Fault Diagnosis of Hydraulic Systems Based on Deep Learning Model With Multirate Data Samples

Keke Huang, *Member, IEEE*, Shujie Wu, Fanbiao Li, *Member, IEEE*, Chunhua Yang, *Senior Member, IEEE*, and Weihua Gui

***Abstract*— Hydraulic systems are a class of typical complex nonlinear systems, which have been widely used in manufacturing, metallurgy, energy, and other industries. Nowadays, the intelligent fault diagnosis problem of hydraulic systems has received increasing attention for it can increase operational safety and reliability, reduce maintenance cost, and improve productivity. However, because of the high nonlinear and strong fault concealment, the fault diagnosis of hydraulic systems is still a challenging task. Besides, the data samples collected from the hydraulic system are always in different sampling rates, and the coupling relationship between the components brings difficulties to accurate data acquisition. To solve the above issues, a deep learning model with multirate data samples is proposed in this article, which can extract features from the multirate sampling data automatically without expertise, thus it is more suitable in the industrial situation. Experiment results demonstrate that the proposed method achieves high diagnostic and fault pattern recognition accuracy even when the imbalance degree of sample data is as large as 1:100. Moreover, the proposed method can increase about 10% diagnosis accuracy when compared with some state-of-the-art methods.**

***Index Terms*— Convolutional neural network (CNN), deep learning (DL), fault diagnosis, hydraulic system, multirate data samples.**

I. Introduction

HYDRAULIC systems are essential in various fields such as manufacturing, metallurgy industry, and energy. In light of some advantages such as high stiffness, high precision, fast response, large driving force, and wide speed range, hydraulic systems are widely adopted as the core component of engineering equipment. However, hydraulic systems are prone to failure because of the harsh environments and unstable working conditions. Furthermore, with the increase of the scale and automation level of hydraulic systems, their operational stability and reliability also need to be improved. Accurate detection and timely treatment of emerging faults can prevent accidents, reduce maintenance costs, and improve productivity [1]–[3]. Therefore, the fault diagnosis of hydraulic systems has important practical significance, and it has received increasing attention.

Nowadays, several studies on fault diagnosis of hydraulic systems have been conducted and many effective algorithms have been developed. Helwig *et al.* [4] developed a condition monitoring system for hydraulic systems based on statistical data evaluation and supervised classification. An effective time-series classification method has been proposed in [5], which employs the Shapelet transformation technique for predictive maintenance of hydraulic systems. Along another technical route, a fault diagnosis method for the hydraulic directional valves has been proposed in [6], which combines the principal component analysis (PCA) with the extreme gradient boosting (XGBoost) algorithm to improve the fault diagnosis accuracy. The methods mentioned above are shallow machine learning (SML) methods. Generally, the SML methods include following steps [7]:

1) First, preprocessing the monitoring data of the system.
2) Then, extracting and selecting statistical features that represent system characteristics.
3) Finally, building a fault diagnosis model based on the obtained features and conduct fault diagnosis.

Although the SML methods have achieved good diagnosis performance, they still suffer from some deficiencies. First, manual feature extraction relies on expert knowledge, so it is subjective and time-consuming to select the most sensitive features in different diagnostic problems, especially in the case of insufficient expertise [8], [9]. Besides, because feature extraction schemes are specific to different monitoring systems, they need to be redesigned for a new system. Second, because the feature extraction and model training are independent in some traditional SML methods, it is difficult to optimize them jointly [10]. Last, because of the shallow architecture of SML methods, they are difficult to mine the essential features of raw data.

Deep learning (DL), as a branch of machine learning theory, can provide an effective way to overcome the above-mentioned shortcomings [11]–[13]. DL can automatically mine the hidden

Manuscript received 21 October 2020; revised 27 February 2021 and 6 April 2021; accepted 20 May 2021. Date of publication 10 June 2021; date of current version 28 October 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 62073340, Grant 61973319, and Grant 61860206014; in part by the National Key Research and Development Program of China under Grant 2019YFB1705300; in part by the Innovation-Driven Plan in Central South University, China, under Grant 2019CX020; in part by the Excellent Youth Natural Science Foundation of Hunan Province under Grant 2019JJ30032; in part by the State Key Laboratory of Robotics and Systems (HIT) under Grant SKLRS-2020-KF-14; and in part by the 111 Project, China, under Grant B17048. *(Corresponding author: Fanbiao Li.)*

The authors are with the School of Automation, Central South University, Changsha 410083, China (e-mail: huangkeke@csu.edu.cn; wushujie@csu.edu.cn; fanbiaoli@csu.edu.cn; ychh@csu.edu.cn; gwh@csu.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TNNLS.2021.3083401.

Digital Object Identifier 10.1109/TNNLS.2021.3083401

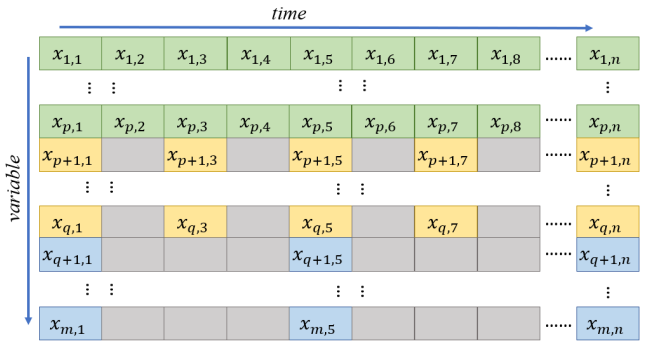

Fig. 1. Multirate data samples.

feature information without expert knowledge, and adaptively integrate feature learning and target prediction into the whole neural network (NN) to optimize them jointly. As one of the most effective algorithms in DL frameworks, convolutional NN (CNN) has achieved remarkable results in multiple detection and recognition problems [10]. Local receptive fields, shared weights, and spatial subsampling endows CNN with the ability to maintain the initial feature despite the shift, scale, and distortion invariance [14], [15]. Based on the characteristics of the industrial vibration signal, a misplaced time-series CNN is proposed in [16]. Wen *et al.* [17] converted signals into 2-D images and proposed a 2-D CNN for fault diagnosis. A general end-to-end monitoring framework based on CNN is introduced in [18] to perform fault diagnosis in different scenarios. Although the above works have achieved good results, they are limited to processing data from a single sensor, so they may not yield good diagnostic performances for complex systems with multisensor data.

Multisensor information fusion technology is the development trend of fault diagnosis, which can produce more comprehensive information than a single sensor to improve the robustness of fault diagnosis. Multisensor information fusion can take place at different levels: data level, feature level, and decision level. Stief *et al.* [19] proposed a sensor fusion approach combining acoustic, electric, and vibration signals to diagnose electrical and mechanical faults in induction motors. Based on the multiscale analysis of motor vibration and stator current signals, Wang *et al.* [20] proposed a multiresolution and multisensor fusion network for motor fault diagnosis. Combining vibration, acoustic, current, and instantaneous angular speed signals, Jing *et al.* [21] proposed a CNN-based method for fault diagnosis of the planetary gearbox. Most of the DL models are aimed at the data of a single sampling rate, whereas in practical applications, the sampling rate of monitoring data of hydraulic systems is sometimes different.

Fig. 1 shows the structure of multirate data samples, where gray squares represent uncollected data. The multirate data samples have the characteristics of incompleteness, regularity, and information asymmetry. Incompleteness reflects in the missing values of the variables with low sampling rate, which are represented by the gray squares in Fig. 1. Regularity means that the variables at each sampling rate are uniform and complete. Information asymmetry indicates that variables with different sampling rates contain asymmetric information. For example, in the industrial process, high sampling rate variables are mostly process variables, which generally contain limited process information, whereas low sampling rate variables are mostly quality-related variables, and the information contained in them may be more valuable [22]. At present, there have been many studies using multirate data samples for condition monitoring and fault detection. These methods can be divided into the down-sampling and up-sampling methods. The down-sampling methods subsample variables with higher sampling rates so that all variables obey the same sampling rate. The up-sampling methods use high sampling rate variables to predict the missing values of low sampling rate variables. A multivariate statistical process control method using soft sensors is proposed in [23], which uses soft sensors to predict variables with low sampling rates. Under multirate sampling conditions, Wu and Luo [24] introduced data fusion technology based on Kalman filter into soft sensor maintenance. Shao *et al.* [25] proposed a Bayesian framework to predict the posterior distribution of unknown states in the observations with low sampling rates. However, both the approaches have inevitable drawbacks. In the down-sampling methods, the information loss caused by the reduction of sampling rate may bring adverse effects, especially when the sampling interval is large. Although in the up-sampling methods, the performance of diagnosis depends on the accuracy of the prediction model, the increase of sampling interval and variables will decrease the prediction accuracy. Therefore, new methods need to be developed to process multirate data samples.

