	64	64.5481	95.37
M11	128	4	128	65.1233	97.22
M12	128	4	256	66.9821	96.29
M13	128	8	64	63.2348	99.07
M14	**128**	**8**	**128**	**64.0623**	**100**
M15	128	8	256	65.7821	98.14
M16	128	16	64	62.3589	97.22
M17	128	16	128	63.2363	98.14
M18	128	16	256	64.4920	96.29
M19	256	4	64	74.1293	92.59
M20	256	4	128	76.8312	93.52
M21	256	4	256	78.0495	90.74
M22	256	8	64	72.8941	93.52
M23	256	8	128	74.3201	95.37
M24	256	8	256	75.8921	94.44
M25	256	16	64	71.2451	93.52
M26	256	16	128	72.8971	94.44
M27	256	16	256	74.0912	94.44

classifying the measured vibration signal. In this experiment, the batch size and epoch are set to 32 and 60, respectively.

It can be observed from Table VII that the classification accuracy of M14 is better than other models. Moreover, the training speed of M14 can reach a higher level. Then, the numbers of filters in convolution layer, the size of kernel in convolution layer, and the numbers of neurons in GRU layer are set to 128, 8, and 128, respectively.

Table VIII shows that the batch size and epoch of the proposed method used for classifying the measured vibration signal. In this experiment, the NFC, SKC, and NNG are set 128, 8, and 128, respectively.

As can be seen from Table VIII, the performance of M32 is superior to other models. Based on the above results, the parameters used in the binary classification are determined. Similarly, the parameters of multiclass classification can also be determined by this method.

2) Experiment 2: Effects of Data Segmentation: The network parameters of the proposed model of Table III is used in this experiment. The measured vibration signal is used to

TABLE VIII
BATCH SIZE AND EPOCH LISTED WITH THE TRAINING TIME AND ACCURACY STATISTICS OF MODEL PERFORMANCE

Model number	Batch size	Epoch	Training Time (s)	Accuracy (%)
M28	16	50	100.9821	94.44
M29	16	60	121.6023	97.22
M30	16	70	142.0226	98.15
M31	32	50	53.3853	95.37
M32	**32**	**60**	**64.0623**	**100**
M33	32	70	74.7394	100
M34	64	50	30.2337	89.81
M35	64	60	35.9012	90.74
M36	64	70	42.0023	90.74

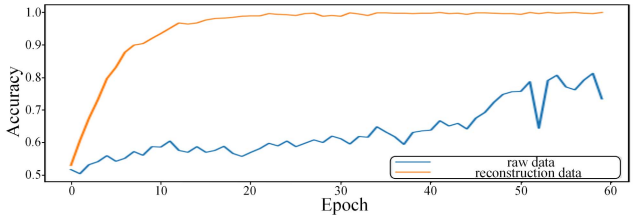


Fig. 14. Curve of training progress for different dataset.

TABLE IX
COST TIME OF ONE EPOCH FOR DIFFERENT DATA

Data	Raw data	Reconstruction data
Time (s)	19.8674	1.0177

confirm the advantages of the data segmentation, which will reconstruct the measured data (size of 1×1200) into new input data (size of 20×60). Then, these datasets are input 1-D CNN-GRU for training, respectively. The accuracy curves with both measured data and reconstruction data are shown in Fig. 14. Under the same conditions, it can be observed that the iteration speed of the measured data is slower than reconstructed data in the training process. Furthermore, 90% accuracy is achieved with only 11 epoch using reconstructed data in the training process. Table IX shows that the time of training one epoch with raw data is nearly 20 times longer than reconstruction data. Therefore, the data segmentation method, which is used to reconstruct vibration data, can improve the convergence speed of the proposed method, its ability for adaptively extracting features, and its higher fault identification accuracy.