In addition to the feature of multiple sampling rates, hydraulic systems have the characteristics of strong fault concealment, highly complex structure, and strong nonlinear time-varying signal, which increases the difficulty of fault diagnosis. Different components of hydraulic systems may fail individually or simultaneously, and it will be more difficult to diagnose multiple components. To the best of our knowledge, the investigation of fault diagnosis of hydraulic systems using multirate data samples is still lacking.

To solve the problem caused by multitimescale properties and synthetic complexity of hydraulic systems data, this work will propose an intelligent fault diagnosis method of hydraulic systems using a DL model with multirate data sample. The main contributions of this work are highlighted as follows:

1) An end-to-end diagnostic method is proposed to process multirate data samples automatically. This method can process data with multiple sampling rates and a high sampling rate ratio.
2) The proposed DL framework can extract features of multirate data samples automatically. Besides, it can extract hidden information that cannot be captured by statistical features, such as features under noise environments.
3) Aiming at the characteristics of complex structure and fault concealment of hydraulic systems, the proposed method can effectively identify the single fault and multiple faults of different components. Compared with some state-of-the-art methods, it has a better fault diagnosis performance.

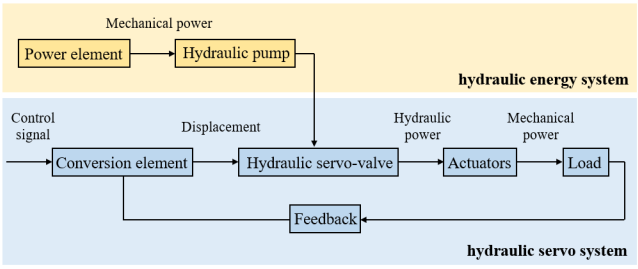

Fig. 2. Schematic of hydraulic systems.

The rest of this article is arranged as follows: Section II details the mechanism of hydraulic systems. A detailed view of the proposed intelligent hydraulic system fault diagnosis method is described in Section III. In Section IV, the fault diagnosis of hydraulic systems is conducted to demonstrate the effectiveness of the proposed method. Finally, Section V provides the conclusion.

II. Analysis of Hydraulic Systems

As a typical hydraulic coupling system, the main function of hydraulic systems is transferring mechanical power to realize mechanical work. Notably, a hydraulic system has coupling characteristics of mechanical, electrical, and hydraulic. As shown in Fig. 2, a hydraulic system is composed of a hydraulic energy and a hydraulic servo system. The hydraulic energy system provides the energy needed to run the whole system. The hydraulic pump converts the mechanical energy transferred by the power element into the pressure energy of the liquid. It is the core element of the hydraulic energy system, and its failure will cause the whole system to break down. Therefore, the fault diagnosis of the hydraulic pump is important for the smooth operation of the hydraulic energy system. Hydraulic servo system can guarantee that the output, such as displacement, speed, or force, automatically follows the changes in the input. The common failure of the hydraulic valve includes nozzle blockage, slide valve sticking, switch performance degradation, and so on [26].

The test bed of a hydraulic project [4] is made up of two hydraulic circuits that are connected by an oil tank, as shown in Fig. 3: one is a primary working circuit [Fig. 3(a)] and the other is a secondary cooling-filtration circuit [Fig. 3(b)]. The red labels represent the key elements of the hydraulic system, such as accumulators, pumps, and the filter, while the blue ones represent the sensors, such as vibration, pressure, and temperature sensors. The primary working circuit consists of main pump MP1, switchable accumulators A1–A4, filter F1, and so on. The proportional pressure relief valve (V11) is utilized to generate different load levels in the working circuit with the main pump (electrical motor power 3.3 kW). The secondary cooling-filtration circuit, which is responsible for cooling and filtering the hydraulic oil, consists of hydraulic pump SP1, a three-way solenoid valve, filter F2, cooler C1, and a variety of sensors. The particle metallic contamination sensor (MCS) detects the pollution degree of hydraulic oil, and the position of the solenoid valve spool is determined according to the result. In Fig. 3, the solenoid valve spool connects to the first interface and the hydraulic oil flow passes through the filter to the condenser. If the solenoid valve spool connects to another interface, the hydraulic oil will flow to the condenser without filtration. Hereafter, the hydraulic project was used as the benchmark to verify the superiority of the proposed method.

TABLE I
Components and Simulated Fault Conditions

Component	Condition	Value	Interpretation
Cooler C1	Cooling power decrease (%)	100	full efficiency
		20	reduced efficiency
		3	close to total failure
Valve V10	Switching degradation (%)	100	optimal switching behavior
		90	small lag
		80	severe lag
		73	close to total failure
Pump MP1	Internal leakage (code)	2	severe leakage
		1	weak leakage
		0	no leakage
Accumulator A1-A4	Gas leakage (bar)	130	optimal pressure
		115	slightly reduced pressure
		100	severely reduced pressure
		90	close to total failure

Several kinds of faults may occur in the hydraulic system, which differ in fault types, severity, and duration. The components and simulated fault conditions are presented in Table I. In cooler C1, different cooling power (CP) degradation degrees are simulated by changing the cooler fan duty cycle. In valve V10, different switching degradation states are simulated by changing the valve current set points. The internal leakage levels of main pump MP1 are caused by cascading three 0.2- and 0.25-mm orifices. Different leakage levels are simulated by changing the precharge pressure steps of accumulators A1–A4. The breakdown performance of different components varies. For example, the fault characteristics of CP reduction and valve switching degradation are obvious, thus the diagnosis of these components is relatively easy. However, the fault characteristics of the hydraulic pump and accumulator are concealed and susceptible to environmental noise. Therefore, it is necessary to adopt advanced diagnostic methods to identify the faults of these components. Besides, these samples are collected under different conditions. Some data samples were sampled under stable conditions (1449 samples), whereas others were recorded when static conditions might not have been reached yet (756 samples). Data from both conditions were used in the experiment, which may cause class imbalances and increase the difficulty of fault detection.

Several sensors are adopted to monitor the conditions of the hydraulic system. As summarized in Table II, the physical sensors include pressure sensors (PS1–PS6), electrical power sensor (EPS1), flow sensors (FS1, FS2), temperature sensors (TS1–TS5), and vibration sensor (VS1). Besides, several virtual sensors are used to provide references of the system,

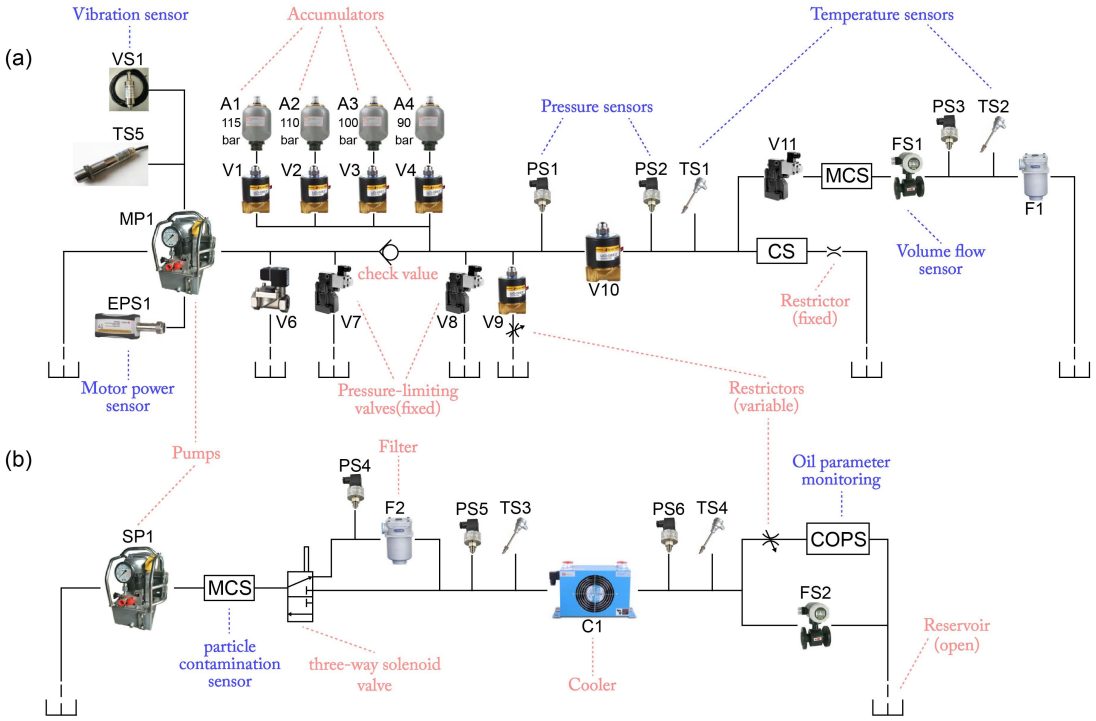

Fig. 3. Structure of hydraulic system: (a) primary working circuit and (b) secondary cooling-filtration circuit.