3) Experiment 3: Effectiveness Evaluation: In experiment 3, in order to evaluate the effectiveness of the proposed method, five classification methods including the proposed method, 1-D CNN, GRU, BPNN, and SVM, were used to diagnose two types of datasets (measured and simulation vibration signal). The input of the proposed method, 1-D CNN, and GRU are the raw vibration data. SVM and BPNN have two inputs. One is the raw data. The other is the frequency domain (FD) features extracted from each sample using variational mode decomposition (VMD) [41]. In the evaluation index of the fault classification method, the classification accuracy, precision,

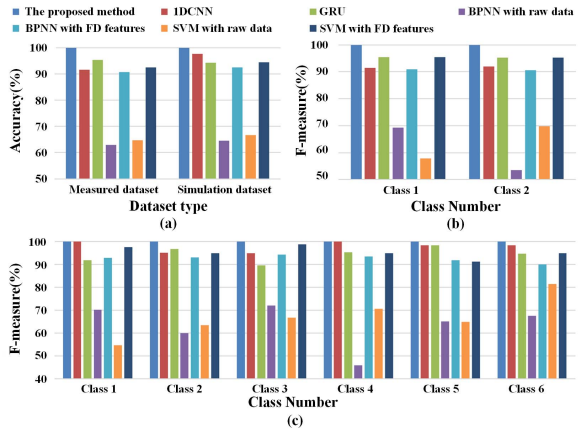


Fig. 15. Comparison results. (a) Accuracy for different datasets with various classifiers. (b) F-measure for measured data with various classifiers. (c) F-measure for simulation data with various classifiers.

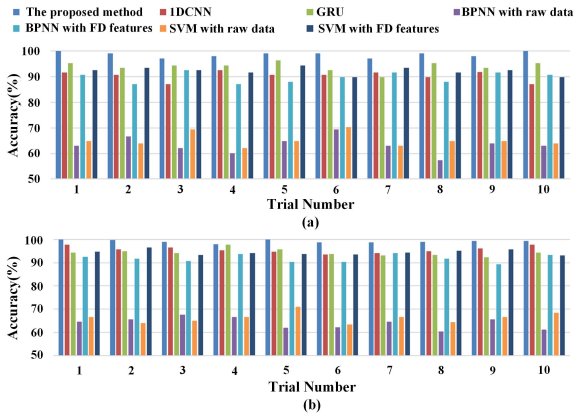


Fig. 16. Comparison results of 10 trials using different methods. (a) Accuracy of ten trials measured data. (b) Accuracy of ten trials simulated data.

and recall rates are considered as the key index. Furthermore, F-measure is another widely used criterion, which contains both the precision rate and recall rate [42]. Therefore, accuracy and F-measure are applied to demonstrate the effectiveness of the proposed method, as shown in Fig. 15.

Based on these results, it can be concluded that:

1) It is shown that the accuracies of deep learning methods are significantly higher than the traditional methods in raw vibration signal. This can be explained that deep learning method can adaptively learn the valuable information from the raw vibration signal. The results also indicate that the accuracy of BPNN and SVM increase greatly after feature extraction, but their results still lower than the proposed method.
2) Compared with 1-D CNN or GRU, the proposed method has higher accuracy and F-measure score. This result indicates that the proposed method can improve the fault diagnosis performance by combining the speed and lightness of convnets with the order-sensitivity of GRU.
3) The accuracy and F-measure score of the proposed method are superior to the other methods in both the binary and multiclass classification. The results can be proved the effectiveness of the proposed method.

4) Experiment 4: Stability Evaluation: In experiment 4, the K-fold cross-validation [43] is applied to verify the stability of the proposed method. In this study, the K value is set to 10, which is a typical 10-fold cross-validation. Ten trials were conducted to compare the diagnosis performance in different datasets using different classification methods. The accuracy of each trail is shown in Fig. 16 and the average accuracy and standard deviation are listed in Table X. Then, the average accuracy and standard deviation are applied to demonstrate the stability of the proposed method.

It can be observed from Table X that the average accuracy and the standard deviation of the proposed method are higher and smaller than other methods in the case of most trails, respectively. Therefore, these results indicate that the stability of the proposed method is superior to the other methods.

TABLE X

AVERAGE ACCURACY AND STANDARD DEVIATION VALUES OF VARIOUS MACHINE LEARNING WITH DIFFERENT DATASET FOR 10 TRIALS

Methods	Average accuracy (Measured)	Standard Deviation (Measured)	Average accuracy (simulation)	Standard Deviation (simulation)
The proposed method	**98.70%**	**0.996**	**99.21%**	**0.647**
1-D CNN	90.38%	1.919	95.71%	1.404
GRU	94.08%	1.873	94.40%	1.529
BPNN with raw data	63.33%	3.230	64.04%	2.481
BPNN with FD features	89.72%	2.068	91.83%	1.592
SVM with raw data	65.18%	2.664	66.28%	2.299
SVM with FD features	92.22%	1.525	94.49%	1.152

From the results mentioned above, it can be demonstrated that the proposed method is more effective and robust for the binary and multi-diagnosis of HGU than other methods.

C. Engineering Practice

At present, the online acquisition system has been installed in four small hydros. The fault diagnosis process in the real industrial environment as follows: First, a vibration signal is randomly selected as test data from data server per hour. Then, this test data will be adopted to the fault diagnosis system consisting of normalized, data segmentation and 1-D CNN-GRU model. Finally, the diagnosis result is recorded on the web browser. Furthermore, if the diagnosis result is judged to abnormal and then this result can be reported to the center console. The diagnostic process based on practical vibration data is shown in Fig. 17.

At present, the proposed method has achieved good results in four small hydro operation monitoring. On September 28, 2018, the fault diagnosis system detected a failure of one of the units in the Xianghu hydropower station. Then multiclass classification algorithm is applied to diagnose the fault type as thrust bearing unevenness. After the on-site evaluation, the fault did occur, but the fault type was that the excessive clearance of bearing bush. That is to say, the binary classification algorithm identified correctly, but the multi-classification algorithm has a misjudgment. The reason is that this is a new type of fault for the sample library, and the fault characteristics of it are similar to the thrust bearing

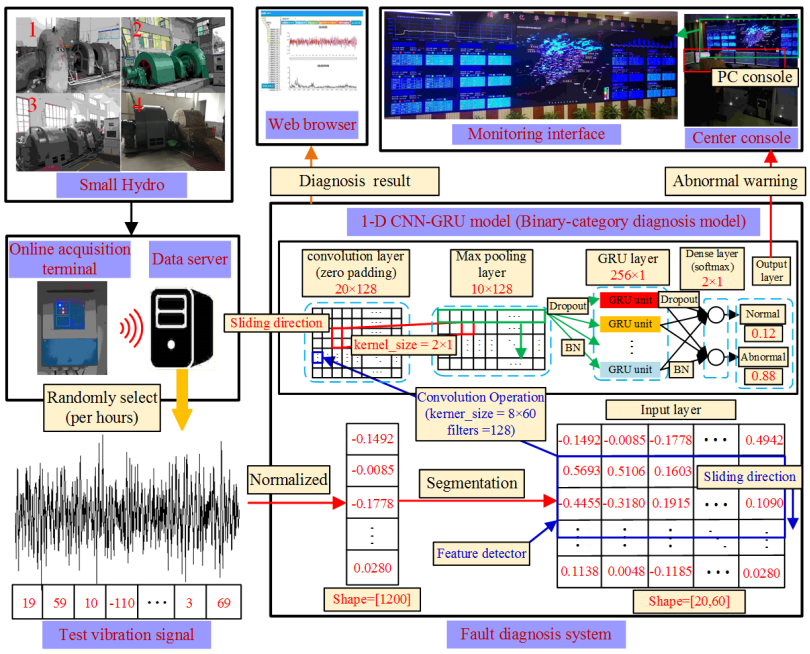


Fig. 17. Fault diagnostic system based on measured vibration data.

unevenness. After updating the fault samples and retraining the model, the new model also shows strong identification accuracy for the new fault.

The significance of this study is that accurate diagnosis can be informed to operators and maintenance personnel, which is very helpful for unattended hydropower systems. In practical applications, as the number of fault samples increases, the fault database will be gradually improved. When a new fault occurs, the diagnostic model will be retrained, the cycle will be repeated, and the diagnostic model is continuously improved. Ultimately, the accuracy of the classification will gradually increase.