TABLE II
MONITORED PARAMETERS OF THE HYDRAULIC SYSTEM

Sensor	Physical quantity	Unit	Sampling rate
PS1-PS6	Pressure	bar	100 Hz
EPS1	Motor power	W	100 Hz
FS1-FS2	Volume flow	l/min	10 Hz
TS1-TS4	Temperature	$^{\circ}C$	1 Hz
VS1	Vibration	mm/s	1 Hz
CE	Cooling efficiency (virtual)	%	1 Hz
CP	Cooling power (virtual)	kW	1 Hz
SE	System efficiency factor (virtual)	%	1 Hz

such as cooling efficiency (CE), CP, and system efficiency factor (SE). The data are collected on a programmable logic controller (PLC) and then transferred to a computer through EtherCAT. The sampling rate of these sensors varies according to the dynamics of the physical qualities. For rapidly changing physical quantities, such as pressure, the sampling rate is 100 Hz. For physical quantities that change slowly, such as temperature, the sampling rate is 1 Hz. To the best of our knowledge, most studies only adopt the data with a sampling rate of 100 Hz to establish the diagnosis model. However, data with low sampling rates also contain tremendous fault information, which will improve the accuracy of fault diagnosis.

The fault diagnosis of hydraulic systems is challenging because it has the following characteristics. First, compared with general electrical systems and mechanical systems, hydraulic systems are airtight and complex, leading to strong fault concealment and vulnerability to interference. Second, the movement of the hydraulic components, such as the reciprocating motion of the cylinder, will generate a large number of excitation sources. Therefore, the monitoring data of hydraulic present nonlinear, time-varying, and impact characteristics [27]. Third, the mapping relationship between fault characteristics and fault cause is complicated. A failure may have multiple causes, so it is hard to locate the fault source. For example, both the oil seal failure and the heavy wear between the valve plate and the cylinder block will cause the hydraulic pump leakage. Finally, the composite failure will increase the difficulty of fault diagnosis. The failures of the hydraulic system may occur simultaneously or separately. When a composite failure occurs, signals transmitted through the complex channels of the hydraulic system may be disturbed, thus affecting the results of fault detection.

Next, an intelligent fault diagnosis method will be proposed for the synthetic complexity diagnosis of the hydraulic system. By integrating features extracted from the multirate data samples, the proposed method can effectively identify the single fault and multiple faults in the hydraulic system.

III. CNN-BASED FAULT DIAGNOSIS METHOD WITH MULTIRATE DATA SAMPLES

Multisensor data fusion is widely used in fault detection because it can fuse various sensors to improve diagnostic performance. However, because of the multitime scale characteristics of monitoring data, this method is not suitable for the hydraulic system. Therefore, it is necessary to process the multirate sampling data. However, the up-sampling and

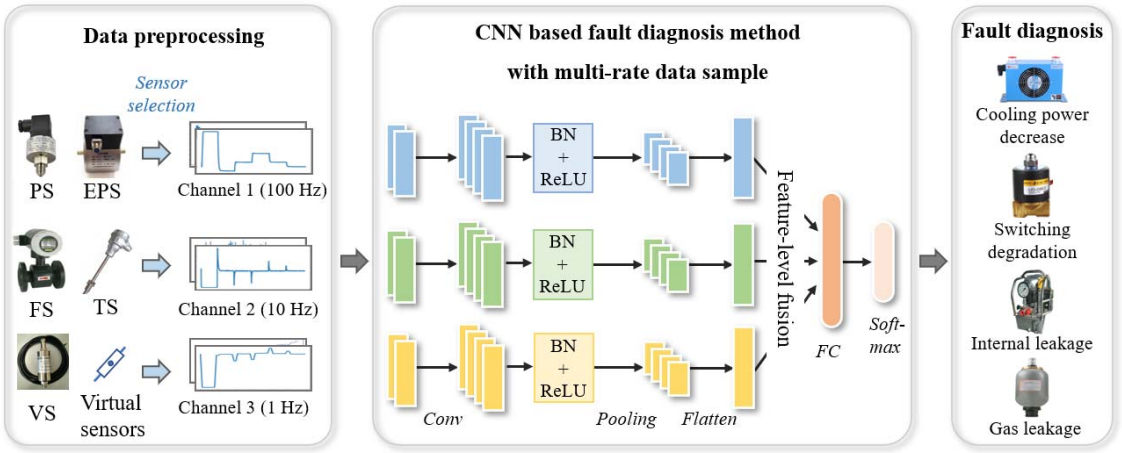

Fig. 4. Framework of the proposed method.

down-sampling methods are not suited to the current application scenario. For example, using down-sampling methods will reduce time resolution and affect fault diagnosis of the valve and accumulator whose detection performance depends on short-time characteristics during signal transients [28]. Given the above considerations, the proposed method introduces a multichannel structure for feature extraction of multirate sampling data and then performs feature-level fusion. This section introduces the details of the proposed CNN-based fault diagnosis method. In general, the proposed method is composed of several main processes: data preprocessing, model training, and fault diagnosis. The flowchart is illustrated in Fig. 4.

A. Data Preprocessing

First, considering that using all variables may introduce redundant information and lead to over-fitting, selecting input variables for the diagnosis model is required. Pearson correlation coefficient is used to measure the correlation between variables at the same sampling rate. The value of the Pearson correlation coefficient is between -1 and 1. The greater the absolute value, the stronger the correlation. Therefore, sensors with high absolute values of correlation coefficients are screened because they have similar information

$$\rho_{x_i,x_j} = \frac{\mathrm{cov}(x_i, x_j)}{\sigma_{x_i}\sigma_{x_j}} \tag{1}$$

where x_i and x_j represent the variables under the same sampling rates, $i, j = 1, 2, \ldots, i \neq j$. $\mathrm{cov}(x_i, x_j)$ is the covariance between the variables x_i and x_j, and σ_{x_i} and σ_{x_j} are the standard deviations of x_i and x_j, respectively.

Then, data samples from different sensors with the same sampling rate are normalized and concatenated at the last dimension to form the input of a channel. The input of channel r can be defined as follows:

$$X^r = \left[x_1^n, x_2^n, \ldots, x_M^n\right], \quad n = 1, 2, 3, \ldots, N \tag{2}$$

where M is the number of sensors under the same sampling rate and N represents the sampling number.

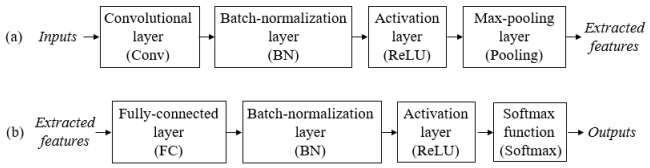

Fig. 5. Structure of proposed model: (a) feature extractor and (b) fault classifier.

B. Model Design

This section will introduce the proposed CNN-based fault diagnosis framework, which consists of the feature extractor and the fault classifier. The feature extractor contains multiple channels that can extract features from raw data with multiple sampling rate characteristics. Then these features are concatenated to form the final representation. Last, the fault classifier can recognize the faults based on these representations.