V. Conclusion

In this paper, a novel HGU fault diagnosis system based on 1-D CNN-GRU is proposed. Firstly, the measured vibration signal under different operational conditions is collected by the online acquisition system, and then the numerical simulation model is used to construct various fault data. Secondly, the raw vibration signal is reconstructed by using the data segmentation method. Finally, the network of 1-D CNN-GRU is used to improve the performance of fault diagnosis.

The experiment results confirm that the proposed method is more effective and robust than the other machine methods in HGU fault diagnosis. Furthermore, the binary fault diagnosis model has successfully applied in real industrial environment. Based on above analysis, the proposed method can be applied on any real systems as a fault diagnosis system based on sequence data. Absolutely, the proposed method has capability to apply on hydroelectricity. The reason is as follows: 1. Since the physical structure of hydroelectricity is similar to that of small hydro, its vibration frequency distribution is also regular; 2. The proposed method can simultaneously implement automatically feature extraction from vibration signal and fault diagnosis, which does not need to rely on rich engineering experience of diagnostic experts and prior knowledge. However, if our work is applied to hydroelectricity, the sensor parameters of the online acquisition system need to be re-selected. Moreover, the vibration signals of hydroelectricity need to be collected to retrain the network model. The authors would research this topic in the further.

References

[1] T. B. Couto and J. D. Olden, "Global proliferation of small hydropower plants–science and policy," *Front. Ecol. Environ.*, vol. 16, no. 2, pp. 91–100, Mar. 2018.

[2] X. Sun, X. Wang, L. Liu, and R. Fu, "Development and present situation of hydropower in China," *Water Policy*, vol. 21, no. 3, pp. 565–581, 2019. doi: 10.2166/wp.2019.206.

[3] M. H. Wang, "Application of extension theory to vibration fault diagnosis of generator sets," *IEE Proc.-Gener., Transmiss. Distrib.*, vol. 151, no. 4, pp. 503–508, Jul. 2004.

[4] R. Saeed, A. Galybin, and V. Popov, "3D fluid-structure modelling and vibration analysis for fault diagnosis of Francis turbine using multiple ANN and multiple ANFIS," *Mech. Syst. Signal Proc.*, vol. 34, nos. 1–2, pp. 259–276, Jan. 2013.

[5] E. A. Simeon, A. J. Áivares, and R. R. Gudwin, "An expert system for fault diagnostics in condition based maintenance," in *Proc. Int. Congr. Mech. Eng.*, vol. 4, Nov. 2009, pp. 304–313.

[6] W. Zhu *et al.*, "A novel KICA–PCA fault detection model for condition process of hydroelectric generating unit," *Measurement*, vol. 58, pp. 197–206, Dec. 2014.

[7] S. Lu, J. Wang, and Y. Xue, "Study on multi-fractal fault diagnosis based on EMD fusion in hydraulic engineering," *Appl. Therm. Eng.*, vol. 103, pp. 798–806, Jun. 2016.

[8] C. Fu *et al.*, "Predictive maintenance in intelligent-control-maintenance-management system for hydroelectric generating unit," *IEEE Trans. Energy Convers.*, vol. 19, no. 1, pp. 179–186, Mar. 2004.

[9] S. Yan, J. Zhou, Y. Zheng, and C. Li, "An improved hybrid backtracking search algorithm based T–S fuzzy model and its implementation to hydroelectric generating units," *Neurocomputing*, vol. 275, pp. 2066–2079, Jan. 2018.

[10] L. Selak, P. Butala, and S. Alojzij, "Condition monitoring and fault diagnostics for hydropower plants," *Comput. Ind.*, vol. 65, no. 6, pp. 924–936, Aug. 2014.

[11] K. Li, P. Chen, and H. Q. Wang, "Intelligent diagnosis method for rotating machinery using wavelet transform and ant colony optimization," *IEEE Sensors J.*, vol. 12, no. 7, pp. 2474–2484, Jul. 2012.