CNN is chosen as the diagnosis model for the following reasons. First, CNN featured by local connection and weight sharing. Therefore, it has fewer trainable parameters and faster training speed than the multilevel perceptual classifier. Second, most industrial monitoring data have local correlation, and the filter of CNN can learn and characterize local relevant information, so it is suitable for industrial fault detection. Last, the industrial environment is often contaminated by noise. CNN can extract translation-invariant features to enhance the robustness and reduce the adverse effects of noise.

1) Feature Extractor: The feature extractor consists of several channels, and each one extracts the features of a single sampling rate data. Different channels have the same structure but different parameters. As shown in Fig. 5(a), they consist of the convolutional layer (Conv), rectified linear unit (ReLU) activation layer (ReLU), batch-normalization (BN) layer, and max-pooling layer (Pooling).

a) Convolutional layer: The convolutional layer is the core element of CNN, which consists of several filters with weighted parameters. Compared with traditional NN that uses different parameters, each filter of CNN shares the same

weights and bias. These filters will convolve with the input to extract translation-invariant features. The design of the convolutional layers should match the characteristics of the underlying data. Therefore, the parameters need to be set according to the following principles. First, the number of filters needs to be proportional to the number of sensors and adjusted according to the different fault detection objectives. In general, the deeper the convolutional layers are, the better the feature extraction ability of the model. For example, in the fault detection of hydraulic pumps and accumulators, the number of filters is set more than that of coolers and valves, because the former is more complicated. Second, the size of the filters is set according to the dimension of the time series. The large filter is used in the first convolutional layer because it has a wide reception area, which allows extracting low-frequency features, such as periodical changes in the signal. The large filters can also suppress high-frequency noise, which is of great significance to the fault detection of practical industrial systems. Besides, smaller convolution filters are needed in the later convolution layers to enable the model to capture local features [29].

The output of the convolutional layer can be defined as

$$C^q = \sum_{p=1}^{P} W^{p,q} * X^p + b^q \tag{3}$$

where C^q is the feature extracted by the convolutional layer, $W^{p,q}$ and b^q denotes the weight and bias of the filter, $p = 1, 2, \ldots, P$, P denotes the number of convolutional channels, $q = 1, 2, \ldots, Q$, Q is the predefined number of the filters. The symbol $*$ is the convolutional operator. The zero-padding method is used to prevent dimension loss of the features.

b) Batch normalization layer: Unlike most models that use activation functions directly on the convolution layer output, the proposed model inputs the extracted features into a batch normalization layer. The batch normalization layer can prevent over-fitting to a certain extent and make the optimization process more smooth. In the training process, as the parameters of the previous layer change, the input distribution of each layer will be different, which makes the deep NN difficult to train. Such a phenomenon is referred to as internal covariate shift. To solve this problem, batch normalization is proposed [30], [31]. Through the normalization of each training batch, the variation of input distribution at each layer can be controlled, and the drift of internal covariable can be reduced. Besides, batch normalization can reparametrize the optimization process to make it more stable and smooth. This means that the gradients used in training will be more predictive, which can accelerate the training process.

Supposing c is the mini-batch of feature C, and $c = c_1, c_2, \ldots, c_t$, then the output h_i of the batch normalization layer can be calculated as

$$\hat{c_i} = \frac{c_i - \mu}{\sqrt{\sigma^2 + \varepsilon}} \tag{4}$$

$$h_i = \gamma \hat{c_i} + \beta \tag{5}$$

where μ and σ^2 are mean and variance of the mini-batch, $\mu = (1/t)\sum_{i=1}^{t} c_i$, $\sigma^2 = (1/t)\sum_{i=1}^{t}(c_i - \mu)^2$. ε is a minimum close to 0, which ensures that the denominator of formula (4) is not 0. Symbols γ and β denote learned parameters.

By using batch normalization, the mini-batch data will affect the parameters γ and β. This function is equivalent to introducing disturbances and noise into training samples, which can be regarded as data augmentation and a method to avoid over-fitting [32].

c) Activation layer: After the batch normalization operation, the results are sent into a nonlinear activation layer, which can learn complex and nonlinear functions. By adding nonlinear activation operations, the network can obtain the nonlinear representation of the features, thus enhancing the representation and recognition ability of the learned features. Without activation function, the model output is just a simple linear function and unable to learn and simulate other complex data, such as images, video, audio, and so on.

Considering the convergence speed as well as the gradient disappearance problem, the ReLUs function is employed as an activation function. Compared with the Sigmoid and tanh activation function, the ReLU function has a faster convergence rate, and thus it can prevent gradient saturation when the model is deep. The ReLU function is defined as follows:

$$T^r = \mathrm{ReLU}(H^r) = \max(H^r, 0) \tag{6}$$

where H^r denotes the output of the batch normalization layer of channel r and T^r is the result of the activation layer.

d) Max-pooling layer: A pooling layer is adopted to handle the output of ReLU activation, which can subsample the outputs of adjacent unit groups in the same feature map to extract the most representative features. The use of the pooling layer will reduce the feature dimension and network parameters, improve the robustness of feature extraction, and accelerate calculation. The max-pooling layer is the most commonly used pooling layer and adopted in this work, which performs a local max operation on the input features. The function of the max-pooling layer can be described as follows, where S denotes the scope of the pooling function:

$$Z^r = \max_{T^r \in S} T^r. \tag{7}$$

Through the feature extractor, the model can automatically learn features from the multirate data samples, thus avoiding manual feature extraction that requires expert knowledge.

2) Modal Recognition and Fault Classification: After feature extraction, the extracted features need to be fused at the feature level to form the global representation. Then, they will be sent into the fault classifier. As seen in Fig. 5(b), the fault classifier consists of the fully connected layer, batch normalization layer, activation layer, and Softmax function.

a) Feature-level fusion and fully connected layer: To form the global representation that incorporates the features learning from multirate data samples, the features learned from all channels are fused. Because the extracted features are 2-D and the full-connected layer can only process 1-D data, the extracted features need to be flattened first. After feature-level fusion, the global representations are fed into a fully connected layer, which is responsible for mapping the multidimensional input to the low-dimensional data. The BN

layer and ReLU activation layer are also adopted to achieve better detection performance. The procedure is defined as follows:

$$Z = \left[Z^1, Z^2, \ldots, Z^r, \ldots, Z^R\right] \tag{8}$$

$$G = \mathrm{ReLU}(\mathrm{BN}(W_g Z + b_g)) \tag{9}$$

where Z^r represent the extracted features of channel r after the flattening process, W_g and b_g denote the weight and bias of the fully-connected layer. ReLU and BN represent the ReLU activation layer and BN layer, respectively. Because of the powerful feature extraction ability of CNN, the redundant information has been well-managed, therefore the global information can be learned by the above steps. In some complex fault detection problems, increasing the number of fully connected layers will improve the diagnosis model.

b) Softmax function: Finally, the Softmax function is used in the proposed model, which can transform the logits of the neurons to make them conform to the probability distribution of the final targets. The output of the Softmax function is a vector in which each number represents the target probability. The Softmax function is defined as follows:

$$O_j = P(y = j \mid G; \theta) = \frac{\exp\left(\theta^{(j)} G\right)}{\sum_{k=1}^{K} \exp\left(\theta^{(k)} G\right)}, \quad j = 1, \ldots, K \tag{10}$$

where G is the input, O_j denotes the estimated probability of category j, $\theta^{(j)}$ are the parameters of the Softmax function, and K is the number of the target categories.

This method can not only avoid the above-mentioned shortcomings of the up-sampling and down-sampling methods, but also effectively fuse the information contained in multirate data samples. Thus, the performance of the fault diagnosis can be improved.

C. Training Method

The proposed method is trained in an end-to-end way. The prediction loss can be calculated through cross-entropy loss of the estimated output and the target class, which is defined as follows:

$$L = -\sum_{j=1}^{K} \hat{y_j} \ln\left(O_j\right) \tag{11}$$

where $\hat{y_j}$ is equal to one when the predicted value equal to the true value, otherwise it is zero.