[12] J. Zhu, Z. Ge, and Z. Song, "HMM-driven robust probabilistic principal component analyzer for dynamic process fault classification," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3814–3821, Jun. 2015.
[13] O. A. Omitaomu, V. A. Protopopescu, and A. R. Ganguly, "Empirical mode decomposition technique with conditional mutual information for denoising operational sensor data," *IEEE Sensors J.*, vol. 11, no. 10, pp. 2565–2575, Oct. 2011.
[14] Q. Fu, B. Jing, P. He, S. Si, and Y. Wang, "Fault feature selection and diagnosis of rolling bearings based on EEMD and optimized Elman_AdaBoost algorithm," *IEEE Sensors J.*, vol. 18, no. 12, pp. 5024–5034, Jun. 2018.
[15] R. Abdelkader, A. Kaddour, A. Bendiabdellah, and Z. Derouiche, "Rolling bearing fault diagnosis based on an improved denoising method using the complete ensemble empirical mode decomposition and the optimized thresholding operation," *IEEE Sensors J.*, vol. 18, no. 17, pp. 7166–7172, Sep. 2018.
[16] Z. Wang *et al.*, "Application of an improved ensemble local mean decomposition method for gearbox composite fault diagnosis," *Complexity*, vol. 2019, May 2019, Art. no. 1564243. doi: 10.1155/2019/1564243.
[17] J. M. Lee, S. J. Qin, and I. B. Lee, "Fault detection of non-linear processes using kernel independent component analysis," *Can. J. Chem. Eng.*, vol. 85, no. 4, pp. 526–536, Aug. 2007.
[18] Z. Wang *et al.*, "A novel method for intelligent fault diagnosis of bearing based on capsule neural network," *Complexity*, vol. 2018, Jun. 2019, Art. no. 6943234. doi: 10.1155/2019/6943234.
[19] I. Ahmed, A. Dagnino, and Y. Ding, "Unsupervised anomaly detection based on minimum spanning tree approximated distance measures and its application to hydropower turbines," *IEEE Trans. Autom. Sci. Eng.*, vol. 16, no. 2, pp. 654–667, Apr. 2019.
[20] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
[21] H. Pan, X. He, S. Tang, and F. Meng, "An improved bearing fault diagnosis method using one-dimensional CNN and LSTM," *J. Mech. Eng.*, vol. 64, nos. 7–8, pp. 443–452, May 2018.
[22] H. Liu, J. Zhou, Y. Zheng, W. Jiang, and Y. Zhang, "Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders," *ISA Trans.*, vol. 77, pp. 167–178, Jun. 2018.
[23] H. Shao, H. Jiang, H. Zhang, W. Duan, T. Liang, and S. Wu, "Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing," *Mech. Syst. Signal Process.*, vol. 100, pp. 743–765, Feb. 2018.
[24] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, "Real-time motor fault detection by 1-D convolutional neural networks," *IEEE Trans. Ind. Electron.*, vol. 63, no. 11, pp. 7067–7075, Jun. 2016.
[25] X. Li, J. Li, Y. Qu, and D. He, "Gear pitting fault diagnosis using integrated CNN and GRU network with both vibration and acoustic emission signals," *Appl. Sci.*, vol. 9, no. 4, p. 768, Feb. 2019.
[26] D. N. Konidaris and J. A. Tegopoulos, "Investigation of oscillatory problems of hydraulic generating units equipped with Francis turbines," *IEEE Trans. Energy Convers.*, vol. 12, no. 4, pp. 419–425, Dec. 1997.
[27] X. L. An, J. Z. Zhou, L. Liu, J. J. Yang, C. S. Li, and X. Q. Xiang, "Vibration fault diagnosis for hydraulic generator units with pattern recognition and cluster analysis," in *Proc. 4th Int. Conf. Wireless Commun., Netw. Mobile Comput.*, Nov. 2008, pp. 1–4.
[28] C. Liu, B. Lung, L. Ye, and G. Morel, "An innovative hydraulic turbine governor for predictive maintenance of hydropower plant," in *Proc. 41st IEEE Conf. Decis. Control*, Dec. 2002, pp. 10–13.
[29] M. Seera, M. L. D. Wong, and A. K. Nandi, "Classification of ball bearing faults using a hybrid intelligent model," *Appl. Soft Comput.*, vol. 57, pp. 427–435, Aug. 2017.
[30] X. An and J. Yang, "A method of eliminating the vibration signal noise of hydropower unit based on NA-MEMD and approximate entropy," *Proc. Inst. Mech. Eng., E, J. Process Mech. Eng.*, vol. 231, no. 2, pp. 317–328, Apr. 2017.
[31] R. K. Monhanta, T. R. Chelliah, S. Allamsetty, A. Akula, and R. Ghosh, "Sources of vibration and their treatment in hydro power stations—A review," *Eng. Sci. Technol. Int. J.*, vol. 20, no. 2, pp. 637–648, Apr. 2017.
[32] A. J. Jerri, "The Shannon sampling theorem—Its various extensions and applications: A tutorial review," *Proc. IEEE*, vol. 65, no. 11, pp. 1565–1596, Nov. 1977.
[33] M. F. Guo, X. D. Zeng, D. Y. Chen, and N. C. Yang, "Deep-learning-based earth fault detection using continuous wavelet transform and convolutional neural network in resonant grounding distribution systems," *IEEE Sensors J.*, vol. 18, no. 3, pp. 1291–1300, Feb. 2017.
[34] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," *J. Mach. Learn. Res.*, vol. 15, no. 4, pp. 315–323, 2011.
[35] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1026–1034.
[36] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen, and J. Wang, "Machine health monitoring using local feature-based gated recurrent unit networks," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1539–1548, Feb. 2018.
[37] S. Han, J. Pool, J. Tran, and W. Dally, "Learning both weights and connections for efficient neural network," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, Dec. 2015, pp. 1135–1143.
[38] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, May 2015, pp. 1–15.
[39] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, Jul. 2015, pp. 448–456.
[40] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, Jun. 2014.
[41] X. An and J. Yang, "Analysis of hydropower unit vibration signals based on variational mode decomposition," *J. Vib. Control*, vol. 23, no. 12, pp. 1938–1953, Jul. 2017.
[42] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Trans. Ind. Electron.*, vol. 63, no. 5, pp. 3137–3147, May 2016.
[43] H. Shao, H. Jiang, H. Zhang, and T. Liang, "Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network," *IEEE Trans. Ind. Electron.*, vol. 65, no. 3, pp. 2727–2736, Mar. 2018.

Guo-Ping Liao was born in Quanzhou, China, in 1995. He received the B.S. degree in electrical engineering from Minnan Normal University, Zhangzhou, China, in 2017. He is currently pursuing the M.S. degree in electrical engineering with Fuzhou University.

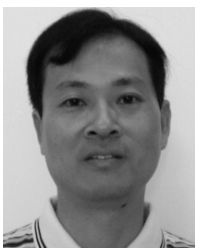
He research interests include advanced signal processing algorithms and hydroelectric generating unit fault diagnostics and its fault forecast.

Wei Gao was born in Fuzhou, China, in 1983. He received the B.S. and M.S. degrees from Fuzhou University, Fuzhou, China, in 2005 and 2008, respectively. He is currently pursuing the Ph.D. degree with the National Taiwan University of Science and Technology, Taiwan.

Since 2008, he has been with Fuzhou University, where he is currently a Lecturer of Electric Power Engineering. His research interests include renewable energy generation technology and application of artificial intelligence in condition-based maintenance of power systems.

Geng-Jie Yang was born in Wuyishan, China, in 1966. He received the B.S. and M.S. degrees in electrical engineering from Fuzhou University, Fuzhou, China, in 1985 and 1988, respectively.

Since 1988, he has been with Fuzhou University, where he is currently a Professor with the Department of Electric Power Engineering. His research interests include power system analysis and control.



Mou-Fa Guo (M'17) was born in Fuzhou, China, in 1973. He received the B.S. and M.S. degrees from Fuzhou University, Fuzhou, in 1996 and 1999, respectively, and the Ph.D. degree from Yuan Ze University, Taiwan, in 2018, all in electrical engineering.

Since 2000, he has been with Fuzhou University, where he is currently a Professor, and the Chairman of the Department of Electric Power Engineering. His research interests include power distribution systems and its automation, and application of artificial intelligence in power distribution systems.