By calculating the partial derivatives of L, the parameters of the model will be updated, and the error will be minimized until the predefined maximum iteration. The refreshment of parameters is expressing as follows:

$$w = w - \alpha \frac{\partial L}{\partial w} \tag{12}$$

where w is the parameter needed to be updated and α is the learning rate that relates to the updating speed. To optimize the updating process, the Adaptive Moment Estimation (Adam) Stochastic optimization algorithm is adopted in this work, which can minimize the cost function in back propagation. The detailed procedures of the proposed method are summarized in the following algorithm:

Algorithm CNN-Based Fault Diagnosis Method With Multirate Data Samples

Input: Multirate data samples and the corresponding labels.
Step 1 (Data preprocessing):
Selecting sensors by Pearson correlation coefficient:

$$\rho_{x_i, x_j} = \frac{\mathrm{cov}(x_i, x_j)}{\sigma_{x_i} \sigma_{x_j}}$$

Constructing inputs for channel r:

$$X^r = \left[X_1^n, X_2^n, \ldots, X_M^n\right], \quad n = 1, 2, 3, \ldots, N$$

Step 2 (Feature extraction):
Constructing the feature extractor with R channels to extract features from X^r. Each channel is composed of convolutional layer, pooling layer, batch normalization layer, and ReLU activation layer, as seen in Fig. 5(a).
Convolutional layer: $C^q = \sum_{p=1}^{P} W^{p,q} * X^p + b^q$
Batch normalization layer: $H = BN(C)$
ReLU activation layer: $T^r = max(H^r, 0)$
Pooling layer: $Z^r = \underset{T^r \in S}{max} T^r$
Step 3 (Modal recognition and fault classification): fusing information learned from all channels at the feature level:

$$Z = \left[Z^1, Z^2, \ldots, Z^r, \ldots, Z^R\right]$$

Constructing the fault classifier, which is composed of fully connected layer, batch normalization layer, activation layer, and Softmax function, as seen in Fig. 5(b).

$$G = ReLU(BN(W_g Z + b_g))$$

$$O_j = P(y = j \mid G; \theta) = \frac{exp(\theta^{(j)} G)}{\sum_{k=1}^{K} exp(\theta^{(k)} G)}$$

Training the model and then conducting fault diagnosis.
Output: the target categories.

IV. Experimental Study

A. Experiment Description

To validate the effectiveness of the proposed method, several experiments are conducted in this section. The experimental data derives from an open-source dataset.[1] detailed in Section II, and more information about the data can be found in [4].

Definition 1: X denotes the multirate data sample composed of multiple single sampling rate data X^r; $X^r = [x_1^n, x_2^n, \ldots, x_M^n]$, $n = 1, 2, \ldots, N$, N represents the sampling number; $r = 1, 2, \ldots, R$, R is the number of the sampling rate; M is the number of sensors under the same sampling rate. Y denotes the corresponding label, using one-hot coding.

Generally, in a given hydraulic system, the fault diagnosis task should be conducted for four components: cooler C1, valve V10, main pump MP1, and accumulators A1–A4. A favorable fault diagnosis method should recognize whether a component has failed and the severity of the failure. In

[1]The dataset has been kindly provided at http://archive.ics.uci.edu/ml/datasets/Condition+monitoring+of+hydraulic+systems

this case, X denotes the multirate data sample defined above, Y denotes the label. To be specific, in the fault diagnosis of the cooler C1, $Y \in$ {full efficiency, reduced efficiency, close to total failure} = {[1,0,0], [0,1,0], [0,0,1]}. In the fault diagnosis of the valve V10, $Y \in$ {optimal switching behavior, small lag, severe lag, close to total failure} = {[1,0,0,0], [0,1,0,0], [0,0,1,0], [0,0,0,1]}. In the fault diagnosis of the main pump MP1, $Y \in$ {no leakage, weak leakage, severe leakage} = {[1,0,0], [0,1,0], [0,0,1]}. In the fault diagnosis of the accumulators A1–A4, $Y \in$ {optimal pressure, slightly reduced pressure, severely reduced pressure, close to total failure} = {[1,0,0,0], [0,1,0,0], [0,0,1,0], [0,0,0,1]}.

Second, to test the diagnostic performance in the noise environment, different degrees of noise are artificially introduced into the data and the same experiments are carried out. In this case, X represents the multirate data sample with noise, Y denotes the corresponding label.

Finally, because the faults of different components in the hydraulic system may occur separately or simultaneously, single fault detection and multiple fault detection experiments are carried out. In the single fault diagnosis experiment, the diagnosis model identifies which of the four components has failed. Specifically, $Y \in$ {faults of the cooler, faults of the main pump, faults of the valve, faults of the accumulator} = {[1,0,0,0], [0,1,0,0], [0,0,1,0], [0,0,0,1]}.

In the multifault diagnosis experiment, the fault of single component or multiple components in the system is identified. In this case, $Y \in$ {normal condition, faults of the main pump, faults of the accumulator, composite fault of the two components} = {[1,0,0,0], [0,1,0,0], [0,0,1,0], [0,0,0,1]}.

B. Fault Diagnosis of Different Components

In the experiment, fault diagnosis is conducted on the four components of the hydraulic system. As illustrated in Table I, there are three conditions of CP decrease, four conditions of valve switching degradation, three conditions of internal pump leakage, and four conditions of accumulator gas leakage. The monitoring data sampled from different sensors are used to construct the developed model. In the data preprocessing, the Pearson correlation coefficient is adopted to measure the correlation of data with the same sampling rate, and the results are represented in the appendix. As mentioned above, sensors with high absolute values of correlation coefficients are screened because they have similar information. Considering the diversity of sensors and the similarity of information, the following sensors are selected. For data with a sampling rate of 100 Hz, the selected sensors are PS1–PS5. For data with a sampling rate of 10 Hz, the selected sensors are FS1 and FS2. For data with a sampling rate of 1 Hz, the selected sensors are TS1, VS, CP, and SE. A total of 2205 instances are used in the experiment: each one is collected in a working cycle lasting 60 s. 80% of the data is randomly divided into the training set and the remaining 20% is divided into the testing set.

Considering that the hydraulic system data have three sampling rates, the number of channels is set to three. Channels 1, 2, and 3 process data with sampling rates of 100, 10, and 1 Hz, respectively. Each channel consists of two sets of structures shown in Fig. 5(a). By setting different parameters, such as the size of the convolution filter, the model can extract features from data at different sampling rates. In the training process, the learning rate of the Adam optimization algorithm is set as 0.001. The number of training iterations is set to vary from 10 to 50 depending on the diagnosis targets. The grid search method is adopted for hyperparameter selection to achieve satisfactory results. However, the time consumption of grid search increases dramatically with the increase of network size and data volume. To make a compromise between training time and diagnosis performance, some hyperparameters are determined by experiences, such as the stride of the convolutional layer, and others are selected by grid search. The hyperparameters of the proposed method in the cooler fault detection are summarized in Table III. Parameters of other fault detection scenarios can be set by referring to this table and the principles in Section III. Note that the proposed method is robust to parameter selection, and the change of a few parameters has little effect on the final results.

TABLE III
PARAMETERS IN FAULT DIAGNOSIS OF THE COOLER

Channel	1	2	3
Input length	6000	600	60
Number of the sensors	5	2	4
Number of the filters (Conv1)	8	4	8
Size of the filters (Conv1)	100	30	7
Strides (Conv1)	8	1	1
Number of the filters (Conv2)	16	8	16
Size of the filters (Conv2)	15	7	3
Strides (Conv2)	3	1	1
Number of the nodes (FC)		64	
Number of the output nodes		3	

To validate the effectiveness of the proposed method, several methods are introduced for comparison. First, to compare the automatic feature extraction capability of the proposed method and the manual feature extraction scheme, linear discriminant analysis (LDA) with handcrafted features are constructed [4]. In LDA, each sample of different sensors is divided into several intervals and feature extraction is performed. The handcrafted features consist of median, variance, skewness, and kurtosis, slope of linear fit, and position of maximum value. Then Pearson correlation coefficient is adopted to analyze the correlation of features and fault; these features will be sorted according to the absolute value and the first few features with the highest correlation will be selected as inputs. Next, Shapelet transform [5] is introduced, which is a feature extraction method based on information acquisition and can effectively classify time-series data. In the experiment, Shapelet transformation converts time-series data to distance data and then performs fault diagnosis. On the other hand, in order to validate the effectiveness of multirate data samples fusion, the proposed method is compared with its baseline method and the CNN method. Here, the CNN method is also a single sampling rate data-based method. The data of sensors with the highest sampling rate are adopted as the input of

TABLE IV
ACCURACY OF FAULT DETECTION OF DIFFERENT COMPONENTS

Component	Method			
	LDA	Shapelet	CNN	**Proposed method**
Cooler	99.89±0.16	98.05±0.66	**100.00±0.00**	**100.00±0.00**
Valve	**99.97±0.31**	90.16±0.96	**100.00±0.00**	**100.00±0.00**
Pump	**99.41±0.31**	93.95±1.51	97.20±0.55	98.98±0.16
Acc.	92.48±1.51	90.06±0.82	97.85±0.55	**99.35±0.38**

CNN, and the parameters are consistent with the first channel of the proposed method.

The experiments are repeated 50 times for all the above methods to avoid particularity and contingency in the fault diagnosis results. The average results are recorded in Table IV in the following format: average testing accuracy (%) ± standard deviation (%). In general, the proposed model achieves the best detection performance. The two CNN-based methods outperform the SML methods in fault detection of cooler, valve, and accumulator fault detection. Especially in the fault diagnosis of accumulator, compared with LDA and Shapelet transform, the accuracy of the proposed method is improved by 7% and 9%. The performance improvement is attributed to CNN's powerful feature extraction ability, and its deep structure can learn more information than the shallow model. Compared with the CNN model using single sampling rate data, the proposed method performs better in the fault detection of hydraulic pump and accumulator with average accuracy reaching 98.98% and 99.35%, respectively. The better performance indicates that multirate data fusion can use more information than single sampling rate data to improve the diagnosis effect. Besides, the variance of the proposed method is smaller than the comparison methods, showing its good stability.

To display the results more visually and give the misclassification analysis, the confusion matrices of the four conditions are presented in Fig. 6. The ordinate axis denotes the true label of samples, and the horizontal axis denotes the predicted label of samples. In fault diagnosis of cooler and valve, all samples are classified into the corresponding categories. Although in the fault detection of the main pump and accumulator, only a few samples are misplaced into adjacent categories. The described misclassification reveals that the similarity of sensor data between consecutive degradation states may cause confusion in the fault recognition [33].

C. Performance Under Noise Environment

In this experiment, the detection accuracy of the proposed method under noisy environments will be discussed. Additive White Gaussian noise with power P_n is added to the original signals to simulate a noisy environment. The power of the noise is proportional to the power of the original signal P_s, and the signal-noise ratio (SNR) is introduced to describe the ratio of P_s and P_n. The calculation of SNR and power is

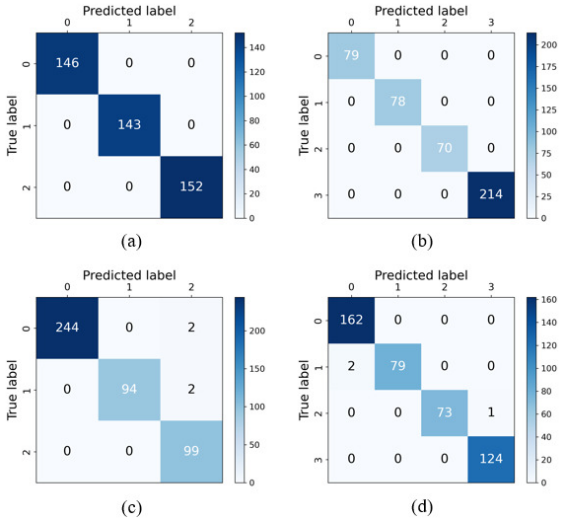

Fig. 6. Confusion matrix of different conditions: (a) cooler condition, (b) valve condition, (c) internal pump leakage, and (d) hydraulic accumulator.

TABLE V
ACCURACY OF SINGLE FAULT DETECTION AND MULTIFAULT DETECTION

Method	Accuracy (%)±standard deviation (%)	
	Single fault	Multi-fault
SVM	95.32±1.84	91.40±2.63
Proposed method	**98.12±0.98**	**99.97±0.18**

displayed as follows:

$$\mathrm{SNR} = 10\lg\left(\frac{P_s}{P_n}\right) \tag{13}$$

$$P = \frac{1}{N}\sum_{i=1}^{N}(x_i)^2 \tag{14}$$

where symbol x is the data of signal or noise and symbol N is the length of the data.

In the experiment, the SNR is set to increase from 20 to 40; in other words, the power of noise is set to increase from 0.01% to 1% of the power of the signal. Fig. 7 displays the pressure sensor signal under a noisy environment. Two SML algorithms, support vector machine (SVM) and random forest (RF), are introduced and compared with the proposed method. The main procedures of the comparative methods are consist with LDA in the previous experiment.

The results are shown in Fig. 8. Generally speaking, the accuracy of fault detection increases with the increase of SNR. The noise environment has different effects on fault diagnosis results of different components. In the fault detection of cooler and valve, the noise has little effect on the results with the diagnosis accuracy of all three methods close to 100.0%. While in the fault diagnosis of pump and accumulator where the fault scenarios are more complicated, the noise environment has more influence on the accuracy. Compared with traditional SML methods, the proposed method is less affected, and its accuracy is all above 90%. With the increase

TABLE VI
VALUE OF PEARSON CORRELATION COEFFICIENT (100 HZ)

Sensors	PS1	PS2	PS3	PS4	PS5	PS6	EPS1
PS1	1.000000	-0.610981	-0.740132	0.010230	-0.034836	-0.034341	**0.969282**
PS2	-0.610981	1.000000	0.733590	-0.002779	-0.020112	-0.019890	-0.568739
PS3	-0.740132	0.733590	1.000000	0.131371	0.202443	0.202142	-0.624139
PS4	0.010230	-0.002779	0.131371	1.000000	0.740596	0.740827	0.151747
PS5	-0.034836	-0.020112	0.202443	0.740596	1.000000	**0.999932**	0.142904
ps6	-0.034341	-0.019890	0.202142	0.740827	**0.999932**	1.000000	0.143361
EPS1	**0.969282**	-0.568739	-0.624139	0.151747	0.142904	0.143361	1.000000

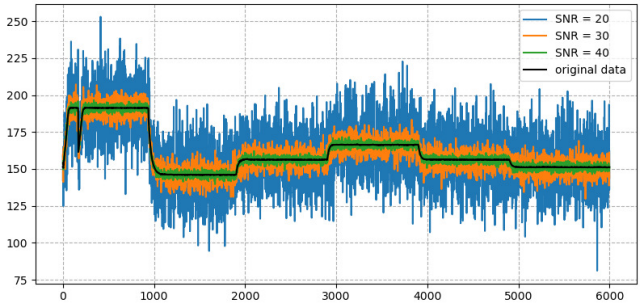

Fig. 7. Pressure sensor signals under noisy environment.

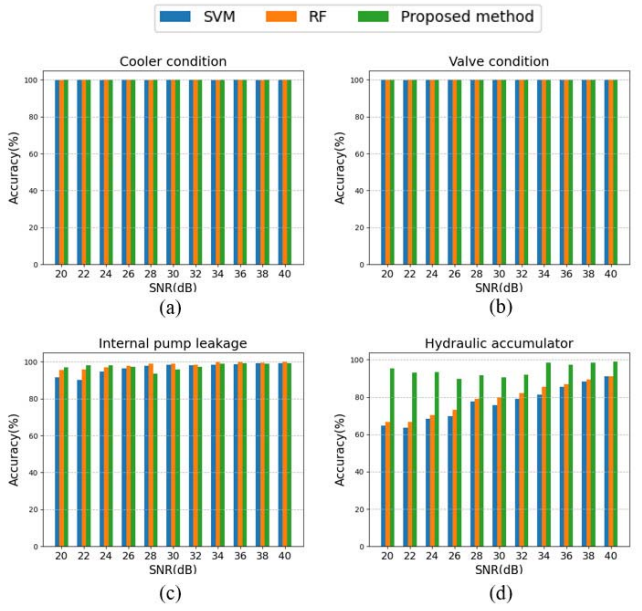

Fig. 8. Accuracy of fault diagnosis under different SNR: (a) cooler condition, (b) valve condition, (c) internal pump leakage, and (d) hydraulic accumulator.

of SNR, the accuracy of the proposed method is close to that in a noiseless environment. In the fault diagnosis of hydraulic accumulators, the gaps of accuracy between the proposed method and the SML methods are particularly obvious, reaching up to 32%, which demonstrates the robustness of the proposed method.

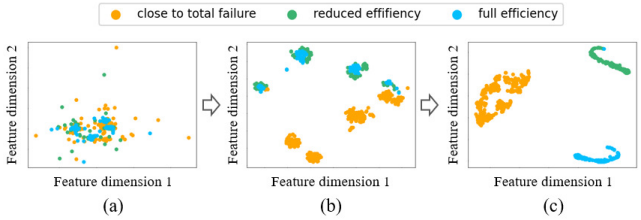

Fig. 9. Visualization of the test samples using t-SNE: (a) raw data, (b) features learned by the feature extractor, and (c) features learned by the classification.

TABLE VII
VALUE OF PEARSON CORRELATION COEFFICIENT (10 HZ)

Sensors	FS1	FS2
FS1	1.0000	0.1747
FS2	0.1747	1.0000

D. Visualization Analysis With t-SNE

To provide an intuitive comprehension of the effectiveness of the proposed method, t-distributed stochastic neighbor embedding (t-SNE) is employed in the experiment to visualize the representation of the test samples in a two-dimension space. t-SNE is a nonlinear dimension reduction technique, which is easy to optimize and particularly suitable for the visualization of high-dimensional datasets. Besides, t-SNE is superior to existing methods in creating a map showing different scale structures. Therefore, it is widely adopted to investigate the feature learning processes [34].

The hydraulic pump is the key component of hydraulic systems, the failure of which may result in the collapse of the whole system. In the experiment, the visual analysis of hydraulic pump fault diagnosis is conducted. The test samples are first fed into the trained model, and the features are obtained from the feature extractor and the classification. Then, t-SNE is adopted to convert the high-dimensional features into 2-D coordinates, the scatter plots are drawn based on these coordinates and displayed in Fig. 9. The first picture demonstrates the distribution of the raw data, the second picture displays the distribution of features learned by the feature extractor of the proposed model, and the third shows the distribution of learned representation from the fully connected layer of the classifier.

TABLE VIII
VALUE OF PEARSON CORRELATION COEFFICIENT (1 HZ)

Sensors	TS1	TS2	TS3	TS4	VS	CE	CP	SE
TS1	1.000000	**0.999189**	**0.999336**	**0.998950**	0.678713	-0.945480	-0.908670	-0.140644
TS2	**0.999189**	1.000000	**0.998829**	**0.998140**	0.672206	-0.945718	-0.905810	-0.135829
TS3	**0.999336**	**0.998829**	1.000000	**0.998315**	0.685130	-0.940870	-0.899737	-0.149874
TS4	**0.998950**	**0.998140**	**0.998315**	1.000000	0.682794	**-0.955705**	-0.923454	-0.145083
VS	0.678713	0.672206	0.685130	0.682794	1.000000	-0.636820	-0.609702	-0.524054
CE	-0.945480	-0.945718	-0.940870	**-0.955705**	-0.636820	1.000000	**0.972298**	0.100326
CP	-0.908670	-0.905810	-0.899737	-0.923454	-0.609702	**0.972298**	1.000000	0.102600
SE	-0.140644	-0.135829	-0.149874	-0.145083	-0.524054	0.100326	0.102600	1.000000

As seen in Fig. 9, the points of raw data from different categories overlap with each other, so they cannot be divided with simple lines. By contrast, the features learned by the feature extractor have better separability. The data samples of close to total failure are separated from the other categories, which demonstrates that the representations learned by the feature extractor are helpful to preliminarily distinguish the fault categories. In the last graph, the data points of different categories are separated from each other, which indicates that the proposed method has strong nonlinear mapping capability and can effectively identify faults.

E. Single Fault and Multifault Detection of Hydraulic Systems

In real hydraulic systems, the faults of the four components may occur individually or simultaneously, and the fault features will become more complicated when a composite failure occurs. To testify the performance of the proposed model in these two cases, single fault diagnosis and multifault diagnosis are performed on hydraulic systems.

The purpose of the single fault detection is to diagnose which component in the hydraulic system has a fault. Specifically, the diagnostic model needs to identify the following four cases: the single fault of the cooler, the main pump, the valve, and the accumulators. The purpose of multifault detection is to identify whether a compound fault occurs in the system. Two components of the hydraulic system are selected to detect the following four cases: no-fault, a single fault of each component, and composite fault of both components. In the experiment, the main pump and accumulator are selected because their fault characteristics are more hidden than those of the cooler and valve, which makes multifault detection more challenging. Therefore, the model needs to identify the following conditions, including normal condition, main pump fault, accumulator fault, and composite fault of the two components. It should be noted that both experiments only identify the faulty components, not measure their severity.

A multivariate time series classification-based diagnosis method is introduced for comparison [35]. It extracts timing characteristics using 1-NN method and trains a SVM to classify multivariate time series of the hydraulic system. The results are displayed in Table V. This method outperforms SVM in both the cases. It achieves higher accuracy and smaller deviation, which shows that the proposed method has better fault recognition ability and is more stable. Especially in the case of multifault detection, the accuracy reaches 99.97%, which is improved by about 8% compared with SVM. The improved recognition performance is attributed to the deep structure of the proposed method, which can excavate more abundant detection information and the hidden coupling relationship from multirate sampling data.

V. CONCLUSION

This article proposed a CNN-based fault diagnosis method of hydraulic systems by designing the DL model with multirate data sample. The channels of the proposed method can automatically extract features from the raw data with different sampling rates. Then these features are concatenated to form the final expression, and the diagnosis results are given by classification. Compared with the data of a single sampling rate, the data based on the multisampling rate can obtain more abundant features. Moreover, compared with artificial feature extraction, this method has stronger feature extraction ability and can extract hidden features of noise. Therefore, expert knowledge is not needed during the fault diagnosis, which can be easily extended to other mechanical systems and more suitable for industrial applications. Besides, the method proposed can effectively diagnose the single fault and multiple faults of different components. Several experiments have been conducted to validate the effectiveness and generalization ability of the proposed method. The experiment results show that compared with Shapelet transformation, the accuracy of fault diagnosis is improved by about 10%. Moreover, in single fault and multifault diagnosis, the proposed method also achieves higher diagnosis performance compared with multiclass SVM.

The limitations of the developed method are summarized as follows. First, the fault patterns of hydraulic systems must be detected and defined before performing fault diagnosis. Second, although CNN is effective in fault diagnosis scenarios, its back-propagation training process is time consuming. Thus, it may not show satisfactory performance when dealing with high-dimensional data and performing real-time diagnosis. In addition, extending the proposed method to detect unknown faults timely of the system with high-dimensional data deserve further research in the future.

APPENDIX

In the data preprocessing, Pearson correlation coefficient is utilized to measure the correlation of data with the same sampling rate, and the true values are represented in Tables VI–VIII.

REFERENCES

[1] C. Keliris, M. M. Polycarpou, and T. Parisini, "An integrated learning and filtering approach for fault diagnosis of a class of nonlinear dynamical systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 4, pp. 988–1004, Apr. 2017.

[2] K. Huang, Y. Wu, C. Wang, Y. Xie, C. Yang, and W. Gui, "A projective and discriminative dictionary learning for high-dimensional process monitoring with industrial applications," *IEEE Trans. Ind. Informat.*, vol. 17, no. 1, pp. 558–568, Jan. 2021.

[3] K. Huang, Y. Wu, C. Yang, G. Peng, and W. Shen, "Structure dictionary learning-based multimode process monitoring and its application to aluminum electrolysis process," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 4, pp. 1989–2003, Oct. 2020.

[4] N. Helwig, E. Pignanelli, and A. Schütze, "Condition monitoring of a complex hydraulic system using multivariate statistics," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I MTC)*, May 2015, pp. 210–215.

[5] Y. Joo, J. Jeong, and S. Bae, "Under sampling bagging shapelet transformation for hydraulic system," in *Proc. Int. Conf. Omni-Layer Intell. Syst.*, May 2019, pp. 13–18.

[6] Y. Lei, W. Jiang, A. Jiang, Y. Zhu, H. Niu, and S. Zhang, "Fault diagnosis method for hydraulic directional valves integrating PCA and XGBoost," *Processes*, vol. 7, no. 9, p. 589, Sep. 2019.

[7] X. Deng, X. Tian, S. Chen, and C. J. Harris, "Nonlinear process fault diagnosis based on serial principal component analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 3, pp. 560–572, Mar. 2018.

[8] G. F. Bin, J. J. Gao, X. J. Li, and B. S. Dhillon, "Early fault diagnosis of rotating machinery based on wavelet packets—Empirical mode decomposition feature extraction and neural network," *Mech. Syst. Signal Process.*, vol. 27, pp. 696–711, Feb. 2012.

[9] H. Shao, H. Jiang, Y. Lin, and X. Li, "A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders," *Mech. Syst. Signal Process.*, vol. 102, pp. 278–297, Mar. 2018.

[10] 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.

[11] T. de Bruin, K. Verbert, and R. Babuska, "Railway track circuit fault diagnosis using recurrent neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 3, pp. 523–533, Mar. 2017.

[12] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mech. Syst. Signal Process.*, vols. 72–73, pp. 303–315, May 2016.

[13] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, Jan. 2019.

[14] J. Kim, A.-D. Nguyen, and S. Lee, "Deep CNN-based blind image quality predictor," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 1, pp. 11–24, Jan. 2019.

[15] W. Shi, Y. Gong, X. Tao, J. Wang, and N. Zheng, "Improving CNN performance accuracies with min–max objective," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 7, pp. 2872–2885, Jul. 2018.

[16] R. Liu, G. Meng, B. Yang, C. Sun, and X. Chen, "Dislocated time series convolutional neural architecture: An intelligent fault diagnosis approach for electric machine," *IEEE Trans. Ind. Informat.*, vol. 13, no. 3, pp. 1310–1320, Jun. 2017.

[17] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5990–5998, Jul. 2018.

[18] Y. Yuan *et al.*, "A general end-to-end diagnosis framework for manufacturing systems," *Nat. Sci. Rev.*, vol. 7, no. 2, pp. 418–429, Feb. 2020.

[19] A. Stief, J. R. Ottewill, J. Baranowski, and M. Orkisz, "A PCA and two-stage Bayesian sensor fusion approach for diagnosing electrical and mechanical faults in induction motors," *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9510–9520, Dec. 2019.

[20] J. Wang, P. Fu, L. Zhang, R. X. Gao, and R. Zhao, "Multilevel information fusion for induction motor fault diagnosis," *IEEE/ASME Trans. Mechatronics*, vol. 24, no. 5, pp. 2139–2150, Oct. 2019.

[21] L. Jing, T. Wang, M. Zhao, and P. Wang, "An adaptive multi-sensor data fusion method based on deep convolutional neural networks for fault diagnosis of planetary gearbox," *Sensors*, vol. 17, no. 2, p. 414, Feb. 2017.

[22] Y. Cong, Z.-Q. Ge, and Z.-H. Song, "Multi-rate principle component analysis for process monitoring," *Shanghai Jiaotong Daxue Xuebao, J. Shanghai Jiaotong Univ.*, vol. 49, pp. 762–767, Jun. 2015.

[23] Y. Masuda, H. Kaneko, and K. Funatsu, "Multivariate statistical process control method including soft sensors for both early and accurate fault detection," *Ind. Eng. Chem. Res.*, vol. 53, no. 20, pp. 8553–8564, May 2014, doi: 10.1021/ie501024w.

[24] Y. Wu and X. Luo, "A novel calibration approach of soft sensor based on multirate data fusion technology," *J. Process Control*, vol. 20, no. 10, pp. 1252–1260, Dec. 2010.

[25] X. Shao, B. Huang, J. M. Lee, F. Xu, and A. Espejo, "Bayesian method for multirate data synthesis and model calibration," *AIChE J.*, vol. 57, no. 6, pp. 1514–1525, Jun. 2011.

[26] R. Zhou, Z. Jiao, and S. Wang, "Current research and developing trends on fault diagnosis of hydraulic systems," *Chin. J. Mech. Eng.*, vol. 30, no. 8, pp. 19–23, 2010.

[27] Z. Zhao, F.-L. Wang, M.-X. Jia, and S. Wang, "Intermittent-chaos-and-cepstrum-analysis-based early fault detection on shuttle valve of hydraulic tube tester," *IEEE Trans. Ind. Electron.*, vol. 56, no. 7, pp. 2764–2770, Jul. 2009.

[28] T. Schneider, N. Helwig, S. Klein, and A. Schütze, "Influence of sensor network sampling rate on multivariate statistical condition monitoring of industrial machines and processes," *Proceedings*, vol. 2, no. 13, p. 781, 2018.

[29] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, p. 425, Feb. 2017.

[30] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, "How does batch normalization help optimization?" in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 2483–2493.

[31] S. Wu *et al.*, "$L1$-norm batch normalization for efficient training of deep neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 7, pp. 2043–2051, Jul. 2019.

[32] L. Chen, G. Xu, S. Zhang, W. Yan, and Q. Wu, "Health indicator construction of machinery based on end-to-end trainable convolution recurrent neural networks," *J. Manuf. Syst.*, vol. 54, pp. 1–11, Jan. 2020.

[33] C. König and A. M. Helmi, "Sensitivity analysis of sensors in a hydraulic condition monitoring system using CNN models," *Sensors*, vol. 20, no. 11, p. 3307, Jun. 2020.

[34] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

[35] X. Zhao, K. Zhang, and Y. Chai, "A multivariate time series classification based multiple fault diagnosis method for hydraulic systems," in *Proc. Chin. Control Conf. (CCC)*, Jul. 2019, pp. 6819–6824.

Keke Huang (Member, IEEE) received the B.A. degree in automatic control from Northeastern University, Shenyang, China, in 2012, and the Ph.D. degree in control science and engineering from Tsinghua University, Beijing, China, in 2017.

He is currently an Associate Professor with Central South University, Changsha, China. His research interests include industrial big data and complex system and network science.

Shujie Wu received the B.A. degree in automation from Central South University, Changsha, China, in 2020, where she is currently pursuing the M.A. degree in control science and engineering.

Her research interests include industrial fault diagnosis and process monitoring.

Fanbiao Li (Member, IEEE) received the B.Sc. degree in applied mathematics from Mudanjiang Normal University, Mudanjiang, China, in 2008, the M.Sc. degree in operational research and cybernetics from Heilongjiang University, Harbin, China, in 2012, and the Ph.D. degree in control theory and control engineering from the Harbin Institute of Technology, Harbin, in 2015.

From December 2013 to April 2015, he was a Joint Training Ph.D. Student with the School of Electrical and Electronic Engineering, The University of Adelaide, Adelaide, Australia, where he was a Research Associate, from April 2015 to February 2016. From April 2017 to March 2018, he was an Alexander von Humboldt Research Fellow with the University of Duisburg-Essen, Duisburg, Germany. In 2016, he joined as an Associate Professor with Central South University, Changsha, China. His research interests include stochastic systems, sliding mode control, and fault diagnosis and identification.

Dr. Li currently serves as an Associate Editor for a number of journals, including IEEE ACCESS and *ICIC-Express Letters*. He is an Associate Editor for the Conference Editorial Board and the IEEE Control Systems Society (Based on document published on September 2, 2019).

Chunhua Yang (Senior Member, IEEE) received the M.Eng. degree in automatic control engineering and the Ph.D. degree in control science and engineering from Central South University, Changsha, China, in 1988 and 2002, respectively.

She was with the Department of Electrical Engineering, Katholieke Universiteit Leuven, Leuven, Belgium, from 1999 to 2001. She is currently a Full Professor with Central South University. Her current research interests include modeling and optimal control of complex industrial processes, intelligent control systems, and fault-tolerant computing of real-time systems.

Weihua Gui received the B.Eng. degree in electrical engineering and the M.S. degree in automatic control engineering from Central South University, Changsha, China, in 1976 and 1981, respectively.

From 1986 to 1988, he was a Visiting Scholar with the University of DuisburgEssen, Duisburg, Germany. Since 2013, he has been an Academician of the Chinese Academy of Engineering, Beijing, China. He is currently with the School of Automation, Central South University. He is an Academician of the Chinese Academy of Engineering. His current research interests include modeling and optimal control of complex industrial processes, fault diagnoses, and distributed robust control.